

Title: Initial development and pragmatic clinical validation of a static disease severity instrument for pyoderma gangrenosum: Investigator Global Assessment for PG (IGAPg).

Michael E. Jacobson, BA^{1,2}, Justin W. Ng, BS², Laura Morales Leon, BS, BA², Jesse J. Keller, MD, MCR², Angelo V. Marzano, MD^{3,4}, William W. Huang, MD, MPH⁵, Robert I Kelly, MBBS⁶, Arash Mostaghimi, MD, MPH⁷, and Alex G. Ortega-Loayza, MD, MCR^{2*}

¹Oregon Health & Science University, School of Medicine, Portland, OR, United States of America.

²Oregon Health & Science University, Department of Dermatology, Portland, OR, United States of America.

³Dermatology Unit, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy.

⁴Department of Pathophysiology and Transplantation, Università Degli Studi di Milano, Milan, Italy

⁵UNC School of Medicine, Department of Dermatology.

⁶Department of Dermatology, St Vincent's Hospital, Melbourne. University of Melbourne.

⁷Department of Dermatology, Brigham and Women's Hospital, Harvard Medical School, Boston, United States of America.

*Twitter handle: @ThePGguy

Corresponding author:

Alex G. Ortega-Loayza, MD, MCR, Oregon Health and Science University, Center for Health and Healing, 3303 SW Bond Ave, CH16D, Portland, OR 97239. ortegalo@ohsu.edu

Funding sources: AOL is supported by National Institutes of Health, National Institute of Arthritis and Musculoskeletal and Skin Diseases, [R01 AR083110](#).

Conflicts of Interest: AVM reports consultancy/advisory boards disease-relevant honoraria from AbbVie, Amgen, Boehringer-Ingelheim, Bristol Myers Squibb, Incyte, Leopharma, Novartis, Pfizer, Sanofi, and UCB. AM has received consulting fees from AbbVie, ACOM Health, Bioniz, Boehringer Ingelheim, Concert, Digital Diagnostics, Eli Lilly, Equillum, Hims & Hers Health, and Pfizer; equity from ACOM Health, Figure-1, and Hims & Hers Health; licensing/royalties from Concert and Pfizer; and research funding from Aclaris, Concert, Eli Lilly, and Incyte. AGOL is the former President of the Pacific Dermatology Association, serves as an associate editor for Dermatology (Karger), and is an editorial board member of the American Journal of Clinical Dermatology. He is also a consultant for Genentec, GuidePoint, InflaRx, Corvus Pharmaceuticals, Castle Biosciences, Clarivate, TFS HealthScience, Otsuka, and Leo Pharma. He is an advisor to Bristol Meyer Squibb, Boehringer Ingelheim, Janssen, and Sanofi. AGOL has received research grants from Lilly, Janssen, and Pfizer. The remaining authors state no conflict of interest. Oregon Health & Science University has filed an application to copyright the IGAPg.

IRB approval status: Reviewed and approved by the Oregon Health & Science University IRB; approval # 27328; 21834

Patient Consent on File: Patient consent: Not applicable.

Reprint requests:

Alex G. Ortega-Loayza, MD, MCR, Oregon Health and Science University, Center for Health and Healing, 3303 SW Bond Ave, CH16D, Portland, OR 97239. ortegalo@ohsu.edu

Manuscript word count: 497 words [excluding capsule summary, abstract, references, figures, tables]

Abstract word count: N/A

Capsule summary word count: N/A

References: 5

Figures: 1

Supplementary figures: 1

Tables: 1

Supplementary tables: 0

Keywords: general dermatology; medical dermatology; pyoderma gangrenosum; clinical research; outcome measures

Abstract

Background: Pyoderma gangrenosum (PG) is a rare neutrophilic ulcerative dermatosis with no FDA-approved therapies and limited validated outcome measures. Investigator's Global Assessments (IGAs) are widely used in dermatology but often lack objective criteria, robust validation, and comparability across studies. There remains a critical need for a PG-specific, standardized, and validated severity instrument.

