



Validation of the Alopecia Areata Patient Priority Outcomes (AAPPO) Questionnaire in Adults and Adolescents with Alopecia Areata

Kathleen W. Wyrwich · Randall Winnette · Randall Bender ·
Kavita Gandhi · Nicole Williams · Nimanee Harris ·
Lauren Nelson

Received: September 28, 2021 / Accepted: November 13, 2021
© The Author(s) 2021, corrected publication 2022

ABSTRACT

Introduction: Individuals with alopecia areata (AA) may experience significant impacts on their health-related quality of life. The novel Alopecia Areata Patient Priority Outcomes (AAPPO) questionnaire has been developed to assess hair loss signs, emotional symptoms, and activity limitations associated with AA. The objective of this study was to evaluate psychometric properties and establish scoring of the AAPPO in adults and adolescents with AA.

Methods: Scoring and measurement properties of the AAPPO were examined using baseline and 2-week follow-up data from a prospective, noninterventive, web-based study of 121 patients with AA (85 adults aged ≥ 18 years, 36

adolescents aged 12–17 years) with Severity of Alopecia Tool (SALT) $\geq 25\%$ scalp hair loss.

Results: Exploratory and confirmatory factor analysis supported four single Hair Loss (HL) items, an Emotional Symptoms domain (ES; 4 items), and an Activity Limitations domain (AL; 3 items). Among all patients, the multi-item ES and AL domains had strong internal consistency ($\alpha \geq 0.87$); all HL items and domain scores had strong test-retest reliability (weighted kappa or intraclass correlation coefficients ≥ 0.78). All HL item scores demonstrated strong construct validity ($r \geq 0.52$) compared with the patient-reported Alopecia Areata Symptom and Impact Scale (AASIS) hair loss subscale score; ES and AL domain scores exhibited strong construct validity ($r \geq 0.66$) compared with the SF-36 Mental Component Summary (MCS) score. Using SALT scores, HL mean item scores were better (lower) in the 25–49% SALT subgroup versus those with highest SALT scores (76–100%); however, ES mean domain scores were better in the SALT 76–100% subgroup in the same comparison ($p < 0.0001$). Using AASIS and MCS score-created subgroups, ES and AL mean domain scores demonstrated hypothesized differences across subgroups (all p values < 0.0001).

Conclusion: The AAPPO questionnaire is a reliable, valid disease-specific measure of hair loss severity and impact in individuals with AA.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s13555-021-00648-z>.

K. W. Wyrwich · R. Winnette
Patient-Centered Outcomes Assessment, Pfizer, New York, NY, USA

R. Bender · N. Williams · N. Harris · L. Nelson (✉)
Patient-Centered Outcomes Assessment, RTI Health Solutions (RTI-HS), 3040 East Cornwallis Road, Research Triangle Park, NC 27709, USA
e-mail: lnelson@rti.org

K. Gandhi
Patient and Health Impact, Pfizer, Collegeville, PA, USA

Keywords: Patient-reported outcome; Psychometric; Hair loss; Symptoms; Impacts; Activity limitations; Emotional symptoms

Key Summary Points

Why carry out this study?

To characterize patients' experiences with alopecia areata (AA), including the psychosocial and functional impacts of the disease, it is important to capture their perspectives using a rigorously developed and validated AA-specific patient-reported outcome measure

What was learned from the study?

This psychometric evaluation demonstrated the disease-specific Alopecia Areata Patient Priority Outcomes (AAPPO) to be reliable and valid in measuring symptom severity and impacts in adults and adolescents with AA

Findings from this study support use of the AAPPO in clinical trials to show treatment benefit from a patient perspective

To fully characterize the patient experience of AA, including the psychosocial and functional impacts of this disease, it is important to capture patients' perspectives directly [6]. Existing AA-specific patient-reported outcome (PRO) measures are missing concepts that are a high priority to individuals with AA or employ response options and recall periods that may not sufficiently capture the impacts of AA. Thus, a novel AA-specific PRO measure, the Alopecia Areata Patient Priority Outcomes (AAPPO) tool, was developed to assess hair loss signs as well as the emotional symptoms and activity limitations from the patient's perspective [7]. Development of the AAPPO met the requirements described in the Food and Drug Administration (FDA) patient-focused guidance and was adherent to the principles of the FDA Patient-Focused Drug Development initiative [6, 8].

The objective of this study was to conduct a noninterventional quantitative evaluation consistent with the requirements of the FDA's patient-focused guidance to determine the optimal structure and scoring algorithm and to assess the psychometric properties of the AAPPO, including reliability and construct validity.

METHODS

Study Design

This study was a prospective, noninterventional, web-based study with two assessment time points, at baseline and follow-up 2 weeks later. A target sample size of 120 patients with a dermatologist-confirmed diagnosis of AA was recruited in the US. Of the target sample, enrollment was planned to include approximately 90 adult patients and 30 adolescent patients. Patients were recruited through dermatology practices that partnered with the Global Perspectives research database organization. Dermatology practices were responsible for identifying potentially eligible patients who had previously agreed to be contacted for studies from their clinical records.

The study was evaluated and deemed exempt from full review by the RTI International Review

INTRODUCTION

Alopecia areata (AA) is an autoimmune condition that targets the hair follicles, with an estimated self-reported point prevalence of approximately 1% [1]. Studies have shown that people living with AA are at a higher risk than the general population of developing depression, anxiety, and social phobia; living with AA has also been associated with much higher levels of body dissatisfaction and concern with general appearance because of the associated perception of hair loss [2–4]. The patient burden of AA, coupled with a lack of highly effective treatment options, represents a significant unmet medical need [5].

Board (IRB; IRB ID MOD00000707 for 20712). All participants provided informed consent.

Study Population

Eligible patients were adults (aged ≥ 18 years) or adolescents (aged 12–17 years) with a dermatologist-confirmed diagnosis of AA and who had experienced at least 6 weeks of hair loss. In addition, recruitment targets were applied to achieve a mix of participants with the following conditions: $\geq 25\%$ scalp hair loss as measured by the Severity of Alopecia Tool (SALT) [9] within the past 30 days; alopecia totalis (AT), defined as complete (100%) scalp hair loss; and alopecia universalis (AU), defined as complete (100%) scalp, facial, and body hair loss [10–12]. Patients were ineligible if they were participating in a clinical trial, undergoing treatment with a Janus kinase (JAK) inhibitor in the past 90 days, or had other forms of alopecia.

Clinical Outcomes Assessment Measures

SALT

The SALT was developed by the National Alopecia Areata Foundation [9, 13] to quantitatively assess AA severity based on terminal scalp hair loss. The dermatologist provided ratings of hair loss in four areas of the scalp: the back, top, and two sides; each area represents a percentage of the total scalp surface area: 24%, 40%, 18%, and 18%, respectively. The SALT total score is the summed percentage of hair loss on the scalp in each of the four areas weighted by their respective surface area.

Dermatologists provided each patient's SALT score assessed within 30 days before the baseline PRO assessments. Participants were classified into three groups (tertiles) based on the SALT total score: 25–49%, 50–75%, and 76–100%.

