UNICA

UNIVERSITÀ
DEGLI STUDI
DI CAGLIARI

IRIS | Università di Cagliari

UNICA IRIS Institutional Research Information System

This is the Author's [*accepted*] manuscript version of the following contribution:

International Expert Consensus for the Standardization of the Physician Global Assessment (PGA) in Systemic Lupus Erythematosus: the PISCOS study

AUTHORS

Matteo PIGA[1], Elisabetta CHESSA[1], Eric F. MORAND[2], Manuel F. UGARTE-GIL[3], Maria TEKTONIDOU[4], Ronald VAN VOLLENHOVEN[5], Michelle PETRI[6], Laurent ARNAUD[7,8] for the PISCOS investigator group.

AFFILIATIONS

1 - Rheumatology Unit, AOU University Clinic and University of Cagliari, Cagliari, Italy

2 – Centre for Inflammatory Diseases, Monash University, Melbourne, Victoria, Australia.

3 – School of Medicine, Universidad Científica del Sur and Rheumatology Department, Hospital Nacional Guillermo Almenara Irigoyen, EsSalud, Lima, Peru.

4 - First Department of Propaedeutic Internal Medicine, Joint Rheumatology Program, Medical School, National and Kapodistrian University of Athens, Laiko General Hospital, Athens, Greece.

5 - Department of Rheumatology, Amsterdam Rheumatology and Immunology Center, Amsterdam, The Netherlands

6 – Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland.

7 - Service de rhumatologie, Hôpitaux Universitaires de Strasbourg, Université de Strasbourg, Strasbourg, France.

8 - Centre National de Référence des Maladies Systémiques et Autoimmunes Rares Est Sud-Ouest (RESO), Strasbourg, France.

COLLABORATOR LIST:

Simone APPENZELLER, Cynthia ARANOW, Anca ASKANASE, Tadej AVCIN, Sang-Cheol BAE, George BERTSIAS, Eloisa BONFA, Ernesto CAIROLI, Mario H. CARDIEL, Ricard CERVERA, François CHASSET, Carlo CHIZZOLINI, Ann E. CLARKE, Fabrizio CONTI, Nathalie COSTEDOAT-CHALUMEAU, László CZIRJÁK, Andrea DORIA, Thomas DÖRNER, Gerard ESPINOSA, Rebecca FISCHER-BETZ, Mercedes GARCÌA, Dafna D. GLADMAN, Luis A. GONZÁLEZ, Iva GUNNARSSON, Laniyati HAMIJOYO, John G. HANLY, Sarfaraz A. HASNI, Frédéric A. HOUSSIAU, Murat INANÇ, Luís S INÊS, David ISENBERG, Soren JACOBSEN, Yeong-Jian JAN WU, Yuko KANEKO, Yasuhiro KATSUMATA, Chak S LAU, Alexandra C. LEGGE, Karoline LERANG, Maarten LIMPER, Worawit LOUTHRENOO, Shue-Fen LUO, António MARINHO, Loreto MASSARDO, Alexis MATHIAN, Marta MOSCA, Mandana NIKPOUR, José M. PEGO-REIGOSA, Christine A. PESCHKEN, Bernardo A. PONS-ESTEL, Guillermo J. PONS-ESTEL, Anisur RAHMAN, Simona REDNIC, Camillo RIBI, Guillermo RUIZ-IRASTORZA, Emilia I. SATO, Amit SAXENA, Matthias SCHNEIDER, Gian Domenico SEBASTIANI, Vibeke STRAND, Elisabet SVENUNGSSON, Yoshiya TANAKA, Zoubida TAZI MEZALEK, Michael L. TEE, Angela TINCANI, Zahi TOUMA, Anne TROLDBORG, Carlos VASCONCELOS, Évelyne VINET, Edward M. VITAL, Alexandre E. VOSKUYL, Anne VOSS, Daniel WALLACE, Michael WARD, Leonid D. ZAMORA.

