Psychometric evaluation of three patient-reported outcome questionnaires assessing the symptoms and impacts of atopic dermatitis in adults and adolescents

Jonathan I Silverberg[1], Eric L Simpson[2], Leighann Litcher-Kelly[3], Jeffrey McDonald[3], Brian M Calimlim[4], Yael A Leshem[5]

[1]Dept of Dermatology, The George Washington University School of Medicine and Health Sciences, Washington DC, USA; [2]Dept of Dermatology, Oregon Health & Science University, Portland OR, USA; [3]Patient-Centered Outcomes, Adelphi Values LLC, Boston MA, USA; [4]AbbVie Inc, North Chicago IL, USA; [5]A Division of Dermatology, Rabin Medical Center, Petah Tikva, Israel; Sackler School of Medicine, Tele Aviv University, Tel Aviv, Israel

Presented at the Revolutionizing Atopic Dermatitis (RAD) Congress • December 13-14, 2020 • Virtual

OBJECTIVE

The objective of this study was to evaluate the psychometric properties of the Atopic Dermatitis Symptom Scale 7-item total symptom score (ADerm-SS TSS-7), the three Atopic Dermatitis Impact Scale (ADerm-IS) domain scores (Sleep, Daily Activities, Emotional State), and the Worst Pruritus Numerical Rating Scale (NRS) score.

INTRODUCTION

- Evaluating the patient perspective for symptoms and impacts of chronic conditions, such as atopic dermatitis (AD), are important for regulated clinical trials.
- Without knowing that patients feel and function better, it is difficult to determine whether a new treatment is truly beneficial to patients' lives.
- Based on best measurement practice summarized in regulatory guidance documents[1-3], three patient-reported outcome (PRO) questionnaires were developed to evaluate the symptoms and impacts of moderate to severe AD experienced by adolescents and adults.
- The ADerm-SS, ADerm-IS, and Worst Pruritus NRS are newly developed PRO questionnaires that assess the symptoms, impacts, and itch of moderate to severe AD, respectively **(Table 1)**
- In order to evaluate treatment benefit, scores from PRO questionnaires should have evidence of reliability and validity. In addition, the meaning of changes in scores, and whether they are important to patients should also be documented.

METHODS

- Data from a global, randomized, double-blind, placebo-controlled multi-center clinical trial of adolescent and adult patients with moderate to severe AD was used for the following analyses:
- Confirmatory factor analyses (CFAs) to evaluate a priori factor structures that were identified with exploratory factor analyses using prior clinical trial data:
  - ADerm TSS-7 was evaluated with a bifactor model to assess essential unidimensionality **(Figure 1)**.
  - A General Factor explains the unidimensionality among the items
  - Specific Factors (i.e., correlated item groups) were constructed to evaluate the variance explained by the General Factor versus the Specific Factors. Greater variance explained by the General Factor suggests unidimensionality:
  - Specific Factor 1 used for daily items (Items 1–3)
  - Specific Factor 2 used for cracking items (Items 4–5)
  - Specific Factor 3 used for dry/flaking skin items (Items 6–7)
  - ADerm-IS was evaluated with a three-factor model **(Figure 2)**; prior exploratory factor analysis results demonstrated the questionnaire includes three domains.
  - Model fit was determined using the Comparative Fit Index (CFI), as well as the proportion of variance explained by the General Factor for unidimensionality (ωh) for the bifactor model.
- Reliability
  - Internal consistency: Cronbach's Alpha (Cr-α)
  - Test-retest reliability for stable patients defined by no change between Baseline and Week 2 on the patient reported global impression of severity and change questionnaires (PGIS and PGIC; **Table 1**): intraclass correlation coefficient (ICC)
- Validity
  - Convergent validity: correlations with concurrent measures (Dermatology Life Quality Index [DLQI], Patient-Oriented Eczema Measure [POEM]; **Table 1**)
  - Known groups validity evaluates whether the scores on the target questionnaires differ between the groups defined by the PGIS response options
- Estimates of minimally important within-person change scores evaluated for potential responder definitions: Anchor-based analyses using PGIS and PGIC