Patient-Reported Outcomes

The administered study PRO assessments at baseline and 2 weeks included the AAPPO, the Alopecia Areata Symptom Impact Scale (AASIS), the 36-item Acute Short Form (SF-36v Acute), the Patient Global Impression of Severity (PGIS)

item, and 2 Patient Global Impression of Change (PGIC) items.

AAPPO The 11-item AAPPO [7] contains four items, categorized as “Hair Loss” from (1) the scalp, (2) eyebrows, (3) eyelashes, and (4) body, and asks the patient to describe the current amount of hair loss using a five-point response scale that ranges from 0 (no hair loss) to 4 (complete hair loss: “I do not have any hair on my [insert hair loss area]”). Four items ask the patient to rate Emotional Symptoms of AA over the past week on a 5-point scale ranging from “Never” to “Always”. Three items ask the patient to rate Activity Limitations on a 5-point scale ranging from “Not at all” to “Completely (did not do any outdoor activities because of hair loss/did not do any physical activities because of hair loss/did not interact with others at all because of hair loss)”.

AASIS The 13-item AASIS asks patients with AA about the severity of their signs and symptoms and how AA interfered with their daily functioning in the past week [14]. For signs and symptom ratings, the measure uses a numeric rating scale of 0 (sign/symptom has not been present) to 10 (the sign/symptom was as “bad as you can imagine it could be”). For the interference with daily functioning ratings, the measure uses a numeric rating scale of 0 (did not interfere) to 10 (interfered completely). The AASIS was designed to enable patients, clinicians, and researchers to make informed decisions about evaluating newer therapies specifically designed for the treatment of AA [14]. Users can calculate a mean total score and four subscale scores (2-item hair loss, 5-item symptoms, 7-item symptoms, 6-item interference), each ranging from 0 to 10 points, with higher scores indicating worse AA-specific health status.

SF-36v2 Acute The Medical Outcomes Study (MOS) SF-36v2 Acute is a generic health status instrument that measures concepts of health-related quality of life over the past week for 8 general health domains: (1) physical functioning, (2) role limitations due to physical health, (3) bodily pain, (4) general health perceptions,

(5) vitality, (6) social functioning, (7) role limitations due to emotional problems, and (8) mental health [15, 16]. These domains can also be summarized as Physical and Mental Component Summary (PCS and MCS) scores. The recommended normed scores were used, ranging from 0 to 100, with higher scores indicating better health status [16].

Global Items: PGIS and PGIC Participants provided an overall assessment of the severity of their hair loss on the PGIS item “I consider my current hair loss to be: [none, mild, moderate, severe, extremely severe].” This single-item assessment was completed by all patients at baseline and at week 2 of the study. Scores ranged from 0 (none) to 4 (extremely severe). Patients also provided an overall retrospective assessment of their AA on the PGIC items. On the baseline questionnaire, they were asked to answer, “In the past 30 days, my alopecia areata has [greatly improved, moderately improved, slightly improved, not changed, slightly worsened, moderately worsened, greatly worsened].” Patients selected one response that best described their experience. On the follow-up questionnaire, they were asked to reply to a different PGIC item: “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 ranged from 1 (greatly improved) to 7 (greatly worsened).

Psychometric Analyses

Prior content validity work in the development of the AAPPO with adults and adolescents indicated that the AAPPO appropriately assesses disease status in both age groups [7]. Therefore, analyses planned to establish the AAPPO scoring algorithm (i.e., response distributions, inter-item correlations, and factor analyses) and assess reliability and construct validity were conducted with data pooled across both age groups using SAS v9.4 for Windows statistical software [17], with sensitivity analyses conducted in the separate adult and adolescent samples.

Distributional characteristics of the AAPPO responses were evaluated for possible response biases, including floor and ceiling effects (overall and by age group). A priori, the threshold for a potentially problematic floor or ceiling effect was set as $\geq 40\%$ of participants (given a uniform distribution) selecting the best (ceiling) or worst (floor) response category [18].

To inform the AAPPO structure and provide scoring recommendations, inter-item polychoric correlations were computed, and a series of factor-analysis models were estimated with mean- and variance-adjusted weighted least squares estimation in Mplus version 7.4 [19]. Exploratory factor analyses (EFAs) were performed on baseline item scores (overall and for adults), and an increasing number of factor solutions were extracted with oblique quartimin rotation for comparison. Based on the EFA results, confirmatory factor analyses (CFAs) were conducted on 2-week follow-up data (overall and for adults), and the results were interpreted using model fit indices, including the root mean square error of approximation [20, 21], comparative fit index [22], Tucker-Lewis Index [23], and standardized and weighted root mean square residual [20, 24, 25] as well as the magnitude and pattern of the factor loadings.

To evaluate the repeatability of scores (i.e., test-retest reliability), weighted kappa and intraclass correlation coefficients (ICCs) were computed using the complete data and for subsets of patients with: (1) PGIS scores that were equal at baseline and 2-week follow-up, (2) PGIC scores that were equal at baseline and 2-week follow-up, and (3) either a 1-point change or no change in the AASIS hair loss subscale at baseline and follow-up. For the AAPPO item-level scores, weighted kappa coefficients were computed using quadratic weights [26–28]. For the AAPPO multi-item domain scores, a two-way mixed-effects analysis of variance model with absolute agreement for single measures was used [29, 30]. According to Landis and Koch [31], kappa coefficients can be interpreted such that ≤ 0 is poor, 0–0.2 indicates slight agreement, 0.21–0.4 indicates fair agreement, 0.41–0.6 indicates moderate agreement, 0.61–0.80 indicates substantial agreement, and 0.81–1.00 indicates

almost perfect agreement. It is generally recommended that ICCs be at least 0.70 for multi-item scales [32, 33].

To evaluate internal consistency reliability, Cronbach's coefficient alpha was computed to evaluate the cohesiveness of the resulting multi-item domains [34]. Cronbach's alpha estimates > 0.70 indicate a set of strongly related items capable of supporting a unidimensional scoring structure [35].

Convergent and discriminant validity analyses aided in the evaluation of relationships among multiple indicators of similar and dissimilar constructs and the degree to which they followed hypothesized patterns. Moderate to strong correlations were anticipated between the AAPPO Hair Loss subscale and the PGIS. Moderate to strong correlations were also hypothesized between the AAPPO Emotional Symptoms and Activity Limitation domain scores and: (1) the AASIS symptoms and interference subscales scores and the AASIS total score; (2) the norm-based SF-36v2 Acute MCS score; and (3) the norm-based SF-36v2 Acute domain scores closely related to MCS (i.e., vitality, emotional functioning, role-emotional, mental health) [16, 36]. Smaller correlations were anticipated between the AAPPO domain scores and the norm-based SF-36v2 Acute PCS score and domain scores closely related to PCS score (i.e., physical functioning, role-physical, bodily pain, general health perceptions) [16, 36]. Correlation coefficients (absolute value) ≥ 0.50 were considered large, 0.30–0.49 were considered moderate, 0.10–0.29 were considered small, and < 0.10 were considered trivial [37].

