



Validation and Interpretation of Short Form 12 and Comparison with Dermatology Life Quality Index in Atopic Dermatitis in Adults

Jonathan I. Silverberg¹, Joel M. Gelfand², David J. Margolis², Mark Boguniewicz³, Luz Fonacier⁴, Mitchell H. Grayson⁵, Peck Y. Ong⁶, Zelma C. Chiesa Fuxench² and Eric L. Simpson⁷

Quality-of-life assessments are not standardized in atopic dermatitis (AD). We sought to determine the validity of the Short Form (SF)-12, a generic quality-of-life assessment, in AD and compare its measurement properties with the Dermatology Life Quality Index (DLQI). A cross-sectional, population-based study of 3,495 adults was performed, including 602 adults who met the modified United Kingdom Working Party Criteria for AD. The SF-12 mental component score and the SF-Six Dimension (SF-6D) had a strong correlation with each other and moderate inverse correlations with the Patient-Oriented Eczema Measure, the Patient-Oriented Scoring Atopic Dermatitis, the Patient-Oriented Scoring Atopic Dermatitis–itch, the Patient-Oriented Scoring Atopic Dermatitis–sleep, and the Numerical Rating Scale of pain (Pearson correlations, $P < 0.0001$ for all). The SF-12 mental component score and the SF-6D showed good discriminant validity as judged by the analysis of variance and receiver operating curves. The SF-12 physical component score had weak correlations with AD severity assessments and poor discriminant validity. The DLQI had better convergent and discriminant validity than the SF-12. The SF-12 and the DLQI showed good internal consistency (Cronbach alpha, 0.89 and 0.94, respectively). Differential item functioning was found for items in the SF-12 and the DLQI. Floor effects were observed for the DLQI but not for the SF-12 mental component score, the SF-12 physical component score, and the SF-6D. Severity thresholds were selected. In conclusion, the SF-12 mental component score and the SF-6D showed good validity in AD but inferior construct validity compared with the DLQI.

Journal of Investigative Dermatology (2019) **139**, 2090–2097; doi:10.1016/j.jid.2019.03.1152

INTRODUCTION

Atopic dermatitis (AD) is a chronic inflammatory skin disease affecting approximately 13% of children (Silverberg and Simpson, 2013, 2014) and 7% of adults in the United States (Hua and Silverberg, 2018; Silverberg et al., 2018b). AD is associated with a wide array of quality-of-life (QOL) impacts, including itch and skin pain, symptoms of anxiety and depression, and impaired physical and social functioning (Brunner et al., 2017; Silverberg et al., 2014, Silverberg et al.,

2018a; Silverberg et al., 2018b; Silverberg et al., 2018c; Silverberg et al., 2019a; Silverberg et al., 2019b; Simpson et al., 2016; Vakharia et al., 2017). All of these can have a profound impact on the overall health of patients with AD. QOL impact should be incorporated into therapeutic decision making for patients with AD (Eichenfield et al., 2014; Simpson et al., 2017). However, there is a lack of specific guidance or standardization as to which QOL assessments should be used in clinical practice or trials (Chalmers et al., 2014, 2016; Gerbens et al., 2016, 2017; Heintz et al., 2016a, 2016b, 2017; Schmitt and Williams, 2010).

Several standardized QOL assessments may provide relevant information about the burden of AD. Patient burden of AD is multifactorial and cannot be assessed by the signs and symptoms of the disease alone (Haack et al., 2012). There are different approaches to assess QOL impairment in AD, including AD-specific, skin-specific, or generic health-related QOL instruments, each with its strengths and weaknesses. The Harmonising Outcome Measures for Eczema (HOME) international consensus group deemed that no QOL instruments met their criteria for recommendation to be included in the AD randomized controlled trial core outcome set (Heintz et al., 2016b, 2017) but that the Dermatology Life Quality Index (DLQI) and the Quality of Life Index for Atopic Dermatitis (QoLIAD) were candidates in adults (Heintz et al., 2016b). The QoLIAD is an example of an AD-specific assessment related to the impact of AD on patient needs (Whalley et al., 2004). The DLQI assesses QOL across dermatological diseases in general (Basra et al., 2008) and

¹Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA; ²University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA; ³National Jewish Health and University of Colorado School of Medicine, Denver, Colorado, USA; ⁴NYU Winthrop Hospital, Mineola, New York, USA; ⁵Nationwide Children's Hospital—The Ohio State University College of Medicine, Columbus, Ohio, USA; ⁶Children's Hospital Los Angeles, University of Southern California, Keck School of Medicine, Los Angeles, California, USA; and ⁷Oregon Health & Science University, Portland, Oregon, USA

Correspondence: Jonathan I. Silverberg, Northwestern University Feinberg School of Medicine, Department of Dermatology, Suite 1600, 676 N. St Clair St., Chicago, IL 60611, USA. E-mail: JonathanSilverberg@gmail.com

Abbreviations: AD, atopic dermatitis; AUC, area under the curve; DIF, differential item functioning; DLQI, Dermatology Life Quality Index; HOME, Harmonising Outcome Measures for Eczema; MCS, mental component score; PCS, physical component score; POEM, Patient-Oriented Eczema Measure; PO-SCORAD, Patient-Oriented Scoring Atopic Dermatitis; QOL, quality of life; QoLIAD, Quality of Life Index for Atopic Dermatitis; SF, Short Form; SF-6D, SF-Six Dimension

Received 6 January 2019; revised 7 March 2019; accepted 25 March 2019; accepted manuscript published online 19 April 2019; corrected proof published online 27 June 2019

is the most frequently used QOL instrument in AD studies (Blome et al., 2016; Chalmers et al., 2016). Although both appear to be valid for assessing AD, they cannot be used to assess QOL impairment in any other disorder (QoLIAD) or outside dermatology (DLQI).

The Short Form (SF)-12 is a generic, multidimensional QOL assessment that is not specific to skin disease and has been studied extensively in many medical disorders (Jenkinson and Layte, 1997; Ware et al., 1996). Although the SF-12 is not specific to AD or skin disease, it captures many QOL domains affected by AD. The SF-12 was not previously validated in AD *per se*; thus, it was not considered by the HOME group because of a lack of existing validation studies. In this study, we sought to determine the construct validity, internal consistency, differential item functioning (DIF), floor or ceiling effects, and interpretation of the SF-12 and compare its measurement properties with the DLQI in US adults with AD.

RESULTS

Patient characteristics

Overall, 2,893 adults were included in this study; 602 met AD criteria (prevalence [95% confidence interval] = 7.39% [5.81%–8.97%]) (Supplementary Figure S1, online). Of these 602 participants with AD, 53.6% were female and 71.9% were Caucasian/white, with mean age $\pm$ standard deviation of 52.0 ± 16.3 years. Among persons with AD, the weighted mean score [95% confidence interval] was 27.5 [25.7–29.3] for the Patient-Oriented Scoring Atopic Dermatitis (PO-SCORAD) index, 7.5 [6.8–8.1] for the Patient-Oriented Eczema Measure (POEM), 45.9 [45.0–46.9] for the SF-12 mental component score (MCS), 53.0 [52.8–53.3] for the SF-12 physical component score (PCS), 0.69 [0.68–0.71] for the SF-Six Dimension (SF-6D), and 4.9 [4.2–5.5] for the DLQI.

Construct validity

Regarding convergent validity, the SF-12 MCS and the SF-6D had strong correlations with each other, and moderate inverse correlations with the POEM, PO-SCORAD, PO-SCORAD–itch, and PO-SCORAD–sleep (Pearson correlations, $P < 0.0001$) (Table 1). The SF-12 PCS had a moderate correlation with the SF-6D and weak correlations with the SF-12 MCS, POEM, PO-SCORAD, PO-SCORAD–itch, PO-SCORAD–sleep, and the Numerical Rating Scale for pain. The DLQI had stronger correlations with the POEM, PO-SCORAD, PO-SCORAD–itch, and the Numerical Rating Scale for pain than with the SF-12 MCS, SF-12 PCS, or SF-6D (z-test, $P < 0.001$ for all).

Regarding discriminative validity, there were significant decreases in the SF-12 MCS and the SF-6D scores, and increases in DLQI scores, with each increasing level of severity for self-reported global AD severity, POEM, PO-SCORAD, PO-SCORAD–itch and PO-SCORAD–sleep, and the Numerical Rating Scale for pain (analysis of variance, $P < 0.0001$ for all) but no significant differences in the SF-12 MCS (Figure 1).

The SF-12 MCS, SF-12 PCS, and SF-6D scores also showed good discriminant validity in predicting overall AD severity and the extent to which AD limits lifestyle (Table 2 and Supplementary Figure S2 online). The area under the curve

(AUC) for the SF-12 MCS and the SF-6D was good for distinguishing severe versus mild disease and poor for distinguishing moderate versus mild or severe versus moderate AD, indicating that they had some ability to distinguish between severity groups. In addition, the AUC for the SF-12 MCS and the SF-6D was good to excellent for distinguishing between quite a bit or a great deal versus not at all for AD limiting lifestyle. Furthermore, the SF-12 PCS had good discrimination between a great deal versus not at all for AD limiting lifestyle but failed to discriminate between AD severity groups.

In contrast, the AUC for the DLQI was excellent for distinguishing severe versus mild disease, and good for distinguishing moderate versus mild or severe versus moderate AD. In addition, the AUC for the DLQI was higher for all other comparisons than the SF-12 MCS, SF-12 PCS, or SF-6D, indicating that the DLQI had better ability to distinguish between severity groups.