Methods: We developed and conducted initial validation of the Investigator Global Assessment for Pyoderma Gangrenosum (IGAPg©), a novel PG-specific IGA. An international multistakeholder panel guided development, with a core team designing the instrument based on prior research identifying key objective clinical features of PG severity. The IGAPg© incorporates ulcer depth, drainage, discoloration, and undermining, with ulcer location and extent used to resolve indeterminate cases. Standardized rater training materials were created. Construct validity was assessed in 36 patients evaluated by an expert PG clinician, with correlations to Patient Global Assessment (PGA), Skindex-Mini, and numeric rating scales (NRS) for 24-hour and 7-day pain. Inter-rater reliability was evaluated in a subset of 26 patients assessed independently by five raters using a two-way random-effects intraclass correlation coefficient (ICC [2,1]) within a linear mixed-effects model.

Results: IGAPg© scores demonstrated strong correlation with PGA when assessed by an expert PG dermatologist (Pearson's $r = 0.73$) and when averaged across all raters ($r = 0.69$). Moderate correlations were observed with Skindex-Mini ($r = 0.49$), 24-hour pain NRS ($r = 0.48$), and 7-day pain NRS ($r = 0.52$), consistent with differing constructs measured. Inter-rater reliability was high (ICC = 0.76). Raters reported the instrument to be comprehensive, comprehensible, and efficient to administer. Reliability and construct validity metrics are summarized in Table 1.

Conclusions: Despite limitations including modest sample size and rater homogeneity, the IGAPg© demonstrated strong construct validity and high inter-rater reliability. Designed for dermatologists and trainees familiar with PG morphology, the IGAPg© represents a promising PG-specific outcome measure for clinical research and therapeutic trials. Future work will focus on refining training materials and expanding validation with structured patient–investigator engagement.

Dear Editor,

Pyoderma Gangrenosum (PG) is a rare neutrophilic ulcerative dermatosis with no FDA-approved therapies and limited validated outcome measures.¹ Investigator's Global Assessments (IGAs) are widely used but often lack objective criteria and robust validation.^{2,3} The use of multiple IGAs for a single condition further limits comparability.² Newer IGAs with objective criteria are being developed across dermatology.³

We report the development and initial validation of a novel IGA for PG, the **IGAPg**© (available <https://data.mendeley.com/datasets/3bbm5cgtgh/1>). An international multistakeholder panel guided its creation. A core team (MEJ, AM, AOL) designed the instrument based on prior research into key clinical signs of PG severity, with critical feedback and input provided by a review team (coauthors).³⁻⁵ While core outcome sets are in development, existing PG assessments remain limited to ulcer size or non-PG-specific measures, falling short of FDA expectations and underscoring the need for a validated PG-specific tool. The **IGAPg**© incorporates objective ulcer features—depth, drainage, discoloration, and undermining—with location and extent used to resolve indeterminate cases. Standardized rater training materials were also developed.

Initial validation assessed construct validity and reliability. The **IGAPg**© was evaluated in 36 patients by an expert clinician (AOL), with patients completing the Patient Global Assessment (PGA), Skindex-Mini, and numeric rating scales (NRS) for 24-hour and 7-day average pain. Pearson correlations were used to assess construct validity (supplementary methods).

A subset of 26 patients was independently evaluated by the expert clinical rater (AOL) and four additional raters (Fig 1). All assessments were conducted during routine clinic visits with raters concurrently present; evaluations were based on direct clinical examination rather than

photographs. Inter-rater reliability was examined using a two-way random-effects intraclass correlation coefficient (ICC [2,1]), implemented within a linear mixed-effects model that included patient ID and rater as random intercepts. This approach appropriately accommodates the unbalanced design, in which not all raters evaluated every patient. Analyses were conducted using Python (v3.11).

IGAPg© scores strongly correlated with the PGA when assessed by a PG expert dermatologist (Pearson's $r = 0.73$; Fig 1). When averaged across all raters, the correlation remained strong ($r = 0.69$; Fig 1). Moderate correlations were observed between IGAPg and Skindex-Mini ($r = 0.49$), 24-hour pain NRS ($r = 0.48$), and 7-day pain NRS ($r = 0.52$), reflecting the difference in constructs being measured (Fig 1). Inter-rater reliability was high, with an ICC of 0.76 (Fig 1). Raters reported that the measure was comprehensive, comprehensible, and quick to use. Reliability and construct validity data are presented in Table 1. The IGAPg© was developed for dermatologists and trainees and requires familiarity with PG-specific ulcer morphology; intentionally designed for use in clinical research and trials.

Despite limitations related to modest sample size and rater homogeneity in training background, the **IGAPg**© demonstrated strong construct validity and reliability. Future directions include refining training materials and assessment content, as well as conducting validation studies that incorporate structured patient-investigator engagement. Overall, the **IGAPg**© represents a promising tool for assessing PG severity and treatment response.

