Quality of Life Research (2023) 32:1319–1327
https://doi.org/10.1007/s11136-022-03170-7

SPECIAL SECTION: METHODOLOGIES FOR MEANINGFUL CHANGE

# Using qualitative methods to establish the clinically meaningful threshold for treatment success in alopecia areata

**Kathleen W. Wyrwich[1] · Helen Kitchen[2] · Sarah Knight[2] · Natalie V. J. Aldhouse[2] · Jake Macey[2] · Natasha Mesinkovska[3] · Justin M. Ko[4] · Brett A. King[5]**

Accepted: 7 June 2022 / Published online: 12 July 2022


## Abstract

**Purpose** Traditionally, appropriate anchors are used to investigate the amount of change on a clinician-reported outcome assessment that is meaningful to individual patients. However, novel qualitative methods involving input from disease state experts together with patients may better inform the individual improvement threshold for demonstrating the clinical benefit of new treatments. This study aimed to establish a clinically meaningful threshold for treatment success for the clinician-reported Severity of Alopecia Tool (SALT) score for patients with alopecia areata (AA).

**Methods** A purposive sample of 10 dermatologists expert in AA and 30 adult and adolescent patients with AA and a history of ≥ 50% scalp hair loss were recruited. Semi-structured interview questions explored the outcome that represented treatment success to clinicians and patients. Findings were analyzed using thematic methods to identify treatment success thresholds.

**Results** Both informant groups confirmed scalp hair amount as the outcome of priority. Most expert clinicians considered a static threshold of 80% ($n = 5$) or 75% ($n = 3$) of the scalp hair as a treatment success. Most patient responses ranged from 70 to 90% (median: 80% of the scalp hair). Subsequently, queried patients confirmed that achieving SALT score ≤ 20 with treatment would be a success, as reflected in the Alopecia Areata Investigator Global Assessment (AA-IGA™). The novel qualitative processes used to inform this meaningful threshold reflects a clinician-then-patient process for: (a) confirmation of the patient outcome of priority; and (b) clinician input on a preliminary treatment success level for independent understanding among patients.

**Conclusion** This qualitative investigation of expert clinicians-then-patients with AA confirmed that achieving an amount of 80% or more scalp hair (SALT score ≤ 20) was an appropriate individual treatment success threshold indicating clinically meaningful improvement for patients with ≥ 50% scalp hair loss. A qualitative investigation of a quantifiable treatment success threshold is possible through a well-designed interview process with expert clinicians and the appropriate patient population.

**Keywords** Alopecia areata · Patient-focused drug development · Meaningful change · Important difference · Qualitative research · Hair loss

✉ Kathleen W. Wyrwich
kathy.wyrwich@gmail.com

Helen Kitchen
hkitchen@clarivate.com

Sarah Knight
sknight@clarivate.com

Natalie V. J. Aldhouse
naldhouse@clarivate.com

Jake Macey
jmacey@clarivate.com

Natasha Mesinkovska
nmesinko@uci.edu

Justin M. Ko
jmko@stanford.edu

Brett A. King
brett.king@yale.edu

[1] St. Louis, USA

[2] Clinical Outcomes Assessment, DRG (Part of Clarivate), Manchester, UK

[3] Department of Dermatology, University of California Irvine, Irvine, CA, USA

[4] Department of Dermatology, Stanford University School of Medicine, Redwood City, CA, USA

[5] Department of Dermatology, Yale University School of Medicine, New Haven, CT, USA

## Introduction

Alopecia areata (AA) is an auto-immune disease, characterized by hair loss, with a devastating psychological and social toll on those affected and their families [1–6]. Contemporary emerging treatments for AA necessitate a clear understanding of patient priorities for treatment and the relevant thresholds for determining treatment success. Qualitative methods provide an opportunity to investigate these key measurement goals.

Great advancements over the past two decades led to the development and the dissemination of dermatologic training for the Severity of Alopecia Tool (SALT), a systematic method for clinician assessment of scalp hair loss on a 0 (= no missing scalp hair) to 100 (= 100% missing; no scalp hair) scale [7, 8]. A publication in 2004 introduced the SALT score and promoted the percent change in SALT score to understand treatment response, with 50% improvement from baseline (i.e., $SALT_{50}$) noted as an acceptable endpoint for trials involving extensive alopecia areata (≥ 50% scalp hair loss at baseline) and systemic agents [7]. Indeed, two landmark proof-of-concept studies investigating the safety and efficacy of oral Janus kinase (JAK) inhibitors in patients with extensive alopecia areata reported the proportion of subjects with ≥ 50% scalp hair regrowth from baseline to end of treatment as the primary endpoint [9] or to define the *strong responder* classification [10].