Known-groups validity examines the ability of the AAPPO scores to discriminate among groups of AA patients who differ on external criteria or known groups. It was hypothesized that AAPPO domain scores would differentiate between patients: (1) with lower SALT scores (SALT 25%–49%) versus those with higher SALT scores (76–100%; greatest scalp hair loss); (2) who reported less hair loss versus those who reported higher levels of hair loss as assessed by the AASIS hair loss subscale items (as defined by AASIS interference subscale scores ≤ 1 and ≥ 5);

and (3) who had higher MCS scores versus those who had lower MCS scores (≤ 30 vs. ≥ 50).

RESULTS

Patient Characteristics

The study population included 121 patients with AA (85 adults aged ≥ 18 years and 36 adolescents aged 12–17 years) (Table 1). A mix of adult and adolescent patients with AA were enrolled: 57.9% (adults, 49; adolescents, 21) had $\geq 25\%$ scalp hair loss (based on dermatologist-confirmed diagnosis of AA), 33.1% (adults, 28; adolescents, 12) had AT, and 9.1% (adults, 8; adolescents, 3) had AU. Furthermore, 37 (30.6%) patients (adults, 31 [36.5%]; adolescents, 6 [16.7%]) were in the SALT 25–49% tertile, 16 (13.2%) patients (adults, 13 [15.3%]; adolescents, 3 [8.3%]) were in the SALT 50%–75% tertile, and 68 (56.2%) patients (adults, 41 [48.2%]; adolescents, 27 [75%]) were in the SALT 76%–100% tertile. The mean number of years since diagnosis of AA was 12 years (adults, 15 years; adolescents, 6 years); the duration since diagnosis ranged from < 1 year to 58 years for adults and < 1 year to 15 years for adolescents.

Of the 121 patients, 88 (72.7%) described themselves as White and 22 (18.2%) described themselves as Black. In the adult cohort, 14 (16.9%) had a high school education or equivalent (e.g., GED), 26 (31.3%) had an undergraduate degree, more than half (57.8%) were employed full time, and 21 (25.3%) were single or never married. In the adolescent cohort, 34 (94.4%) were students, with the majority (88.6%) not yet having completed high school.

Item-Level Distribution

As expected and given study inclusion criteria, item-level floor effects ($\geq 40\%$ at the worst health level) were observed on the AAPPO Item 1 assessing scalp hair loss (i.e., “a great deal” [46%] or “complete” [42%]) (Table S1, Supplementary Material). At baseline, 33.9% of patients reported complete hair loss of the

Table 1 Patient demographics and characteristics

Patient-reported survey item	Overall <i>n</i> = 121	Adult <i>n</i> = 85	Adolescent <i>n</i> = 36
Age, mean (SD) [range], years	32.3 (15.72) [12–69]	40.1 (12.08) [18–69]	14.0 (1.68) [12–17]
Female, <i>n</i> (%)	91 (75.2)	75 (88.2)	16 (44.4)
Hair loss classification, <i>n</i> (%)			
AA with 25–99% scalp hair loss	70 (57.9)	49 (57.6)	21 (58.3)
AT (complete loss of hair on the scalp)	40 (33.1)	28 (32.9)	12 (33.3)
AU (complete loss of hair on the scalp, face, and body)	11 (9.1)	8 (9.4)	3 (8.3)
SALT score			
Mean (SD)	72.1 (30.04)	67.9 (31.33)	82.0 (24.35)
Median	90.0	69.0	90.0
Min, max	25.0, 100.0	25.0, 100.0	25.0, 100.0
SALT score categories, <i>n</i> (%)			
25–49%	37 (30.6)	31 (36.5)	6 (16.7)
50–75%	16 (13.2)	13 (15.3)	3 (8.3)
76–100%	68 (56.2)	41 (48.2)	27 (75.0)
Years since alopecia areata diagnosis ^a			
Mean (SD)	12 (12.90)	15 (14.27)	6 (4.20)
Median	8	14	5
Min, max	< 1, 58	< 1, 58	< 1, 15
Region, <i>n</i> (%)			
Northeast	22 (18.3)	13 (15.3)	9 (25.7)
Midwest	27 (22.5)	17 (20.0)	10 (28.6)
South	49 (40.8)	38 (44.7)	11 (31.4)
West	22 (18.3)	17 (20.0)	5 (14.3)
Prefer not to answer	1 (0.8)	0 (0.0)	1 (2.8)
Employment status, <i>n</i> (%)			
Employed full time	48 (40.3)	48 (57.8)	0 (0.0)
Employed part time	12 (10.1)	12 (14.5)	0 (0.0)
Not currently employed	14 (11.8)	12 (14.5)	2 (5.6)
Student	36 (30.3)	2 (2.4)	34 (94.4)
Retired	6 (5.0)	6 (7.2)	0 (0.0)
Disabled/unable to work	3 (2.5)	3 (3.6)	0 (0.0)

Table 1 continued

Patient-reported survey item	Overall <i>n</i> = 121	Adult <i>n</i> = 85	Adolescent <i>n</i> = 36
Prefer not to answer	2 (1.7)	2 (2.4)	0 (0.0)
Education level, <i>n</i> (%)			
Less than high school	34 (28.8)	3 (3.6)	31 (88.6)
High school or equivalent (e.g., GED)	17 (14.4)	14 (16.9)	3 (8.6)
Some college but no degree	22 (18.6)	22 (26.5)	0 (0.0)
Undergraduate degree	27 (22.9)	26 (31.3)	1 (2.9)
Advanced degree or professional degree	18 (15.3)	18 (21.7)	0 (0.0)
Prefer not to answer	3 (2.5)	2 (2.4)	1 (2.8)
Race/ethnicity (select all that apply), <i>n</i> (%)			
White	88 (72.7)	58 (68.2)	30 (83.3)
Black	22 (18.2)	19 (22.4)	3 (8.3)
Asian or Pacific Islander	5 (4.1)	4 (4.7)	1 (2.8)
Native American or Alaskan native	2 (1.7)	1 (1.2)	1 (2.8)
Mixed racial background	2 (1.7)	1 (1.2)	1 (2.8)
Other race	0 (0.0)	0 (0.0)	0 (0.0)
Prefer not to answer	1 (0.8)	1 (1.2)	0 (0.0)

AA alopecia areata, *AT* alopecia totalis, *AU* alopecia universalis, *GED* General Education Development test, *Min*, *Max* minimum, maximum, *SALT* Severity of Alopecia Tool, *SD* standard deviation

^aYears since diagnosis calculated from year of diagnosis to year of baseline assessment. "Prefer not to answer" is not included in the percentages

eyebrows (Item 2), 29.8% reported complete hair loss of the eyelashes (Item 3), and 25.6% reported complete hair loss on the body (Item 4). The Emotional Symptoms items revealed an age group split, with adult endorsement levels of the most severe category ("always") considerably higher than those of the adolescent group: self-conscious (Item 5; 42 vs. 17%), embarrassed (Item 6; 33 vs. 17%), sad (Item 7; 33 vs. 14%), and frustrated (Item 8; 38 vs. 17%). Finally, ceiling effects ($\geq 40\%$ at the best health level) were observed in both the adult and adolescent responses related to limitations due to hair loss in outdoor activities (Item 9), exercise (Item 10), and interaction with others (Item 11). For example, at baseline 40.0% of adults and 63.9% of adolescents responded "not at all"

to limitations in outdoor activities because of hair loss (Item 9).

