Received: 17 October 2022 | Accepted: 15 February 2023

DOI: 10.1111/jdv.18999

JEADV JOURNAL OF THE EUROPEAN ACADEMY OF DERMATOLOGY & VENEREOLOGY

ORIGINAL ARTICLE

# Psychometric validation of the generalized pustular psoriasis physician global assessment (GPPGA) and generalized pustular psoriasis area and severity index (GPPASI)

A. David Burden[1] | Robert Bissonnette[2] | Mark G. Lebwohl[3] | Tristan Gloede[4] | Milena Anatchkova[5] | Ismail Budhiarso[5] | Na Hu[6] | Christian Thoma[7] | Anne M. Skalicky[5] | Hervé Bachelez[8]

[1]School of Infection and Immunity, University of Glasgow, Glasgow, UK

[2]Innovaderm Research Inc., Montréal, Québec, Canada

[3]Icahn School of Medicine at Mount Sinai, New York, New York, USA

[4]Boehringer Ingelheim Pharmaceuticals, Inc., Ridgefield, Connecticut, USA

[5]Evidera, Bethesda, Maryland, USA

[6]Boehringer Ingelheim (China) Investment Co. Ltd, Shanghai, China

[7]Boehringer Ingelheim International GmbH, Biberach, Germany

[8]Service de Dermatologie, Assistance Publique–Hôpitaux de Paris Hôpital Saint-Louis, and INSERM Unité 1163, Imagine Institute of Genetic Diseases, Université Paris Cité, Paris, France

**Correspondence**
A. David Burden, Nuffield Health Glasgow Hospital, 25 Beaconsfield Rd, G12 0PJ Glasgow, UK.
Email: david.burden@glasgow.ac.uk

**Funding information**
Boehringer Ingelheim

**Abstract**

**Background:** Generalized pustular psoriasis (GPP) is a rare and life-threatening skin disease often accompanied by systemic inflammation. There are currently no standardized or validated GPP-specific measures for assessing severity.

**Objective:** To evaluate the reliability, validity and responder definitions of the Generalized Pustular Psoriasis Physician Global Assessment (GPPGA) and Generalized Pustular Psoriasis Area and Severity Index (GPPASI).

**Methods:** The GPPGA and GPPASI were validated using outcome data from Week 1 of the Effisayil™ 1 study. The psychometric analyses performed included confirmatory factor analysis, item-to-item/item-to-total correlations, internal consistency reliability, test–retest reliability, convergent validity, known-groups validity, responsiveness analysis and responder definition analysis.

**Results:** Using data from this patient cohort (N = 53), confirmatory factor analysis demonstrated unidimensionality of the GPPGA total score (root mean square error of approximation <0.08), and GPPGA item-to-item and item-to-total correlations ranged from 0.58 to 0.90. The GPPGA total score, pustulation subscore and GPPASI total score all demonstrated good test–retest reliability (intraclass correlation coefficient: 0.70, 0.91 and 0.95 respectively), and good evidence of convergent validity. In anchor-based analyses, all three scores were able to detect changes in symptom and disease severity over time; reductions of −1.4, −2.2 and − 12.0 were suggested as clinically meaningful improvement thresholds for the GPPGA total score, GPPGA pustulation subscore and GPPASI total score respectively. Anchor-based analyses also supported the GPPASI 50 as a clinically meaningful threshold for improvement.

**Conclusions:** Overall, our findings indicate that the GPPGA and GPPASI are valid, reliable and responsive measures for the assessment of GPP disease severity, and support their use in informing clinical endpoints in trials in GPP.

## INTRODUCTION

Generalized pustular psoriasis (GPP) is a rare autoinflammatory skin disease characterized by widespread eruptions of sterile, neutrophilic pustules,[1] often accompanied by systemic inflammation. Due to the rarity of the disease, there are no validated assessment measures for GPP severity, which presents a significant challenge to monitoring the efficacy of new therapies in controlled trials. Evaluating the psychometric properties of clinical outcomes assessments is

**Linked article:** P. Wolf. J Eur Acad Dermatol Venereol 2023; 37:1260–1261. https://doi.org/10.1111/jdv.19178.

particularly challenging in rare diseases; statistical tests are often underpowered due to small sample sizes and sparsity of clinical data.[2]

To date, trials in patients with GPP have mostly used scoring methods developed for plaque psoriasis, such as the Physician Global Assessment (PGA) and Psoriasis Area Severity Index (PASI).[3,4] Although the reliability of these measures has been widely investigated in plaque psoriasis, they do not assess skin pustulation (a key clinical manifestation of GPP), and are unsuitable for assessing GPP severity. To accurately assess disease severity and monitor treatment outcomes, Boehringer Ingelheim, in collaboration with leading global experts in GPP, developed the GPP Physician Global Assessment (GPPGA) and GPP Area Severity Index (GPPASI). The GPPGA and GPPASI were adapted from the established PGA and PASI, respectively, replacing the induration component with a pustulation component. To ensure consistent evaluation across assessors, hundreds of photographs from patients with GPP were reviewed and discussed by experts. A pocket guide, which includes photographs and descriptions for each score, was developed to train investigators in clinical trials in GPP.[5]

The GPPGA and GPPASI were used to define the primary and secondary endpoints in the Effisayil™ 1 study (NCT03782792),[6,7] a randomized, placebo-controlled trial of spesolimab, an anti-interleukin 36 receptor monoclonal antibody, in patients with GPP.[7] However, these novel clinician-reported outcomes (clinROs) are yet to be formally validated, as recommended by the US Food and Drug Administration (FDA).[8]

Here, we use data from Effisayil™ 1 to evaluate the reliability, validity and responder definitions of the GPPGA total score, GPPGA pustulation subscore and GPPASI as clinROs for GPP, and confirm their suitability to assess GPP severity.