Currently, there are no regulatory-approved treatments for AA. As noted by the U.S. Food and Drug Administration (FDA), "an important aspect of medical product development is the demonstration of clinical benefit and how that benefit is measured" [11], with similar expectations from other regulatory authorities for registration studies (e.g., European Medicines Agency, Japan's Pharmaceuticals and Medical Devices Agency). The primary, co-primary, or pre-specified secondary endpoints in registration trials used to support medical product approval and labeling claims and other communications of clinical benefit are often clinical outcome assessments (COAs), which include patient-reported outcome (PRO), clinician-reported outcome (ClinRO), observer-reported outcome (ObsRO), performance outcome (PerfO) tools, and certain COAs derived from health technologies [11]. The FDA defines *clinical benefit* as "a positive clinically meaningful effect of an intervention on how an individual feels, functions, or survives" and per the FDA, "the process of selecting or developing a COA for use in a medical product development program depends on having adequately characterized the disease or condition, defined the target context of use, and conceptualized a concept of interest that represents clinical benefit" [11]. Moreover, "when a clinical benefit is demonstrated, a description of that benefit can be provided in the regulators' approved labeling or approved communications of the concept or outcome measured (i.e., the aspect of an individual's clinical, biological, physical, functional state, or experience that the assessment is intended to capture)" [11].

With these key directives for registration trial endpoints, it is interesting to consider the 50% improvement in SALT score from baseline ($SALT_{50}$) that has been suggested to define responders in many AA studies. First, if a patient with no scalp hair (SALT score 100) enrolled in a treatment study and achieved this responder status of 50% improvement over time (i.e., achieved SALT score 50), the patient would continue to have extensive scalp hair loss after treatment. Would this patient (or their provider) consider achieving $SALT_{50}$ status to be a *clinical benefit* with "a positive clinically meaningful effect of an intervention on how an individual feels, functions, or survives"? Additionally, a $SALT_{50}$ endpoint assumes that scalp hair regrowth is the most important and meaningful treatment outcome for patients. For all stakeholders, it therefore became critical to understand whether scalp hair regrowth was indeed the most important and meaningful treatment outcome for patients with AA versus hair restoration at other locations (e.g., eyebrows and eyelashes). Furthermore, if this key treatment outcome concept could be soundly established, what would be the best estimate for the improvement needed to achieve a clinical benefit at the individual level?

Estimations of meaningful change thresholds on COAs have traditionally been derived through established quantitative methods with anchor-based methods considered the 'gold standard' and distribution-based methods, e.g., half standard deviation (0.5 SD) and standard error of measurement (SEM) considered supportive [11–15]. Recently, qualitative methods have emerged as a complementary endeavor [14] to answer this fundamentally patient-centered question of 'What is a meaningful change for patients?' and there are several examples of interview studies, clinical trial exit interviews, and Delphi studies that have addressed this question [16–20]. Patient perspectives permit further contextualization of the quantitative metric that is derived from anchor- and distribution-based methods, allowing us to understand *why* the score change is important [21].

This emergent qualitative methodology has mostly focused on within-patient (i.e., individual level) change to understand meaningful improvement and clinical benefit [11, 21]. The focus on exploring within-patient change thresholds has likely been driven in part by regulatory need in drug development (i.e., the need to define and understand a responder definition) and in part by practicalities involved in COA development; patients and/or caregivers are typically interviewed individually to report on personal experience and not to report on differences between groups

of patients or treatment groups [21]. There is also a theoretical justification for the focus on individual-level perception of meaningful differences which is congruent with the epistemological foundation of qualitative methods based in grounded theory, such as thematic analysis [22], and the phenomenological interpretative approach which seeks to understand the multiple realities of participants rather than one 'true' reality, and focuses on the perceptions and lived experiences of individuals [23].

This study utilized a clinicians-then-patients qualitative interview methodology to derive insights into a clearer understanding of: the disease; conceptualization of the most important treatment need; and a categorization of the SALT score into an Investigator Global Assessment (IGA) that could detect clinically meaningful improvement for patients with extensive AA. This paper details the novel qualitative method used to establish a clinically meaningful threshold for treatment success in AA.

## Methods

### Overview

A novel clinician-then-patient qualitative interview process was created to investigate the key objectives of: (1) understanding patient priorities for treatment, and (2) the relevant threshold to declare treatment success for adult and adolescent patients with AA and ≥ 50% scalp hair loss. This process began with a label review of primary endpoints in recent (2015–2017) FDA dermatology product approvals, followed by in-depth clinician interviews to gain relevant knowledge of clinicians' perceptions, in particular of treatment success. The process continued with in-depth patient concept elicitation and cognitive debriefing interviews, with discussion topics informed by the clinicians' detailed input. This interview flow allowed the outcome of priority to emerge within each informant group, in addition to informed estimates of treatment success levels.

### Clinician Interviews

Clinician Interviews detailed the diagnosis, management, and treatment of patients with AA, and the AA measurement tools that clinicians used. These interviews also provided detailed insight on clinician perceptions of the importance of hair loss in specific locations (e.g., scalp, eyebrows, and eyelashes).

