




The Validated Investigator Global Assessment for Atopic Dermatitis (vIGA-AD™): a clinical outcome measure for the severity of atopic dermatitis

Eric L. Simpson,¹ Robert Bissonnette,² Amy S. Paller,³ Brett King ,⁴ Jonathan I. Silverberg ,⁵ Kristian Reich ,⁶ Jacob P. Thyssen,⁷ Helen Doll,⁸ Luna Sun,⁹ Amy M. DeLozier,⁹ Fabio P. Nunes¹⁰ and Lawrence F. Eichenfield¹¹

¹Oregon Health and Science University, Portland, OR, USA

²Innovaderm Research Inc., Montreal, QC, Canada

³Northwestern University Feinberg School of Medicine, Chicago, IL, USA

⁴Yale School of Medicine, New Haven, CT, USA

⁵George Washington University School of Medicine and Health Sciences, Washington, DC, USA

⁶Center for Translational Research in Inflammatory Skin Diseases, Institute for Health Services Research in Dermatology and Nursing, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

⁷Department of Dermatology, Bispebjerg Hospital, University of Copenhagen, Copenhagen, Denmark

⁸Clinical Outcomes Solutions, Folkestone, UK

⁹Eli Lilly and Company, Indianapolis, IN, USA

¹⁰Janssen Global Services, LLC, PA; this work performed previously while at Eli Lilly and Company, Indianapolis, IN, USA

¹¹University of California San Diego School of Medicine, La Jolla, CA, USA

Abstract

Correspondence

Eric L. Simpson.

Email: simpson@ohsu.edu

Accepted for publication

14 April 2022

DOI 10.1111/bjd.21615

Background The validated Investigator Global Assessment for Atopic Dermatitis (vIGA-AD™) is a standardized severity assessment for use in clinical trials and registries for atopic dermatitis (AD).

Objectives To investigate the reliability, validity, responsiveness and within-patient meaningful change of the vIGA-AD.

Methods Data were analysed from adult patients with moderate-to-severe AD in the BREEZE-AD1 (N = 624 patients; NCT03334396), BREEZE-AD2 (N = 615; NCT03334422) and BREEZE-AD5 (N = 440; NCT03435081) phase III baricitinib clinical studies.

Results Across studies, test–retest reliability for stable patients showed moderate-to-good agreement [range of Kappa values for Patient Global Impression of Severity–Atopic Dermatitis (PGI-S-AD), 0.516–0.639; for Eczema Area and Severity Index (EASI), 0.658–0.778]. Moderate-to-large correlations between vIGA-AD and EASI or body surface area (range at baseline, 0.497–0.736; Week 16, 0.716–0.893) supported convergent validity. Known-groups validity was demonstrated vs. EASI and PGI-S-AD (vIGA-AD for severe vs. moderate EASI categories at baseline, $P < 0.001$). Responsiveness was demonstrated vs. EASI ($P < 0.001$ for much improved vs. improved and improved vs. stable). Anchor- and distribution-based methods supported a vIGA-AD change of –1.0 as clinically meaningful. These findings are limited to populations defined by the studies' inclusion and exclusion criteria.

Conclusions The vIGA-AD demonstrated sufficient reliability, validity, responsiveness and interpretation standards for use in clinical trials.

What is already known about this topic?

- A description of the development of the validated Investigator Global Assessment for Atopic Dermatitis (vIGA-AD™) has been published previously.

What does this study add?

- The current study validates the vIGA-AD by demonstrating appropriate test–retest reliability, convergent validity, known-groups validity and responsiveness across three baricitinib clinical studies.
- In addition, a 1-point change was identified as a clinically meaningful patient-perceived change minimal clinically important difference in the vIGA-AD.

What are the clinical implications of the work?

- The vIGA-AD is a measure for investigator assessment of atopic dermatitis suitable for use in clinical research.

Atopic dermatitis (AD) is a common chronic, inflammatory skin disease with significant unmet medical need and several novel treatments under development.^{1,2} At present, numerous outcome measures are used in clinical trials of AD,^{3,4} however, in a recent systematic review, only two of these instruments, the Eczema Area and Severity Index (EASI)⁵ and the Scoring Atopic Dermatitis (SCORAD) index⁶ were considered adequately validated and recommended for use in clinical trials.⁴ As US regulators require a single-item investigator global assessment (IGA) among clinical trial endpoints for AD, a validated measure suitable for use across AD clinical trials is needed.