CORRESPONDING AUTHOR: Prof. Laurent ARNAUD, Service de Rhumatologie, Centre National de Référence des Maladies Auto-immunes et Systémiques Rares, Hôpital de Hautepierre, 1 Avenue Molière BP 83049, 67098 Strasbourg Cedex, France. Tel: 03 88 12 84 74; Fax: 03 88 12 82 90; E-mail: laurent.arnaud@chru-strasbourg.fr

2

ABSTRACT

Objectives
The PISCOS study aimed at obtaining an evidence- and expert-based consensus standardization of the Physician Global Assessment (PGA) scoring of disease activity in Systemic Lupus Erythematosus (SLE).

Methods
An international panel of 79 SLE experts participated. In a 3-round Delphi consensus, 41 statements related to the PGA in SLE were rated, using a 0 (strongly disagree) to 10 (strongly agree) numeric rating scale. Statements with agreement ≥75% were selected and further validated by the expert panel.

Results
Consensus was reached on 27 statements, grouped in 14 recommendations, for the use of the PGA in SLE, design of the PGA scale, practical considerations for PGA scoring, and the relationship between PGA values and levels of disease activity. Among these recommendations, the expert panel agreed that the PGA should consist of a 0-3 visual analogue scale for measuring disease activity in SLE patients in the preceding month. The PGA is intended to rate the overall disease activity taking into account the severity of active manifestations and clinical laboratory results, but excluding organ damage, serology, and subjective findings unrelated to disease activity. The PGA scale ranges from "no disease activity" to the "most severe disease activity" and incorporates the values 1 and 2 as inner markers to categorize disease activity as mild (≥0.5-1), moderate (>1 and ≤2) and severe (>2 to 3).

Conclusions
The PISCOS results will allow for increased homogeneity and reliability of PGA ratings in routine clinical practice, definitions of remission and low disease activity, as well as in future SLE trials.

Keywords: systemic lupus erythematosus, disease activity, physician global assessment, PGA, low disease activity, remission.

3

INTRODUCTION

Systemic lupus erythematosus (SLE) is an autoimmune systemic disease characterised by a wide variability of clinical manifestations and by periods of remission interspersed with unpredictable flares (1-2). Disease activity is a significant predictor of organ damage and mortality (3) while conversely, attainment and maintenance of Lupus Low Disease Activity State (LLDAS) or remission is associated with reduced risk of damage and mortality (4-9). Because of SLE complexity, the assessment of disease activity, both in clinical trials and routine clinical practice, is particularly challenging (10). Numerous instruments have been derived to assess disease activity in SLE, with the Physician Global Assessment (PGA) being the only one included in the updated EULAR recommendations for the management of SLE (11). The PGA was developed on a 0 to 3 scale as part of the Lupus Activity Index [12] and later incorporated into the SLE Responder Index used in the belimumab clinical trials [13], along with the SLEDAI and BILAG. As the initial step of the International Expert Consensus for the Standardization of the Physician Global Assessment in Systemic Lupus Erythematosus (PISCOS) research programme, we conducted a systematic literature review (SLR) (published in 2020 [14]), in which the PGA was shown to be a valid, responsive, and feasible instrument, though its reliability was affected by the scale adopted, further supporting the need for standardization of its scoring. Data from this SLR were used as a valuable way to understand the psychometric properties of the various formats of PGA available in the literature, but also to identify the main areas of PGA rating which needed to be standardized.

Currently there is no consensus on the optimal way to assess the PGA in SLE, such as the visual analogue scale (VAS) scoring and the time frame of assessment. Also, recent studies showed that laboratory results may influence PGA rating and it remains unclear whether lupus serology should be incorporated in its use [15, 16]. The aim of the PISCOS study was to standardize PGA ratings in SLE, using the results from the SLR, a formal consensus methodology (the Delphi method) and a large panel of international lupus experts from five continents.

METHODS

Study design and Delphi survey

The PISCOS study was headed by a steering committee composed of rheumatologists specialized in care of SLE patients (M. Piga, M. Petri & L. Arnaud) and a fellow in rheumatology (E. Chessa).