Internal consistency

The SF-12 items showed good internal consistency among adults both with and without AD (Cronbach alpha, 0.89 and 0.89, respectively), with weak to moderate interitem correlations (Supplementary Table S1 online). Similarly, the DLQI showed very good internal consistency among adults with AD (0.94), with moderate to strong interitem correlations.

Differential item functioning

Significant uniform and nonuniform DIF was found for multiple SF-12 items between AD versus non-AD. Among respondents with AD, uniform and nonuniform DIF was found for multiple SF-12 and DLQI items by sex, race/ethnicity, and age (Supplementary Table S2 online).

Floor or ceiling effects

The proportions of patients with the lowest and highest values for the SF-12 MCS (0.2% or 1.2%), SF-12 PCS (0.2% or 0.3%), and SF-6D (0.4% or 9.2%) were below 15%, indicating there were no floor or ceiling effects overall. However, >15% of patients reported the lowest or highest value for individual SF-12 questions 2, 3, 8, 11, and 12; Q4, Q5, Q6, and Q7 have binary response options (Supplementary Table S3 online). In contrast, the DLQI showed floor effects (17.6%) but not ceiling effects (1.2%) overall, with >15% of patients reporting the lowest values for individual DLQI items.

Interpretation

Severity thresholds were selected based on maximizing the concordance probability in receiver operating characteristic analysis. The optimal thresholds for the SF-12 MCS were 46.6 for moderate AD (sensitivity, 60.2%; specificity, 62.6%) and 42.1 for severe AD (sensitivity, 70.6%; specificity, 56.7%) (Table 3), whereas the optimal thresholds for the SF-6D were 0.67 for moderate AD (sensitivity, 67.3%; specificity, 66.7%) and 0.61 for severe AD (sensitivity, 69.2%; specificity, 52.8%). In addition, the thresholds for the SF-12 and the SF-6D ranged from 51.5 and 0.66, respectively, for slight to 38.5 and 0.48, respectively, for a great deal of AD limiting lifestyle, respectively. The thresholds identified for the DLQI were similar to those previously reported (Rogers et al., 2012; Vakharia et al., 2018a, 2018b).

Table 1. Pearson Correlation of SF-12 Mental and Physical Component Scores and SF-6D with AD Symptoms and Severity and DLQI

QOL Assessment	Pearson r							
	POEM	PO-SCORAD	PO-SCORAD-Itch	PO-SCORAD-Sleep	NRS-Pain	SF-12 MCS	SF-12 PCS	SF-6D
SF-12 MCS	−0.32 ¹	−0.47 ¹	−0.34 ¹	−0.50 ¹	−0.42 ¹	1.00	—	—
SF-12 PCS	−0.15 ¹	−0.12 ¹	−0.14 ¹	−0.20 ¹	−0.20 ¹	0.21 ¹	1.00	—
SF-6D	−0.30 ¹	−0.44 ¹	−0.31 ¹	−0.41 ¹	−0.44 ¹	0.79 ¹	0.45 ¹	1.00
DLQI	0.62 ¹	0.71 ¹	0.48 ¹	0.43 ¹	0.67 ¹	−0.44 ¹	−0.06	−0.42 ¹

Boldface indicates statistical significance.

Abbreviations: AD, atopic dermatitis; DLQI, Dermatology Life Quality Index; MCS, mental component score; NRS, Numerical Rating Scale; PCS, physical component score; POEM, Patient-Oriented Eczema Measure; PO-SCORAD, Patient-Oriented Scoring Atopic Dermatitis; SF, Short Form; QOL, quality of life.

¹P < 0.0001.

DISCUSSION

This study demonstrated that the SF-12 MCS and the SF-6D had convergent validity, discriminative validity for AD severity and QOL impact, and good internal consistency, with no observed floor or ceiling effects for total scores. In contrast, the DLQI correlated better with AD severity. The DLQI had better discriminative validity than the SF-12 MCS or the SF-6D, particularly at distinguishing between moderate versus mild and severe versus moderate AD. However, floor effects were observed for total DLQI scores, which may be an indication that it is not comprehensive. Multiple items from the SF-12 and the DLQI had floor effects, which may be related to individual baseline characteristics and not lack of conceptual validity in AD. Individual baseline characteristics were previously demonstrated to impact the burden of itch (Silverberg et al., 2018d). For example, patients who are usually sedentary and/or do not participate in sports are unlikely to report that itch limits these activities in the SF-12 or the DLQI. This likely occurs for items in the SF-12 and the DLQI in patients with AD as well.

Although the DLQI better reflects the burden of AD, as would be expected with a skin-specific instrument, the SF-12 offers several distinct advantages over the DLQI. The SF-12 is a generic QOL instrument that has been validated and widely used across a multitude of health conditions, whereas the DLQI only measures skin-related QOL. The results of generic QOL instruments allow comparison across disease states. SF-12 items can be used to estimate the MCS and PCS, which are comparable to the US population with a mean score of 0.5 and standard deviation of 0.1. Furthermore, SF-12 items can be used to estimate the SF-6D health utility score, which in turn can be used to estimate quality-adjusted life years lost from a disorder and the cost-effectiveness of interventions. However, the SF-12 has complex scoring approaches using different weights to estimate the MCS, PCS, and SF-6D. Hand scoring may not be feasible in clinical practice, although electronic scoring can easily be incorporated into most clinical and research settings. We did not formally assess the time to complete the SF-12 and the DLQI in this study. However, previous studies found the median time to completion was 1–3 minutes, with most or all completed in ≤5 minutes (Elman et al., 2010; Hahn et al., 2001; Patel et al., 2018; Ware et al., 1998). Both instruments would be considered “acceptable” or “adequate” for use in clinical practice (Schmitt et al., 2007). The results suggest that the

SF-12 has sufficient validity to be used as an assessment of burden in AD. Future studies should validate the SF-12 for the study of AD in clinical and other settings. We did not compare the measurement properties of the SF-12 against other generic QOL assessments, such as EuroQOL Five Dimensions (Brooks, 1996; EuroQol Group, 1990). Future studies should compare the measurement properties of other generic QOL assessments and determine whether they outperform the SF-12 in AD.

The SF-12 and the DLQI had multiple items that suffered from uniform and/or nonuniform DIF between those with AD versus those without AD, as well as age, sex, and race/ethnicity within the AD population. Items with DIF may perform differently across various demographic subgroups. Thus, scores for these different patient groups should neither be combined nor compared during analysis of clinical trials or research studies.

These results are consistent with previous findings of strong correlations of the SF-12 with SCORAD (Coutanceau and Stalder, 2014), POEM, and Numerical Rating Scale for itch (Patel et al., 2018); moderate correlations with the Eczema Area and Severity Index (Patel et al., 2018), multiple SF-36 domains, and the objective-SCORAD (Holm et al., 2006); and moderate inverse correlations with the Patient Generated Index (Herd et al., 1997), but only low-to-moderate correlations with the Nottingham Health Profile domains (Spearman correlation coefficients ranged from 0.12 to 0.32) (Badia et al., 1999) in AD. The DLQI was found to be internally consistent in AD (Badia et al., 1999; Patel et al., 2018).

QOL is considered a core domain for clinical trials in AD by the HOME international consensus group (Schmitt et al., 2011). The HOME group deemed that no QOL instruments met their criteria for recommendation to be included in AD trials (Heinl et al., 2016b, 2017) but that the DLQI and the QoLIAD were candidates in adults (Heinl et al., 2016b). To reduce respondent burden, we did not assess both the DLQI and the QoLIAD. We selected the DLQI because it is the most commonly used QOL assessment in dermatology overall and clinical trials in AD (Blome et al., 2016; Chalmers et al., 2016) and may be more practical for use in clinical practice. This study provides further support of the convergent and discriminative validity, consistency, and interpretation of the DLQI in AD. Further studies are needed to validate the QoLIAD and compare its measurement properties with