US dermatologists were identified by Eli Lilly and Company scientists for their contemporary expertise in the diagnosis and treatment of patients with AA, and were recruited through email invitations that outlined the scope of their participation. One-on-one telephone interviews lasting approximately 60 minutes were conducted by experienced qualitative interviewers trained in COA development techniques (HK or NVJA) and using a semi-structured interview script to systematically explore topics in depth with each clinician while providing consistency across the interviews.

Using the recommendations that emerged from the Clinician Interviews, a Small Panel of two expert clinicians who participated in the interviews was convened to review the quotes and explanations provided in the Clinician Interviews data. The Small Panel was able to incorporate the data from the 10 clinician informants to permit draft COA wordings for review by patients.

### Patient Interviews

#### Concept elicitation and cognitive debriefing

The Patient Interviews study protocol was approved by Western Institutional Review Board (ref. #20171820). The interviews were conducted to solicit open-ended patient input to understand the signs and symptoms of AA, the associated impacts (concept elicitation) and the thresholds for meaningful change that patients considered a treatment success. The content validity of newly developed PRO and ClinRO measures were evaluated and documented (cognitive debriefing). The categories of the newly developed IGA were of particular interest to gain patient insights into the categories that would represent a meaningful change. While all patients participated in the concept elicitation and cognitive debriefing portions of each interview, the latter was abbreviated for some patients due to time considerations.

Learnings from the Clinician Interviews informed the semi-structured Patient Interviews guide. One-on-one face-to-face interviews lasting 90 minutes—all conducted by the same trained interviewer (NVJA)—offered the opportunity to systematically explore topics in depth while providing consistency across the interviews.

To recruit a patient sample that reflected the range of clinical and demographic characteristics representative of the AA patient population, purposive sampling was used [24]. Minimal sampling targets were used to recruit patients within key demographic and socioeconomic status subgroups. It was important to include patients who had been treated successfully with JAK inhibitors to understand patient perception of their changes in hair growth and to enable comparisons of key concepts and assessment of patients' perceptions of clinically meaningful change with JAK inhibitor naïve patients. Additionally, it was important to understand the key concepts and clinically meaningful improvement from the perspectives of patients with eyebrow and/or eyelash involvement in addition to scalp hair involvement; therefore, this patient group was purposely oversampled. Patients were recruited from two US clinical

sites: University of California-Irvine and Yale University in Connecticut.

### Coding and analysis

Clinical experience (clinicians) and demographic and clinical characteristics (patients) were collected and summarized. Interviews were audio-recorded with the permission of the interviewee, transcribed, and then examined via thematic analysis assisted by ATLAS.ti Version 7.5 software for the coding and organization of the interview data [25]. During the coding process, all identifying information (e.g., name and specific location) was removed from the transcripts. Employing a phenomenological interpretative approach, the thematic analysis sought understandings of participants' multiple realities, focused on the individual interviewee's feelings, perceptions, and lived experiences [23].

The following steps for thematic analysis were followed to explore the open-ended concept elicitation interview data:

1. Familiarizations: the lead analysts read the transcripts to identify overarching ideas.
2. Generating codes: within the transcripts, descriptive codes were generated and then assigned to interviewee quotes.
3. Searching for themes: using the descriptive codes, potential themes were collated.
4. Reviewing, defining, and naming themes: these themes were then compared and contrasted in order to assess any relationships between then, both within and between participants.
5. Reporting: key concepts and themes were identified within each interview and across the respective samples (clinicians and patients), and supportive quotes extracted and reported.

Data obtained during the cognitive debriefing review of draft items were subject to framework analysis whereby a pre-defined code list was applied to identify the relevance and appropriateness of item wordings, response options, and recall periods. As emergent data were expected in the debriefing/item review discussion, iterative codes were also applied.

## Results

### Dermatology product labels endpoint review

The FDA-labeled dermatology products endpoint review concluded that, in general, ordinal, static investigator global assessment (IGA) scales informed the interpretation of clinical trial results in approved dermatologic conditions (2015–2017). Nearly all the IGAs used in recent dermatology product labeling had 4 or 5 levels and the IGA levels demonstrated distinct, non-overlapping and clinically relevant gradations of the specific disease/condition, with the highest level (0 or none) representing true clinical absence of the disease/condition. A static (versus dynamic) IGA assesses the clinician's current impression of disease severity and does not depend on the clinician's recollection of baseline disease severity, which could introduce recall bias [26]. Interestingly, achievement of the top two levels (0 or 1) with at least a 2-level change from baseline on the reviewed IGAs was often required to establish individual patient treatment success.

### Clinician Interviews

Ten dermatologists, expert in the diagnosis and treatment of AA from across the US, participated in qualitative telephone interviews in July 2017. On average, these clinicians had been treating or managing patients with AA for 21.2 years (range 6–38 years).

The clinicians overwhelmingly described scalp hair loss both as the most common chief complaint for patients with AA and the primary sign of AA. All interviewed clinicians assessed patients' scalp hair loss in clinical practice, and used either the SALT to evaluate a patient's change over time or made a visual assessment of whether significant improvement had been achieved since a prior visit. Additionally, two of the clinicians made use of photography to monitor the progress of their patients with AA. All clinicians were familiar with the SALT, and nine of the 10 clinicians had used the SALT in AA clinical studies. All clinicians agreed that assessments of regrown hair should include only terminal hair (not vellus hair), although the latter may be an early indicator of eventual terminal hair growth.