To address this need, the validated Investigator Global Assessment for AD (vIGA-ADTM) was developed.⁷ The vIGA-AD is a clinician-rated scale to assess the overall severity of AD lesions at a given timepoint. It is scored from 0 (clear) to 4 (severe) based on four clinical features of AD lesions: erythema, induration/papulation, lichenification and oozing/crusting, and takes extent of disease into account. Expert consensus was previously established around a 5-point IGA scale including morphological descriptions, content validity, strong interrater reliability and agreement, and the development of a training module and certification.⁷ Uptake of the vIGA-AD as a clinical outcome measure has been rapid: by January 2020, over 4500 investigators from 48 countries had been trained on the vIGA-AD, and it had been adopted by 13 sponsors for use in 38 clinical trials.⁷

In order to support the psychometric validity and interpretation of the vIGA-AD, data from three phase III clinical studies of baricitinib monotherapy for moderate-to-severe AD in adults were used to assess the measure's psychometric properties (test–retest reliability, validity and responsiveness) and to provide an estimate of within-patient clinically meaningful change, the meaningful change threshold.

Materials and methods

Patients

BREEZE-AD1 (clinicaltrials.gov NCT03334396), BREEZE-AD2 (NCT03334422) and BREEZE-AD5 (NCT03435081) were phase III, multicentre, randomized, double-blind, placebo-

controlled, parallel-group, monotherapy studies evaluating the efficacy and safety of baricitinib treatment for AD compared with placebo worldwide. All studies evaluated patients aged ≥ 18 years, diagnosed with moderate-to-severe AD for ≥ 12 months, who had responded inadequately to, or who were intolerant of, topical therapy. Results for the BREEZE-AD1 (N = 624 patients) and -AD2 (N = 615) studies have been reported previously.⁸ Both had a 16-week double-blinded treatment period. Patients were randomized 2 : 1 : 1 : 1 to receive placebo once daily (QD) or baricitinib 1 mg QD, baricitinib 2 mg QD or baricitinib 4 mg QD. BREEZE-AD5 (N = 440) had a 104-week double-blinded treatment period. Patients were randomized 1 : 1 : 1 to receive placebo QD, baricitinib 1 mg QD or baricitinib 2 mg QD.

The clinical studies were performed in compliance with the International Conference on Harmonization Good Clinical Practice Guidelines. All patients provided written informed consent, and institutional review boards or ethics committees approved the study protocol before each study started.

Analytical methods

The psychometric evaluation within the BREEZE-AD1, -AD2 and -AD5 studies employed blinded data. All analyses were conducted on the full intent-to-treat population, or the relevant intent-to-treat-derived population as specified in each analysis. The performance of the vIGA-AD was evaluated with respect to prespecified thresholds for acceptability. The vIGA-AD was completed at every study visit associated with clinician-rated outcomes (ClinROs).

Instruments included the following ClinROs and patient-reported outcomes (PROs):

Body surface area (BSA): The BSA measurement is a clinician-rated assessment of the percentage of BSA involved with AD; it does not incorporate lesion severity.

Eczema Area and Severity Index (EASI): The EASI is a validated, clinician-rated scoring system that grades the area and severity of eczema across four body regions, with a total range of 0–72.⁵

Patient-Oriented Eczema Measure (POEM): The POEM⁹ is a seven-item, patient-reported questionnaire assessing AD/

eczema-specific symptoms over the last week; each item is scored 0 (no days) to 4 (every day) based on the number of days affected.

Dermatology Life Quality Index (DLQI): The DLQI¹⁰ is a 10-item questionnaire assessing health-related quality of life over the last week in patients with dermatological symptoms, with each item scored for impact from not at all (0) to very much (3).

Patient Global Impression of Severity – Atopic Dermatitis (PGI-S-AD): The PGI-S-AD is a single-item question asking the patient how they would rate their overall AD symptoms over the past 24 h. The five categories of responses range from ‘no symptoms’ to ‘severe’.

Psychometric analyses

Reliability (test–retest)

Test–retest reliability of the vIGA-AD was assessed in patients who were stable during the interval between baseline and Week 1, and between Weeks 4 and 8. Although there are no perfect timepoints to assess test–retest reliability in a clinical trial, based on phase II data, we anticipated few changes occurring within the first week of the study, or between Weeks 4 and 8.¹¹ Test–retest reliability was assessed using Cohen’s Kappa statistic with quadratic weighting.¹² Kappa was evaluated as: ≤ 0.21 , low; $0.21–0.40$, fair; $0.41–0.60$, moderate; $0.61–0.80$, good; and ≥ 0.81 , excellent.¹³ Stable patients were defined in two ways: no change in single-timepoint assessment of PGI-S-AD between visits; and change in EASI score < 6.6 points [the minimal clinically important difference (MCID) of the measure]¹¹ between visits.