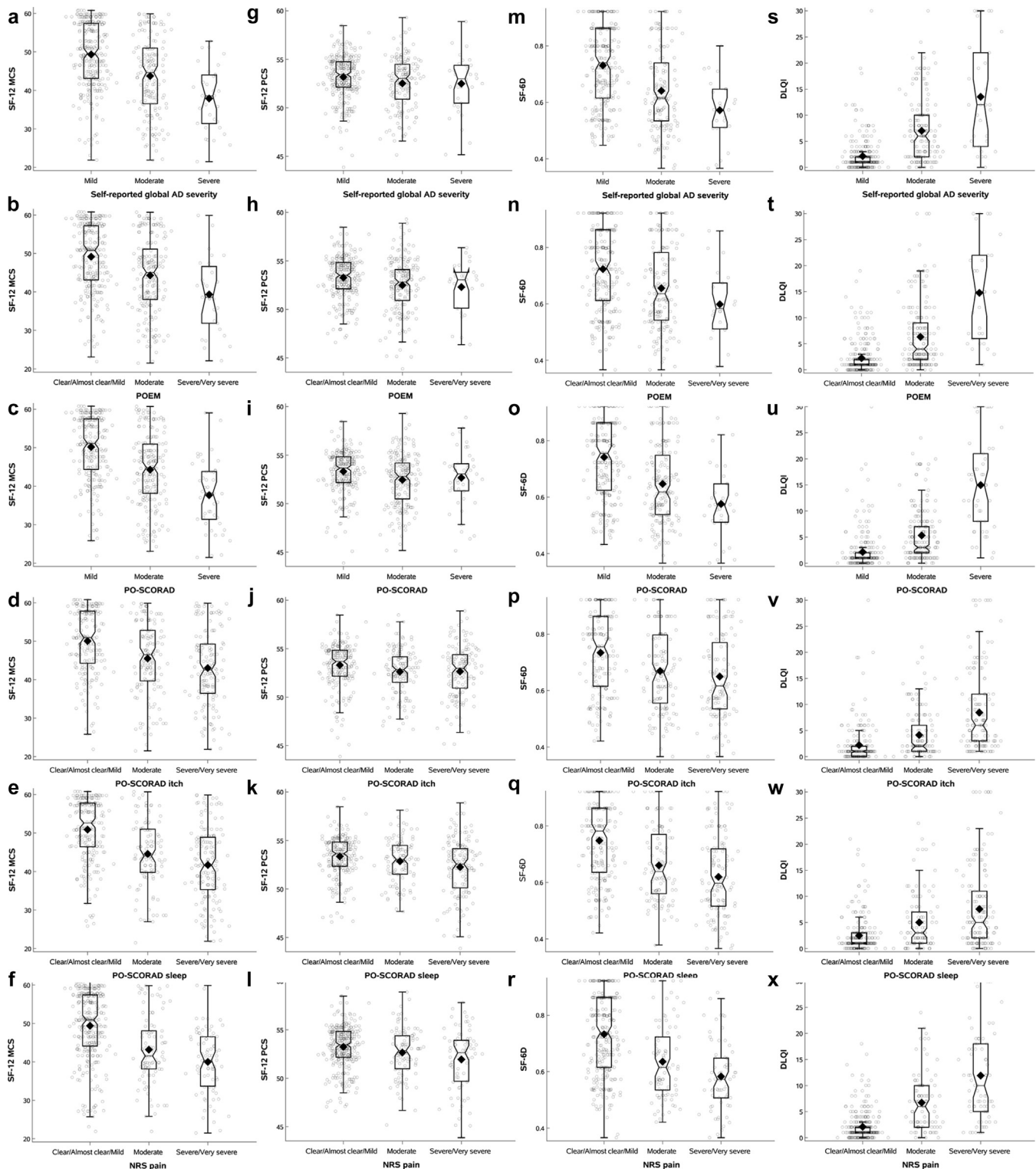


Figure 1. Discriminant validity of SF-12 scores. (a–f) Box-whisker plots and overlaid jitter plots of SF-12 MCS, (g–l) PCS, (m–r) SF-6D, and (s–x) DLQI scores are presented for each severity level of self-reported global AD severity (a, g, m, and s), POEM (b, h, n, and t), PO-SCORAD (c, i, o, and u), PO-SCORAD itch (d, j, p, and v), PO-SCORAD sleep (e, k, q, and w), and NRS worst pain (f, l, r, and x). AD, atopic dermatitis; DLQI, Dermatology Life Quality Index; MCS, mental component score; NRS, Numerical Rating Scale; PCS, physical component score; POEM, Patient-Oriented Eczema Measure; PO-SCORAD, Patient-Oriented Scoring Atopic Dermatitis; SF, Short Form.

other QOL assessments. We also found that the SF-12 has sufficient validity and may be considered as an additional option for assessing QOL impact in patients with AD when a generic QOL instrument is needed or deemed more feasible.

This study has several strengths, including use of a large, nationally representative cohort (as previously described in [Chiesa Fuxench et al., 2019](#)) and good representation across gender, race/ethnicity, and AD severity. However, there are

Table 2. Area Under the ROC Curve for SF-12 MCS, PCS, SF-6D, and DLQI as Predictors of Quality of Life Impact and Severity of AD

Groups	AUC, P-Value			
	SF-12 MCS	SF-12 PCS	SF-6D	DLQI
<i>Global severity</i>				
Moderate vs. mild	0.67³	0.57²	0.68³	0.79³
Severe vs. mild	0.83³	0.43	0.81³	0.90³
Severe vs. moderate	0.67¹	0.50	0.62¹	0.70³
<i>AD limits lifestyle</i>				
Slightly vs. not at all	0.63³	0.58¹	0.66³	0.78³
Moderately vs. not at all	0.78³	0.63¹	0.76³	0.94³
Quite a bit vs. not at all	0.85³	0.48	0.82³	0.94³
A great deal vs. not at all	0.88³	0.81³	0.91³	0.99²

Boldface indicates statistical significance.

Abbreviations: AD, atopic dermatitis; AUC, area under the curve; DLQI, Dermatology Life Quality Index; MCS, mental component score; PCS, physical component score; ROC, receiver operating characteristic; SF, Short Form.

¹P < 0.05.

²P < 0.01.

³P < 0.0001.

some limitations. AD classification was not confirmed by physician diagnosis. We did not assess responsiveness, test-retest reliability, or content validity, although we expect the SF-12 to have good content validity in AD. Although self-reported global AD severity was previously validated (Silverberg et al., 2018a), the assessment of the impact of AD on lifestyle was not. This study was limited to adults. Future studies are needed to validate generic QOL assessments in pediatric AD.

In conclusion, the SF-12 MCS and the SF-6D have good construct validity, with no overall floor or ceiling effects. The SF-12 MCS and the SF-6D may be useful for the assessment of the burden of AD in clinical trials and practice.

METHODS

Data source

Data were obtained from the AD in America survey whose population was sampled from the long-standing GfK Knowledge Panel.

The AD in America study was designed to examine multiple facets of the patient burden, comorbidities, and resource utilization of US adults with AD. The GfK Knowledge Panel is the largest and oldest probability-based web panel in the US and contains between 40,000 and 50,000 adult panel members at any given time. The GfK web panel was initially constructed from a national address-based sample of households in the US who were recruited to join in and receive small incentives for participating in web surveys on a regular basis. This approach uses a single sampling frame via the Delivery Sequence File of the US Postal Service to provide a statistically valid representation of the US population, as well as many difficult-to-survey populations. The GfK web panel also provides internet access to households without existing internet access. This web-based panel has been previously used in other large, epidemiologic studies and has been shown to be representative of the US population (Chiesa Fuxench et al., 2019; Hays et al., 2015; Heckman et al., 2013; Rowen et al., 2016; Silverberg et al., 2018b). The survey questionnaire and protocol were approved by the ICF Institutional Review Board.

Study design

This was a cross-sectional study involving a two-stage sampling process. Stage 1 was designed to determine the prevalence of AD in US adults. In this stage, an initial cross-sectional sample of 2,137 adults from the existing GfK Knowledge web panel was invited to participate in the survey. The focus of the survey was not disclosed in the invitation to members of the web panel to avoid biasing participation based on respondent interest or disinterest in the subject. A total of 1,278 adults completed the survey (completion rate, 60.2%). Although this sample provided a precise estimate of the prevalence of AD among the adult population, it did not yield a large enough sample of patients with AD to investigate differences between different levels of disease severity. In stage 2, an additional sample of 13,713 adults from the GfK knowledge panel were invited, and 8,217 adults completed screening (completion rate, 59.9%) to identify an additional group of 509 adults (6.2% prevalence) with AD and 1,106 without AD (Supplementary Figure S1 online). Using data from the US Census Bureau, sample weights were created that adjusted for age, gender, race, ethnicity, education level, census region, household income, home ownership status, and metropolitan area using an iterative proportional fitting

Table 3. Optimal Cut Points and their Sensitivity and Specificity for SF-12 MCS, SF-6D, and DLQI as Predictors of Quality of Life Impact and Severity of AD

Groups	SF-12 MCS		SF-6D		DLQI	
	Threshold	Sensitivity, %	Specificity, %	Threshold	Sensitivity, %	Specificity, %
<i>Global severity</i>						
Moderate vs. mild	46.6	60.2	62.6	0.67	67.3	66.7
Severe vs. moderate	42.1	70.6	56.7	0.61	69.2	52.8
<i>AD limits lifestyle</i>						
Slightly vs. not at all	51.5	76.6	47.5	0.66	59.4	69.4
Moderately vs. slightly	46.6	79.2	48.4	0.57	56.3	70.9
Quite a bit vs. moderately	41.6	70.4	51.4	0.52	40.0	76.6
A great deal vs. quite a bit	38.5	69.2	55.6	0.48	62.5	80.0

Abbreviations: AD, atopic dermatitis; DLQI, Dermatology Life Quality Index; MCS, mental component score; SF, Short Form.

procedure. Sample weights were included in all analyses to allow for representative estimates of the US population.

AD, AD severity, and atopic comorbidities

An adaptation of the United Kingdom Working Party criteria was selected by the AD in America advisory committee as the screening tool for patient eligibility (Williams et al., 1994). This included all aspects of the United Kingdom Working Party criteria of having an itchy skin condition during the past 12 months and three or more of the following: (1) a history of skin crease involvement; (2) a personal history of asthma or hay fever; (3) a history of general dry skin during the past year; and (4) onset under the age of 2 years, except assessment of visible flexural eczema by a clinician was not performed. Self-reported lifetime history of asthma or hay fever and 1-year history of food allergy were assessed.