Inter-Item Correlations

In general, inter-item correlations were positive and strong in magnitude ($|r| \geq 0.50$) (Table 2). The inter-item correlations were positive and strong among the four Hair Loss items (Items 1–4) for adults, ranging from 0.73 to 0.92, and moderate to strong ($|r| \geq 0.30$) for adolescents, ranging from 0.31 to 0.94 (Table S2, Supplementary Material). The inter-item correlations were also positive and strong in magnitude between the respective Emotional Symptoms and Activity Limitations items (Items 5–8 and 9–11), ranging from 0.70 to 0.98.

Table 2 AAPPO inter-item correlations at baseline: overall ($n = 121$)

AAPPO item	Polychoric correlation, r									
	AAPPO Item									
	1	2	3	4	5	6	7	8	9	10
Hair loss										
Item 1. Hair loss on scalp	–									
Item 2. Hair loss on eyebrows	<i>0.67</i>	–								
Item 3. Hair loss on eyelashes	<i>0.61</i>	<i>0.93</i>	–							
Item 4. Hair loss on body	<i>0.71</i>	<i>0.87</i>	<i>0.85</i>	–						
Emotional symptoms										
Item 5. Self-conscious about hair loss	– 0.30	– 0.35	– 0.26	– 0.36	–					
Item 6. Embarrassed about hair loss	– 0.29	– 0.33	– 0.26	– 0.34	<i>0.95</i>	–				
Item 7. Sad about hair loss	– 0.30	– 0.29	– 0.19	– 0.26	<i>0.86</i>	<i>0.85</i>	–			
Item 8. Frustrated about hair loss	– 0.37	– 0.26	– 0.17	– 0.27	<i>0.87</i>	<i>0.86</i>	<i>0.87</i>	–		
Activity limitations										
Item 9. Limit participation in outdoor activities because of hair loss	– 0.17	– 0.10	– 0.06	– 0.09	<i>0.69</i>	<i>0.77</i>	<i>0.69</i>	<i>0.72</i>	–	
Item 10. Limit exercise because of hair loss	– 0.16	– 0.11	– 0.08	– 0.13	<i>0.69</i>	<i>0.71</i>	<i>0.58</i>	<i>0.68</i>	<i>0.84</i>	–
Item 11. Limit interaction with others because of hair loss	– 0.13	– 0.14	– 0.08	– 0.17	<i>0.72</i>	<i>0.78</i>	<i>0.66</i>	<i>0.67</i>	<i>0.78</i>	<i>0.73</i>

Italics indicates inter-item correlations within a given domain
 AAPPO Alopecia Areata Patient Priority Outcomes

The observed negative correlations between the Hair Loss items (Items 1–4) and the Emotional Symptoms and Activity Limitations items (Items 5–11), ranging from -0.06 to -0.37 , were not expected (Table S2, Supplementary Material). This finding may suggest the potential adaptation to the effects of hair loss in the more severe hair loss cases.

Exploratory and Confirmatory Factor Analyses

The EFA results using baseline data supported a three-factor solution for the AAPPO (Table S3, Supplementary Material), and CFAs fitted to the follow-up overall data confirmed this structure

(Table 3). Although the CFA results provided support for consideration of an overall hair loss subscale (Items 1–4), patients with AA present clinically with hair loss on the scalp and/or any hair-bearing area on the body [38]. Moreover, the qualitative evidence obtained during the AAPPO development process demonstrated that not all patients experienced hair loss in each measured location (i.e., scalp, eyebrows, eyelashes, and the body) or prioritized hair loss from each area equally [7]. Therefore, the decision was made to score the four individual Hair Loss items separately and not as a four-item summed domain score.

Table 3 CFA results at follow-up: overall sample

AAPPO item	Factor loading		
	3-Factor solution Hair loss	3-Factor solution Emotional symptoms	3-Factor solution Activity limitations
Item 1. Hair loss on scalp	<i>0.708 (0.049)</i>	–	–
Item 2. Hair loss on eyebrows	<i>0.968 (0.014)</i>	–	–
Item 3. Hair loss on eyelashes	<i>0.960 (0.017)</i>	–	–
Item 4. Hair loss on body	<i>0.880 (0.026)</i>	–	–
Item 5. Self-conscious about hair loss	–	<i>0.960 (0.012)</i>	–
Item 6. Embarrassed about hair loss	–	<i>0.949 (0.014)</i>	–
Item 7. Sad about hair loss	–	<i>0.960 (0.011)</i>	–
Item 8. Frustrated about hair loss	–	<i>0.953 (0.013)</i>	–
Item 9. Limit participation in outdoor activities because of hair loss	–	–	<i>0.969 (0.019)</i>
Item 10. Limit exercise because of hair loss	–	–	<i>0.937 (0.024)</i>
Item 11. Limit interaction with others because of hair loss	–	–	<i>0.911 (0.028)</i>
Quartimin factor correlations	F1–F2: – 0.302 F1–F3: – 0.053 F2–F3: 0.856		
Fit indices			
Root mean square error of approximation (90% CI)	0.063 (0.023–0.095)		
Comparative fit index	0.998		
Tucker-Lewis index	0.997		
Weighted root mean square residual	0.528		
Chi-square test	60.62 (<i>df</i> = 41) <i>p</i> = 0.0247		

Italics indicates factor loadings within a given domain

AAPPO Alopecia Areata Patient Priority Outcomes, CFA confirmatory factor analysis, CI confidence interval, *df* degree of freedom

Description of Recommended AAPPO Domain Scoring

Based on the content validity results identifying the 11 AAPPO items as distinct and important concepts, as well as the overall pattern of inter-item and construct validity correlations and the EFA and CFA results, all 11 items were retained,

with 6 independent AAPPO scores: (1) Hair Loss on the Scalp (Item 1); (2) Hair Loss on the Eyebrows (Item 2); (3) Hair Loss on the Eyelashes (Item 3); (4) Hair Loss on the Body (Item 4); (5) Emotional Symptoms domain computed as the mean of Items 5–8, with the requirement that at least 2 domain items have nonmissing

Table 4 Six AAPPO domain scores: test-retest reliability

AAPPO domain score	Test-retest reliability coefficients (95% CI), <i>n</i>			
	All patients	PGIS _{BL} = PGIS _{FU}	PGIC _{FU} = Not changed	Stable AASIS Hair loss ^a
Overall				
Hair loss on scalp	0.79 (0.70–0.89), 121	0.82 (0.72–0.93), 85	0.78 (0.68–0.88), 87	0.94 (0.90–0.99), 74
Hair loss on eyebrows	0.95 (0.92–0.99), 121	0.95 (0.90–0.99), 85	0.98 (0.96–0.99), 87	0.96 (0.91–1.00), 74
Hair loss on eyelashes	0.95 (0.93–0.98), 121	0.96 (0.94–0.98), 85	0.96 (0.93–0.98), 87	0.97 (0.96–0.99), 74
Hair loss on body	0.87 (0.82–0.93), 121	0.87 (0.80–0.94), 85	0.86 (0.79–0.93), 87	0.92 (0.89–0.96), 74
Emotional symptoms	0.90 (0.84–0.93), 121	0.91 (0.84–0.95), 85	0.91 (0.84–0.95), 87	0.91 (0.85–0.94), 74
Activity limitations	0.82 (0.75–0.87), 121	0.85 (0.78–0.90), 85	0.84 (0.77–0.90), 87	0.85 (0.78–0.90), 74