When queried about their perspectives on AA "treatment success," clinicians noted several factors influencing their judgment of this goal. The *quantity* of scalp hair growth was described as the primary aim of treatment by all clinicians, with other features noted by a smaller number of clinicians, including location/pattern of regrowth and hair density.

When asked to describe the amount/percentage of scalp hair they would consider a treatment success, the predominant response was 80% of the scalp hair ($n=5$), followed by 75% ($n=3$) and 90% ($n=1$) (Fig. 1). One clinician did not report a static scalp hair amount, choosing "at least 50% improvement" as the treatment success response.

Although patient input on the key construct and the appropriate level for that construct to indicate a treatment success/clinical benefit was still needed, clinicians are the primary reporter of scalp hair loss in clinical practice and research. Therefore, clinicians reviewed and iteratively developed the IGA for AA scalp hair loss, with a focus on

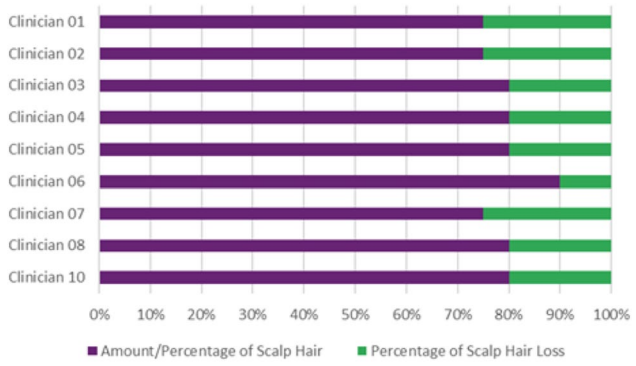


**Fig. 1** Clinician treatment success thresholds. (Color figure online)

ensuring distinct and clinically relevant gradations of scalp hair loss.

The iterative process used in these IGA discussion is detailed elsewhere [27], and summarized here. While consensus was not reached on the exact cut points for each level for an AA scalp hair loss IGA, the Small Panel incorporated the 10 clinicians' reported perceptions to finalize the draft AA-IGA™, subsequently reviewed during the patient interviews. Using a top level of 0 (None) representing the absence of scalp hair loss (SALT score 0), the next level (1 = Limited) included SALT score 1–20, with the SALT score 20 upper bound representing the clinician's most commonly reported treatment success level (Fig. 1). The fourth level (3 = Severe) of the proposed IGA initiated at SALT score 50, and aligned with the lower limit for extensive scalp hair loss [7, 28]. Consequently, the third level (2 = Moderate; SALT score 21–49) was sandwiched between Limited and Severe. Achieving clinician consensus on the draft IGA's fifth level at the highest end of the extensive scalp hair loss spectrum was a challenge; nonetheless, with careful review of all de-identified Clinician Interviews responses by the Small Panel, a relevant description of the 5th level reflecting the clinicians' learnings from patients in this scalp hair loss category (4 = Very Severe; SALT score 95–100) was created to capture patients with nearly complete or complete scalp hair loss. In due course, the draft IGA was reviewed during the Patient Interviews.

## Patient Interviews

Thirty patients with AA and a history of ≥ 50% scalp hair loss were interviewed in October 2017. The patients' demographic and clinical characteristics at the time of the interviews are detailed elsewhere [29] and summarized here. Five of the patient interviewees were adolescents (ages 15–17 years old; 3 females/2 males) and 25 of the patient interviewees were adults (ages 18–72 years old; 14 females/11 males). Nine patients were non-Caucasian (Asian, Black, Other). Sixty percent of the adolescents and 84% of the adult patient interviewees had experienced some eyebrow and/or eyelash hair loss, meeting the recruitment goal (80% overall) to oversample these patients with AA. On average, these patients had been diagnosed with AA for 11.4 years (range 1–46 years). The most recent clinician-assessed SALT scores for these patients ranged of 0–100 (mean = 57.9), reflecting the inclusion of patients who had experienced improvement with treatment (60% were currently or previously treated with JAK inhibitors), which was felt to be critically important for understanding clinically meaningful change/clinical benefit to patients.

### Concept elicitation

The Patient Interviews commenced with discussions of the signs and symptoms of AA, previous treatments and the impacts that AA had on each patient's everyday life and well-being. These discussions were powerful and informative, and the results directly informed a new conceptual model for AA detailing the signs and symptoms, physical, emotional, and functional impacts of AA, including stigmatization, relationship, and social impacts that further elucidated the outcome of priority's impact on patients' lives [6].