Following the SLR, 41 research questions were selected by the steering committee and refined through a 3-round Delphi process **(supplementary document #1**) to achieve consensus. A Delphi survey is a systematic process involving a series of questionnaires and rounds of item generation, data collection and analysis, to derive expert consensus on a topic. This interactive process involved an international panel of 79 SLE experts (see **supplementary document #2**) who were asked to rate a list of statements using a 0 (strongly disagree) to 10 (strongly agree) Numeric Rating Scale (NRS) during 3 consecutive rounds, based on their expertise and the evidence from the SLR. At each Delphi round, anonymized and aggregated responses of prior rounds were provided and consensus for each statement was defined as agreement by at least 75% of respondents [17]. Those statements that reached the Delphi's agreement were grouped in different sections and then summarized into final remarks according to content affinity.

Commentato [MP1]: Should we insert here that we asked to revote the 27 statements for the final level of agreement ?

or should we now ask the expert to vote their level of agreement on the 14 final remarks ? This will be time consuming but I think it might improve the manuscript quality and clearness so **increasing the chances for acceptance in ARD.**

Statistical analyses

Qualitative data were expressed as numbers and percentages, and quantitative data as median and interquartile range (IQR). According to methodological criteria for reporting of Delphi studies [17], statements which were scored 7-10 on the NRS by ≥75% of the experts were selected while those scored 0-3 by ≥75% were rejected. Statements which did not reach these pre-specified levels of agreement were conditionally included in the subsequent Delphi round together with the comments from the panel, to be re-voted and eventually rejected if ≥75% agreement was not reached. Statistical analyses were performed using the software JMP version 13 (SAS institute, Cary, NC, USA).

RESULTS

All experts completed all of the three Delphi rounds. Consensus at pre-specified levels of agreement was reached for sixteen statements in round 1, nine statements in round 2, and two more in round 3 (**supplementary document #3).** This yielded a total of 27 statements grouped into 4 sections: overarching principles for scoring the PGA, design of the PGA scale, practical considerations for PGA scoring, and relationship between PGA values and levels of disease activity (**Table 1**). Finally, the statements were summarized in 14 remarks according to content affinity .

> **Commentato [MP2]:** I would suggest asking the 79 experts to vote the 14 final remarks (obtained summarizing the 27 statements with Delphi agreement >75%). It could be annoying and time-consuming but, from a methodological point of view, it would be more acceptable. Once this is done, we could report in the manuscript both the Delphi agreement for the individual statements (currently in the supplementary files) and the LoA for final remarks (deleting the LoA for the individual statements).

Overarching principles

The PGA is an outcome instrument consisting of a VAS for measuring disease activity in SLE (statements 1.1 and 1.2). The panel agreed that the PGA should meet the OMERACT criteria (truth or validity, discrimination, feasibility) for outcome measures [18]. Those criteria have been mainly assessed and confirmed by studies using a VAS to score disease activity as a continuous range of values [19-42]. Although no studies have examined which measurement tool performs better for the assessment of PGA, the experts agreed to rely on the use of a VAS given the quantity and quality of available evidence.

The PGA rating of overall disease activity should reflect the severity of active manifestations, but not organ damage (statements 1.3, 1.4, and 1.5). A large amount of evidence demonstrated the ability of the PGA to measure disease activity in comprehensive (content validity) and accurate (face validity) ways as well as to capture the severity of active disease manifestations (criterion validity) [19-52]. Evidence in support of the criterion validity of the PGA includes the association with treatment changes [53], response to treatment [35] and irreversible organ damage accrual [54]. In order to maintain the instrument's face validity, the experts agreed that manifestations related to organ damage should not be considered in PGA scoring.

The PGA reflects the clinician's judgment about disease activity in SLE and should be scored independently of pre-specified thresholds used in other scores (statements 1.6 and 1.7). The PGA was

originally proposed to capture manifestations not covered by the SLEDAI (face validity) [19]. Since it does not provide a predefined or limited list of disease manifestations or organ systems, the PGA enables the capture of all aspects of SLE disease activity (content validity) without referring to definition or thresholds proposed in glossary-based indices (e.g., in SELENA-SLEDAI, leukopenia only results in a score if $<3,000/mm^3$).

The PGA should assess SLE disease activity according to the physician's perspective and should be scored only by a medical doctor with expertise in SLE (statements 1.8, 1.9, and 1.10). The panel debated on whether the PGA could be scored by professionals other than physicians (e.g. nurses, trainees) and finally agreed that the instrument face validity and discrimination (i.e. reliability and responsiveness) are better preserved if the PGA is scored by medical doctors with expertise in SLE [15, 28].