Weighted kappa coefficients were computed for AAPPO Hair Loss item scores; intraclass correlation coefficients were computed for AAPPO Emotional Symptoms and Activity Limitations subscale scores

AAPPO Alopecia Areata Patient Priority Outcomes, AASIS Alopecia Areata Symptom Impact Scale, BL baseline, CI confidence interval, FU follow-up, PGIC patient global impression of change, PGIS patient global impression of severity
^aStudy participants with a 1-point change or no change in the AASIS hair loss subscale and its component items at baseline (test) and 2-week follow-up (retest)

responses; (6) Activity Limitations domain computed as the mean of Item 9–11, with the requirement that at least 2 domain items have nonmissing responses. Each domain score ranges from 0 to 4, and a total score from the 11 items is not recommended.

Reliability

Using baseline (test) and follow-up (retest) data, ICC values estimating the test-retest reliability for the six AAPPO scores were acceptable (≥ 0.78) for the full sample as well as for the patient subgroups selected a priori to demonstrate stability over 2 weeks using the PGIS, PGIC, or the AASIS hair loss subscale assessments (Table 4). Similar ICC results were observed within each age group (Table S4, Supplementary Material). Internal consistency reliability was also strong for the two multi-item domain scores, Emotional Symptoms and Activity Limitations, with Cronbach's alpha ranging from 0.87–0.96 at baseline and at week 2 (Table S5, Supplementary Material). Although the alpha > 0.90 levels may indicate redundancy [27], patients provided differentiation between and the importance of each of the

Emotional Symptoms and Activity Limitations domain items during cognitive debriefing interviews; therefore, no items were removed [7].

Validity

The four AAPPO Hair Loss item scores demonstrated moderate to strong construct validity ($r \geq 0.34$) compared with the AASIS hair loss subscale score, with similar moderate to strong associations with the PGIS ($r \geq 0.41$) (Table 5). Moreover, the four Hair Loss item scores had notably weaker relationships with the PCS score ($|r| \leq 0.06$) and with each of the PCS-related domains of the SF-36v2 Acute ($|r| \leq 0.14$). As hypothesized, Emotional Symptoms and Activity Limitations domain scores were strongly correlated with the SF-36v2 Acute MCS score ($|r| \geq 0.58$), AASIS interference subscale and total scores ($r \geq 0.68$), and AASIS symptoms subscales scores ($r \geq 0.51$) and moderately to strongly correlated with MCS-related SF-36v2 Acute domain scores ($|r| \geq 0.44$). The Emotional Symptoms and Activity Limitations domain scores had much weaker correlation with the AASIS hair loss subscale score ($r \leq 0.19$) and the

Table 5 Six AAPPO domain scores: construct validity results at baseline

AAPPO items	Correlation coefficient					
	Hair loss items				Emotional symptoms	Activity limitations
	Scalp	Eyebrows	Eyelashes	Body		
Overall (<i>n</i> = 121)						
PGIS	0.59*	0.41*	0.44*	0.48*	− 0.06	0.01
AASIS scales						
Total	− 0.19*	− 0.02	− 0.02	− 0.07	0.68*	0.77*
Hair loss	0.34*	0.55*	0.56*	0.52*	<i>0.03</i>	<i>0.19*</i>
5 symptoms	− 0.35*	− 0.18	− 0.21*	− 0.25*	0.59*	0.65*
7 symptoms	− 0.14	0.10	0.08	0.02	0.51*	0.63*
Interference	− 0.21*	− 0.12	− 0.10	− 0.14	0.73*	0.78*
SF-36v2 Acute						
Physical functioning	<i>0.00</i>	− <i>0.02</i>	− <i>0.01</i>	<i>0.02</i>	− 0.18*	− 0.24*
Role limitations due to physical health	<i>0.05</i>	<i>0.06</i>	<i>0.07</i>	<i>0.05</i>	− 0.26*	− 0.33*
Bodily pain	<i>0.03</i>	− <i>0.01</i>	− <i>0.02</i>	<i>0.01</i>	− 0.32*	− 0.35*
General health perceptions	<i>0.12</i>	<i>0.09</i>	<i>0.13</i>	<i>0.14</i>	− 0.42*	− 0.36*
Vitality	0.32*	0.25*	0.23*	0.33*	− 0.54*	− 0.44*
Social functioning	0.25*	0.21*	0.25*	0.23*	− 0.50*	− 0.59*
Role limitations due to emotional problems	0.20*	0.13	0.20*	0.16	− 0.51*	− 0.60*
Mental health	0.29*	0.20*	0.22*	0.23*	− 0.59*	− 0.59*
PCS	− 0.06	− 0.05	− 0.06	− 0.03	− 0.10	− 0.10
MCS	0.30*	0.22*	0.26*	0.26*	− 0.58*	− 0.60*

Bold indicates correlations expected to demonstrate convergent validity, and italics indicate discriminant validity correlations

AAPPO Alopecia Areata Patient Priority Outcomes, MCS Mental Component Summary, PCS Physical Component Summary, PGIS Patient Global Impression of Severity

**p* < 0.05 test for H0: $\rho = 0$

PCS score ($|r| = 0.10$). The PCS-related domains of the SF-36v2 Acute demonstrated generally moderate relationships ($0.18 \leq |r| \leq 0.42$) with the Emotional Symptoms and Activities Limitations domain scores (Table S4, Supplementary Material). These trends were similar for both adults and adolescents (Table S6, Supplementary Material).

Results from the first set of known-groups validity analyses to confirm whether the hypothesized difference between groups known to differ on a key variable of interest (scalp hair loss) provided important insights. As predicted, the four AAPPO Hair Loss item mean scores were better (lower) for patients in the 25–49% SALT tertile compared with those in the highest

Table 6 Six AAPPO domain scores: known-groups validity at baseline by SALT subgroup