**Ranking exercise** During concept elicitation discussions, the interviewer noted the signs and symptoms mentioned by the patients, and saturation of physical signs and symptoms was achieved [29]. All 30 patients named scalp hair loss as a key sign/symptom. After elicitation of the signs and symptoms experienced, each patient was asked to rank (first/most, second, third) their most bothersome signs and symptoms of AA. Scalp hair loss was named as the most bothersome sign/symptom by 77% of the sample (100% of adolescents/72% of adults). Four adults (16%) named eyebrow hair loss as the most bothersome sign/symptom; eyelash, nose, and body hair loss each received the most bothersome ranking from one adult patient [29]. The results from this patient ranking exercise confirmed scalp hair loss as the key concept, despite oversampling patients with eyebrow/eyelash hair loss.

**Meaningful treatment success** All 30 patients were asked to discuss their ideal treatment experience, including both the amount, quality and the time to achieve the hair growth that they would deem clinically meaningful. When patients were asked to propose the percentage (amount) of scalp hair coverage—short of 100%—that they would need for a treatment to be considered successful, 4 patients were initially unable to answer this question, as they experienced some difficulty in discussing scalp hair coverage in terms of percentages. Of the 26 patients (4 adolescents and 22 adults) who were comfortable answering the question, the majority ($n = 20$)

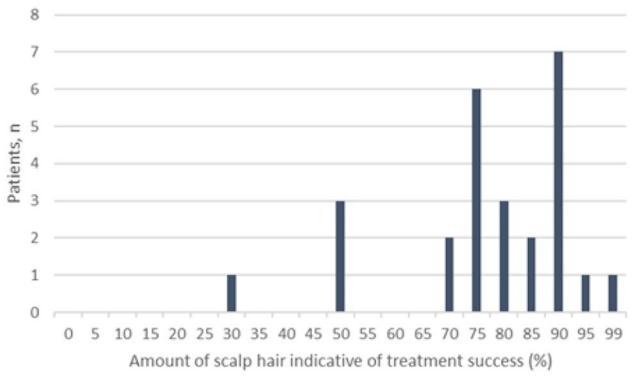


**Fig. 2** Patient treatment success thresholds

provided answers within the range of 70–90% scalp hair (median: 80% of scalp hair) (Fig. 2), which was generally similar to the Clinician Interviews results (Fig. 1; median: 80%). Moreover, these results were similar for patients with and without JAK inhibitor treatment experience (median: 75% and 85%, respectively).

To understand the 'why' behind the treatment success metric, patients were asked how their desired treatment success threshold would impact them. Patients explained how achieving the reported amount of scalp hair would improve their emotional/psychological wellbeing by increasing their confidence levels, reducing stress, and feeling more comfortable around other people. Some improvements to daily life were also predicted as a result of feeling more comfortable around others, by being able to work more sociable hours and/or live a more active lifestyle by attending the gym/swimming pool (Table 1).

Patients noted that a treatment would be successful even if the scalp hair grown was not the exact same color, quality, or thickness as their hair before AA. In fact, most patients thought their hair might grow back differently. Scalp location of remaining missing hair was not a predominant factor [27].

### Cognitive debriefing

During the cognitive debriefing, patient input was solicited on the relevance, appropriateness, and importance of the draft IGA developed during the Clinician Interviews. All 30 patients confirmed agreement with the proposed IGA measure, and no further changes to the IGA wording or response levels were suggested. Due to interview time limitations, only nine patients were asked about their perception of meaningful change as measured by the draft IGA. All nine respondents noted that achieving the Limited (SALT score 1–20) level after 9 months would indicate treatment success, with affirming quotes such as "That would be great. That would be fantastic." and "I think that would be a win if you got to Limited."

These results confirmed the content validity of the AA-IGA™ as a ClinRO conceptualizing of the most important clinical need that reflects and detects clinically meaningful improvement for patients with extensive AA. The final AA-IGA™ is published elsewhere [27].

## Discussion

This systematic qualitative investigation of expert clinicians and patients with AA confirmed that the amount of scalp hair is the outcome of priority, and that achievement of 80% or more scalp hair (SALT score ≤ 20) is an appropriate treatment success threshold, reflecting clinically meaningful improvement for patients with ≥ 50% scalp hair loss. This qualitative investigation of a quantifiable treatment success threshold was possible through the input of

**Table 1** Example patient interview quotes describing the meaningful impact of achieving the desired scalp hair amount

| |
|---|
| I would say 80–90%, because then you could still wear a hat and go swimming and do things like that and cover it up a little bit. (27-F-A-100-N) |
| It would probably be more of a confidence booster. (28-M-A-100-N) |
| Well, it's kind of hard to say. I mean, for the most part, I guess I would change my hours of operation. Instead of working at nighttime, I'd work in the daytime. That's what would change my life at least…being around people, yeah. (01-M-A-100-N) |
| I think just like not being stressed about it because like right now I feel like I can cover it. I don't have to wear a wig or anything like that but and it still is stressful to like worry about like oh it's like exposed or something like that. […] I'm guessing if it was 80 percent I would be able to do like different hairstyles […] Also between the two that would be nice and then just generally like not worrying about it as much. I feel like 80 percent I wouldn't like worry about oh exposed or like my hair what is that going to be like? So I think just in terms of like stress and like variety that would be nice. (20-F-P-60-JAK) |
| Because it's difficult [currently] to, like I said, go swimming, and you can't—the hardest thing for me is like I mean I used to live such an active lifestyle. So I don't like to go to the gym in a scarf, because it's just so hot and sweaty. And I don't like to go to the pool with a bald head, because everyone like stares at you and then the sun. So I would say that however long it took I would be happy with that to get my regular life back. (27-F-A-100-N) |