# MATERIALS AND METHODS

## Study design

Details on the Effisayil™ 1 study design and outcomes have been published in full.[6,7] Briefly, patients presenting with a GPP flare were randomized (2:1) to receive a single intravenous dose of spesolimab (900 mg) or placebo. The primary endpoint was a GPPGA pustulation subscore of 0 at Week 1; the key secondary endpoint was a GPPGA total score of 0 or 1 at Week 1. Patients were followed for 12 weeks, and GPPGA and GPPASI scores were assessed throughout.

## Baseline demographics and clinical characteristics

Baseline demographic characteristics are presented descriptively (mean, standard deviation [SD] and range for quantitative variables; frequencies and percentages for categorical variables).

## GPPGA and GPPASI

The GPPGA assesses the severity of erythema, pustules and scaling, each scored using a 5-point scale. The GPPGA total score is an average of the three subscores, rounded to the nearest integer.[5]

The GPPASI comprises a lesion severity score and a score assessing area of involvement.[9] The severity score assesses erythema, pustulation and scaling severity on the patient's head, trunk and limbs. The area of involvement score measures the area of these body parts affected by lesions using a 7-point scale. The GPPASI total score is the sum of the severity and area of involvement scores. Full details on the GPPGA and GPPASI are given in Appendix S1.

## Evaluation of measurement properties

Psychometric validation analyses evaluate the measurement properties of an assessment, that is, they help determine the extent to which an assessment measures what it is supposed to measure. In dermatology, psychometric validation can provide important guidance on the development of new assessments, for example, measuring the severity of disease signs and symptoms. Psychometric validation can inform how well the new assessment reflects the signs and symptoms it was designed to capture, and identify potential needs for adjustment.

Psychometric validation of the GPPGA and GPPASI was conducted using data from the Effisayil™ 1 study. Outcome measures used for validation include the dermatology-specific, clinician-reported Japanese Dermatological Association Generalized Pustular Psoriasis Severity Index (JDA-GPPSI); the generic, clinician-reported Clinical Global Impression-Improvement (CGI-I); the dermatology-specific, patient-reported Dermatology Life Quality Index (DLQI); and the generic, patient-reported measure EQ-5D-5L.[3,10–12] Full details on the JDA-GPPSI, CGI-I, DLQI and EQ-5D-5L are given in Appendix S1. Table S1 presents descriptive data for patient-reported outcomes (PROs).

Owing to small sample sizes and limited variance for some analyses at Week 4, only analyses using Week 1 data are presented. For anchor-based analyses, where groups had small sample sizes, adjacent groups were collapsed into combined categories (e.g. 'improved' rather than 'minimal,' 'moderate,' or 'large' improvement).

### Confirmatory factor analysis

To explore how well the clinical data support the factor structure of the GPPGA and GPPASI scores, confirmatory factor analysis was performed using constrained and unconstrained models. For the constrained model, the three GPPGA rating scales were restricted to load equally on the GPPGA total score; for the unconstrained model, all scales on the GPPGA factor were freely estimated and allowed to differ.

A model that explains the data well should have a comparative fit index ≥0.9.[13] The standardized root mean residual (SRMR) measures the mean absolute difference between observed and model-implied correlations. The root mean square error of approximation (RMSEA) is a measure of fit assessing the discrepancy between predicted and observed data per degrees of freedom. An SRMR <0.1 and an RMSEA <0.08 are considered acceptable.

## Item-to-item and item-to-total correlations

Pearson correlations were performed to assess the extent to which GPPGA items correlate with each other and with the GPPGA total score. Correlations were assessed at Week 1. Optimal items should demonstrate moderate correlation (>0.4); a correlation coefficient >0.8 indicates item redundancy.[14] As the GPPASI measures the severity and extent of lesions in different body areas, the relationship between scores on items is not relevant, and inter-item correlations were not performed.

## Internal consistency reliability

Internal consistency reliability assesses agreement between multiple items in an instrument. The agreement of GPPGA items, GPPASI severity items for head, trunk, upper limb and lower limb, and GPPASI area of involvement items were assessed using Cronbach's alpha (α) coefficient.[15] Internal consistency was assessed using data at Week 1, and was not analyzed for the GPPASI as no correlation is expected between severity items and area of involvement items.

## Test–retest reliability

Test–retest reliability reflects the ability of an instrument to give reproducible results over time when the patient's clinical state is stable.[16] Intraclass correlation coefficients (ICCs) and difference scores (evaluated by paired t-test) were calculated between Days 3 and 4, using the JDA-GPPSI part A assessment of skin symptoms to define a stable population. An ICC ≥0.7 is considered acceptable for establishing test–retest reliability.[16]

## Convergent validity

Convergent validity evaluates the correlation between similar measures. Instruments used for validation should, therefore, measure related concepts. The convergent validity of the GPPGA total score, GPPGA pustulation subscore and GPPASI total score were examined by correlating them with the CGI-I, DLQI total score, DLQI item 1 (itchy, sore, painful and stinging skin), DLQI item 2 (embarrassed/self-conscious about skin), EQ-5D pain/discomfort item and EQ-VAS score. DLQI items 1 and 2 were selected as anchors as they assess skin symptoms and their direct impact on the patient. Convergent validity was assessed at Week 1 using Spearman's rank order correlation coefficients.

