Further information available on Figshare for Burden AD, *et al*. ‘The Generalized Pustular Psoriasis Physician Global Assessment (GPPGA) score: online assessment and validation study of a specific measure of GPP disease activity’

Contents

[The GPPGA score 2](#_Toc110258051)

[What is known and what this study adds 3](#_Toc110258052)

[Plain language summary 4](#_Toc110258053)

[Excerpted pages from the GPPASI and GPPGA pocket guide 5](#_Toc110258054)

[Patient images and online portal 10](#_Toc110258055)

[Intra- and inter-rater statistical analyses 12](#_Toc110258056)

[Intra- and inter-rater reliability among dermatologists 13](#_Toc110258057)

# The GPPGA score

*Development*

Boehringer Ingelheim collaborated with leading global experts in generalized pustular psoriasis (GPP) to agree on the three key manifestations of generalized pustular psoriasis (GPP) that should be assessed, and developed the Generalized Pustular Psoriasis Physician Global Assessment (GPPGA) to define how these three skin components should be assessed, namely pustulation, erythema, and scaling. GPPGA is a composite score adapted from the Physician Global Assessment (PGA), which is widely accepted by health authorities as a suitable primary or co-primary endpoint for the assessment of psoriasis. The GPPGA score replaces the PGA induration component, which is absent or marginal in most cases of GPP, with pustulation.

*Clinical use*

The GPPGA score has been successfully used in a small, open-label, Phase I proof-of-concept trial (clinicaltrials.gov identifier NCT02978690) [Bachelez H, *et al. N Engl J Med* 2019; **380**:981–3] and a Phase II trial (Effisayil™ 1; clinicaltrials.gov identifier NCT03782792 [Bachelez H, *et al. N Engl J Med* 2021; **385**:2431–40]) of spesolimab, a novel anti-interleukin-36 receptor antibody currently in development for GPP. In addition, it is an endpoint in an ongoing, Phase II trial in GPP (Effisayil™ 2; clinicaltrials.gov identifier NCT04399837) and a long-term extension study in GPP (clinicaltrials.gov identifier NCT03886246), and an outcome measure in the International Rare And Severe Psoriasis Expert Network (IRASPEN) registry (clinicaltrials.gov identifier NCT04359394).

# What is known and what this study adds

*What’s already known about this topic?*

* The Generalized Pustular Psoriasis Physician Global Assessment (GPPGA) was developed for dermatologists to assess the severity of generalized pustular psoriasis (GPP), owing to an absence of validated GPP-specific assessment tools.
* It has been used successfully in two completed trials and as an endpoint in additional ongoing trials. To ensure that the GPPGA is a reliable and consistent measure of disease severity, validation is required.

*What does this study add?*

* This study demonstrates the reproducibility and reliability of GPPGA as a physician-based assessment of GPP severity. GPPGA was validated through intra-rater (consistency of scoring by the same assessor over time) and inter-rater (consistency of scoring among different assessors) reliability.
* Dermatologists with experience in GPPGA were able to consistently score GPP severity, supporting use of GPPGA as a suitable endpoint for GPP trials and as a standard tool in clinical practice.

# Plain language summary

*A new method for measuring the severity of generalized pustular psoriasis*

Generalized pustular psoriasis (GPP) is a rare skin disease, which means that it affects few people in the population. When a severe GPP attack (“flare”) happens, it can be life-threatening. At present, there are only a few medicines for treating flares. Testing new medicines is also difficult because there is no agreed way to measure how severe a flare is, or how well the medicine is working.

Our study tested a new score for assessing how severe the skin symptoms of GPP are. We took photographs of skin from patients with GPP and asked skin specialists (known as dermatologists) to give each photograph a score based on how severe they thought the disease was. We compared how similar a specialist’s scores were for the same photograph when they looked at it 10 to 14 days apart. We also compared how similar scores for the same photograph were between different specialists.

We found that dermatologists gave the photographs similar scores each time they were shown them. We also found that the dermatologists scored the photographs at the same level of severity as each other. This means that the new method for scoring GPP that we tested is reliable and could be used by doctors to measure how severe GPP is. The method could also be used to investigate the magnitude of the impact of medicines on GPP.

Excerpted pages from the GPPASI and GPPGA pocket guide  
Excerpted pages from the *Generalized Pustular Psoriasis Severity Scoring (GPPASI and GPPGA)* pocket guide provided toinvestigators of the clinical trials of spesolimab in GPP (clinicaltrials.gov identifiers NCT03782792 and NCT04399837) as well as to participants of this study.

Graphical user interface, text, application

Description automatically generated

Graphical user interface

Description automatically generated with medium confidence

A picture containing chart

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

GPP, generalized pustular psoriasis; GPPASI; Generalized Pustular Psoriasis Area Severity Index; GPPGA, Generalized Pustular Psoriasis Physician Global Assessment.