Convergent and divergent validity

Convergent and divergent validity between the vIGA-AD and the EASI, BSA, POEM and DLQI were assessed using polyserial correlations. Analyses were conducted on baseline scores, with additional analyses conducted at Week 16. Cohen’s conventions were used to interpret the absolute value of the correlation results, where > 0.7 is large; $0.4–0.7$ is moderate; and < 0.4 is small.^{14–18} Relatively strong correlations with the clinician-reported EASI and BSA would support convergent validity, while weaker correlations with the PROs (POEM and DLQI), which measure concepts more distally related to AD symptoms, would support divergent validity.

Known-groups validity

Known-groups (discriminant) validity was evaluated between subgroups defined by the EASI and PGI-S-AD at baseline. The following groups were prespecified: EASI ‘moderate to severe’ (scores > 7.0 to ≤ 50.0) and ‘very severe’ (scores > 50.0);¹⁹ and PGI-S-AD ‘no symptoms to mild symptoms’ (1–3) and ‘moderate to severe’ (4–5). Due to the inclusion criteria of

each study, no patients were classified as ‘clear to mild’ (≤ 7.0) on the EASI.

Both the ClinRO EASI and PGI-S-AD single-timepoint assessment met the criteria for inclusion in this analysis of correlation with the vIGA-AD above 0.35.²⁰ EASI or PGI-S-AD severity groups and vIGA-AD of 3 vs. 4 were evaluated using two-by-two crosstabulation tables with the χ^2 -test to assess known-groups validity.

Responsiveness

Responsiveness was evaluated using nonparametric methods to assess shift in vIGA-AD from baseline to Week 4 and baseline to Week 16 within the four previously published change categories of the EASI (point change: < -13.2 , much improved; -13.2 to ≤ -6.6 , improved; -6.6 to < 6.6 , stable; and ≥ 6.6 , declined).¹⁹ Differences in vIGA-AD change scores between groups were tested using pairwise two-sample Wilcoxon comparisons.

Clinically meaningful change

The PGI-S-AD was used as an anchor as the primary method to interpret the meaningful change threshold in the vIGA-AD score. Mean changes in vIGA-AD were compared with change in the ‘minimal (–1)’, ‘marked (–2)’ and ‘very marked (–3 to –4)’ improvement groups as defined by the uncollapsed PGI-S-AD anchor groups at Week 16. Distribution-based methods were used as a supportive measure.

Results

Baseline characteristics are shown by study for the overall study populations in Table 1. The mean age across the studies ranged from 34.7 to 39.5 years and the majority of patients in each study were male. Over half of the patients in each study were White (57.3–68.5%), with Asian patients being the second largest racial group (18.5–30.4% across studies). Black/African American patients were not substantially represented in the BREEZE-AD1 or BREEZE-AD2 studies but made up 18.3% of the population in BREEZE-AD5. The mean time since AD diagnosis ranged across studies from 23.6 to 25.7 years.

Reliability (test–retest)

Results assessing test–retest reliability based on Kappa values are shown in Table 2. Across the BREEZE-AD1, BREEZE-AD2 and BREEZE-AD5 studies, for patients reporting no changes in the PGI-S-AD, Kappa values ranged from 0.516 to 0.623 for baseline to Week 1, and from 0.546 to 0.639 from Week 4 to Week 8 (both indicating moderate-to-good agreement). For patients with EASI change < 6.6 (the MCID), Kappa statistics ranged from 0.658 to 0.703 from baseline to Week 1 and from 0.673 to 0.778 from Week 4 to Week 8, indicating good agreement for both analyses.