Self-assessments of AD severity included the previously validated self-reported global AD severity question "Would you describe your atopic dermatitis or eczema as mild, moderate, or severe?" (Vakharia et al., 2018c), as well as the PO-SCORAD index (range, 0–103; 3-day recall period), PO-SCORAD itch and sleep subscores (range, 0–10; 3-day recall period) (Stalder et al., 2011), and POEM (7 questions; range, 0–28; 1-week recall period) (Charman et al., 2013). A higher score for all of these indicates more severe AD. These patient-reported outcomes were selected a priori based on their feasibility, being preferred measures by international consensus of the HOME group (Chalmers et al., 2018; Spuls et al., 2017), and/or previously being validated (Chalmers et al., 2018; Silverberg et al., 2018; Spuls et al., 2017; Vakharia et al., 2018c). Scores were divided into three categories (clear/almost clear/mild, moderate, and severe/very severe) using previously reported severity strata (Charman et al., 2013; Rogers et al., 2012).

QOL assessments

The SF-12 Health Survey version 1.0 was used to assess health status (12 items, 8 domains, 4-week recall period) (Jenkinson and Layte, 1997; Ware et al., 1996). Different weights are applied to each of the items to estimate the MCS and the PCS, which are comparable to the US population with a mean score of 0.5 and a standard deviation of 0.1. Lower scores indicate poorer overall health. The SF-6D is a preference-based single index measure for health that was derived from responses to the SF-12 using the Brazier scoring method and US population-based weights (Brazier et al., 2002). SF-6D scores range from 0.0 (worst health) to 1.0 (best health). The DLQI was used to assess dermatology-related QOL (10 items; range, 0–30; 1-week recall period) (Blome et al., 2016; Chalmers et al., 2016). A higher score indicates poorer health.

Statistical analysis

Summary statistics were estimated for baseline population characteristics. Convergent construct validity of the SF-12 MCS and PCS and the SF-6D were established using Pearson correlations. Correlation coefficients scores of ≥ 0.70 or ≤ -0.70 were considered strong, 0.50–0.69 or -0.69 to -0.50 moderate, and 0.30–0.49 or -0.49 to -0.30 weak. Discriminant construct validity was determined using Wilcoxon rank sum tests. Monotonic and statistically significant between-group differences support discriminant validity. Analyses were done as per the COnsensus-based Standards for the selection of health status Measurement INstruments criteria (Mokkink et al., 2010). In addition, logistic regression models were constructed with self-reported global AD severity (moderate vs. mild, severe vs. moderate) and the extent to which AD limits lifestyle

as the binary dependent variables (slightly vs. not at all, moderately vs. slightly, quite a bit vs. moderately, a great deal vs. quite a bit). The independent variables were SF-12 MCS, SF-12 PCS, or SF-6D scores. Receiver operating characteristic curves were employed to determine optimal thresholds among severity groups. Separate receiver operating characteristic curves were developed for each severity cut point. Discriminant validity was determined using AUC. AUC scores of ≥ 0.90 were considered excellent, 0.80–0.89 good, 0.70–0.79 fair, <0.70 poor, and <0.60 fail (Metz, 1978). Optimal thresholds for each receiver operating characteristic curve were determined using the maximum conditional probability function c , which is equal to the product of sensitivity and specificity. Sensitivity and specificity are the probabilities of correctly being assigned and not being assigned to a particular severity group, respectively. A previous simulation study showed that the concordance probability method outperformed other commonly used methods (Rota and Antolini, 2014).

Internal consistency was determined using Cronbach alpha and Pearson correlation between individual questions. Internal consistency measures the extent to which questions in the assessment measure the same construct. Cronbach alpha <0.70 or >0.95 indicate inadequate internal consistency.

Floor or ceiling effects of total scores and individual items were considered present if 15% of responses fell in the lowest or highest scores (Lim et al., 2015; Terwee et al., 2007). The above statistical analyses were performed in SAS version 9.4.3, and sample weights were incorporated into all analyses (SAS Institute, Cary, IN). Complete case analysis was performed. A two-sided P -value of 0.05 was considered statistically significant.

DIF was assessed across all respondents by AD versus no AD, and across patients with AD by sex (male/female), race (white/black/Asian/other), and age ($<50/\geq 50$ years). DIF occurs when respondents from different groups have differing probabilities of success on an item, after controlling for overall ability based on their total scores. If DIF is present, then item responses may be related to both the underlying trait being measured as well as other factors. DIF was detected for items with multiple response options, using an iterative hybrid of ordinal logistic regression and item response theory (Choi et al., 2011). Uniform and nonuniform DIF are present when differences in the probabilities of answering an item are or are not constant across ability levels, respectively. Uniform and/or nonuniform DIF were indicated by P -values of <0.05 for each question (lordif-package, R version 3.5.1, Vienna, Austria) (Choi et al., 2011).

Data availability statement

Data are available upon request from the Allergy and Asthma Foundation of America.

ORCIDs

Joel M. Gelfand: <https://orcid.org/0000-0003-3480-2661>
Mitchell H. Grayson: <https://orcid.org/0000-0002-4673-2122>
Luz Fonacier: <https://orcid.org/0000-0002-4381-1861>
David J. Margolis: <https://orcid.org/0000-0002-0506-8085>
Peck Y. Ong: <https://orcid.org/0000-0002-6664-463X>
Eric L. Simpson: <https://orcid.org/0000-0002-8793-7087>
Jonathan I. Silverberg: <https://orcid.org/0000-0003-3686-7805>
Mark Boguniewicz: <https://orcid.org/0000-0002-1015-4439>
Zelma C. Chiesa Fuxench: <https://orcid.org/0000-0003-0023-4176>

CONFLICT OF INTEREST

JIS served as a consultant and/or advisory board member for Abbvie, AnaplysBio, Asana, Arena, Boehringer-Ingelheim, Dermavant, Eli Lilly, Galderma, GlaxoSmithKline, Glenmark, Kiniksa, Leo, Menlo, Novartis, Pfizer,

Regeneron-Sanofi, and Realm, receiving honoraria; was a speaker for Regeneron-Sanofi; and received research grants from GlaxoSmithKline. JMG served as a consultant for BMS, Boehringer Ingelheim, GlaxoSmithKline, Janssen Biologics, Novartis Corp, UCB (DSMB), Sanofi, and Pfizer Inc, receiving honoraria; receives research grants (to the Trustees of the University of Pennsylvania) from Abbvie, Janssen, Novartis Corp, Celgene, Ortho Dermatologics, and Pfizer Inc; and received payment for continuing medical education work related to psoriasis that was supported indirectly by Lilly, Ortho Dermatologics and Novartis. JMG is a co-patent holder of resiquimod for the treatment of cutaneous T-cell lymphoma. JMG is a Deputy Editor for the *Journal of Investigative Dermatology* receiving honoraria from the Society for Investigative Dermatology. ZCCF has served as a consultant for the National Eczema Association and the Asthma and Allergy Foundation, receiving honoraria; receives or has received research grants (to the Trustees of the University of Pennsylvania) from Regeneron, Sanofi, Tioga and Vanda and Menlo pharmaceuticals and Realm Therapeutics for work in atopic dermatitis; and has received payment for continuing medical education work related to atopic dermatitis that was supported indirectly by Regeneron and/or Sanofi. DJM is the chair of the data monitoring committee for all Sanofi clinical trials of dupilumab and has received independent research funding to his institution from the National Institute of Health and Valeant to study atopic dermatitis. He is a consultant for Pfizer. MB has received research funding from Regeneron and consulted for Regeneron, Sanofi-Genzyme, and Eli Lilly. LF has served as consultant and advisory board member for Regeneron, receiving honoraria; and received research and educational grants from Genentech, Regeneron, Baxter, and Pfizer. MHG is a board member of AAFA and chair for the AAFA Medical Scientific Council and has served on advisory boards for AstraZeneca, Genentech, and Novartis. ELS is an investigator for Eli Lilly Co, Galderma, Leo Pharmaceutical Co, Merck, Pfizer, and Regeneron Pharmaceuticals, Inc and a consultant with honorarium for AbbVie Inc, Boehringer-Ingelheim, Dermavant, Eli Lilly Co, Incyte, Leo Pharmaceutical Co, Pfizer, Pierre Fabre Dermo Cosmetique, Regeneron Pharmaceuticals, Inc, and Sanofi Genzyme. PYO is a co-investigator of Atopic Dermatitis Research Network and has consulted for Pfizer and Theravance and has received research funding from Regeneron.

ACKNOWLEDGMENTS

Atopic Dermatitis in America Study is an independent research project of the Asthma and Allergy Foundation of America in partnership with the National Eczema Association (NEA) and sponsored by Sanofi Genzyme and Regeneron.

Disclaimer

The sponsor had no role in the design and conduct of the study; the collection, management, analysis, and interpretation of data; the preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication.

AUTHOR CONTRIBUTIONS

Conceptualization: JIS; Data Curation: JIS, JMG, DJM, LF, MB, ELS, MHG, PYO, ZCCF; Formal Analysis: JIS; Funding Acquisition: Allergy and Asthma Foundation of America; Investigation: JIS, JMG, DJM, LF, MB, ELS, MHG, PYO, ZCCF; Methodology: JIS, JMG, DJM, LF, MB, ELS, MHG, PYO, ZCCF; Project Administration: Allergy and Asthma Foundation of America; Resources: Allergy and Asthma Foundation of America; Writing - Original Draft Preparation: JIS; Writing - Review and Editing: JIS, JMG, DJM, LF, MB, ELS, MHG, PYO, ZCCF

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at www.jidonline.org, and at <https://doi.org/10.1016/j.jid.2019.03.1152>.