## Known-groups validity

Known-groups validity assesses whether a measure can distinguish between distinct groups of patients. Known-groups validity was examined by grouping subjects into varying levels of disease severity, using data from Week 1; groupings are summarized in Table S2. Group differences were determined using analysis of variance. F-statistics of the group difference were examined; when F-statistics were significant, a post-hoc test was performed using Scheffé's method to assess multiple pairs of means comparisons across the group.[17]

## Responsiveness/ability to detect change

Responsiveness, or the ability to detect change, refers to whether a measure is sensitive to true change in health status.[18] The ability of the GPPGA total score, GPPGA pustulation subscore and GPPASI total score to detect change between baseline and Week 1 was evaluated using the DLQI, EQ-5D pain/discomfort, EQ-VAS and CGI-I as anchors. Responder groups for each anchor are defined in Table S2.

Correlations between change scores were calculated before the responsiveness analyses to ensure that they were sufficiently large ($|r| \geq 0.30$).[19] Analysis of covariance (ANCOVA) was used to test differences in mean scores for the GPPGA total score, pustulation subscore and GPPASI total score. F-statistics of the group difference were examined; when F-statistics were significant, a post-hoc test was performed using Scheffé's method to assess multiple pairs of means comparisons across the group.[17] Effect size statistics were analyzed for each group of subjects; the effect sizes were calculated as mean change/baseline SD.[20]

## Responder definition

Responder definition refers to the magnitude of change in a measure that is clinically important to the patient. To evaluate the responder definition threshold, anchor-based estimates were calculated, as recommended by the FDA.[8] The anchor groups used in the analysis are summarized in Table S2. Differences in mean change scores between the clinROs and the anchor categories were assessed by ANCOVA, adjusted by baseline scores.

### Clinical meaningfulness of the GPPASI percentage improvement threshold

To evaluate the clinical relevance of the GPPASI 50 threshold, GPPASI percentage improvement categories (GPPASI <50%; GPPASI 50% to <75%; and GPPASI ≥75%) were compared with mean changes in the anchor scores (EQ-VAS; DLQI total score) from baseline to Week 1 using ANCOVA.

## RESULTS

### Baseline demographics and clinical characteristics

The mean (SD) age of patients enrolled in the study was 43.0 years (10.9) (Table 1), and there were more female patients than male (67.9% vs. 32.1%). Geographically, almost half the patients (47.2%) attended study sites in Asia (excluding Japan), 30.2% in Europe, 13.2% in Africa, 5.7% in the United States and 3.8% in Japan.

**TABLE 1** Baseline demographics and clinical characteristics.

| Parameter | *N* = 53 |
|---|---|
| Age, years | |
| Mean | 43.0 |
| Sex, *n* (%) | |
| Male | 17 (32.1) |
| Female | 36 (67.9) |
| Race, *n* (%) | |
| White | 24 (45.3) |
| Asian | 29 (54.7) |
| Study site, *n* (%) | |
| US | 3 (5.7) |
| Japan | 2 (3.8) |
| Asia (excluding Japan) | 25 (47.2) |
| Europe | 16 (30.2) |
| Africa | 7 (13.2) |
| South America | 0 |
| BMI class, *n* (%) | |
| <25 kg/m$^2$ | 24 (45.3) |
| 25 to <30 kg/m$^2$ | 16 (30.2) |
| ≥30 kg/m$^2$ | 13 (24.5) |
| Smoking status, *n* (%) | |
| Never | 38 (71.7) |
| Former | 4 (7.5) |
| Current | 11 (20.8) |

Abbreviation: BMI, body mass index.

### Psychometric validation

#### Confirmatory factor analysis and inter-item correlations

Confirmatory factor analysis demonstrated the unidimensionality of the GPPGA total score. Comparative fit index estimates were 1.00 for both the constrained and unconstrained models, surpassing the threshold of ≥0.9 for an acceptable fit. Using both models, SRMR estimates and RMSEA values fell within the prespecified thresholds for acceptable fit (SRMR <0.1; RMSEA <0.08).

Item-to-item correlations for the GPPGA were all statistically significant at Week 1 ($p < 0.0001$), with coefficients of 0.58 (between the erythema and scaling items), 0.61 (between erythema and pustulation) and 0.68 (between pustulation and scaling). There was also significant correlation between the GPPGA items and the total score ($p < 0.0001$), with item-to-total coefficients of 0.80, 0.90 and 0.81 for erythema, pustulation and scaling respectively.

#### Internal consistency reliability and test–retest reliability

The GPPGA total score showed good internal consistency at Week 1, with a Cronbach's α reliability coefficient of 0.81, falling within the acceptable range (>0.70). The GPPGA total score, GPPGA pustulation subscore and GPPASI total score all demonstrated good test–retest reliability from Day 3 to Day 4. When using the JDA GPP part A assessment of skin symptoms to define a stable population, ICC values for the GPPGA total score, pustulation score and GPPASI total score were 0.70, 0.91 and 0.95, respectively, reaching the acceptable threshold for good test–retest reliability (≥0.70).