Table 1 Baseline characteristics of patients by study

	BREEZE-AD1 N = 624	BREEZE-AD2 N = 615	BREEZE-AD5 N = 440
Age, mean (SD)	35.6 (12.8)	34.7 (12.8)	39.5 (16.1)
Sex, n (%) female	233 (37.3)	234 (38.0)	216 (49.1)
Race, n (%)			
White	366 (58.9)	421 (68.5)	251 (57.3)
Asian	189 (30.4)	183 (29.8)	81 (18.5)
Black/African American	2 (0.3)	0	80 (18.3)
American Indian/Alaska Native	30 (4.8)	0	6 (1.4)
Native Hawaiian or Other Pacific Islander	0	1 (0.2)	2 (0.5)
Multiple	34 (5.5)	10 (1.6)	18 (4.1)
Other/missing	3	0	2
Years since diagnosis, mean (SD; min–max)	25.7 (15.1; 1–76)	24.2 (13.9; 1–72)	23.6 (16.7; 1–76)

Table 2 Test–retest reliability of the validated Investigator’s Global Assessment for atopic dermatitis by study

	Baseline to Week 1		Weeks 4–8	
	N	Kappa (95% CI) ^a	N	Kappa (95% CI) ^a
BREEZE-AD1				
PGI-S-AD	347	0.623 (0.551–0.696)	294	0.612 (0.541–0.682)
EASI	376	0.699 (0.636–0.761)	402	0.777 (0.733–0.821)
BREEZE-AD2				
PGI-S-AD	289	0.516 (0.432–0.600)	275	0.639 (0.570–0.707)
EASI	313	0.658 (0.582–0.735)	382	0.778 (0.730–0.826)
BREEZE-AD5				
PGI-S-AD	193	0.594 (0.493–0.695)	190	0.546 (0.456–0.636)
EASI	270	0.703 (0.624–0.783)	280	0.673 (0.607–0.740)

CI, confidence interval; EASI, Eczema Area and Severity Index; N, number of total patients; PGI-S-AD, Patient Global Impression of Severity–Atopic Dermatitis ^aCohen’s Kappa was evaluated as: ≤ 0.21 , low; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, good; and ≥ 0.81 , excellent.

Convergent and divergent validity

Results for convergent and divergent validity of the vIGA-AD relative to other clinical outcome measures are shown in Table 3. At baseline, correlations with the vIGA-AD were moderate to-large for EASI (range 0.689–0.736) and moderate for BSA (0.497–0.567). At Week 16, correlations were large for both EASI (range 0.826–0.893) and BSA (0.716–0.745); these findings support convergent validity of the vIGA-AD. In support of divergent validity were the small correlations (range 0.297–0.365) found between the vIGA-AD and PRO assessments (DLQI and POEM) at baseline, with moderate correlations found at Week 16 (range 0.429–0.647).

Known-groups validity

Known-groups validity was assessed based on the ability of the vIGA-AD to discriminate between subgroups of patients with different underlying disease severity as measured by the EASI and PGI-S-AD (Table 4). At baseline in all three studies, patients with a vIGA score of 4 were significantly more likely to have categorically worse disease severity based on either

the EASI or PGI-S-AD vs. patients with a vIGA of 3 (all $P < 0.01$).

Responsiveness

In all three studies, the magnitude of improvement in the vIGA-AD increased with greater improvement in the EASI, demonstrating responsiveness of the vIGA-AD (Table 5 and Table S1; see Supporting Information). In each study, for both measures at Weeks 4 and 16, the ‘Much Improved’ group differed at $P < 0.0001$ from the ‘Improved’ group, and the ‘Improved’ group differed at $P < 0.001$ from the ‘Stable’ group.

Estimate of meaningful change

Anchor-based estimates compared changes in the vIGA-AD to changes considered meaningful in the PGI-S-AD, using the uncollapsed PGI-S-AD anchor groups at Week 16 (Table S2; see Supporting Information). The overall clinical threshold for minimal meaningful change was -1.00 , for moderate change -1.25 or -1.50 , and for large change -1.75 or -2.00 ,

Table 3 Convergent and divergent validity – correlation of clinician- and patient-reported measures with the validated Investigator's Global Assessment for atopic dermatitis (vIGA-AD)

	BREEZE-AD1 ^a		BREEZE-AD2 ^a		BREEZE-AD5 ^a	
	Baseline	Week 16	Baseline	Week 16	Baseline	Week 16
Clinician-reported outcomes						
EASI	0.708	0.870	0.689	0.826	0.736	0.893
BSA	0.497	0.716	0.555	0.735	0.567	0.745
Patient-reported outcomes						
DLQI	0.305	0.439	0.297	0.429	0.307	0.555
POEM	0.304	0.499	0.365	0.542	0.311	0.647

BSA, body surface area; DLQI, Dermatology Quality of Life Index; EASI, Eczema Area and Severity Index; POEM, Patient-Oriented Eczema Measure ^aPolyserial correlation coefficients were calculated as correlations between vIGA-AD and continuous reference measures EASI, BSA, DLQI and POEM. Concurrent validity was small if the resulting coefficient was < 0.4, moderate if the coefficient was >0.4 to 0.7, and large if the coefficient was > 0.7.