METHODS, CONTINUED

Table 1. Questionnaires included in the analyses

Target questionnaires	
ADerm-SS	• The ADerm-SS questionnaire includes 11 items to assess signs and symptoms of moderate to severe AD, including itch, skin pain, skin cracking and pain associated with cracking, and dry and flaking skin; • Each item uses a 0–10 NRS; all items use a recall period of "Past 24 hours" • Items 1–3 are completed daily and Items 4–11 are completed weekly • Although the ADerm-SS includes 11 items, the first 7 items of the questionnaire are summed to calculate the ADerm-SS TSS-7 score.† – ADerm-SS TSS-7 score ranges from 0–70 – Higher scores mean more severe symptoms
ADerm-IS	• The ADerm-IS questionnaires include 10 items to assess the sleep impacts, daily activity limitations, and emotional impacts associated with moderate to severe AD. • Each item uses a 0–10 NRS. • Items 1–3 use a recall period of "Past 24 hours," and are completed daily – ADerm-IS sleep is the sum of Items 1–3, and scores range from 0–30 – Higher scores mean more sleep impact • Items 4–10 use a recall period of "Past 7 days," and are completed weekly – ADerm-IS Daily Activities is the sum of Items 4–7, and scores range from 0–40 – ADerm-IS Emotional State is the sum of Items 8–10, and scores range from 0–30 – Higher scores mean more impact on daily activities and emotional state
Worst Pruritus NRS	• Single-item questionnaire to assess worst pruritus severity assessed on a 0–10 NRS • Questionnaire uses a recall period of "Past 24 hours," and is completed daily • Worst Pruritus NRS Weekly Average is calculated for each week, and scores range from 0–10 – Higher scores mean worse pruritus
Supportive questionnaires	
Dermatology Life Quality Index (DLQI)	• DLQI is a 10-item questionnaire to measure impact of skin problems • Items use a verbal response scale (VRS) • Items use a recall period of "Over the last week" • Questionnaire is completed at Baseline and select post-treatment clinic visits • DLQI Total Score is the sum of the 10 items; score ranges from 0–30 • Higher scores mean more severe impact
Patient-Oriented Eczema Measure (POEM)	• The POEM includes 7 items to measure symptoms using a 4-point VRS • Questionnaire uses a recall period of "Over the past week" • Questionnaire is completed at Baseline and select post-treatment clinic visits • POEM total score is the sum of the 7 items and ranges from 0–28 – Higher scores mean more severe symptoms
Patient Global Impression of Severity (PGIS)	• Single-item questionnaire to assess AD severity using a 7-point VRS – 0=Absent, no symptoms to 6=Very Severe • Questionnaire uses a recall period of "Right now" • Questionnaire is completed at Baseline and select post-treatment clinic visits • PGIS scores ranges from 0–6 – Higher scores mean more severe symptoms
Patient Global Impression of Change (PGIC)	• Single-item questionnaire to measure change in AD on a 7-point VRS – 1=Very much improved to 7=Very much worse • Questionnaire uses a recall period of "Since the start of treatment" • Questionnaire is completed at post-treatment clinic visits • PGIC scores ranges from 1–7 – Higher scores mean worsening of symptoms

† To minimize participant burden and avoid repeated measurement of concepts potentially assessed by other instruments in clinical trials, the ADerm-SS TSS-7 was developed by summing items that assess concepts not measured by clinician-reported questionnaires, specifically: itch while asleep, itch while awake, skin pain, skin cracking, pain caused by skin cracking, dry skin, and skin flaking.