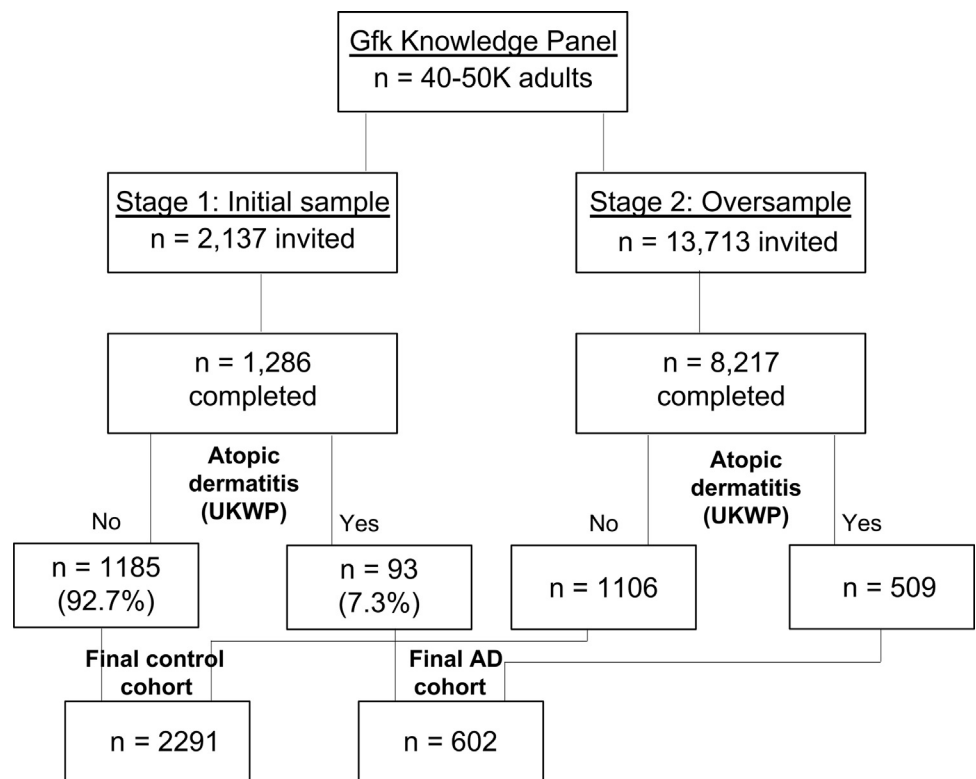
REFERENCES

- Badia X, Mascaró JM, Lozano R. Measuring health-related quality of life in patients with mild to moderate eczema and psoriasis: clinical validity, reliability and sensitivity to change of the DLQI. *The Cavide Research Group*. *Br J Dermatol* 1999;141:698–702.
- Basra MK, Fenech R, Gatt RM, Salek MS, Finlay AY. The Dermatology Life Quality Index 1994–2007: a comprehensive review of validation data and clinical results. *Br J Dermatol* 2008;159:997–1035.
- Blome C, Radtke MA, Eissing L, Augustin M. Quality of life in patients with atopic dermatitis: disease burden, measurement, and treatment benefit. *Am J Clin Dermatol* 2016;17:163–9.

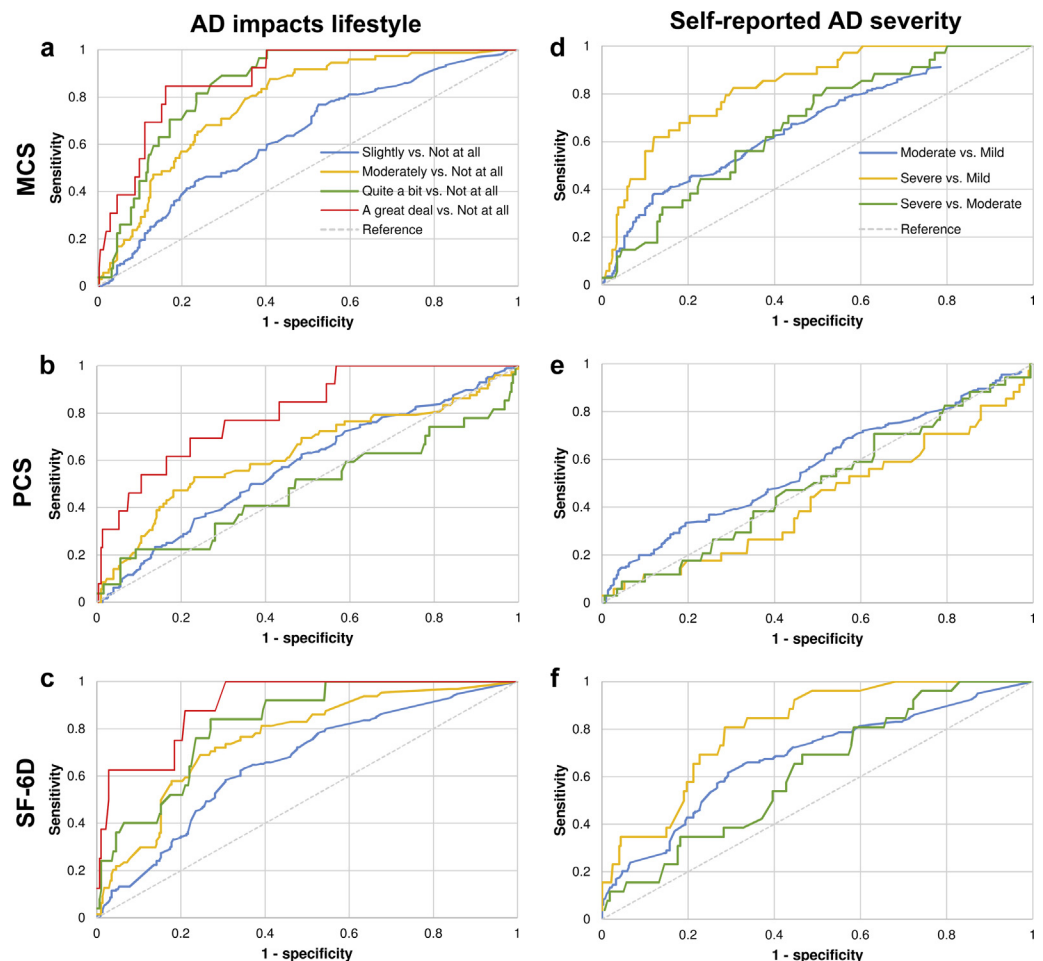
- Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ* 2002;21:271–92.
- Brooks R. EuroQol: the current state of play. *Health Policy* 1996;37:53–72.
- Brunner PM, Silverberg JL, Guttman-Yassky E, Paller AS, Kabashima K, Amagai M, et al. Increasing comorbidities suggest that atopic dermatitis is a systemic disorder. *J Invest Dermatol* 2017;137:18–25.
- Chalmers JR, Schmitt J, Apfelbacher C, Dohil M, Eichenfield LF, Simpson EL, et al. Report from the third international consensus meeting to harmonise core outcome measures for atopic eczema/dermatitis clinical trials (HOME). *Br J Dermatol* 2014;171:1318–25.
- Chalmers JR, Simpson E, Apfelbacher CJ, Thomas KS, von Kobyletzki L, Schmitt J, et al. Report from the fourth international consensus meeting to harmonize core outcome measures for atopic eczema/dermatitis clinical trials (HOME initiative). *Br J Dermatol* 2016;175:69–79.
- Chalmers JR, Thomas KS, Apfelbacher C, Williams HC, Prinsen CA, Spuls PI, et al. Report from the fifth international consensus meeting to harmonize core outcome measures for atopic eczema/dermatitis clinical trials (HOME initiative). *Br J Dermatol* 2018;178:e332–41.
- Charman CR, Venn AJ, Ravenscroft JC, Williams HC. Translating Patient-Oriented Eczema Measure (POEM) scores into clinical practice by suggesting severity strata derived using anchor-based methods. *Br J Dermatol* 2013;169:1326–32.
- Chiesa Fuxench ZC, Block JK, Boguniewicz M, Boyle J, Fonacier L, Gelfand JM, et al. Atopic dermatitis in America study: a cross-sectional study examining the prevalence and disease burden of atopic dermatitis in the US adult population. *J Invest Dermatol* 2019;139:583–90.
- Choi SW, Gibbons LE, Crane PK. Iordif: an R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *J Stat Softw* 2011;39:1–30.
- Coutanceau C, Stalder JF. Analysis of correlations between patient-oriented SCORAD (PO-SCORAD) and other assessment scores of atopic dermatitis severity and quality of life. *Dermatology* 2014;229:248–55.
- Eichenfield LF, Tom WL, Chamlin SL, Feldman SR, Hanifin JM, Simpson EL, et al. Guidelines of care for the management of atopic dermatitis: section 1. Diagnosis and assessment of atopic dermatitis. *J Am Acad Dermatol* 2014;70:338–51.
- Elman S, Hynan LS, Gabriel V, Mayo MJ. The 5-D itch scale: a new measure of pruritus. *Br J Dermatol* 2010;162:587–93.
- EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16:199–208.
- Gerbens LA, Chalmers JR, Rogers NK, Nankervis H, Spuls PI. Harmonising Outcome Measures for Eczema (HOME) initiative. Reporting of symptoms in randomized controlled trials of atopic eczema treatments: a systematic review. *Br J Dermatol* 2016;175:678–86.
- Gerbens LA, Prinsen CA, Chalmers JR, Drucker AM, von Kobyletzki LB, Limpens J, et al. Evaluation of the measurement properties of symptom measurement instruments for atopic eczema: a systematic review. *Allergy* 2017;72:146–63.
- Haec IM, ten Berge O, van Velsen SG, de Bruin-Weller MS, Bruijnzeel-Koomen CA, Knol MJ. Moderate correlation between quality of life and disease activity in adult patients with atopic dermatitis. *J Eur Acad Dermatol Venereol* 2012;26:236–41.
- Hahn HB, Melfi CA, Chuang TY, Lewis CW, Gonin R, Hanna MP, et al. Use of the Dermatology Life Quality Index (DLQI) in a midwestern US urban clinic. *J Am Acad Dermatol* 2001;45:44–8.
- Hays RD, Liu H, Kapteyn A. Use of Internet panels to conduct surveys. *Behav Res Methods* 2015;47:685–90.
- Heckman CJ, Darlow S, Manne SL, Kashy DA, Munshi T. Correspondence and correlates of couples' skin cancer screening. *JAMA Dermatol* 2013;149:825–30.
- Heinl D, Chalmers J, Nankervis H, Apfelbacher CJ. Eczema trials: quality of life instruments used and their relation to patient-reported outcomes. A systematic review. *Acta Derm Venereol* 2016a;96:596–601.
- Heinl D, Prinsen CA, Deckert S, Chalmers JR, Drucker AM, Ofenloch R, et al. Measurement properties of adult quality-of-life measurement instruments for eczema: a systematic review. *Allergy* 2016b;71:358–70.