Table 4 Known-groups validity of the vIGA-AD based on EASI and PGI-S-AD subgroups at baseline

	EASI		PGI-S-AD	
	> 7 to ≤ 50 (moderate to severe)	> 50 (very severe)	1–3 (none-to-mild symptoms)	4 or 5 (moderate-to-severe symptoms)
BREEZE-AD1				
Sample size	562	59	168	436
vIGA-AD = 3 at baseline, n (%)	357 (63.5)	3 (5.1)	126 (75.0)	224 (51.4)
vIGA-AD = 4 at baseline, n (%)	205 (36.5)	56 (94.9)	42 (25.0)	212 (48.6)
Innovaderm Research Inc., Montreal ^a		74.84		27.77
Montreal χ^2 -test ^a				
P-value ^a	–	< 0.0001	–	< 0.0001
BREEZE-AD2				
Sample size	527	85	151	439
vIGA-AD = 3 at baseline, n (%)	299 (56.7)	4 (4.7)	102 (67.5)	195 (44.4)
vIGA-AD = 4 at baseline, n (%)	228 (43.3)	81 (95.3)	49 (32.5)	244 (55.6)
χ^2 -test ^a		79.27		24.05
P-value ^a	–	< 0.0001	–	< 0.0001
BREEZE-AD5				
Sample size	412	25	104	309
vIGA-AD = 3 at baseline, n (%)	254 (61.7)	0 (0)	75 (72.1)	167 (54.0)
vIGA-AD = 4 at baseline, n (%)	158 (38.3)	25 (100)	29 (27.9)	142 (46.0)
χ^2 -test ^a		36.81		10.47
P-value ^a	–	< 0.0001	–	0.0012

EASI, Eczema Area and Severity Index; PGI-S-AD, Patient Global Impression of Severity – Atopic Dermatitis; vIGA-AD, validated Investigator's Global Assessment for AD. ^aBetween-group comparisons based on χ^2 -test analysis of two-by-two crosstabulation tables.

indicating that a reduction of 1.0 in the vIGA-AD is consistent with a small but perceptible change in patient-perceived severity. Distribution-based methods gave estimates of -0.25 (0.5 baseline SD) and -0.65 (minimal detectable change with 95% confidence) (data not shown), indicating that a change of -1.0 was above the measurement error. Thus, -1.0 was determined to be an estimate of minimal clinically meaningful change. This meaningful change threshold can be used as a responder definition in clinical trial responder analyses to

determine the difference in meaningful, patient-perceived improvement between treatment arms.

Discussion

IGAs are relatively easy to complete, holistic measures of disease severity.²¹ IGAs have served as primary endpoints for AD in clinical trials, including registrational studies, and may be required by regulatory agencies in support of other validated

Table 5 Within-group change scores for responsiveness of the validated Investigator's Global Assessment for atopic dermatitis (vIGA-AD) to change on the Eczema Area and Severity Index (EASI) between baseline and Week 16

	EASI groups at Week 16 (point change)			
	Much improved (< -13.2)	Improved (-13.2 to ≤ -6.6)	Stable (-6.6 to < 6.6)	Declined (≥ 6.6)
BREEZE-AD1				
Sample size	320	125	106	16
Mean (SD) change in vIGA-AD	-1.39 (0.830)	-0.58 (0.721)	-0.23 (0.484)	0.38 (0.500)
Median change in vIGA-AD	-1.00	-1.00	0.00	0.00
Between-group comparisons ^a	-	< 0.0001	0.0002	0.0002
BREEZE-AD2				
Sample size	338	99	96	21
Mean (SD) change in vIGA-AD	-1.43 (0.870)	-0.62 (0.634)	-0.20 (0.609)	0.10 (0.539)
Median change in vIGA-AD	-1.00	-1.00	0.00	0.00
Between-group comparisons ^a	-	< 0.0001	< 0.0001	0.1876
BREEZE-AD5				
Sample size	195	86	68	18
Mean (SD) change in vIGA-AD	-1.59 (0.944)	-0.52 (0.608)	-0.06 (0.596)	0.06 (0.539)
Median change in vIGA-AD	-2.00	-1.00	0.00	0.00
Between-group comparisons ^a	-	< 0.0001	< 0.0001	0.8771