RESULTS

- Adolescents and adults were included in these analyses **(Table 2)**
- Distribution of scores on the items of the three target questionnaires showed no floor or ceiling effects (results not shown)

Table 2. Sample demographics

Characteristic		Adults (n=769)	Adolescents (n=113)
Age	Mean (SD)	36.9 (14.1)	15.5 (1.7)
	Range	18.0–75.0	12.0–18.0
Gender, N (%)	Male	295 (38.4%)	51 (45.1%)
	Female	474 (61.6%)	62 (54.9%)
Race, N (%)	Asian	166 (21.6%)	18 (15.9%)
	Black/African American	39 (5.1%)	10 (8.8%)
	White	546 (71.0%)	84 (74.3%)
	Multiple/Other	18 (2.3%)	1 (0.8%)

RESULTS, CONTINUED

- Confirmatory Factor Analysis (CFA) supports the factor structure for a single domain of the ADerm-SS TSS-7 and three domains for the ADerm-IS **(Table 3)**

Table 3. Confirmatory factor analysis model fit statistics

Measure	Time point	Adults		Adolescents	
		CFI	ω_h	CFI	ω_h
ADerm-SS TSS-7 One Factor (Figure 1)	Baseline	0.968	0.776	0.970	0.872
	Week 4	0.976	0.861	0.969	0.857
ADerm-IS Three Factors (Figure 2)	Baseline	0.970	N/A	0.970	N/A
	Week 4	0.981	N/A	0.978	N/A

N/A= Not applicable because no General Factor (ω_h) was included in the ADerm-IS models

Figure 1. Hypothesized factor structure of ADerm-SS TSS-7

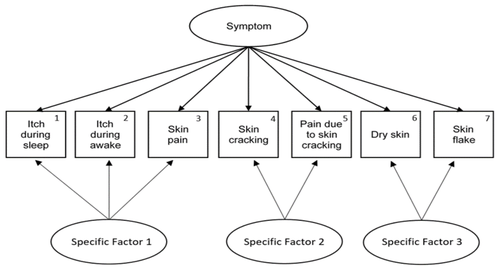
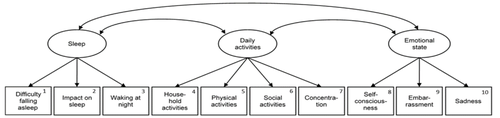


Figure 2. Hypothesized factor structure of ADerm-IS



- ADerm-SS TSS-7 and the three ADerm-IS domain scores demonstrated adequate internal consistency reliability (for both adults and adolescents; Cr-α in **Table 4**).
- The ADerm-SS TSS-7, ADerm-IS domains, and Worst Pruritus NRS scores demonstrated adequate test-retest reliability in adults (ICC presented in **Table 5**).
  - Test-retest reliability for adolescents could not be accurately measured due to the small number of adolescent patients with no change between Baseline and Week 2 on the PGIS and PGIC (results not shown).

Table 4. Internal consistency reliability at Baseline

Target measure/scores		Adults (n=769)	Adolescents (n=113)
ADerm-SS	TSS-7	0.92	0.91
ADerm-IS	Sleep	0.98	0.96
	Daily Activities	0.92	0.90
	Emotional State	0.90	0.89

Table 5. Test-retest reliability for adults at Baseline and Week 2

Target measure/scores		PGIC-stability (n=165)	PGIS-stability (n=117)
ADerm-SS	TSS-7	0.59	0.58
ADerm-IS	Sleep	0.73	0.80
	Daily Activities	0.67	0.71
	Emotional State	0.65	0.68
Worst Pruritus NRS		0.57	0.66

- Results supported the validity of the scores for the three target questionnaires.
- Convergent validity correlations are presented in **Table 6**
- Known groups results for adults are presented in **Table 7**
  - Similar results were seen for the adolescent sample (results not shown).