The PGA should not take into account subjective findings (e.g. headache, fatigue, arthralgia,...) not clearly related to disease activity (statement 1.11). The expert panel agreed that the PGA should capture only disease activity attributable to SLE (face and content validity), thus, potentially responsive to treatment (criterion validity and responsiveness). Pisetsky et al. [55] distinguished between manifestations strictly related to SLE disease activity (type 1) such as arthritis, vasculitis, nephritis, versus those related to damage or non-specific consequences of chronic disease (type 2) including fatigue, myalgia, depression, which are typically unresponsive to immunosuppressive therapy [55-58]. Arthralgia may be difficult to differentiate from inflammatory arthritis and, besides the physician's expertise, the use of advanced imaging might help in the attribution process in doubtful cases [59].

The PGA should be scored taking into account common clinical laboratory parameters such as urinalysis, serum creatinine level and blood cell count (statement 1.12). The PGA scoring was originally proposed with the inclusion of laboratory tests and lupus-related serology abnormalities (complement fractions and anti-dsDNA levels) [19, 60] but subsequently underwent an adaptation [1, 20, 43]. A single-clinician pilot study suggested significant variability between the PGA scored before and after knowledge of laboratory results including serology abnormalities [21]. A recent study conducted among a group of international lupus experts confirmed that knowledge of laboratory data (including

serology abnormalities) increased the construct validity [15] but not the reliability [15-16] of the PGA. Starting from this evidence, the panel reached a substantial agreement in defining the need for PGA taking into account common clinical laboratory parameters, such as urinalysis, serum creatinine level, and blood cell count, to satisfactorily assess disease activity, especially in renal and hematologic systems. However, consensus was not achieved for taking into account serology abnormalities such as complement and anti-dsDNA levels in PGA scoring.

PGA scale design

The PGA is rated on a VAS anchored by 0 and 3, incorporating the values 1 and 2 as inner markers (statements 2.1 and 2.2).

Our recent systematic literature review highlighted that the PGA measurements were rated on a scale from 0 to 3 in 54 out of 81 studies, a 0–10 scale in 12 studies, a 0–100 scale in 9 studies, as well as other less common scales [14]. Moreover, the composite SELENA-SLEDAI flare index [61], SLE responder index-4 (SRI-4) [13], BILAG-Based Composite Lupus Assessment (BICLA) [62], and the definitions of LLDAS [63] and DORIS remission [64, 65] include PGA as a 0-3 scale. Because most scientific evidence on the validity and psychometric properties of the PGA were based on a 0-3 VAS with anchored values (0, 1, 2, 3), the expert panel recommended that disease activity should be measured on such a scale (**Figure 1A**). No consensus was achieved on the physical representation of the scale, such as its actual length if scored on paper or an electronic device.

Practical considerations for PGA scoring

Preliminary training on a set of "training cases" is mandatory before a physician can rate the PGA (statement 3.1). According to previous evidence about a significantly lower inter-rater reliability (discrimination) of the PGA when assessed by untrained physicians than by trained investigators [28], preliminary training is recommended for trainees or non-experienced physicians before they independently rate the PGA.

For a given patient, the PGA should be preferably scored by the same rater at each visit (statement 3.2). Studies on PGA reproducibility showed a good-to-excellent ability to provide consistent scores among two or more physicians who evaluate disease activity at the same time (inter-rater reliability coefficients ranging from 0.67 to 0.96), even retrospectively [28, 40, 51, 66-69]. However, Gladman et al. demonstrated a better reproducibility when PGA was consecutively assessed by the same rater [70]. Therefore, the expert panel agreed that it is preferable, but not mandatory, to have the PGA scored by the same physician at consecutive visits when feasible.

The PGA should preferably be scored during the consultation and, if needed, amended as soon as all elements for PGA rating are available (statement 3.3). The PGA was initially validated to be performed at the end of a patient encounter and then amended when laboratory results including serology abnormalities became available [12]. Intending to reinforce the concept expressed in statement 1.12 and decrease variability, the panel agreed that the PGA should preferably be scored during the consultation. However, as in many clinical settings the laboratory results are only available later, the PGA can be revised once all elements required for rating are obtained.