^aThe P-value for the pairwise comparisons between consecutive severity groups was derived from pairwise Wilcoxon comparisons assessing differences in score change between adjacent groups (Improved vs. Much improved; Stable vs. Improved; Declined vs. Stable).

measures such as the EASI.⁷ However, prior to the development of the vIGA-AD, no single global scale had been adequately validated to encourage widespread adoption across study sponsors, and thus allow for harmonization across clinical development programmes. A prior report described the development and initial validation of the vIGA-AD, indicating strong inter- and intrarater reliability, and the measure has been widely adopted in clinical studies of AD across multiple compounds and sponsors.⁷ The present results, based on data from three phase III clinical trials in adult patients with moderate-to-severe AD, support this by demonstrating that the single-item vIGA-AD has appropriate reliability, validity and responsiveness for use as a primary endpoint in randomized clinical trials in moderate-to-severe AD in adults.

In the present study, reliability was confirmed by the moderate-to-good agreement found among stable patients using both 1-week and 4-week intervals. Validity of the vIGA-AD was demonstrated by confirmation of the a priori hypotheses of convergent validity, as correlations with clinician-reported EASI and BSA were moderate-to-large at baseline and large at Week 16, and divergent validity, as smaller correlations were observed with PROs (POEM and DLQI), demonstrating uniqueness of the vIGA-AD concept in comparison with these measures. Validity of the vIGA-AD was also demonstrated by the degree to which the vIGA-AD distinguished groups under a priori hypotheses for known groups such that patients with higher EASI and PGI-S-AD scores had significantly worse overall disease severity (higher average vIGA-AD) in the respective severe categories compared with moderate categories. Responsiveness was demonstrated by the finding of statistically significant differences when comparing

mean changes in vIGA-AD between 'Much Improved' and 'Improved' and 'Improved' and 'Stable' groups based on EASI. Finally, anchor- and distribution-based analyses demonstrated that a 1-point reduction for the vIGA-AD would be an appropriate criterion to interpret treatment benefit in patients with AD.

IGAs represent a measure of disease severity at a single timepoint and are relatively easy to complete; however, they do not necessarily incorporate extent, which is an important consideration in assessing AD. While the vIGA-AD does suggest that extent be used to differentiate between severity scores in cases where morphology is intermediate between categories, other clinical assessments such as EASI may incorporate this information in greater detail. Some differences do exist between AD-specific clinical measures, and not all morphological descriptors or manifestations of AD are fully captured in any assessment; for example, manifestations of scratching are captured in the EASI but not the vIGA-AD, which is a limitation.

The present results are limited to the study populations assessed here, which included only adult patients with moderate-to-severe AD. Additional studies would be needed to validate the vIGA-AD for use in children and adolescents, and in patients with mild AD. A single-timepoint comparison of measures including the vIGA-AD has been reported in children; the results included a strong correlation between vIGA-AD and BSA, EASI and SCORAD, with even stronger correlations shown for those measures with the multiplied product of vIGA by BSA.²² An additional limitation of our study is that a patient-reported assessment (PGI-S-AD) was used for calculation of meaningful change, while a clinician-reported measure

would arguably be more appropriate given that the measure is a ClinRO. Nevertheless, this use of a patient-reported assessment to assess meaningful change provides the level of change which is meaningful from the patient's perspective. In addition, prior validation work of MCID for clinician-reported assessments of skin inflammation, such as EASI, have used different IGAs in their validation, which could introduce bias when using meaningful changes in EASI to validate those for the vIGA-AD. Bias may also be introduced based on the order in which the disease severity assessments are completed in a study. In the BREEZE-AD programme, investigators were required to complete the vIGA-AD assessment first, before proceeding to assessment of disease severity by EASI and SCORAD. Moreover, while the acknowledged regulatory endpoint is vIGA-AD (0, 1), with at least a 2-point improvement, the results of this study suggest that a smaller change (i.e. a 1-point change from a score of 3 or 4) is meaningful from the patient's perspective.

In summary, the evidence provided demonstrates that the vIGA-AD has sufficient reliability, validity, responsiveness and interpretation standards to be considered a well-defined and reliable clinician-reported instrument. These findings indicate that it is suitable to be used in clinical trials, and to evaluate labelling claims in patients with moderate-to-severe AD.

Acknowledgments

Medical writing support was provided by Thomas Melby, and editorial support was provided by Antonia Baldo, both of Syneos Health, funded by Eli Lilly and Company.

Funding sources

Eli Lilly and Company.