AAPPO Items	Six AAPPO Domain Scores, Mean (SD) [Median], <i>n</i>					
	Hair Loss Items				Emotional Symptoms	Activity Limitations
	Scalp	Eyebrows	Eyelashes	Body		
Overall (n = 121)						
SALT subgroups						
25%–49%	↓ 2.7 (0.47) [3.0], 37	↓ 0.9 (1.14) [1.0], 37	↓ 0.7 (0.97) [0.0], 37	↓ 0.9 (1.05) [1.0], 37	↑ 3.2 (0.86) [3.5], 37	↑ 1.2 (0.98) [1.0], 37
50%–75%	↓ 3.1 (0.44) [3.0], 16	↓ 1.9 (1.50) [2.0], 16	↓ 1.6 (1.59) [1.5], 16	↓ 1.4 (1.21) [1.5], 16	↑ 2.7 (1.24) [2.9], 16	↑ 1.4 (1.15) [1.0], 16
76%–100%	↓ 3.7 (0.58) [4.0], 68	↓ 3.0 (1.26) [4.0], 68	↓ 2.7 (1.49) [3.0], 68	↓ 3.0 (1.21) [3.0], 68	↑ 1.9 (1.37) [2.0], 68	↑ 0.8 (0.99) [0.3], 68
ANOVA results						
Overall, <i>F</i> ; <i>p</i> value	43.38; < 0.0001	34.85; < 0.0001	27.04; < 0.0001	38.89; < 0.0001	14.65; < 0.0001	3.25; 0.0422
Pairwise, <i>t</i> ; <i>p</i> value						
25%–49% vs. 50%–75%	-2.41; 0.0514	-2.54; 0.0371	-2.25; 0.0761	-1.41; 0.4112	1.42; 0.4020	-0.64; 0.8929
25%–49% vs. 76%–100%	-9.14; < 0.0001	-8.26; < 0.0001	-7.28; < 0.0001	-8.43; < 0.0001	5.31; < 0.0001	1.93; 0.1579
50%–75% vs. 76%–100%	-4.12; 0.0002	-3.34; 0.0034	-2.92; 0.0125	-4.68; < 0.0001	2.38; 0.0564	2.11; 0.1075
Lower SALT score = Better (lower) AAPPO hair loss score				Lower SALT score = Higher (worse) AAPPO Emotional Symptoms and Activity Limitations score		

SALT scores provided in the eCRF were within 30 days

AAPPO Alopecia Areata Patient Priority Outcomes, ANOVA analysis of variance, eCRF electronic case report form, SALT Severity of Alopecia Tool, SD standard deviation

SALT tertile (76–100%; $p < 0.0001$). However, AAPPO Emotional Symptoms and Activity Limitations domain mean scores tended to be worse (higher) for participants in the 25–49% SALT tertile compared with those who had higher SALT scores (Table 6; Table S7, Supplementary Material). Additional known group analyses for the Emotional Symptoms and Activity Limitations comparing the adult and adolescent groups with: (1) higher versus lower AASIS interference scores and (2) higher versus lower MCS scores (≤ 30 or ≥ 50) confirmed known group expectations for these two AAPPO domains in each age group (Table 7). Patients with lower (better) AASIS Interference scores had lower (better) AAPPO Emotional Symptoms and Activity Limitations mean domain scores compared with the subgroup with higher AASIS Interference scores ($p < 0.0001$). Similarly, the

subgroup with higher (better) MCS scores had lower (better) AAPPO Emotional Symptoms and Activity Limitations mean domain scores compared with the subgroup of patients with lower MCS scores ($p < 0.0001$), with similar relationships demonstrated for both the adults and the adolescents (Table 7).

DISCUSSION

The performance of the AAPPO was evaluated using standard psychometric methods on data collected in the context of a prospective, non-interventional, web-based study. A total of 121 adults ($n = 85$) and adolescents ($n = 36$) with a dermatologist-confirmed diagnosis of AA were recruited in the US. A mix of patients with AA were enrolled: 37 (30.6%) had 25–49% scalp

Table 7 Six AAPPO domain scores: known group validity at baseline by AASIS interference and MCS

Sample/known group	Mean (SD) [median], <i>n</i>	
	Emotional symptoms	Activity limitations
Overall		
AASIS Interference subscale		
≤ 1	1.0 (1.08) [0.8], 42	0.2 (0.38) [0.0], 42
≥ 5	3.5 (0.51) [3.8], 31	2.1 (0.94) [2.3], 31
ANOVA overall, <i>F</i> ; <i>p</i> value	145.02; < 0.0001	141.78; < 0.0001
MCS		
≤ 30	3.2 (0.96) [3.5], 34	1.8 (1.02) [1.7], 34
≥ 50	1.2 (1.38) [0.8], 34	0.3 (0.72) [0.0], 34
ANOVA overall, <i>F</i> ; <i>p</i> value	49.25; < 0.0001	46.64; < 0.0001
Adults		
AASIS Interference subscale		
≤ 1	1.4 (1.30) [0.8], 20	0.2 (0.48) [0.0], 20
≥ 5	3.5 (0.53) [3.8], 28	2.1 (0.93) [2.3], 28
ANOVA overall, <i>F</i> ; <i>p</i> value	64.73; < 0.0001	70.80; < 0.0001
MCS		
≤ 30	3.4 (0.63) [3.5], 24	2.1 (0.89) [2.2], 24
≥ 50	1.6 (1.52) [0.9], 22	0.5 (0.86) [0.0], 22
ANOVA overall, <i>F</i> ; <i>p</i> value	28.65; < 0.0001	39.21; < 0.0001
Adolescents		
AASIS Interference subscale		
≤ 1	0.7 (0.74) [0.6], 22	0.1 (0.26) [0.0], 22
≥ 5	3.6 (0.38) [3.5], 3	1.4 (0.84) [1.3], 3
ANOVA overall, <i>F</i> ; <i>p</i> value	42.50; < 0.0001	37.15; < 0.0001
MCS		
≤ 30	2.8 (1.45) [3.4], 10	1.1 (1.02) [1.0], 10
≥ 50	0.5 (0.66) [0.1], 12	0.1 (0.21) [0.0], 12
ANOVA overall, <i>F</i> ; <i>p</i> value	24.82; < 0.0001	11.49; 0.0029

AAPPO Alopecia Areata Patient Priority Outcomes, *AASIS* Alopecia Areata Symptom Impact Scale, *ANOVA* analysis of variance, *MCS* Mental Component Summary score, *SD* standard deviation

hair loss, 16 (13.2%) had 50–75% scalp hair loss, and 68 (56.2%) had 76–100% scalp hair loss based on their SALT total scores.

Reflecting the distribution of the SALT scalp hair loss scores provided by their clinicians, the majority of adults and adolescents considered

their scalp hair loss as severe or extremely severe at the baseline and week 2 assessments, resulting in anticipated floor effects for the AAPPO Hair Loss items. Descriptive statistics also revealed ceiling effects (no limitation reported) for some AAPPO Emotional Symptoms and Activity Limitations items, most notably for the adolescent group. Taking into account the item-level correlations and the factor analysis, as well as the qualitative research conducted in the development of the AAPPO, six AAPPO scores are recommended to reflect their unique content: Hair Loss on the Scalp, Hair Loss on the Eyebrows, Hair Loss on the Eyelashes, Hair Loss on the Body, Emotional Symptoms domain, and Activity Limitations domain. A mean scoring algorithm is proposed for each domain ranging from 0 to 4, with higher scores indicating greater impacts.

The test-retest reliability coefficients (≥ 0.78) were adequate for demonstrating the reproducibility of the six scores, and internal consistency results (Cronbach's alpha ≥ 0.87) were supportive of the two multi-item domains. Strong convergent and discriminant validity correlations and several known-group analyses provide additional empirical evidence that the AAPPO domains were measuring what they were intended to measure.