- Heinl D, Prinsen CAC, Sach T, Drucker AM, Ofenloch R, Flohr C, et al. Measurement properties of quality-of-life measurement instruments for infants, children and adolescents with eczema: a systematic review. *Br J Dermatol* 2017;176:878–89.
- Herd RM, Tidman MJ, Ruta DA, Hunter JA. Measurement of quality of life in atopic dermatitis: correlation and validation of two different methods. *Br J Dermatol* 1997;136:502–7.
- Holm EA, Wulf HC, Stegmann H, Jemec GB. Life quality assessment among patients with atopic eczema. *Br J Dermatol* 2006;154:719–25.
- Hua T, Silverberg JI. Atopic dermatitis in US adults: epidemiology, association with marital status, and atopy. *Ann Allergy Asthma Immunol* 2018;121:622–4.
- Jenkinson C, Layte R. Development and testing of the UK SF-12 (short form health survey). *J Health Serv Res Policy* 1997;2:14–8.
- Lim CR, Harris K, Dawson J, Beard DJ, Fitzpatrick R, Price AJ. Floor and ceiling effects in the OHS: an analysis of the NHS PROMs data set. *BMJ Open* 2015;5:e007765.
- Metz CE. Basic principles of ROC analysis. *Semin Nucl Med* 1978;8:283–98.
- Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737–45.
- Patel KR, Singam V, Vakharia PP, Chopra R, Sacotte R, Patel N, et al. Measurement properties of three assessments of burden used in atopic dermatitis in adults [e-pub ahead of print]. *Br J Dermatol* 2018. <https://doi.org/10.1111/bjd.17243> (accessed 8 May 2019).
- Rogers A, DeLong LK, Chen SC. Clinical meaning in skin-specific quality of life instruments: a comparison of the Dermatology Life Quality Index and Skindex banding systems. *Dermatol Clin* 2012;30(2):333–42, x.
- Rota M, Antolini L. Finding the optimal cut-point for Gaussian and gamma distributed biomarkers. *Comp Stat Data Anal* 2014;69:1–14.
- Rowen TS, Gaither TW, Awad MA, Osterberg EC, Shindel AW, Breyer BN. Pubic hair grooming prevalence and motivation among women in the United States. *JAMA Dermatol* 2016;152:1106–13.
- Schmitt J, Langan S, Stamm T, Williams HC. Harmonizing Outcome Measurements in Eczema (HOME) Delphi panel. Core outcome domains for controlled trials and clinical recordkeeping in eczema: international multiperspective Delphi consensus process. *J Invest Dermatol* 2011;131:623–30.
- Schmitt J, Langan S, Williams HC, European Dermato-Epidemiology Network. What are the best outcome measurements for atopic eczema? A systematic review. *J Allergy Clin Immunol* 2007;120:1389–98.
- Schmitt J, Williams H, HOME Development Group. Harmonising outcome measures for eczema (HOME). Report from the First International Consensus Meeting (HOME 1), 24 July 2010, Munich, Germany. *Br J Dermatol* 2010;163:1166–8.
- Silverberg J, Garg N, Silverberg NB. New developments in comorbidities of atopic dermatitis. *Cutis* 2014;93:222–4.
- Silverberg JI, Chiesa Fuxench ZC, Gelfand JM, Margolis DJ, Boguniewicz M, Fonacier L, et al. Content and construct validity, predictors, and distribution of self-reported atopic dermatitis severity in US adults. *Ann Allergy Asthma Immunol* 2018a;121:729–34.e4.
- Silverberg JI, Gelfand JM, Margolis DJ, Boguniewicz M, Fonacier L, Grayson MH, et al. Patient burden and quality of life in atopic dermatitis in US adults: a population-based cross-sectional study. *Ann Allergy Asthma Immunol* 2018b;121:340–7.
- Silverberg JI, Gelfand JM, Margolis DJ, Boguniewicz M, Fonacier L, Grayson MH, et al. Association of atopic dermatitis with allergic, autoimmune, and cardiovascular comorbidities in US adults. *Ann Allergy Asthma Immunol* 2018c;121:604–12.e3.
- Silverberg JI, Kantor RW, Dalal P, Hickey C, Shaunfield S, Kaiser K, et al. A comprehensive conceptual model of the experience of chronic itch in adults. *Am J Clin Dermatol* 2018d;19:759–69.
- Silverberg JI, Gelfand JM, Margolis DJ, Boguniewicz M, Fonacier L, Grayson MH, et al. Measurement properties of the hospital anxiety and depression scale used in atopic dermatitis in adults. *J Invest Dermatol* 2019a;139:1388–91.
- Silverberg JI, Gelfand JM, Margolis DJ, Boguniewicz M, Fonacier L, Grayson MH, et al. Symptoms and diagnosis of anxiety and depression in atopic dermatitis in U.S. adults. *Br J Dermatol*. 2019b: <https://doi.org/10.1111/bjd.17683> (accessed 8 May 2019).
- Silverberg JI, Simpson EL. Association between severe eczema in children and multiple comorbid conditions and increased healthcare utilization. *Pediatr Allergy Immunol* 2013;24:476–86.
- Silverberg JI, Simpson EL. Associations of childhood eczema severity: a US population-based study. *Dermatitis* 2014;25:107–14.
- Simpson EL, Bieber T, Eckert L, Wu R, Ardeleanu M, Graham NM, et al. Patient burden of moderate to severe atopic dermatitis (AD): insights from a phase 2b clinical trial of dupilumab in adults. *J Am Acad Dermatol* 2016;74:491–8.
- Simpson EL, Bruin-Weller M, Flohr C, Arden-Jones MR, Barbarot S, Deleuran M, et al. When does atopic dermatitis warrant systemic therapy? Recommendations from an expert panel of the International Eczema Council. *J Am Acad Dermatol* 2017;77:623–33.
- Spuls PI, Gerbens LAA, Simpson E, Apfelbacher CJ, Chalmers JR, Thomas KS, et al. Patient-Oriented Eczema Measure (POEM), a core instrument to measure symptoms in clinical trials: a Harmonising Outcome Measures for Eczema (HOME) statement. *Br J Dermatol* 2017;176:979–84.
- Stalder JF, Barbarot S, Wollenberg A, Holm EA, De Raevae L, Seidenari S, et al. Patient-Oriented SCORAD (PO-SCORAD): a new self-assessment scale in atopic dermatitis validated in Europe. *Allergy* 2011;66:1114–21.
- Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34–42.
- Vakharia PP, Cella D, Silverberg JI. Patient-reported outcomes and quality of life measures in atopic dermatitis. *Clin Dermatol* 2018a;36:616–30.
- Vakharia PP, Chopra R, Sacotte R, Patel N, Immaneni S, White T, et al. Severity strata for five patient-reported outcomes in adults with atopic dermatitis. *Br J Dermatol* 2018b;178:925–30.
- Vakharia PP, Chopra R, Sacotte R, Patel KR, Singam V, Patel N, et al. Burden of skin pain in atopic dermatitis. *Ann Allergy Asthma Immunol* 2017;119:548–52.e3.
- Vakharia PP, Chopra R, Sacotte R, Patel N, Immaneni S, White T, et al. Validation of patient-reported global severity of atopic dermatitis in adults. *Allergy* 2018c;73:451–8.
- Ware J Jr, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care* 1996;34:220–33.
- Ware JE Jr, Kosinski M, Keller SD. SF-12: how to score the SF-12 physical and mental health summary scales. Boston, MA: The Health Institute; 1998.
- Whalley D, McKenna SP, Dewar AL, Erdman RA, Kohlmann T, Niero M, et al. A new instrument for assessing quality of life in atopic dermatitis: international development of the Quality of Life Index for Atopic Dermatitis (QoLIAD). *Br J Dermatol* 2004;150:274–83.
- Williams HC, Burney PG, Pembroke AC, Hay RJ. The U.K. Working Party's Diagnostic Criteria for Atopic Dermatitis. III. Independent hospital validation. *Br J Dermatol* 1994;131:406–16.