Conflicts of interest

E.L.S. reports grants and fees for participation as a consultant and principal investigator from Eli Lilly and Company, LEO Pharma, Pfizer and Regeneron; grants for participation as a principal investigator from Galderma and Merck & Co.; and fees for consultant services from AbbVie, Boehringer Ingelheim, Dermavant Sciences, Incyte, FortéBio, Pierre Fabre and Sanofi-Genzyme. R.B. is an investigator, consultant, advisory board member, speaker for and/or receives honoraria from AbbVie, Antiobix, Aquinox Pharmaceuticals, Asana, Astellas, Boehringer Ingelheim, Brickell Biotech, Dermavant Sciences, Dermira, Dignity Sciences, Eli Lilly and Company, Galderma, GlaxoSmithKline-Stiefel, Glenmark, Hoffman-LaRoche Ltd, Incyte, Kiniksa, LEO Pharma, Neokera, Pfizer, Regeneron, Relaxer, Sanofi, Sienna and Vitae; and is an employee and shareholder of Innovaderm Research. A.S.P. has been an investigator for AbbVie, AnaptysBio, Eli Lilly and Company, Incyte, LEO Pharma and Regeneron, a consultant with honorarium from AbbVie, Almirall, AnaptysBio, Asana Bio, Boehringer

Ingelheim, Dermavant Sciences, Dermira, Eli Lilly and Company, Forté, Incyte, LEO Pharma, Novartis, Regeneron and Sanofi, and on a data safety monitoring board with fees from Galderma. B.K. has served on advisory boards and/or is a consultant and/or is a clinical trial investigator for AbbVie, Aclaris Therapeutics, AltruBio, Almirall, AnaptysBio, Arena Pharmaceuticals, Bioniz Therapeutics, Bristol Meyers Squibb, Concert Pharmaceuticals, Dermavant Sciences, Horizon Therapeutics, Eli Lilly and Company, Incyte Corp, LEO Pharma, Otsuka/Visterra, Pfizer, Regeneron, Sanofi-Genzyme, TWi Biotechnology and Viela Bio. He is on speaker bureaus for Pfizer, Regeneron and Sanofi-Genzyme. J.I.S. reports honoraria for consultant and advisory board services and for participation as an investigator from Eli Lilly and Company. K.R. has served as advisor and/or paid speaker for and/or participated in clinical trials sponsored by AbbVie, Affibody, Almirall, Amgen, Avillion, Biogen, Boehringer Ingelheim, Bristol Myers Squibb, Celgene, Centocor, Covagen AG, Eli Lilly and Company, Forward Pharma, Fresenius Medical Care, Galapagos, GlaxoSmithKline, Janssen-Cilag, Kyowa Kirin, LEO Pharma, Medac, Merck Sharp & Dohme, Miltenyi Biotec, Novartis, Ocean Pharma, Pfizer, Regeneron, Samsung Bioepis, Sanofi, Sun Pharma, Takeda, UCB, Valeant and Xenoport. J.P.T. is an advisor for AbbVie, Almirall, Arena Pharmaceuticals, Aslan Pharmaceuticals, Coloplast, Eli Lilly & Co, LEO Pharma, OM Pharma, Pfizer, Regeneron, Sanofi-Genzyme and Union Therapeutics; a speaker for AbbVie, Almirall, Eli Lilly & Co, LEO Pharma, Pfizer, Regeneron and Sanofi-Genzyme; and received research grants from Pfizer, Regeneron and Sanofi-Genzyme. H.D. was an employee of Clinical Outcome Solutions, contracted by Eli Lilly and Company. L.S., A.M.DeL. and F.P.N. are employees of and stockholders in Eli Lilly and Company. L.F.E. reports grants and fees for participation as a consultant and/or investigator from AbbVie, Almirall, Aslan, Dermavant, Dermira, Eli Lilly and Company, Forté Biosciences, Galderma, Incyte, LEO Pharma, Novartis, Pfizer, Regeneron and Sanofi-Genzyme; and honoraria and fees from Asana and Glenmark for data safety monitoring board services.

Data availability statement

Lilly provides access to all individual participant data collected during the trial, after anonymization, with the exception of pharmacokinetic or genetic data. Data are available to request 6 months after the indication studied has been approved in the USA and EU and after primary publication acceptance, whichever is later. No expiration date of data requests is currently set once data are made available. Access is provided after a proposal has been approved by an independent review committee identified for this purpose and after receipt of a signed data-sharing agreement. Data and documents, including the study protocol, statistical analysis plan, clinical study report, blank or annotated case report forms, will be provided in a secure data-sharing environment. For details on submitting a request, see the instructions provided at www.vivli.org.