# Patient images and online portal

*Patient images*

Photographs were taken using 1st or 2nd generation models of the 2D DermaViz, according to settings and instructions provided by QuantifiCare (Biot, France). Settings for the D5100 Nikon camera (1st generation model) were ISO 400, aperture F20, shutter speed 1/200 and manual focus; settings for the D5600 Nikon camera (2nd generation model) were ISO 125, shutter speed 1/200 and manual focus. Photographs of subjects were taken in a standing position, against a blue background, at a distance determined by dual LED beam pointers and with the camera held perpendicular to the area being photographed. Images were uploaded to QuantifiCare’s central server and image quality verified with respect to framing, focus and photo classification. Two expert dermatologists with clinical experience of assessing and treating generalized pustular psoriasis (GPP) shortlisted 75 photographs from a total of 120 images selected to represent different disease severity. Two other expert dermatologists independently reviewed the shortlist. From the shortlist, 16 images from 14 patients were selected, reflecting a full spectrum of disease severity levels for all skin disease components, for use in an online assessment (**Figure**).

**Figure:** Images used for assessment on the online portal. The 16 images were from 14 patients enrolled on the Effisayil™ 1 trial (Bachelez H, *et al. N Engl J Med* 2021; **385**:2431–40). The portal enabled zooming and panning of the images to facilitate detailed on-screen examination.

A collage of a person's legs

Description automatically generated with low confidence

*Online portal*

The 16 images were presented in high resolution on an online portal developed by Trifecta Clinical (Indianapolis, USA), which included the ability to zoom and pan within the images to facilitate detailed on-screen examination. Twenty-six GPP-experienced dermatologists were recruited to independently assess and score the 16 images via the online portal. Participants were required to score each of the three components of the Generalized Pustular Psoriasis Physician Global Assessment (GPPGA) score (erythema, scaling and pustulation) individually, according to the instructions provided within the online portal.

# Intra- and inter-rater statistical analyses

*Intra-rater analyses*

Intra-rater analyses were performed using the Generalized Pustular Psoriasis Physician Global Assessment (GPPGA) total score and component subscores from the same rater across two time points to assess the reliability. Calculation of the intra-class correlation coefficient (ICC) for intra-rater analyses was estimated using a two-way mixed-effects analysis of variance (ANOVA) model for absolute agreement ICC (2,1) (Koo TK, Li MY. *J Chiropr Med* 2016; **15**:155–63). This model considers the effect of repeated measurements from the same rater. The ICC was calculated as:

MSR, mean square for images (or rows); MSE, mean square for error; MSC, mean square for repeated measurements within a core rater (two time points); n, number of images; r, number of repeated measurements

*Inter-rater analyses*

For inter-rater reliability of GPPGA, ICCs were provided with corresponding 95% confidence internals (CIs). All raters scored all images for each component at the same single time point: the two-way random-effects model for single measurement ICC (2,1) was used for the analysis. ICC for absolute agreement was calculated.

MSR, mean square for images (or rows); MSE, mean square for error; MSC, mean square for raters (or columns); n, number of images; k, number of raters

For both intra-rater and inter-rater assessments, ICC values range from 0 to 1.0, with higher scores indicating a more stable instrument. Cicchetti characterised ICC values <0.40 as poor, 0.40–0.59 as fair, 0.60–0.74 as good and 0.75–1.00 as excellent (Cicchetti DV. *Psychological Assessment* 1994; **6**:284–90). To assess the level of reliability, 95% CIs for ICC estimates were calculated using the approach proposed by Shrout and Fleiss (Shrout PE, Fleiss JL. *Psychol Bull* 1979; **86**:420–8).

# Intra- and inter-rater reliability among dermatologists

*Intra-rater reliability*

**Intra-rater reliability among dermatologists (n=20 dermatologists; n=16 photographs at each time point)**

|  | **ICC\* absolute agreement** | | | |
| --- | --- | --- | --- | --- |
| **Mean (SD)** | **Median** | **Minimum** | **Maximum** |
| GPPGA total score | 0.90 (0.06) | 0.91 | 0.78 | 1.00 |
| Erythema | 0.89 (0.07) | 0.91 | 0.76 | 1.00 |
| Pustules | 0.90 (0.07) | 0.92 | 0.75 | 1.00 |
| Scaling | 0.87 (0.09) | 0.87 | 0.66 | 1.00 |

\*ICC using a two-way mixed-effects model for absolute agreement. ICC <0.40 = poor; 0.40–0.59 = fair; 0.60–0.74 = good; 0.75–1.00 = excellent.

GPPGA, Generalized Pustular Psoriasis Physician Global Assessment; ICC, intra-class correlation coefficient; SD, standard deviation.

**Mean scaling scores for the 16 images, by dermatologist (black bars, score at first review; grey bars, score at second review).**

Chart, bar chart

Description automatically generated

\*Intra-rater ICC <0.75 (0.66, 95% confidence interval: 0.29, 0.85). All other ICC values ≥0.75.

*Inter-rater reliability*

**Inter-rater reliability for all raters using the first time point**

|  |  |
| --- | --- |
| **Dermatologists (n=26)** | **ICC\* (2,1) absolute agreement [95% CI]** |
| GPPGA total score | 0.82 [0.73, 0.91] |
| Erythema | 0.82 [0.73, 0.91] |
| Pustules | 0.78 [0.67, 0.88] |
| Scaling | 0.76 [0.65, 0.87] |

\*ICC using a two-way random-effects model for absolute agreement. ICC <0.40 = poor;   
0.40–0.59 = fair; 0.60–0.74 = good; 0.75–1.00 = excellent.

CI, confidence interval; GPPGA, Generalized Pustular Psoriasis Physician Global Assessment; ICC, intra-class correlation coefficient.