References

1. Lu JD, Hobbs MM, Huang WW, Ortega-Loayza AG, Alavi A. Identification and evaluation of outcome measurement instruments in pyoderma gangrenosum: a systematic review. *Br J Dermatol*. 2020;183(5):821-828. doi:10.1111/bjd.19027
2. Futamura M, Leshem YA, Thomas KS, Nankervis H, Williams HC, Simpson EL. A systematic review of Investigator Global Assessment (IGA) in atopic dermatitis (AD) trials: Many options, no standards. *J Am Acad Dermatol*. 2016;74(2):288-294. doi:10.1016/j.jaad.2015.09.062

3. Simpson E, Bissonnette R, Eichenfield LF, et al. The Validated Investigator Global Assessment for Atopic Dermatitis (vIGA-AD): The development and reliability testing of a novel clinical outcome measurement instrument for the severity of atopic dermatitis. *J Am Acad Dermatol*. 2020;83(3):839-846. doi:10.1016/j.jaad.2020.04.104
4. Pimentel MA, Li MM, Noe MH, et al. Features that define clinical severity of ulcerative pyoderma gangrenosum: a Delphi consensus study of experts and patients on behalf of the US Medical Dermatology Society. *Br J Dermatol*. 2023;188(4):566-568. doi:10.1093/bjd/ljac137
5. Haddadin OM, Ortega-Loayza AG, Marzano AV, et al. An approach to diagnosis and management of patients with pyoderma gangrenosum from an international perspective: results from an expert forum. *Arch Dermatol Res*. 2024;316(3):89. Published 2024 Feb 24. doi:10.1007/s00403-024-02829-0

Table 1. Construct validity and inter-rater reliability of the IGAPg morphological tool. Correlations between IGAPg scores and patient-reported outcomes (PGA, Skindex-Mini, and average pain) demonstrate moderate-to-strong construct validity. Values are shown for all raters, Rater 1 alone, and Raters 2-5. Inter-item PRO correlations confirm internal consistency. Pearson correlations between Rater 1 and Raters 2-5 are reported with shared patient counts and p-values ($p < 0.05$ was considered significant). ICC(2,1) indicated good inter-rater reliability with most score variability attributable to differences between patients with minimal variability due to rater differences and a moderate residual component. Inter-item patient reported outcomes correlations confirm difference in measured constructs.