Patient IDs: order of interview—sex—adult/pediatric—SALT score—previous JAKi treatment. For example, patient 28-M-A-100-N was interviewed 28th, is a male adult with 100% hair loss and not previously treated with JAKi

clinicians-then-patients who generously shared their perspectives during one-on-one discussions.

"Hearing" the patient voice required first listening closely to expert dermatologists to understand clinically meaningful measurement and categorization of COA scores. This resulted in a categorization of SALT scores into the draft AA-IGA™ that represented distinct gradations of AA scalp hair loss severity; all patients independently confirmed agreement with the proposed IGA measure.

Patient input is fundamental in determining within-patient meaningful change/treatment success thresholds [11, 21]. However, clinicians are also in a unique position to provide treatment success input, given their interactions/relationships with patients and knowledge of clinical practice. For these reasons, we interviewed both groups of informants to capture both perspectives to inform the treatment success threshold, and the results were similar. Consequently, the AA-IGA™ and the SALT score ≤ 20 threshold serves as a patient-informed ClinRO reflecting a clinical benefit for patients with ≥ 50% scalp hair loss at the individual patient level [27]. This treatment success threshold exceeds the $SALT_{50}$ improvement threshold [7] and provides a key aid to understand the clinical benefit of new treatments for AA for all stakeholders [30].

Key limitations of this study are as follows: (1) patient perceptions of successful hair growth may be influenced by cultural and societal factors, and the results of this study of US informants may not be applicable in other countries/cultures; and (2) only nine patients were debriefed about their perceptions of meaningful change using the AA-IGA™ 0 or 1 category due to time limitations. To address these concerns, similar qualitative interviews were conducted in Japan in 2018–19 that confirmed scalp hair loss was the most important sign/symptom of AA and the greatest treatment priority [31]. Moreover, during cognitive debriefings, all expert dermatologists ($n = 6/6$; 100%) and most patients with AA ($n = 11/15$;73%) reported that achieving ≤ 20% scalp hair loss using the AA-IGA™ categories 0 or 1 would be treatment success for patients with ≥ 50% scalp hair loss. These results increase confidence in this COA's cultural acceptability and the threshold for achieving a clinically meaningful benefit for patients.

Currently, in 2021, SALT score ≤ 20 is the primary endpoint responder definition in several ongoing Phase 2 or Phase 3 clinical trial programs for the treatment of AA (e.g., NCT03570749, NCT03732807, and NCT04518995). As this threshold for interpreting within-patient meaningful change was derived qualitatively, it is important that the validity of this clinically meaningful threshold also be empirically investigated in the emerging study data.

As experts in what it is like to live with their condition, patients are uniquely positioned to inform the understanding of the therapeutic context for drug development and evaluation [32], and as shown here, informed a ClinRO measure to identify responders according to patients' needs and expectations [33]. Indeed, with this qualitative exploration of a quantitative responder threshold, we can now understand not only the within-patient change in SALT score that is meaningful but also *why* it is meaningful.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s11136-022-03170-7.

**Acknowledgements** The authors appreciate the wisdom and courage of each interviewee; the careful recruitment and patient access by clinic site staff and coordinators at the Department of Dermatology, University of California, Irvine, (UCI) and Yale University School of Medicine; and the editorial assistance and review by Emily Heredia and Diane Stothard at Eli Lilly and Company in the preparation of this manuscript.

**Funding** Eli Lilly and Company.

## Declarations

**Conflict of interest** Kathleen W. Wyrwich is a former employee and stockholder at Eli Lilly and Company and Pfizer Inc. She is currently employed at Bristol Myers Squibb Company. Helen Kitchen, Sarah Knight, Natalie V. J. Aldhouse and Jake Macey are employees and stockholders of Clarivate, a health economic and outcomes research consultancy that consults with various pharmaceutical companies. Natasha Mesinkovska has received honoraria/fees for advisory work with Eli Lilly and Company, Concert Pharmaceuticals Inc, Arena Pharmaceuticals, Nutrafol. Justin M. Ko has served on advisory boards and is a consultant and clinical investigator for Eli Lilly and Company. He has served as a clinical investigator and/or consultant for AbbVie, Sanofi, Regeneron, Dermira, BMS and Arena Pharmaceuticals. He has received consulting fees from Eli Lilly and Company and Arena Pharmaceuticals. Brett King has served on advisory boards and/or is a consultant and/or is a clinical trial investigator for Aclaris Therapeutics Inc, Almirall, Arena Pharmaceuticals, Bristol-Meyers Squibb, Concert Pharmaceuticals Inc, Dermavant Sciences Inc, Eli Lilly and Company, Incyte Corp, Pfizer Inc, TWi Biotechnology Inc, and Viela Bio. He is on speaker bureaus for Eli Lilly, Pfizer Inc, Regeneron and Sanofi Genzyme.