The timeframe to consider when rating the PGA should be explicitly stated and is defined as the preceding month (statements 3.4, 3.5). The timeframe for disease activity assessment using PGA varies among studies. Following the Delphi rounds, and in line with many studies and the 30-day window for scoring in instruments such as SLEDAI and BILAG, the consensus process defined the preceding month as the timeframe to consider when scoring disease activity using the PGA.

Using a printed sheet, the PGA should be rated by putting a vertical tick on the scale, but web or app-based scales might also be used and scores can be reported as a continuous measure with one decimal (statements 3.6, 3.7, 3.8). Several smart health technologies have been recently developed [71], and implementing a digital version of the PGA in clinical practice is conceivable. As required so far in trials by the EMA and FDA, the PGA should be rated using a vertical tick on the VAS when using a printed version (**Figure 1B**). The expert panel agreed that the PGA score should be reported as a

number with one decimal, yielding a score with 30 degrees of freedom, with satisfactory inter-rater (0.67 to 0.98) and intra-rater (0.55 to 0.88) reliability in different settings [12, 44, 66, 69].

Relationship between PGA values and severity of disease activity

On the PGA scale, the anchor "0" refers to "No disease activity" and the anchor "3" to "Most severe disease activity" (statements 4.1, 4.2). Several different definitions have been used for the anchors at the two extremities of the VAS [14]. To standardize the actual wording of the PGA anchors, the expert panel defined them as "no disease activity" and "most severe disease activity", respectively. When referring to the most severe disease activity, the PGA assessor should refer to the overall clinical spectrum of SLE rather than to the most severe disease experienced by the individual patient currently being assessed. The anchor definitions should not be reported alongside the VAS in the PGA graph (**Figure 1**).

A PGA score ≥ 0.5 and ≤1 reflects mild disease activity, while values >1 and ≤2 reflect moderate disease activity, and values >2 and ≤3 reflect severe disease activity (statements 4.4, 4.5, 4.6). Different categorizations of disease activity according to the PGA were retrieved through the systemic literature review [14]. The most frequent is that proposed by Barr et al., defining mild activity between "0" and "≤1", moderate activity between ">1" and "≤2", and severe activity between ">2" and "3". In the PISCOS study, the experts agreed to implement this definition, including PGA <0.5 points [64, 65] as in the DORIS-definition of remission and the cut-off of PGA ≤1 as adopted for the prospectively validated definition of LLDAS [4]. The definition of "severe disease activity" as a continuous interval, rather than an anchor as proposed in the past, will allow differentiation among patients with high levels of disease activity and minimize the risk of a ceiling effect.

10

Discussion

The PISCOS study focused on the standardization of PGA scoring in the assessment of disease activity in SLE, and followed evidence-based and consensus-based methodology involving a large international panel of experts. An extensive preliminary literature review and analysis of the psychometric properties of the PGA, including truth, discrimination, and feasibility, were successively integrated within a Delphi consensus. This combined methodology was essential to standardize the PGA while maintaining its previously ascertained psychometric properties, enabling its use as a reliable tool for disease activity monitoring in SLE, alone or in combination with other glossary-based indices [11]. Our results are especially important given that the PGA is included among indices currently used as endpoints in SLE clinical trials (i.e. SRI, BICLA) and in the definitions of LLDAS and remission [39, 62-65].

Whether subjective findings and lupus serology should be taken into account in PGA scoring were important issues addressed by the expert panel. Type 2 symptoms such as fatigue, myalgia, depression, and anxiety are frequently reported by patients as a major source of quality of life impairment [55-58, 72-74]. Nevertheless, these symptoms are poorly associated with active SLE and typically unresponsive to standard immunosuppressive therapy [55]. Avoiding the potential overestimation of PGA scores by considering only findings clearly related to disease activity is therefore an important statement which was agreed upon during the Delphi process. This in no way diminishes the importance of the symptoms of SLE as they are reported by the patient. The central issue is that the physician must determine to what extent these symptoms are reflective of lupus activity rather than consequences of damage, comorbidity or medication side-effects.