#### Convergent validity and known-groups validity

The GPPGA total score, GPPGA pustulation subscore and GPPASI total score demonstrated good evidence of convergent validity (Table 2). Correlations between the GPPGA total score and GPPGA pustulation subscore were moderate for most anchors, with coefficients ranging from −0.47 to 0.54 for the GPPGA total score and −0.47 to 0.48 for the GPPGA pustulation subscore (Table 2).

Overall, all three scores demonstrated the ability to differentiate between select known groups, measuring different levels of symptom or disease severity (Table 3). For the GPPGA total score, mean scores were significantly higher for patients with higher severity scores for the JDA-GPPSI, JDA GPP erythema area with pustules, JDA GPP part A oedema area, CGI-I, DLQI total score and EQ-VAS ($p < 0.0001$–0.0270) (Table 3). Similarly, mean GPPGA pustulation subscores were significantly higher for patients with higher severity according to the JDA-GPPSI, JDA GPP erythema

**TABLE 2** Convergent validity of the GPPGA total score, GPPGA pustulation subscore and GPPASI total score with selected anchors.

| PRO/clinRO variables | Correlations[a] | | |
|---|---|---|---|
| | GPPGA total score | GPPGA pustulation subscore | GPPASI total score |
| CGI-I | 0.45* | 0.48** | 0.24 |
| DLQI total score | 0.36* | 0.33* | 0.14 |
| DLQI item 1: How itchy, sore, painful or stinging has your skin been? | 0.49** | 0.45** | 0.37* |
| DLQI item 2: How embarrassed or self-conscious have you been because of your skin? | 0.39* | 0.30* | 0.25 |
| EQ-5D pain/discomfort | 0.54*** | 0.47** | 0.46** |
| EQ-VAS score | −0.47** | −0.47** | −0.40* |

Abbreviations: CGI-I, Clinical Global Impression – Improvement; clinRO, clinician-reported outcome; DLQI, Dermatology Life Quality Index; EQ-VAS, EuroQol-visual analogue scale; GPPASI, Generalized Pustular Psoriasis Area and Severity Index; GPPGA, Generalized Pustular Psoriasis Physician Global Assessment; PRO, patient-reported outcome.

[a]Spearman's rank order correlation, Correlation interpretation: <0.3 = weak; 0.3–0.7 = moderate; 0.7–0.9 = strong; >0.9 = very strong.

Significance levels for correlations *p*-values are: *$p < 0.05$, **$p < 0.001$, ***$p < 0.0001$.

area with pustules, CGI-I, DLQI total score and EQ-VAS ($p < 0.0001$–$0.0110$) (Table 3). The GPPASI total scores were significantly higher in patients with more severe disease by JDA-GPPSI, JDA GPP erythema area with pustules, JDA GPP part A oedema area and EQ-VAS ($p < 0.0001$–$0.0444$), but not by CGI-I or DLQI total score (Table 3).

### Responsiveness and responder definitions

Overall, the GPPGA total score, GPPGA pustulation subscore and GPPASI total score demonstrated the ability to detect changes in symptom and disease severity over time. All three scores showed statistically significant least squares mean change scores from baseline between improved and not improved patient subgroups according to CGI-I ($p = 0.0250$ for GPPGA total score; $p < 0.0001$ for GPPGA pustulation subscore and GPPASI total), DLQI item 1 ($p = 0.0009$ for GPPGA total score; $p = 0.0008$ for GPPGA pustulation subscore; and $p < 0.0001$ for GPPASI total) and EQ-5D pain/discomfort ($p = 0.0032$ for GPPGA total score; $p < 0.0001$ for GPPGA pustulation subscore and GPPASI total) (Table 4).

According to anchor-based estimates, clinically meaningful change scores ranged from −1.34 to −1.56 for the GPPGA total score; −2.11 to −2.30 for the GPPGA pustulation subscore; and −10.82 to −12.65 for the GPPASI total score (Table 5). Therefore, average reductions of −1.4, −2.2 and −12.0 (to one decimal place) were suggested as thresholds for clinically meaningful improvement for the three scores respectively.

Psychometric analyses supported the GPPASI 50 as a clinically meaningful threshold for improvement. The differences in mean change scores in the EQ-VAS and DLQI total between GPPASI improvement categories (<50%; 50% to <75%; ≥75%) were statistically significant, with *p*-values of <0.0001 and 0.0252 respectively (Table S3). Post-hoc pairwise comparisons revealed significantly lower EQ-VAS change scores for patients with a GPPASI <50% relative to GPPASI 50% to <75% ($p < 0.0001$) and GPPASI ≥75% ($p = 0.0200$). DLQI total scores were not significantly different between GPPASI improvement groups; however, mean change scores of −1.28 for GPPASI <50%, −6.65 for GPPASI 50% to <75%, and −7.2 for GPPASI ≥75% indicated a trend in the right direction.