**Ethical approval** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee (Western Institutional Review Board [ref #20171820]) and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

**Consent to participate** Informed consent was obtained from all individual participants included in the study.

## References

1. Liu, L. Y., King, B. A., & Craiglow, B. G. (2016). Health-related quality of life (HRQoL) among patients with alopecia areata (AA): A systematic review. *Journal of the American Academy of Dermatology, 75*(4), 806–12.e3.
2. Rencz, F., Gulácsi, L., Péntek, M., Wikonkál, N., Baji, P., & Brodszky, V. (2016). Alopecia areata and health-related quality

of life: A systematic review and meta-analysis. *British Journal of Dermatology, 175*(3), 561–571.
3. Davis, D. S., & Callender, V. D. (2018). Review of quality of life studies in women with alopecia. *International Journal of Women's Dermatology, 4*(1), 18–22.
4. Wolf, J. J., & Hudson, B. P. (2019). Alopecia areata: Factors that impact children and adolescents. *Journal of Adolescent Research., 34*(3), 282–301.
5. Davey, L., Clarke, V., & Jenkinson, E. (2019). Living with alopecia areata: An online qualitative survey study. *British Journal of Dermatology., 180*(6), 1377–1389.
6. Aldhouse, N. V. J., Kitchen, H., Knight, S., Macey, J., Nunes, F. P., Dutronc, Y., Mesinkovska, N., Ko, J. M., King, B. A., & Wyrwich, K. W. (2020). "'You lose your hair, what's the big deal?' I was so embarrassed, I was so self-conscious, I was so depressed:" A qualitative interview study to understand the psychosocial burden of alopecia areata. *Journal of Patient-Reported Outcomes, 4*(1), 76.
7. Olsen, E. A., Hordinsky, M. K., Price, V. H., Roberts, J. L., Shapiro, J., Canfield, D., Duvic, M., King, L. E., Jr., McMichael, A. J., Randall, V. A., Turner, M. L., Sperling, L., Whiting, D. A., Norris, D., National Alopecia Areata Foundation. (2004). Alopecia areata investigational assessment guidelines—Part II. National Alopecia Areata Foundation. *Journal of the American Academy of Dermatology, 51*(3), 440–447.
8. Wambier, C. G., & King, B. A. (2019). Rule of thumb: A simple tool to estimate 1% scalp surface area. *Journal of the American Academy of Dermatology, 81*(2), 630–631.
9. Mackay-Wiggan, J., Jabbari, A., Nguyen, N., Cerise, J. E., Clark, C., Ulerio, G., Furniss, M., Vaughan, R., Christiano, A. M., & Clynes, R. (2016). Oral ruxolitinib induces hair regrowth in patients with moderate-to-severe alopecia areata. *JCI Insight, 1*(15), e89790.
10. Kennedy Crispin, M., Ko, J. M., Craiglow, B. G., Li, S., Shankar, G., Urban, J. R., Chen, J. C., Cerise, J. E., Jabbari, A., Winge, M. C., Marinkovich, M. P., Christiano, A. M., Oro, A. E., & King, B. A. (2016). Safety and efficacy of the JAK inhibitor tofacitinib citrate in patients with alopecia areata. *JCI Insight, 1*(15), e89776.
11. FDA. (2018). Discussion document for the patient-focused drug development public workshop on guidance 3: Select, develop or modify fit-for-purpose clinical outcome assessments. Prepared for a patient-focused drug development guidance public workshop held on October 15–16, 2018. https://www.fda.gov/media/116277/download
12. McLeod, L. D., Coon, C. D., Martin, S. A., Fehnel, S. E., & Hays, R. D. (2011). Interpreting patient-reported outcome results: US FDA guidance and emerging methods. *Expert Review of Pharmacoeconomics & Outcomes Research, 11*(2), 163–169.
13. Wyrwich, K. W., Norquist, J. M., Lenderking, W. R., & Acaster, S. (2013). Methods for interpreting change over time in patient-reported outcome measures. *Quality of Life Research, 22*(3), 475–483.
14. Coon, C. D., & Cappelleri, J. C. (2016). Interpreting change in scores on patient-reported outcome instruments. *Ther Innov Regul Sci., 50*(1), 22–29.
15. Guidance for industry. (2006). patient-reported outcome measures: Use in medical product development to support labeling claims: Draft guidance. *Health and Quality of Life Outcomes, 11*(4), 79.
16. Gelhorn, H. L., Kulke, M. H., O'Dorisio, T., Yang, Q. M., Jackson, J., Jackson, S., Boehm, K. A., Law, L., Kostelec, J., Auguste, P., & Lapuerta, P. (2016). Patient-reported symptom experiences in patients with carcinoid syndrome after participation in a study of telotristat etiprate: A qualitative interview approach. *Clinical Therapeutics, 38*(4), 759–768.