No consensus was achieved on whether lupus serology should be taken into account when rating the PGA. A large number of lupus experts concurred that the PGA should not be influenced by the presence of lupus serologic abnormalities or by their levels, to avoid the risk of overestimation of disease activity and subsequently overtreatment. Behind this statement lies the evidence that some patients with SLE have clinically quiescent disease despite persistently increased anti-dsDNA antibody

levels and low complement [75], and do not require immunosuppressive therapy escalation [76,77]. Future studies could address the association of inclusion of serological results in the PGA with hard outcomes such as damage accrual, so that any future revision of the current recommendation would be evidence-based.

Limitations of the current study include those intrinsic to expert consensus processes in general, as they are based on levels of agreement rather than empirical data. Nonetheless, a standardized methodology incorporating contemporary recommendations for such studies was applied and the Delphi process was informed by a systematic literature review. The lack of consensus regarding the optimal length of the physical representation of the PGA scale is considered a minor limitation which would be operationally less relevant than the consensus reached for the use a 0-3 anchored VAS. Also, it is important to underline that the PGA is a widely used instrument in SLE. Therefore the aim of the PISCOS study was not to derive a new instrument but on the contrary standardize the use of the PGA in SLE.

In conclusion, the standardization of PGA use in the assessment of SLE will help increase homogeneity and reliability of PGA rating of disease activity across centers, in routine clinical practice, clinical research and clinical trials, while maintaining the validity, responsiveness and feasibility of the instrument.

References

1. Barr SG, Zonana-Nacach A, Magder LS, Petri M. Patterns of disease activity in systemic lupus erythematosus. Arthritis Rheum. 1999;42(12):2682–8.

2. Zen M, Bassi N, Nalotto L, Canova M, Bettio S, Gatto M, et al. Disease activity patterns in a monocentric cohort of SLE patients: a seven-year follow-up study. Clin Exp Rheumatol. 2012;30(6):856-63.

3. Keeling SO, Vandermeer B, Medina J, et al Measuring Disease Activity and Damage with Validated Metrics: A Systematic Review on Mortality and Damage in Systemic Lupus Erythematosus. J Rheumatol. 2018;45(10):1448-1461.

4. Golder V, Kandane-Rathnayake R, Huq M, et al. Lupus low disease activity state as a treatment endpoint for systemic lupus erythematosus: a prospective validation study. The Lancet Rheumatol 2019;1(2):e95-e102.

5. Piga M, Floris A, Cappellazzo G, Chessa E, Congia M, Mathieu A, et al. Failure to achieve lupus low disease activity state (LLDAS) six months after diagnosis is associated with early damage accrual in Caucasian patients with systemic lupus erythematosus. Arthritis Res Ther. 2017;19:247.

6. Zen M, Iaccarino L, Gatto M, Bettio S, Nalotto L, Ghirardello A, Punzi L, Doria A. Prolonged remission in Caucasian patients with SLE: prevalence and outcomes. Ann Rheum Dis. 2015 Dec;74(12):2117-22.

7. Petri M, Magder LS. Comparison of Remission and Lupus Low Disease Activity State in Damage Prevention in a United States Systemic Lupus Erythematosus Cohort. Arthritis Rheumatol. 2018 Nov;70(11):1790-1795.

8. Ugarte-Gil MF, Wojdyla D, Pons-Estel GJ, Catoggio LJ, Drenkard C, Sarano J, et al. Remission and Low Disease Activity Status (LDAS) protect lupus patients from damage occurrence: data from a multiethnic, multinational Latin American Lupus Cohort (GLADEL). Ann Rheum Dis. 2017 Dec;76(12):2071-2074.

13

9. Floris A, Piga M, Perra D, Chessa E, Congia M, Mathieu A, et al. Treatment Target in Newly Diagnosed Systemic Lupus Erythematosus: The Association of Lupus Low Disease Activity State and Remission With Lower Accrual of Early Damage. Arthritis Care Res (Hoboken). 2020 Dec;72(12):1794-1799.

10. Piga M, Arnaud L. The Main Challenges in Systemic Lupus Erythematosus: Where Do We Stand? J Clin Med. 2021 Jan 11;10(2):243.

11. Fanouriakis A, Kostopoulou M, Alunno A, Aringer M, Bajema I, Boletis JN, et al. 2019 update of the EULAR recommendations for the management of systemic lupus erythematosus. Ann Rheum Dis. 2019 Jun 1;78(6):736–45.