## DISCUSSION

The GPPGA and GPPASI were developed to address the need for appropriate outcomes that accurately capture GPP severity. These novel, GPP-specific clinROs were used to define the primary and secondary endpoints of the Effisayil™ 1 study, and are being used in ongoing clinical trials in GPP (NCT04399837; NCT03886246). To assess their suitability as clinical endpoints in patients with GPP, we have formally evaluated the psychometric properties of the GPPGA and GPPASI, as recommended by FDA guidelines.[21]

In our analyses, the GPPGA total score, GPPGA pustulation subscore and GPPASI total score demonstrated good test–retest reliability, good evidence of convergent validity and differentiated between patients with different levels of disease severity. Reductions of −1.4, −2.2 and −12.0 were suggested as thresholds for clinically meaningful improvement for the scores respectively. Our analyses also support the GPPASI 50% as a meaningful threshold for improvement.

These findings support the use of the GPPGA and GPPASI as endpoints for use in ongoing and future trials in GPP, and demonstrate their potential to become standard assessments for GPP severity. Our study complements a recent analysis of the intra-rater and inter-rater reliability of the GPPGA, which was shown to be a robust and reproducible instrument for assessing GPP severity.[5] Our data further support the clinical use of the GPPGA and GPPASI by demonstrating that they correlate with established measures of disease severity, and can reliably detect clinically meaningful change.

There is a need for objective, standardized outcomes for assessing GPP severity. Currently, the variation of outcome measures across trials precludes the comparison of results.[3,4,22,23] The implementation of standardized measures, such as the GPPGA and GPPASI, would facilitate the robust evaluation of GPP treatments, allow for meta-analyses and aid future trial design, regardless of the therapeutic intervention.[9] The ability to compare results across different studies is particularly pertinent for rare diseases like GPP, where patient numbers are small. Standardized assessments and thresholds for meaningful change could also

**TABLE 3** Known-groups validity at Week 1 by JDA, DLQI, CGI-I and EQ-VAS anchor categories.

| | GPPGA total score | | | GPPGA pustulation subscore | | | GPPASI total score | | |
|---|---|---|---|---|---|---|---|---|---|
| **Anchor categories** | ***N*** | **Mean (SD)** | **Overall *F*-test (*p*-value)** | ***N*** | **Mean (SD)** | **Overall *F*-test (*p*-value)** | ***N*** | **Mean (SD)** | **Overall *F*-test (*p*-value)** |
| JDA-GPPSI, 2-category | | | | | | | | | |
| 0–6 (mild) | 34 | 1.8 (1.0) | 0.0010 | 34 | 1.0 (1.4) | 0.0009 | 34 | 12.9 (8.7) | <0.0001 |
| 7–17 (moderate/severe) | 17 | 2.8 (0.8) | | 17 | 2.4 (1.2) | | 17 | 27.2 (11.1) | |
| JDA GPP erythema area with pustules, 2-category | | | | | | | | | |
| 0 (none) | 22 | 1.2 (0.4) | <0.0001 | 22 | 0 | <0.0001 | 22 | 11.6 (6.6) | 0.0005 |
| 1–3 (mild/moderate/severe) | 30 | 2.9 (0.8) | | 30 | 2.7 (0.9) | | 30 | 22.8 (12.7) | |
| JDA GPP part A oedema area, 2-category | | | | | | | | | |
| 0 (none) | 28 | 1.9 (1.0) | 0.0270 | 28 | 1.3 (1.5) | 0.1324 | 28 | 14.8 (9.3) | 0.0301 |
| 1–3 (mild/moderate/severe) | 24 | 2.5 (1.0) | | 24 | 1.9 (1.4) | | 24 | 21.9 (13.6) | |
| CGI-I, 3-category | | | | | | | | | |
| Very much improved | 19 | 1.6 (1.0) | 0.0071 | 19 | 0.7 (1.4) | 0.0019 | 19 | 14.5 (8.3) | 0.3340 |
| Much improved | 12 | 2.1 (1.0) | | 12 | 1.5 (1.2) | | 12 | 19.6 (14.2) | |
| Worsened/no change/ minimally improved | 18 | 2.7 (0.9) | | 18 | 2.3 (1.2) | | 18 | 19.6 (13.0) | |
| DLQI total score, 3-category | | | | | | | | | |
| 0–10 (no/small/moderate effect) | 18 | 1.9 (0.9) | 0.0101 | 18 | 1.4 (1.6) | 0.0110 | 18 | 17.4 (11.9) | 0.7231 |
| 11–20 (very large effect) | 17 | 1.8 (1.0) | | 17 | 0.9 (1.3) | | 17 | 16.8 (12.1) | |
| 21–30 (extremely large effect) | 17 | 2.8 (1.0) | | 17 | 2.4 (1.2) | | 17 | 19.9 (12.2) | |
| EQ-VAS score, 3-category | | | | | | | | | |
| Bad to very bad (score 0–65) | 24 | 2.6 (1.0) | 0.0091 | 24 | 2.3 (1.2) | 0.0009 | 24 | 22.3 (12.7) | 0.0444 |
| Moderate (score 66–85) | 17 | 1.8 (0.9) | | 17 | 0.8 (1.2) | | 17 | 15.3 (11.5) | |
| Good to very good (score 86–100) | 11 | 1.7 (1.0) | | 11 | 0.9 (1.6) | | 11 | 12.9 (7.4) | |

*Note*: Data were missing for two patients for the JDA-GPPSI; one patient for the JDA GPP score for erythema area with pustules; one patient for the JDA GPP score for part A oedema area; four patients for the CGI-I; one patient for the DLQI total score; and one patient for the EQ-VAS score.