Table 6. Convergent validity correlations of target and concurrent measures at Week 2 for adults (n=765) and adolescents (n=112)

Target measures/scores		Concurrent measures			
		DLQI		POEM	
		Adults	Adolescents	Adults	Adolescents
ADerm-SS	TSS-7	0.69	0.58	0.75	0.70
ADerm-IS	Sleep	0.63	0.70	0.61	0.52
	Daily Activities	0.78	0.71	0.65	0.61
	Emotional State	0.74	0.68	0.63	0.50
Worst Pruritus NRS		0.66	0.67	0.71	0.61

- Based on adult (n=637) change score distributions by PGIS **(Figure 3)** and PGIC **(Figure 4)**, the following ranges characterize a minimally important within-person change:
  - ADerm-SS TSS-7: 19–29 pts.
  - ADerm-IS Sleep: 8–13 pts.
  - ADerm-IS Daily Activities: 10–16 pts.
  - ADerm-IS Emotional State: 8–12 pts.
  - Worst Pruritus NRS: 3–4 pts.

Table 7. Known-groups validity for ADerm-SS, ADerm-IS, and Worst Pruritus NRS for adults (n=765) at Week 2

PGIS group	n	Mean (SD)	Median	p-value
ADerm-SS TSS-7				
Absent	23	6.2 (10.4)	2.0	
Minimal	176	10.6 (9.6)	7.0	
Mild	168	19.3 (11.6)	17.0	
Moderate	176	30.5 (14.3)	29.0	<0.001
Moderately severe	124	36.0 (13.0)	35.0	
Severe	58	49.8 (11.4)	50.0	
Very Severe	20	56.7 (12.1)	60.5	
ADerm-IS Sleep				
Absent	23	3.9 (6.6)	1.4	
Minimal	176	4.6 (5.3)	2.6	
Mild	168	8.4 (6.3)	7.8	
Moderate	176	12.1 (7.5)	12.3	<0.001
Moderately severe	124	13.9 (6.5)	14.1	
Severe	58	20.7 (6.0)	20.8	
Very Severe	20	23.4 (5.9)	23.2	
ADerm-IS Daily Activities				
Absent	23	2.8 (6.1)	0.0	
Minimal	176	3.7 (5.3)	1.0	
Mild	168	8.3 (7.5)	7.0	
Moderate	177	14.1 (9.4)	13.0	<0.001
Moderately severe	124	19.2 (9.0)	20.0	
Severe	58	27.5 (8.8)	28.5	
Very Severe	20	32.8 (6.4)	32.0	
ADerm-IS Emotional State				
Absent	23	2.4 (4.6)	0.0	
Minimal	176	3.8 (4.8)	2.0	
Mild	168	8.3 (6.6)	7.0	
Moderate	177	13.0 (8.1)	12.0	<0.001
Moderately severe	124	16.2 (7.3)	17.0	
Severe	58	22.1 (6.0)	22.5	
Very Severe	20	25.1 (5.1)	25.0	
Worst Pruritus NRS				
Absent	23	1.8 (2.1)	1.2	
Minimal	176	2.5 (1.7)	2.1	
Mild	168	3.8 (1.8)	3.9	
Moderate	176	5.2 (1.9)	5.3	<0.001
Moderately severe	124	5.9 (1.7)	6.0	
Severe	58	7.6 (1.3)	7.7	
Very Severe	20	8.2 (1.6)	8.0	

Figure 3. 95%confidence intervals and mean change scores on the target questionnaires, by PGIS change score between Baseline and Week 16

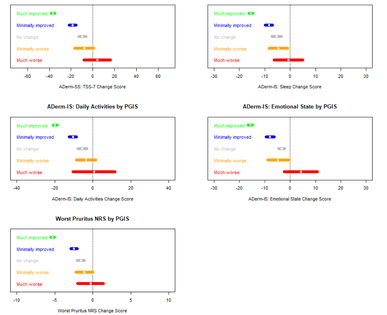
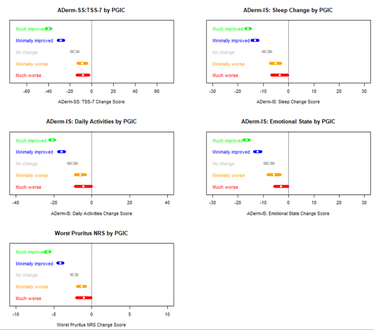