Supplementary Figure S1. Design of the AD in America study. AD, atopic dermatitis.



Supplementary Figure S2. Discriminant validity of SF-12 scores using ROC curves. (a–c) ROC curves of the relationship between SF-12 MCS, PCS and SF-6D scores with the impact of AD on lifestyle. (d–f) ROC curves of the relationship between SF-12 MCS, PCS and SF-6D scores with the impact of AD on self-reported overall severity. AD, atopic dermatitis; MCS, mental component score; PCS, physical component score; ROC, receiver operating characteristic; SF, Short Form.



Supplementary Table S1. Interitem Correlations for SF-12 and DLQI in US Adults with AD

SF-12 Item	Pearson r											
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12
Q1	1.00	0.39	0.48	0.42	0.40	0.29	0.27	0.53	0.34	0.46	0.31	0.37
Q2	—	1.00	0.77	0.54	0.62	0.49	0.40	0.61	0.14	0.24	0.31	0.51
Q3	—	—	1.00	0.56	0.64	0.43	0.35	0.61	0.20	0.30	0.33	0.45
Q4	—	—	—	1.00	0.65	0.48	0.40	0.52	0.21	0.36	0.27	0.38
Q5	—	—	—	—	1.00	0.46	0.45	0.61	0.22	0.28	0.26	0.39
Q6	—	—	—	—	—	1.00	0.67	0.40	0.29	0.26	0.44	0.46
Q7	—	—	—	—	—	—	1.00	0.39	0.26	0.25	0.45	0.43
Q8	—	—	—	—	—	—	—	1.00	0.28	0.37	0.31	0.53
Q9	—	—	—	—	—	—	—	—	1.00	0.66	0.35	0.38
Q10	—	—	—	—	—	—	—	—	—	1.00	0.24	0.35
Q11	—	—	—	—	—	—	—	—	—	—	1.00	0.54
Q12	—	—	—	—	—	—	—	—	—	—	—	1.00
DLQI Item	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	
Q1	1.00	0.57	0.52	0.50	0.54	0.51	0.37	0.49	0.52	0.47	0.49	—
Q2	—	1.00	0.76	0.70	0.78	0.68	0.37	0.55	0.66	0.56	0.67	—
Q3	—	—	1.00	0.74	0.86	0.79	0.49	0.43	0.74	0.64	0.72	—
Q4	—	—	—	1.00	0.75	0.68	0.37	0.44	0.65	0.58	0.66	—
Q5	—	—	—	—	1.00	0.79	0.49	0.55	0.72	0.67	0.72	—
Q6	—	—	—	—	—	1.00	0.55	0.51	0.75	0.67	0.67	—
Q7	—	—	—	—	—	—	1.00	—	0.50	0.49	0.50	—
Q8	—	—	—	—	—	—	—	1.00	0.61	0.52	0.44	—
Q9	—	—	—	—	—	—	—	—	1.00	0.82	0.73	—
Q10	—	—	—	—	—	—	—	—	—	1.00	0.69	—
Q11	—	—	—	—	—	—	—	—	—	—	1.00	—

$P < 0.0001$ for all correlations.

Abbreviations: AD, atopic dermatitis; DLQI, Dermatology Life Quality Index; Q, question; SF, Short Form.

Supplementary Table S2. DIF by AD, Sex, Race/Ethnicity, and Age of Items in SF-12 and DLQI

		Adults with AD ($n = 602$)							
		DIF by AD ($n = 2893$)		DIF by Sex		DIF by Race/Ethnicity		DIF by Age	
	Item	Uniform	Nonuniform	Uniform	Nonuniform	Uniform	Nonuniform	Uniform	Nonuniform
SF-12	Q1	<0.0001	0.89	0.07	0.005	0.24	0.71	0.16	0.002
	Q2	<0.0001	0.58	0.84	0.93	0.09	0.42	0.33	0.24
	Q3	<0.0001	0.72	0.28	0.7	0.45	0.42	0.0004	0.29
	Q4	<0.0001	0.0004	0.61	0.3	0.16	0.06	0.09	0.76
	Q5	<0.0001	0.02	0.54	0.53	0.51	0.77	0.69	0.99
	Q6	<0.0001	0.48	0.7	0.97	0.05	0.46	0.14	0.15
	Q7	<0.0001	0.04	0.67	0.32	0.28	0.92	<0.0001	0.77
	Q8	<0.0001	0.68	0.08	0.004	0.05	0.20	0.03	0.89
	Q9	0.003	0.33	0.19	0.07	0.37	0.34	<0.0001	0.35
	Q10	<0.0001	0.0002	0.02	0.31	0.03	0.18	0.01	0.002
	Q11	<0.0001	0.0008	0.37	0.26	0.58	0.36	<0.0001	0.14
	Q12	<0.0001	0.001	0.36	0.03	0.49	0.53	<0.0001	0.19
DLQI	Q1	—	—	0.91	0.07	0.0001	0.24	0.28	0.52
	Q2	—	—	0.63	0.75	0.02	0.04	0.0001	0.008
	Q3	—	—	0.46	0.99	0.20	0.12	0.18	0.005
	Q4	—	—	0.002	0.10	0.47	0.23	0.75	0.55
	Q5	—	—	0.21	0.22	0.97	0.54	0.97	0.45
	Q6	—	—	0.81	0.03	0.0001	0.03	0.66	0.92
	Q7	—	—	0.67	0.59	0.96	0.26	0.36	0.003
	Q8	—	—	0.12	0.58	0.47	0.13	0.03	0.07
	Q9	—	—	0.06	0.62	0.03	0.36	0.008	0.44
	Q10	—	—	0.04	0.25	0.18	0.66	0.02	0.15
	Q11	—	—	0.76	0.10	0.001	0.56	0.03	0.33

Boldface indicates statistically significant items with DIF.

Abbreviations: AD, atopic dermatitis; DIF, differential item functioning; DLQI, Dermatology Life Quality Index; Q, question; SF, Short Form.

Supplementary Table S3. Distribution of Responses for SF-12 and DLQI Items

Items		Frequency (%)				
SF-12 Item	Excellent	Very Good	Good	Fair	Poor	—
Q1	30 (5.0%)	171 (28.4%)	245 (40.7%)	126 (20.9%)	30 (5.0%)	—
	No, not limited at all	Yes, limited a little	Yes, limited a lot	—	—	—
Q2	322 (53.5%)	187 (31.1%)	93 (15.5%)	—	—	—
Q3	272 (45.2%)	192 (31.9%)	137 (22.8%)	—	—	—
	No	Yes	—	—	—	—
Q4	293 (48.7%)	309 (51.3%)	—	—	—	—
Q5	327 (54.3%)	274 (45.5%)	—	—	—	—
Q6	348 (57.8%)	254 (42.2%)	—	—	—	—
Q7	418 (69.4%)	183 (30.4%)	—	—	—	—
	Not at all	A little bit	Moderately	Quite a bit	Extremely	—
Q8	146 (24.3%)	194 (32.2%)	123 (20.4%)	97 (16.1%)	42 (7.0%)	—
	All of the time	Most of the time	A good bit of the time	Some of the time	A little of the time	None of the time
Q9	34 (5.7%)	158 (26.3%)	122 (20.3%)	150 (24.9%)	121 (20.1%)	17 (2.8%)
Q10	16 (2.7%)	87 (14.5%)	125 (20.8%)	164 (27.2%)	140 (23.3%)	70 (11.6%)
Q11	36 (6.0%)	47 (7.8%)	76 (12.6%)	127 (21.1%)	186 (30.9%)	130 (21.6%)
	None of the time	A little of the time	A good bit of the time	Most of the time	All of the time	—
Q12	198 (32.9%)	205 (34.1%)	103 (17.1%)	60 (10.0%)	36 (6.0%)	—
DLQI	Not at all/Not relevant	A little	A lot	Very much	—	—
Q1	130 (21.6%)	341 (56.7%)	92 (15.3%)	38 (6.3%)	—	—
Q2	377 (62.7%)	155 (25.8%)	45 (7.5%)	24 (4.0%)	—	—
Q3	494 (82.1%)	69 (11.5%)	24 (4.0%)	15 (2.5%)	—	—
Q4	402 (66.9%)	117 (19.5%)	49 (8.2%)	33 (5.5%)	—	—
Q5	472 (78.5%)	86 (14.3%)	27 (4.5%)	16 (2.7%)	—	—
Q6	489 (81.5%)	64 (10.7%)	32 (5.3%)	15 (2.5%)	—	—
	Not relevant/No	Yes	—	—	—	—
Q7	574 (95.7%)	26 (4.3%)	—	—	—	—
	Not at all	A little	A lot	—	—	—
Q8	302 (72.8%)	107 (25.8%)	6 (1.5%)	—	—	—
	Not at all/Not relevant	A little	A lot	Very much	—	—
Q9	486 (81.0%)	77 (12.8%)	21 (3.5%)	16 (2.7%)	—	—
Q10	507 (84.5%)	58 (9.7%)	17 (2.8%)	18 (3.0%)	—	—
Q11	447 (74.4%)	112 (18.6%)	26 (4.3%)	16 (2.7%)	—	—

Missing values were present in ≤ 3 subjects for any SF-12 or DLQI questions and 0 for sex, race/ethnicity, and age.
Abbreviations: DLQI, Dermatology Life Quality Index; Q, question; SF, Short Form.