Abbreviations: CGI-I, Clinical Global Impression – Improvement; DLQI, Dermatology Life Quality Index; EQ-VAS, EuroQol-visual analogue scale; GPP, Generalized Pustular Psoriasis; GPPASI, Generalized Pustular Psoriasis Area and Severity Index; GPPGA, Generalized Pustular Psoriasis Physician Global Assessment; GPPSI, Generalized Pustular Psoriasis Severity Index; JDA, Japanese Dermatological Association; SD, standard deviation.

support improved management of patients with GPP[24,25] Furthermore, the GPPGA and GPPASI have the advantage of being straightforward for dermatologists to interpret, as they are based on the widely used PGA and PASI.[5] The use of both endpoints together may allow for the most complete assessment of patients as, while the GPPGA is simpler for dermatologists to score, the GPPASI is able to capture the extent of the disease by body surface area affected.

A key advantage of our analyses is the use of global data from Effisayil™ 1, the largest clinical study in patients with GPP to date; the results presented here reflect a more diverse patient population than previous studies, which predominantly have Japanese patient cohorts.[3,4,22,26] However, our analyses are limited by the lack of a gold-standard instrument for assessing GPP that can be used as an anchor. We have, however, addressed this by using a combination of clinROs previously used in GPP studies, and PROs that are widely used to measure patient-reported symptoms and quality of life (QoL) in other dermatologic conditions. Of note, while the DLQI and EQ-5D-5L measure patient QoL over the previous week, GPPGA and GPPASI scores reflect severity at the time of assessment; this difference in timing could impact correlations between these instruments. Furthermore, unlike the JDA-GPPSI, which incorporates laboratory findings including C-reactive protein and white blood cell levels, neither the GPPGA nor GPPASI capture the systemic manifestations of GPP. Therefore, the extent to which they correlate with the JDA-GPPSI may be limited, as they do not capture the same disease components. Our analyses were also limited by small sample sizes, a common challenge in rare diseases.[2] These limitations highlight the need for reliable, standardized tools to assess GPP severity; future research should replicate our analyses in larger populations.

Our findings demonstrate for the first time that the GPPGA and GPPASI, both specifically designed to assess

**TABLE 4** ClinRO responsiveness by DLQI item 1, EQ-5D pain/discomfort and CGI-I.

| Change score from baseline to Week 1 | DLQI item 1 change score, 2-category | | | | Overall *F*-test | | |
|---|---|---|---|---|---|---|---|
| | Worsened/no change | | Improved | | | | |
| | *N* | LS mean (SE) | *N* | LS mean (SE) | *F*-test | *p*-value[b] | Effect size[a] |
| GPPGA total score | 28 | −0.57 (0.16) | 23 | −1.56 (0.18) | 8.21 | 0.0009 | −2.5428 |
| GPPGA pustulation subscore | 28 | −1.02 (0.26) | 23 | −2.24 (0.28) | 8.26 | 0.0008 | −2.0947 |
| GPPASI total score | 28 | −3.88 (2.05) | 23 | −12.64 (2.27) | 25.69 | <0.0001 | −0.5596 |
| **Change score from baseline to Week 1** | **EQ-5D pain/discomfort change score, 2-category** | | | | **Overall *F*-test** | | |
| | Worsened/no change | | Improved | | | | |
| | *N* | LS mean (SE) | *N* | LS mean (SE) | *F*-test | *p*-value[b] | Effect size[a] |
| GPPGA total score | 16 | −0.36 (0.23) | 36 | −1.34 (0.15) | 6.49 | 0.0032 | −2.6095 |
| GPPGA pustulation subscore | 16 | −0.51 (0.34) | 36 | −2.11 (0.22) | 11.46 | <0.0001 | −2.1508 |
| GPPASI total score | 16 | 0.30 (2.48) | 36 | −11.88 (1.65) | 34.56 | <0.0001 | −0.5841 |
| **Change score from baseline to Week 1** | **CGI-I change score, 3-category** | | | | | | **Overall F-test** | | |
| | Worsened/no change/ minimally improved | | Much improved | | Very much improved | | | | |
| | N | LS mean (SE) | N | LS mean (SE) | N | LS mean (SE) | *F*-test | *p*-value[b] | Effect size[a] |
| GPPGA total score | 18 | −0.53 (0.22) | 12 | −1.08 (0.27) | 19 | 1.50 (0.21) | 3.42 | 0.0250 | −2.7871 |
| GPPGA pustulation subscore | 18 | −0.74 (0.30) | 12 | −1.66 (0.36) | 19 | 2.46 (0.29) | 9.13 | <0.0001 | −2.1977 |
| GPPASI total score | 18 | −2.45 (2.63) | 12 | −5.64 (3.03) | 19 | 13.36 (2.54) | 14.79 | <0.0001 | −0.5796 |

*Note*: Data were missing for two patients for the DLQI item 1 change score; one patient for the EQ-5D pain/discomfort change score; and four patients for the CGI-I change score.

Abbreviations: ANCOVA, analysis of covariance; CGI-I, Clinical Global Impression – Improvement; clinRO, clinician-reported outcome; DLQI, Dermatology Quality of Life Index; GPPASI, Generalized Pustular Psoriasis Area and Severity Index; GPPGA, Generalized Psoriasis Physician Global Assessment; LS, least squares; SE, standard error.

[a]Effect size (ES) calculated using Cohen's *d*, a calculation of the difference of the means divided by the standard deviation at baseline/Day 1. ES: small ES = 0.20, moderate ES = 0.50 and large ES = 0.80. Pairwise comparisons calculated only if ≥5 patients in each group.