Figure 4. 95%confidence intervals and mean change scores on the target questionnaires, by PGIC score at Week 16



REFERENCES

1. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Center for Devices and Radiological Health. Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Silver Spring, MD: Office of Communications, Division of Drug Information;2009.
2. Patrick DL, Burke LB, Gwaltney CJ, et al. Content Validity-Establishing and Reporting the Evidence in Newly Developed Patient-Reported Outcomes (PRO) Instruments for Medical Product Evaluation: ISPOR PRO Good Research Practices Task Force Report: Part 1-Eliciting Concepts for a New PRO Instrument. *Value in Health*. 2011;14(8):967-977.
3. Patrick DL, Burke LB, Gwaltney CJ, et al. Content Validity-Establishing and Reporting the Evidence in Newly Developed Patient-Reported Outcomes (PRO) Instruments for Medical Product Evaluation: ISPOR PRO Good Research Practices Task Force Report: Part 2-Assessing Respondent Understanding. *Value in Health*. 2011;14(8):978-988.

ACKNOWLEDGEMENTS

AbbVie Inc., funded this study and participated in the study design; study research; collection, analysis and interpretation of data; and writing, reviewing and approving of this publication. All authors had access to the data, and participated in the development, review, and approval, and in the decision to submit this publication. No honoraria or payments were made for authorship.

The authors thank the following individuals who supported this work: Ethan Arenson, Roger E Lamoureux, Shirley Li, Paolo Medrano, Sylvia Su, Xiaowu Sun, and Christine Yip. These individuals are employees of Adelphi Values and were part of the research team for this project. This support was funded by AbbVie.

CONCLUSIONS

- Results demonstrate the reliability, convergent validity, and meaning of change for the ADerm SS TSS-7, three ADerm-IS domain scores (Sleep, Daily Activity, Emotional State), and Worst Pruritus NRS score, and support their use to assess the symptoms and impacts experienced by adults and adolescents with moderate to severe AD.
- The potential responder definitions for minimally important within-person change may be used to evaluate treatment effects in future clinical trials.

DISCLOSURES

JIS received honoraria for advisory board, speaker, and consultant services from AbbVie, Asana, Bluefin, Boehringer-Ingelheim, Celgene, Dermavant, Dermira, Eli Lilly, Galderma, GlaxoSmithKline, Glenmark, Incyte, Kiniksa, Leo Pharma, Novartis, Pfizer, Regeneron, Sanofi, and research grants for investigator services from Galderma.
ELS reports grants, personal fees, and non-financial support from Eli Lilly; grants and personal fees from Anacor Pharma, GlaxoSmithKline, Regeneron Pharmaceuticals, Sanofi Genzyme, Pfizer, Leo Pharma, Eli Lilly, and Valeant Pharmaceuticals; personal fees from AbbVie, Celgene Corporation, Dermira, Galderma, Genentech, Leo Pharma, Menlo Therapeutics; and grants from MedImmune, Novartis, Roivant Sciences, Tioga Pharmaceuticals, and Vanda Pharmaceuticals.
LL-K & JM are employed by Adelphi Values LLC, which received payment from AbbVie Inc. to support the research activities presented in this publication.
BMC is an employee of AbbVie and may own AbbVie stock or stock options.
YL has received honoraria or fees as a consultant from AbbVie, Sanofi, and Genentech and as an advisory board member from Sanofi and Regeneron Pharmaceuticals, Pfizer, and Dexcel Pharma, an independent research grant from Abbvie, and has, without personal compensation, provided investigator services for Eli Lilly, Pfizer, and AbbVie.