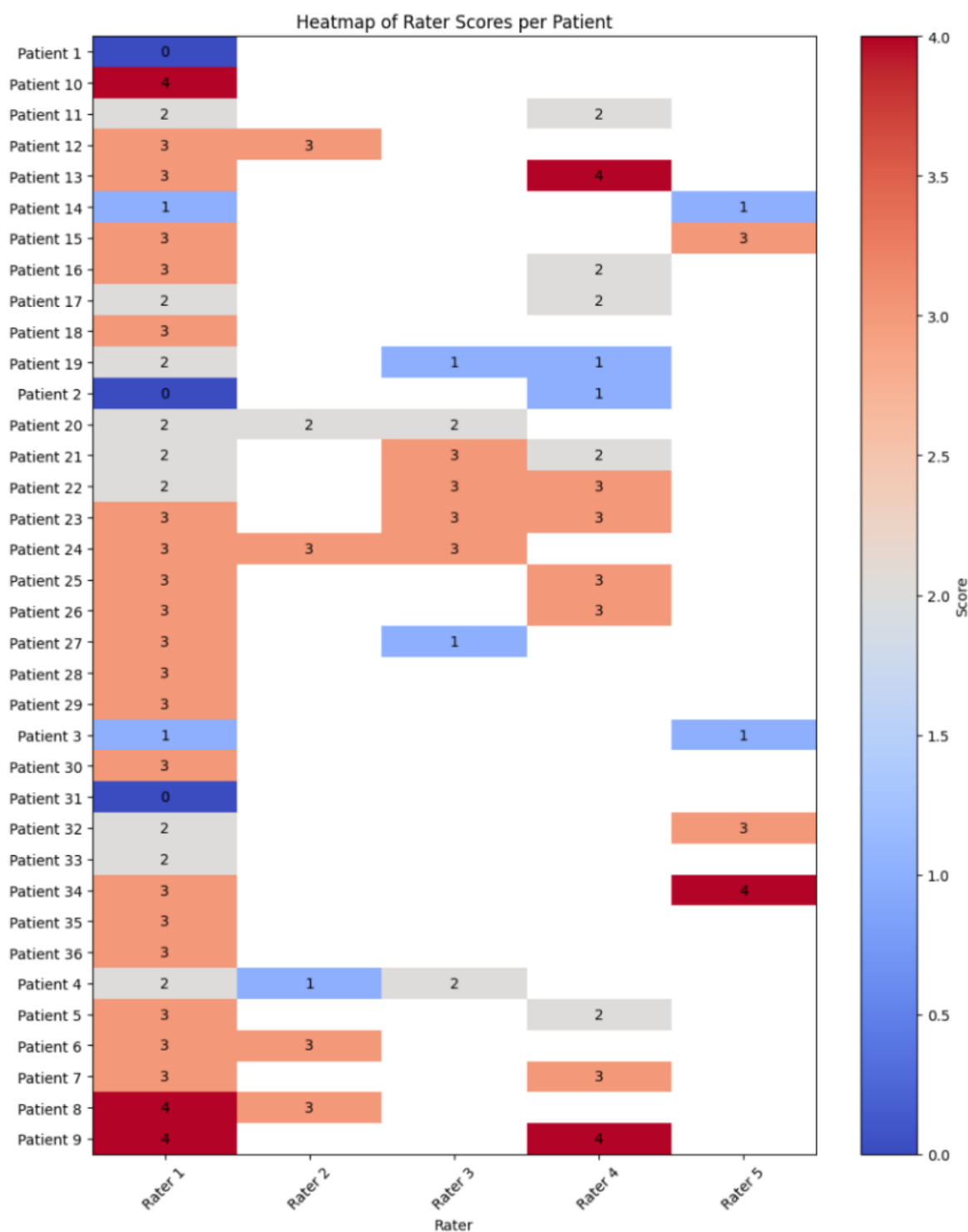
Instrument	All Patients N=36	Rater 1 N=36	Raters 2-5 n=26
Patient Global Assessment	0.75	0.73	0.58
Skindex-Mini	0.49	0.45	0.48
24hr Average Pain	0.48	0.45	0.53
7-day Average Pain	0.52	0.52	0.54

Instrument	Patient Global Assessment	Skindex-Mini	24 hr. Average Pain	7-day Average Pain
Patient Global Assessment	1.0	0.57	0.65	0.59
Skindex-Mini	0.57	1.0	0.83	0.78
24 hr. Average Pain	0.65	0.83	1.0	0.92
7-day Average Pain	0.59	0.78	0.92	1.0

Rater	n Shared Patients	Pearson r	p-value
Rater 2	6	0.79	0.06
Rater 3	8	0.078*	0.85
Rater 4	14	0.74	0.003
Rater 5	5	0.93	0.02

ICC (2,1) n=33	σ^2 Patient	σ^2 Rater	σ^2 Residual
0.76	0.99	0.06	0.25

*Rater 3's ratings differed by only 0–1 points from other raters; however, Pearson correlation is sensitive to restricted score ranges and low variance, which were present for these patients. The low correlation likely reflects statistical instability rather than true clinical disagreement.



Characteristic	All Patients N=36	Reliability Cohort N=26
<i>Sex, n (%)</i>		
Female	24 (66.7)	16 (61.5)
Male	12 (33.3)	10 (38.5)
<i>Race, n (%)</i>		
American Indian or Alaskan Native	2 (5.6)	1 (3.8)
White	35 (91.7)	24 (92.3)
Not disclosed	1 (2.8)	1 (3.8)
<i>Ethnicity, n (%)</i>		
Hispanic or Latino	2 (5.6)	1 (3.8)
Not Hispanic or Latino	29 (80.6)	21 (80.8)
Not disclosed	5 (13.9)	4 (15.4)
<i>Age, n (%)</i>		
18-39 years	3 (8.3)	2 (7.7)
40-59 years	13 (36.1)	9 (34.6)
60-79	16 (44.4)	12 (46.2)
≥80 years	4 (11.1)	3 (11.5)

Figure 1. **Top**-Heatmap of clinical rater IGAPg scores and rater characteristics. Warmer colors indicate higher severity scores and cooler colors indicate lower scores. Rater 1: Dermatologist instrument creator; Raters 2, 4, 5: Dermatology trainees; Rater 3: Attending dermatologist. **Bottom**-Pyoderma Gangrenosum patient participant demographics. The reliability cohort includes patients assessed by two or more raters.