[b]ANCOVA adjusted by baseline/Day 1 score and anchor change score.

**TABLE 5** Anchor-based estimates for clinically meaningful improvement.

| Mean score change from baseline to Week 1 | Anchor | | | |
|---|---|---|---|---|
| | DLQI item 1 | EQ-5D pain/ discomfort | EQ-VAS | CGI-I |
| GPPGA total score | −1.56 | −1.34 | −1.45 | −1.36 |
| GPPGA pustulation subscore | −2.24 | −2.11 | −2.30 | −2.17 |
| GPPASI total score | −12.64 | −11.88 | −12.65 | −10.82 |

Abbreviations: CGI-I, Clinical Global Impression – Improvement; DLQI, Dermatology Quality of Life Index; EQ-VAS, EuroQol-visual analogue scale; GPPASI, Generalized Pustular Psoriasis Area and Severity Index; GPPGA, Generalized Psoriasis Physician Global Assessment.

GPP severity, are valid, reliable and sensitive. These measures are an important advance towards standardization of endpoints for clinical trials in GPP and have the potential to become valuable tools for physicians to support improved patient care.

## ACKNOWLEDGEMENTS

Isabella Goldsbrough, PhD, of OPEN Health Communications (London, United Kingdom) provided writing, editorial and formatting support, which was contracted and funded by Boehringer Ingelheim.

## FUNDING INFORMATION

The study was supported and funded by Boehringer Ingelheim.

## CONFLICT OF INTEREST STATEMENT

The authors met criteria for authorship as recommended by the International Committee of Medical Journal Editors (ICMJE). The authors did not receive payment related to the development of this manuscript. Boehringer Ingelheim was given the opportunity to review the manuscript for medical and scientific accuracy as well as intellectual property considerations. **ADB** declares paid consulting activities for AbbVie, Almirall, Boehringer Ingelheim, Bristol Myers Squibb, Celgene, Eli Lilly, Janssen, LEO Pharma, Novartis and UCB. **RB** is an advisory board member, consultant, speaker and/or investigator for and received honoraria and/or grants from, AbbVie, Alumis, Almirall, Amgen, AnaptysBio, Arcutis, Aristea, Bausch Health, Boehringer Ingelheim, Boston,

Bristol Myers Squibb, Dermavant Sciences, Eli Lilly, Escalier, Janssen, Kyowa Kirin, LEO Pharma, Nimbus, Novartis, Pfizer, Regeneron, Sienna, UCB, VentyxBio and Xencor; and an employee and shareholder of Innovaderm Research. **MGL** is an employee of Mount Sinai and has received research funds from AbbVie, Amgen, Arcutis, Avotres, Boehringer Ingelheim, Cara Therapeutics, Dermavant Sciences, Eli Lilly, Incyte, Janssen Research & Development, LLC, Ortho Dermatologics, Regeneron and UCB, and is a consultant for Aditum Bio, Almirall, AltruBio, AnaptysBio, Arcutis, Arena Pharmaceuticals, Aristea Therapeutics, Arrive Technologies, Avotres Therapeutics, BiomX, Boehringer Ingelheim, Brickell Biotech, Bristol Myers Squibb, Cara Therapeutics, Castle Biosciences, CorEvitas, Dermavant Sciences, Dr. Reddy's Laboratories, Evelo Biosciences, Evommune, Facilitation of International Dermatology Education, Forte Biosciences, Foundation for Research and Education in Dermatology, Helsinn Therapeutics, Hexima, LEO Pharma, Meiji Seika Pharma, Mindera, Pfizer, Seanergy and Verrica. **TG, NH** and **CT** are employees of Boehringer Ingelheim. **MA**, **IB** and **AMS** are employees of Evidera, which was contracted by Boehringer Ingelheim for the purposes of this study. **HB** declares paid consulting activities for AbbVie, Almirall, AnaptysBio, Aristea Therapeutics, BIOCAD, Boehringer Ingelheim, Bristol Myers Squibb, Dermavant Sciences, Eli Lilly, Janssen, Kyowa Kirin, LEO Pharma, Novartis, UCB and Xion Pharmaceuticals; grant support from Boehringer Ingelheim, Bristol Myers Squibb, Janssen, LEO Pharma, Novartis and Pfizer; and participation in a data safety monitoring board/advisory board for Avillion.

## DATA AVAILABILITY STATEMENT

To ensure independent interpretation of clinical study results and enable authors to fulfil their role and obligations under the ICMJE criteria, Boehringer Ingelheim grants all external authors access to clinical study data pertinent to the development of the publication. In adherence with the Boehringer Ingelheim Policy on Transparency and Publication of Clinical Study Data, scientific and medical researchers can request access to clinical study data when it becomes available on Vivli Center for Global Clinical Research Data, and earliest after publication of the primary manuscript in a peer-reviewed journal, regulatory activities are complete, and other criteria are met. Please visit Medical & Clinical Trials | Clinical Research | MyStudyWindow for further information.