17. Peay, H., Kennedy, A., Fischer, R., Bronson, A., & Furlong, P. (2016). Promoting meaningful clinical trial outcome measures for Duchenne muscular dystrophy. *Neuromuscular Disorders., 26*, S187.
18. McGraw, S., Qian, Y., Henne, J., Jarecki, J., Hobby, K., & Yeh, W. S. (2017). A qualitative study of perceptions of meaningful change in spinal muscular atrophy. *BMC Neurology, 17*(1), 68.
19. Arbuckle, R., Staunton, H., Sully, K., Tomkins, S., Khindri, S., Svedsater, H., & Nelsen, L. (2019). Use of both qualitative and quantitative methods to estimate meaningful change thresholds for key endpoints in pediatric asthma trials. *Value in Health, 22*(3), 340–347.
20. Kitchen, H., Seitz, C., Trigg, A., Aldhouse, N., Willgoss, T., Schmitz, H., Gater, A., Gerlinger, C., & Haberland, C. (2021). Patients' and clinicians' perspectives on item importance, scoring, and clinically meaningful differences for the Endometriosis Symptom Diary (ESD) and Endometriosis Impact Scale (EIS). *Health and Quality of Life Outcomes, 19*(1), 7.
21. Staunton, H., Willgoss, T., Nelsen, L., Burbridge, C., Sully, K., Rofail, D., & Arbuckle, R. (2019). An overview of using qualitative techniques to explore and define estimates of clinically important change on clinical outcome assessments. *Journal of Patient-Reported Outcomes, 3*(1), 16.
22. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology., 3*(2), 77–101.
23. Guest, G., MacQueen, K. M., & Namey, E. E. (2011). *Applied thematic analysis*. Sage.
24. Wyrwich, K. W., Kitchen, H., Knight, S., Aldhouse, N. V. J., Macey, J., Nunes, F. P., Dutronc, Y., Mesinkovska, N., Ko, J. M., & King, B. A. (2020). Development of clinician-reported outcome (ClinRO) and patient-reported outcome (PRO) Measures for eyebrow, eyelash and nail assessment in alopecia areata. *American Journal of Clinical Dermatology, 21*(5), 725–732.
25. Friese S. ATLAS.ti 7 User Manual Copyright ©2014 by ATLAS.ti Scientific Software Development GmbH, Berlin. All rights reserved. Manual Version: 190.20140902. Updated for program version: 7.5 2014 Contract No.: Document Number|.
26. Langley, R. G., Feldman, S. R., Nyirady, J., van de Kerkhof, P., & Papavassilis, C. (2015). The 5-point Investigator's Global Assessment (IGA) Scale: A modified tool for evaluating plaque psoriasis severity in clinical trials. *The Journal of Dermatological Treatment, 26*(1), 23–31.
27. Wyrwich, K. W., Kitchen, H., Knight, S., Aldhouse, N. V. J., Macey, J., Nunes, F. P., Dutronc, Y., Mesinkovska, N., Ko, J. M., & King, B. A. (2020). The Alopecia Areata Investigator Global Assessment scale: A measure for evaluating clinically meaningful success in clinical trials. *British Journal of Dermatology, 183*(4), 702–709.
28. Olsen, E. A. (2011). Investigative guidelines for alopecia areata. *Dermatologic Therapy, 24*(3), 311–319.
29. Wyrwich, K. W., Kitchen, H., Knight, S., Aldhouse, N. V. J., Macey, J., Nunes, F., Dutronc, Y., Mesinkovska, N. A., Ko, J. M., & King, B. A. (2020). The role of patients in alopecia areata endpoint development: understanding physical signs and symptoms. *The Journal of Investigative Dermatology Symposium Proceedings, 20*(1), S71–S77.
30. Wyrwich, K. W., Kitchen, H., Knight, S., Aldhouse, N. V. J., Macey, J., Nunes, F. P., Dutronc, Y., Mesinkovska, N., Ko, J. M., & King, B. A. (2020). Development of the Scalp Hair Assessment PRO™ measure for alopecia areata. *British Journal of Dermatology, 183*(6), 1065–1072.
31. Macey, J., Kitchen, H., Aldhouse, N. V. J., Burge, R. T., Edson-Heredia, E., McCollam, J. S., Isaka, Y., & Torisu-Itakura, H. (2021). Dermatologist and patient perceptions of treatment

success in alopecia areata and evaluation of clinical outcome assessments in Japan. *Dermatol Ther (Heidelb), 11*(2), 433–447.
32. Chalasani, M., Vaidya, P., & Mullin, T. (2018). Enhancing the incorporation of the patient's voice in drug development and evaluation. *Research Involvement and Engagement, 4*, 10.
33. FDA. (2018). *Patient-focused drug development: Collecting comprehensive and representative input*. Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders. https://www.fda.gov/media/139088/download