Although it was anticipated that greater hair loss severity, as indicated by the highest SALT tertile (76–100%), would yield the highest mean scores on the Emotional Symptoms and Activity Limitations domains, the pattern of these domain scores across the SALT tertiles was reversed. Specifically, Emotional Symptoms and Activity Limitations scores tended to be worse (higher) for those in the 25–49% SALT tertile than in the most severe SALT tertile (76–100%). These results demonstrated a greater emotional and activity impact of patients with 25–49% scalp hair loss compared with those with greater scalp hair loss and the pressing need for safe and efficacious treatments for patients at this moderate hair loss level [39] to alleviate their burden. Although a possible explanation for this finding is adaptation to life with AA by the patients in the highest SALT tertile, our preliminary analyses of the impact of years since diagnosis on the relationship between SALT

tertiles and the AAPPO Emotional Symptoms and Activity Limitations domain scores did not reveal statistically significant trends in mean differences ($p > 0.05$). One exception to this conclusion was the trend for higher (worse) AAPPO mean Activity Limitations scores for the binary subgroup of patients with ≤ 10 years since diagnosis compared with patients with > 10 years since diagnosis ($p = 0.0494$; analyses available on request).

Another plausible explanation for this unexpected SALT 25–49% tertile finding is the greater daily emotional and activity-limiting burden to cosmetically conceal and manage smaller patchy areas of hair loss compared with patients with far greater or complete scalp hair loss (AT/AU). The latter group (SALT 75–100%) may, in general, focus less on these concealment challenges and instead sport a bald or prosthesis-covered scalp, thus reducing: (1) the amount of time focused on daily concealing activities [40], (2) the emotional concerns of being “found out” if the concealment is imperfect or becomes disrupted, and (3) activity limitations necessary to avoid water, sweat, and/or wind that could disrupt cosmetic concealment of scalp hair loss. These possible explanations elucidate the need for future research to better understand this finding of greatest emotional and activity impact in the 25–49% SALT tertile.

An additional interesting finding in these analyses was the trivial-to-small relationships of the Emotional Symptoms and Activity Limitations domain scores with the PCS score ($r = -0.10$; Table 5); these trivial and small PCS relationships differing in magnitude from the generally moderate correlations observed for PCS-related domains of the SF-36v2 Acute demonstrated a relationship ($0.18 \leq |r| \leq 0.42$; Table 5). Because the PCS score calculation is computed using all eight SF-36v2 Acute domain scores with positive weighting for physical domains and negative factor weights for mental domains, its relationship to the Emotional Symptoms and Activity Limitations domain scores is more complex than a simple examination of the correlations of the SF-36v2 Acute domains considered related to the PCS. This known challenge for best understanding the PCS score has been reported by others [41, 42].

In addition to the AAPPO, other AA-specific patient-reported measures are available, including the AASIS [14], the Alopecia Areata Quality of Life Index [43], and the Alopecia Areata Patients' Quality of Life instrument [44], although these measures have not been extensively validated or used frequently in studies evaluating HRQoL in patients with AA [45]. The AAPPO has established content validity, reflects the symptoms and impacts that qualitative research has shown matter most to patients, and has been rigorously evaluated for reproducibility and cross-sectional measurement properties.

Limitations of this study include (1) the modest sample size ($n = 121$), (2) an adult sample that was primarily composed of females, (3) a greater proportion of patients with AU/AT than is reflected in a recent study of the AA population in the US [1], potentially limiting generalizability, and (4) a lack of longitudinal analyses to investigate the AAPPO domain scores' ability to detect change over time and to explore meaningful within-patient change thresholds. Nonetheless, the AAPPO is currently being administered in a longitudinal, interventional study to investigate meaningful within-patient change thresholds [46, 47], providing the opportunity to investigate and understand these important measurement properties in the AAPPO domain scores.

CONCLUSION

The AAPPO is a novel, AA-specific PRO measure with domains that capture the outcomes of importance to patients with AA. This psychometric evaluation demonstrated the reliability and validity of the AAPPO to measure symptom severity and impacts in adults and adolescents with AA, supporting its use in clinical trials to show treatment benefit from a patient perspective.

ACKNOWLEDGEMENTS

Funding. This study was developed under a research contract between RTI Health Solutions and Pfizer. This study was sponsored by Pfizer, and Pfizer is funding all publication fees.

Medical writing, editorial, and other assistance. Medical writing was provided by Kate Lothman of RTI Health Solutions and funded by Pfizer. Manuscript formatting support was provided by Linda Cirella at Engage Scientific Solutions and funded by Pfizer; no contribution was made to editorial content.

Authorship. All named authors meet the International Committee of Medical Journal Editors (ICMJE) criteria for authorship for this article, take responsibility for the integrity of the work as a whole, and have given their approval for this version to be published.

Author contributions. Concept and design: Kathleen W. Wyrwich, Randall Winnette, Randall Bender, Kavita Gandhi, Lauren Nelson; Acquisition, analysis, or interpretation of data: All authors; Drafting of the manuscript: All authors; Critical revision of the manuscript for important intellectual content: All authors; Statistical analysis: Randall Bender, Nicole Williams, Lauren Nelson; Obtained funding: Kathleen W. Wyrwich, Randall Winnette, Kavita Gandhi; Administrative, technical, or material support: Kathleen W. Wyrwich, Randall Winnette, Kavita Gandhi, Nimanee Harris; Supervision: All authors. Kathleen W. Wyrwich, Randall Winnette, Kavita Gandhi, Lauren Nelson.

Disclosures. Kathleen W. Wyrwich (currently unaffiliated) and Kavita Gandhi (currently of Janssen Pharmaceutical Companies of Johnson and Johnson) were salaried employees of Pfizer when this study was conducted and may hold stock and/or stock options in Pfizer. Randall Winnette is a salaried employee of Pfizer and holds stock and/or stock options in Pfizer. Randall Bender, Nicole Williams, Nimanee Harris, and Lauren Nelson are employees of

RTI Health Solutions and were paid consultants to Pfizer in connection with the development of this manuscript.

Compliance with ethics guidelines. The study was evaluated and deemed exempt from full review by the RTI International Review Board (IRB; IRB ID MOD00000707 for 20712). All participants provided informed consent.

Data availability. The datasets generated during and analyzed during the current study are not publicly available in order to protect participant confidentiality.