Ethics statement

The clinical studies were performed in compliance with the International Conference on Harmonization Good Clinical Practice Guidelines. All patients provided written informed consent, and institutional review boards or ethics committees approved the study protocol before each study started.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Table S1 Within-group change scores for responsiveness of the vIGA-AD to change on the EASI between baseline and Week 4.

Table S2 Within-group meaningful change by study for vIGA at Week 16 (uncollapsed).

References

- Bieber T. Atopic dermatitis. *N Engl J Med* 2008; **358**:1483–94.
- Paller AS, Kabashima K, Bieber T. Therapeutic pipeline for atopic dermatitis: end of the drought? *J Allergy Clin Immunol* 2017; **140**:633–43.
- Schmitt J, Langan S, Williams HC. What are the best outcome measurements for atopic eczema? A systematic review. *J Allergy Clin Immunol* 2007; **120**:1389–98.
- Schmitt J, Langan S, Deckert S *et al.* Assessment of clinical signs of atopic dermatitis: a systematic review and recommendation. *J Allergy Clin Immunol* 2013; **132**:1337–47.
- Hanifin JM, Thurston M, Omoto M *et al.* The Eczema Area and Severity Index (EASI): assessment of reliability in atopic dermatitis. EASI Evaluator Group. *Exp Dermatol* 2001; **10**:11–18.
- [No authors listed]. Severity scoring of atopic dermatitis: the SCORAD index. Consensus Report of the European Task Force on Atopic Dermatitis. *Dermatology* 1993; **186**:23–31. <https://doi.org/10.1159/000247298>.
- Simpson E, Bissonnette R, Eichenfield LF *et al.* The validated Investigator Global Assessment for Atopic Dermatitis (vIGA-AD): the development and reliability testing of a novel clinical outcome measurement instrument for the severity of atopic dermatitis. *J Am Acad Dermatol* 2020; **83**:839–46.
- Simpson EL, Lacour JP, Spelman L *et al.* Baricitinib in patients with moderate-to-severe atopic dermatitis and inadequate response to topical corticosteroids: results from two randomized monotherapy phase III trials. *Br J Dermatol* 2020; **183**:242–55.
- Charman CR, Venn AJ, Williams HC. The Patient-Oriented Eczema Measure: development and initial validation of a new tool for measuring atopic eczema severity from the patients' perspective. *Arch Dermatol* 2004; **140**:1513–19.
- Finlay AY, Khan GK. Dermatology Life Quality Index (DLQI) – a simple practical measure for routine clinical use. *Clin Exp Dermatol* 1994; **19**:210–16.
- Schram ME, Spuls PI, Leeftang MM *et al.* EASI, (objective) SCORAD and POEM for atopic eczema: responsiveness and minimal clinically important difference. *Allergy* 2012; **67**:99–106.
- Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 1968; **70**:213–20.
- Altman D. *Practical Statistics for Medical Research*. London: Chapman and Hall, 1991.
- Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd edn. Hillsdale: Lawrence Erlbaum Associates, 1988.
- Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951; **16**:297–334.
- Nunnally JC. The assessment of reliability. In: *Psychometric Theory* (Bernstein I, ed.). New York: McGraw Hill, 1994; 248–92.
- Litwin MS. *How to Measure Survey Reliability and Validity*. London: Sage Publications, 1995.
- Clinical Outcomes Solutions. Outcomes Psychometric Summit: Consensus Panel. C-Path PRO Consortium Partner-led Meeting; 2015; Tucson, AZ, USA. See https://www.clinoutsolutions.com/publications_presentations/
- Leshem YA, Hajar T, Hanifin JM, Simpson EL. What the Eczema Area and Severity Index score tells us about the severity of atopic dermatitis: an interpretability study. *Br J Dermatol* 2015; **172**:1353–7.
- Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 2008; **61**:102–9.
- Futamura M, Leshem YA, Thomas KS *et al.* A systematic review of Investigator Global Assessment (IGA) in atopic dermatitis (AD) trials: many options, no standards. *J Am Acad Dermatol* 2016; **74**:288–94.
- Suh TP, Ramachandran D, Patel V *et al.* Product of Investigator Global Assessment and Body Surface Area (IGAxBSA): a practice-friendly alternative to the Eczema Area and Severity Index to assess atopic dermatitis severity in children. *J Am Acad Dermatol* 2020; **82**:1187–94.