## ORCID

*Robert Bissonnette* https://orcid.org/0000-0001-5927-6587

## REFERENCES

1. Navarini AA, Burden AD, Capon F, Mrowietz U, Puig L, Köks S, et al. European consensus statement on phenotypes of pustular psoriasis. J Eur Acad Dermatol Venereol. 2017;31(11):1792–9.
2. Benjamin K, Vernon MK, Patrick DL, Perfetto E, Nestler-Parr S, Burke L. Patient-reported outcome and observer-reported outcome assessment in rare disease clinical trials: an ISPOR COA emerging good practices task force report. Value Health. 2017;20(7):838–55.
3. Morita A, Yamazaki F, Matsuyama T, Takahashi K, Arai S, Asahina A, et al. Adalimumab treatment in Japanese patients with generalized pustular psoriasis: results of an open-label phase 3 study. J Dermatol. 2018;45(12):1371–80.
4. Saeki H, Nakagawa H, Ishii T, Morisaki Y, Aoki T, Berclaz PY, et al. Efficacy and safety of open-label ixekizumab treatment in Japanese patients with moderate-to-severe plaque psoriasis, erythrodermic psoriasis and generalized pustular psoriasis. J Eur Acad Dermatol Venereol. 2015;29(6):1148–55.
5. Burden AD, Bachelez H, Choon SE, Marrakchi S, Tsai T-F, Morita A, et al. Validation of the Generalized Pustular Psoriasis Physician Global Assessment (GPPGA) scale for assessing severity of GPP. 2022 American Academy of Dermatology Annual Meeting; March 25-29, 2022; Boston, Massachusetts.
6. Choon SE, Lebwohl MG, Marrakchi S, Burden AD, Tsai TF, Morita A, et al. Study protocol of the global Effisayil 1 phase II, multicentre, randomised, double-blind, placebo-controlled trial of spesolimab in patients with generalized pustular psoriasis presenting with an acute flare. BMJ Open. 2021;11(3):e043666.
7. Bachelez H, Choon SE, Marrakchi S, Burden AD, Tsai TF, Morita A, et al. Trial of spesolimab for generalized pustular psoriasis. N Engl J Med. 2021;385(26):2431–40.
8. Guidance for industry: patient-reported outcome measures, use in medical product development to support labeling claims. Fed Reg. 2009;74(235):65132–3.
9. Burden AD, Choon SE, Gottlieb AB, Navarini AA, Warren RB. Clinical disease measures in generalized pustular psoriasis. Am J Clin Dermatol. 2022;23(Suppl 1):39–50.
10. Busner J, Targum SD. The clinical global impressions scale: applying a research tool in clinical practice. Psychiatry (Edgmont). 2007;4(7):28–37.
11. Finlay AY, Khan GK. Dermatology Life Quality Index (DLQI)—a simple practical measure for routine clinical use. Clin Exp Dermatol. 1994;19(3):210–6.
12. Herdman M, Gudex C, Lloyd A, Janssen MF, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011;20(10):1727–36.
13. Hu L-t, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Model Multidiscip J. 1999;6(1):1–55.
14. Nunnally JC. Psychometric theory. New York, NY: Tata McGraw-Hill Education; 1994.
15. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951;16(3):297–334.
16. Aaronson N, Alonso J, Burnam A, Lohr KN, Patrick DL, Perrin E, et al. Assessing health status and quality-of-life instruments: attributes and review criteria. Qual Life Res. 2002;11(3):193–205.
17. Scheffé H, Wiley J, Sons. The analysis of variance. New York: Wiley; 1959.
18. Revicki DA, Hays R. Reliability and validity (including reponsiveness). Ass Qual Life Clin Trial Meth Practice. 2005;2:25–39.
19. 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(2):102–9.
20. Cohen J. Statistical power analysis for the behavioral sciences. New York, NY: Academic Press; 1977.
21. Administration USFaD. Considerations for clinical outcome assessment development. Administration USFaD. https://www.fda.gov/media/142629/download
22. Yamasaki K, Nakagawa H, Kubo Y, Ootaki K. Efficacy and safety of brodalumab in patients with generalized pustular psoriasis and psoriatic erythroderma: results from a 52-week, open-label study. Br J Dermatol. 2017;176(3):741–51.
23. Sano S, Kubo H, Morishima H, Goto R, Zheng R, Nakagawa H. Guselkumab, a human interleukin-23 monoclonal antibody in Japanese patients with generalized pustular psoriasis and erythrodermic psoriasis: efficacy and safety analyses of a 52-week, phase 3, multicenter, open-label study. J Dermatol. 2018;45(5):529–39.
24. Komine M, Morita A. Generalized pustular psoriasis: current management status and unmet medical needs in Japan. Expert Rev Clin Immunol. 2021;17(9):1015–27.

25. Gooderham MJ, Van Voorhees AS, Lebwohl MG. An update on generalized pustular psoriasis. Expert Rev Clin Immunol. 2019;15(9):907–19.
26. Imafuku S, Honma M, Okubo Y, Komine M, Ohtsuki M, Morita A, et al. Efficacy and safety of secukinumab in patients with generalized pustular psoriasis: a 52-week analysis from phase III open-label multicenter Japanese study. J Dermatol. 2016;43(9):1011–7.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Burden AD, Bissonnette R, Lebwohl MG, Gloede T, Anatchkova M, Budhiarso I, et al. Psychometric validation of the generalized pustular psoriasis physician global assessment (GPPGA) and generalized pustular psoriasis area and severity index (GPPASI). J Eur Acad Dermatol Venereol. 2023;37:1327–1335. https://doi.org/10.1111/jdv.18999