Supplementary Table S1. Interitem Correlations for SF-12 and DLQI in US Adults with AD

SF-12 Item	Pearson r											
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12
Q1	1.00	0.39	0.48	0.42	0.40	0.29	0.27	0.53	0.34	0.46	0.31	0.37
Q2	—	1.00	0.77	0.54	0.62	0.49	0.40	0.61	0.14	0.24	0.31	0.51
Q3	—	—	1.00	0.56	0.64	0.43	0.35	0.61	0.20	0.30	0.33	0.45
Q4	—	—	—	1.00	0.65	0.48	0.40	0.52	0.21	0.36	0.27	0.38
Q5	—	—	—	—	1.00	0.46	0.45	0.61	0.22	0.28	0.26	0.39
Q6	—	—	—	—	—	1.00	0.67	0.40	0.29	0.26	0.44	0.46
Q7	—	—	—	—	—	—	1.00	0.39	0.26	0.25	0.45	0.43
Q8	—	—	—	—	—	—	—	1.00	0.28	0.37	0.31	0.53
Q9	—	—	—	—	—	—	—	—	1.00	0.66	0.35	0.38
Q10	—	—	—	—	—	—	—	—	—	1.00	0.24	0.35
Q11	—	—	—	—	—	—	—	—	—	—	1.00	0.54
Q12	—	—	—	—	—	—	—	—	—	—	—	1.00
DLQI Item	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	
Q1	1.00	0.57	0.52	0.50	0.54	0.51	0.37	0.49	0.52	0.47	0.49	—
Q2	—	1.00	0.76	0.70	0.78	0.68	0.37	0.55	0.66	0.56	0.67	—
Q3	—	—	1.00	0.74	0.86	0.79	0.49	0.43	0.74	0.64	0.72	—
Q4	—	—	—	1.00	0.75	0.68	0.37	0.44	0.65	0.58	0.66	—
Q5	—	—	—	—	1.00	0.79	0.49	0.55	0.72	0.67	0.72	—
Q6	—	—	—	—	—	1.00	0.55	0.51	0.75	0.67	0.67	—
Q7	—	—	—	—	—	—	1.00	—	0.50	0.49	0.50	—
Q8	—	—	—	—	—	—	—	1.00	0.61	0.52	0.44	—
Q9	—	—	—	—	—	—	—	—	1.00	0.82	0.73	—
Q10	—	—	—	—	—	—	—	—	—	1.00	0.69	—
Q11	—	—	—	—	—	—	—	—	—	—	1.00	—

$P < 0.0001$ for all correlations.

Abbreviations: AD, atopic dermatitis; DLQI, Dermatology Life Quality Index; Q, question; SF, Short Form.

Supplementary Table S2. DIF by AD, Sex, Race/Ethnicity, and Age of Items in SF-12 and DLQI

	Item	DIF by AD (<i>n</i> = 2893)		Adults with AD (<i>n</i> = 602)					
				DIF by Sex		DIF by Race/Ethnicity		DIF by Age	
		Uniform	Nonuniform	Uniform	Nonuniform	Uniform	Nonuniform	Uniform	Nonuniform
SF-12	Q1	<0.0001	0.89	0.07	0.005	0.24	0.71	0.16	0.002
	Q2	<0.0001	0.58	0.84	0.93	0.09	0.42	0.33	0.24
	Q3	<0.0001	0.72	0.28	0.7	0.45	0.42	0.0004	0.29
	Q4	<0.0001	0.0004	0.61	0.3	0.16	0.06	0.09	0.76
	Q5	<0.0001	0.02	0.54	0.53	0.51	0.77	0.69	0.99
	Q6	<0.0001	0.48	0.7	0.97	0.05	0.46	0.14	0.15
	Q7	<0.0001	0.04	0.67	0.32	0.28	0.92	<0.0001	0.77
	Q8	<0.0001	0.68	0.08	0.004	0.05	0.20	0.03	0.89
	Q9	0.003	0.33	0.19	0.07	0.37	0.34	<0.0001	0.35
	Q10	<0.0001	0.0002	0.02	0.31	0.03	0.18	0.01	0.002
	Q11	<0.0001	0.0008	0.37	0.26	0.58	0.36	<0.0001	0.14
	Q12	<0.0001	0.001	0.36	0.03	0.49	0.53	<0.0001	0.19
DLQI	Q1	—	—	0.91	0.07	0.0001	0.24	0.28	0.52
	Q2	—	—	0.63	0.75	0.02	0.04	0.0001	0.008
	Q3	—	—	0.46	0.99	0.20	0.12	0.18	0.005
	Q4	—	—	0.002	0.10	0.47	0.23	0.75	0.55
	Q5	—	—	0.21	0.22	0.97	0.54	0.97	0.45
	Q6	—	—	0.81	0.03	0.0001	0.03	0.66	0.92
	Q7	—	—	0.67	0.59	0.96	0.26	0.36	0.003
	Q8	—	—	0.12	0.58	0.47	0.13	0.03	0.07
	Q9	—	—	0.06	0.62	0.03	0.36	0.008	0.44
	Q10	—	—	0.04	0.25	0.18	0.66	0.02	0.15
	Q11	—	—	0.76	0.10	0.001	0.56	0.03	0.33

Boldface indicates statistically significant items with DIF.

Abbreviations: AD, atopic dermatitis; DIF, differential item functioning; DLQI, Dermatology Life Quality Index; Q, question; SF, Short Form.

Supplementary Table S3. Distribution of Responses for SF-12 and DLQI Items

Items	Frequency (%)					
SF-12 Item	Excellent	Very Good	Good	Fair	Poor	—
Q1	30 (5.0%)	171 (28.4%)	245 (40.7%)	126 (20.9%)	30 (5.0%)	—
	No, not limited at all	Yes, limited a little	Yes, limited a lot	—	—	—
Q2	322 (53.5%)	187 (31.1%)	93 (15.5%)	—	—	—
Q3	272 (45.2%)	192 (31.9%)	137 (22.8%)	—	—	—
	No	Yes	—	—	—	—
Q4	293 (48.7%)	309 (51.3%)	—	—	—	—
Q5	327 (54.3%)	274 (45.5%)	—	—	—	—
Q6	348 (57.8%)	254 (42.2%)	—	—	—	—
Q7	418 (69.4%)	183 (30.4%)	—	—	—	—
	Not at all	A little bit	Moderately	Quite a bit	Extremely	—
Q8	146 (24.3%)	194 (32.2%)	123 (20.4%)	97 (16.1%)	42 (7.0%)	—
	All of the time	Most of the time	A good bit of the time	Some of the time	A little of the time	None of the time
Q9	34 (5.7%)	158 (26.3%)	122 (20.3%)	150 (24.9%)	121 (20.1%)	17 (2.8%)
Q10	16 (2.7%)	87 (14.5%)	125 (20.8%)	164 (27.2%)	140 (23.3%)	70 (11.6%)
Q11	36 (6.0%)	47 (7.8%)	76 (12.6%)	127 (21.1%)	186 (30.9%)	130 (21.6%)
	None of the time	A little of the time	A good bit of the time	Most of the time	All of the time	—
Q12	198 (32.9%)	205 (34.1%)	103 (17.1%)	60 (10.0%)	36 (6.0%)	—
DLQI	Not at all/Not relevant	A little	A lot	Very much	—	—
Q1	130 (21.6%)	341 (56.7%)	92 (15.3%)	38 (6.3%)	—	—
Q2	377 (62.7%)	155 (25.8%)	45 (7.5%)	24 (4.0%)	—	—
Q3	494 (82.1%)	69 (11.5%)	24 (4.0%)	15 (2.5%)	—	—
Q4	402 (66.9%)	117 (19.5%)	49 (8.2%)	33 (5.5%)	—	—
Q5	472 (78.5%)	86 (14.3%)	27 (4.5%)	16 (2.7%)	—	—
Q6	489 (81.5%)	64 (10.7%)	32 (5.3%)	15 (2.5%)	—	—
	Not relevant/No	Yes	—	—	—	—
Q7	574 (95.7%)	26 (4.3%)	—	—	—	—
	Not at all	A little	A lot	—	—	—

Items	Frequency (%)					
Q8	302 (72.8%)	107 (25.8%)	6 (1.5%)	—	—	—
	Not at all/Not relevant	A little	A lot	Very much	—	—
Q9	486 (81.0%)	77 (12.8%)	21 (3.5%)	16 (2.7%)	—	—
Q10	507 (84.5%)	58 (9.7%)	17 (2.8%)	18 (3.0%)	—	—
Q11	447 (74.4%)	112 (18.6%)	26 (4.3%)	16 (2.7%)	—	—

Missing values were present in ≤ 3 subjects for any SF-12 or DLQI questions and 0 for sex, race/ethnicity, and age.

Abbreviations: DLQI, Dermatology Life Quality Index; Q, question; SF, Short Form.