Open Access. This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

- Benigno M, Anastassopoulos KP, Mostaghimi A, Udall M, Daniel SR, Cappelleri JC, et al. A large cross-sectional survey study of the prevalence of alopecia areata in the United States. *Clin Cosmet Investig Dermatol*. 2020;13:259–66.
- Cash TF. The psychological effects of androgenetic alopecia in men. *J Am Acad Dermatol*. 1992;26(6): 926–31.
- 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.
- Ruiz-Doblado S, Carrizosa A, Garcia-Hernandez MJ. Alopecia areata: psychiatric comorbidity and adjustment to illness. *Int J Dermatol*. 2003;42(6): 434–7.
- Skogberg G, Jackson S, Astrand A. Mechanisms of tolerance and potential therapeutic interventions in Alopecia Areata. *Pharmacol Ther*. 2017;179: 102–10.
- Food and Drug Administration (FDA). Patient-focused drug development: methods to identify what is important to patients. Draft guidance for industry, Food and Drug Administration staff, and other stakeholders. 2019. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-focused-drug-development-methods-identify-what-important-patients-guidance-industry-food-and>. Accessed 2 Aug 2020.
- Winnette R, Martin S, Harris N, Deal L. Development of the alopecia areata patient priority outcomes instrument, a new patient reported outcome measure. *J Am Acad Dermatol*. 2019;81(4 (suppl 1)): AB46.
- Food and Drug Administration (FDA). Guidance for industry. Patient-reported outcome measures: use in medical product development to support labeling claims. 2009. <https://www.fda.gov/downloads/drugs/guidances/ucm193282.pdf>. Accessed 29 Mar 2020.
- Olsen EA, Hordinsky MK, Price VH, Roberts JL, Shapiro J, Canfield D, et al. Alopecia areata investigational assessment guidelines—part II. National Alopecia Areata Foundation. *J Am Acad Dermatol*. 2004;51(3):440–7.
- Buckley J, Rapini RP. Totalis alopecia. *Treasure Island: StatPearls*; 2020.
- Burroway B, Griggs J, Tosti A. Alopecia totalis and universalis long-term outcomes: a review. *J Eur Acad Dermatol Venereol*. 2020;34(4):709–15.
- National Alopecia Areata Foundation. What you need to know about the different types of alopecia areata. 2021. <https://www.naaf.org/alopecia-areata/types-of-alopecia-areata>. Accessed 3 Feb 2021.
- Olsen E, Hordinsky M, McDonald-Hull S, Price V, Roberts J, Shapiro J, et al. Alopecia areata investigational assessment guidelines. National Alopecia Areata Foundation. *J Am Acad Dermatol*. 1999;40(Pt 1):242–6.
- Mendoza TR, Osei J, Duvic M. The utility and validity of the alopecia areata symptom impact

- scale in measuring disease-related symptoms and their effect on functioning. *J Investig Dermatol Symp Proc.* 2018;19(1):S41–6.
15. Ware JEJ. How to score the revised MOS short-form health scales. Boston: Institute for the Improvement of Medical Care and Health; 1988.
 16. Maruish MEE. User's manual for the SF-36v2 Health Survey. 3rd ed. Lincoln: QualityMetric Incorporated; 2011.
 17. SAS Institute I. SAS proprietary software, version 9.4. Cary (NC); 2012.
 18. Dean K, Walker Z, Jenkinson C. Data quality, floor and ceiling effects, and test-retest reliability of the Mild Cognitive Impairment Questionnaire. *Patient Relat Outcome Meas.* 2018;9:43–7.
 19. Muthén LK, Muthén BO. Mplus user's guide. 7th ed. Los Angeles: Muthén & Muthén; 1998-2015.
 20. Browne MW, Cudeck R. Alternative ways of assessing model fit. In: Bollen KA, Long JS, editors. *Testing structural equation models.* Newbury Park: Sage; 1993.
 21. Steiger JH. Structural model evaluation and modification: an interval estimation approach. *Multivar Behav Res.* 1990;25(2):173–80.
 22. Bentler PM. EQS structural equations program manual. Los Angeles: BMDP Statistical Software; 1989.
 23. Tucker LR, Lewis C. A reliability coefficient for maximum likelihood factor analysis. *Psychometrika.* 1973;38(1):1–10.
 24. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model.* 1999;6(1):1–55.
 25. Schumacker RE, Lomax RG. A beginner's guide to structural equation modeling. Mahwah: Erlbaum; 1996.
 26. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas.* 1973;33: 613–9.
 27. Streiner DL, Norman GR, Cairney J. *Health Measurement Scales: a practical guide to their development and use.* 5th ed. Oxford: Oxford University Press; 2015.
 28. Warrens M. Some paradoxical results for the quadratically weighted kappa. *Psychometrika.* 2012;77(2):315–23.
 29. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods.* 1996;1:30–46.
 30. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull.* 1979;86: 420–8.
 31. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159–74.
 32. Nunnally JC, Bernstein IH. *Psychometric theory.* 3rd ed. New York: McGraw-Hill; 1994.
 33. Food and Drug Administration (FDA). Methods to identify what is important to patients and select, develop or modify fit-for-purpose clinical outcomes assessments. Patient-focused drug development guidance public workshop. October 15–16 2018. <https://www.fda.gov/media/116281/download>. Accessed 4 Feb 2020.
 34. Cronbach L. Coefficient alpha and the internal structure of tests. *Psychometrika.* 1951;16:297–334.
 35. Cappelleri JC, Zou KH, Bushmakina AG, Alvir JM, Alemayehu D, Symonds T. Patient-reported outcomes: measurement, implementation and interpretation. Cambridge: CRC Press; 2013.
 36. Hays R, Prince-Embury S, Chen H. HIS: RAND-36 health status inventory. San Antonio: The Psychological Corporation; 1998.
 37. Cohen J. A power primer. *Psychol Bull.* 1992;112(1):155–9.
 38. Madani S, Shapiro J. Alopecia areata update. *J Am Acad Dermatol.* 2000;42(4):549–66.
 39. Wyrwich KW, Kitchen H, Knight S, Aldhouse NVJ, Macey J, Nunes FP, et al. Development of the Scalp Hair Assessment PRO measure for alopecia areata. *Br J Dermatol.* 2020;183(6):1065–72.
 40. Mesinkovska N, King B, Mirmirani P, Ko J, Cassella J. Burden of illness in Alopecia Areata: a cross-sectional online survey study. *J Investig Dermatol Symp Proc.* 2020;20(1):S62–8.
 41. Taft C, Karlsson J, Sullivan M. Do SF-36 summary component scores accurately summarize subscale scores? *Qual Life Res.* 2001;10(5):395–404.
 42. Laucis NC, Hays RD, Bhattacharyya T. Scoring the SF-36 in orthopaedics: a brief guide. *J Bone Jt Surg Am.* 2015;97(19):1628–34.
 43. Fabbrocini G, Panariello L, De Vita V, Vincenzi C, Lauro C, Nappo D, et al. Quality of life in alopecia

-
- areata: a disease-specific questionnaire. *J Eur Acad Dermatol Venereol.* 2013;27(3):e276–81.
44. Endo Y, Miyachi Y, Arakawa A. Development of a disease-specific instrument to measure quality of life in patients with alopecia areata. *Eur J Dermatol.* 2012;22(4):531–6.
45. Chernyshov PV, Tomas-Aragones L, Finlay AY, Manolache L, Marron SE, Sampogna F, et al. Quality of life measurement in alopecia areata. Position statement of the European Academy of Dermatology and Venereology Task Force on Quality of Life and Patient Oriented Outcomes. *J Eur Acad Dermatol Venereol.* 2021;35(8):1614–21.
46. ClinicalTrials.gov. NCT3732807: PF-06651600 for the Treatment of Alopecia Areata (ALLEGRO-2b/3). 2020. <https://www.clinicaltrials.gov/ct2/show/NCT03732807>. Accessed 3 Feb 2021.
47. ClinicalTrials.gov. NCT04006457: long-term PF-06651600 for the Treatment of Alopecia Areata (ALLEGRO-LT). 2020. <https://www.clinicaltrials.gov/ct2/show/NCT04006457>. Accessed 3 Feb 2021.