12. Petri M, Hellmann D, Hochberg M. Validity and reliability of lupus activity measures in the routine clinic setting. J Rheumatol. 1992;19(1):53–9

13. Furie RA, Petri MA, Wallace DJ, Ginzler EM, Merrill JT, Stohl W, et al. Novel evidence-based systemic lupus erythematosus responder index. Arthritis Care Res. 2009;61(9):1143–51

14. Chessa E, Piga M, Floris A, Devilliers H, Cauli A, Arnaud L. Use of Physician Global Assessment in systemic lupus erythematosus: a systematic review of its psychometric properties. Rheumatology (Oxford). 2020;59(12):3622-3632.

15. Aranow C, Askanase A, Oon S, Huq M, Calderone A, Morand EF, et al. Laboratory investigation results influence Physician's Global Assessment (PGA) of disease activity in SLE. Ann Rheum Dis. 2020;79(6):787-792.

16. Chessa E, Piga M, Arnaud L. Physician global assessment in systemic lupus erythematosus: can we rely on its reliability? Ann Rheum Dis. 2020 May 27. doi: 10.1136/annrheumdis-2020-217632.

17. Diamond IR, Grant RC, Feldman BM, Pencharz PB, Ling SC, Moore AM, et al. Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies. J Clin Epidemiol. 2014;67(4):401-9.

18. Beaton DE, Maxwell LJ, Shea BJ, Wells GA, Boers M, Grosskleg S, et al. Instrument Selection Using the OMERACT Filter 2.1: The OMERACT Methodology. J Rheumatol. 2019 Aug;46(8):1028–35.

19. Petri M, Genovese M, Engle E, Hochberg M. Definition, incidence, and clinical description of flare in systemic lupus erythematosus. A prospective cohort study. Arthritis Rheum. 1991;34(8):937–44

20. Györi N, Giannakou I, Chatzidionysiou K, Magder L, van Vollenhoven RF, Petri M. Disease activity patterns over time in patients with SLE: analysis of the Hopkins Lupus Cohort. Lupus Sci Med. 2017;4(1):e000192.

21. Aranow C. A pilot study to determine the optimal timing of the Physician Global Assessment (PGA) in patients with systemic lupus erythematosus. Immunol Res. 2015;63(1–3):167–9.

22. Jesus D, Rodrigues M, Matos A, Henriques C, Pereira da Silva JA, Inês LS. Performance of SLEDAI-2K to detect a clinically meaningful change in SLE disease activity: a 36–month prospective cohort study of 334 patients. Lupus. 2019;28(5):607–12.

23. Jesus D, Matos A, Henriques C, Zen M, Larosa M, Iaccarino L, et al. Derivation and validation of the SLE Disease Activity Score (SLE-DAS): a new SLE continuous measure with high sensitivity for changes in disease activity. Ann Rheum Dis. 2019;78(3):365–71

24. Parodis I, Emamikia S, Gomez A, Gunnarsson I, van Vollenhoven RF, Chatzidionysiou K. Clinical SLEDAI-2K zero may be a pragmatic outcome measure in SLE studies. Expert Opin Biol Ther. 2019;19(2):157–68

25. Merrill JT, Petri MA, Buyon J, Ramsey-Goldman R, Kalunian K, Putterman C, et al. Erythrocyte-bound C4d in combination with complement and autoantibody status for the monitoring of SLE. Lupus Sci Med. 2018;5(1):e000263.

26. Fanouriakis A, Adamichou C, Koutsoviti S, Panopoulos S, Staveri C, Klagou A, et al. Low disease activity—irrespective of serologic status at baseline—associated with reduction of corticosteroid dose and number of flares in patients with systemic lupus erythematosus treated with belimumab: A real-life observational study. Semin Arthritis Rheum. 2018;48(3):467–74.

27. Gandhi N, Arora S, Sengupta M, Sequeira W, Jolly MM, Block JA. Validation of SIMPLE Index for Lupus Disease Activity. J Clin Rheumatol. 2018;24(6):313–8

28. Askanase AD, Nguyen SC, Costenbader K, Lim SS, Kamen D, Aranow C, et al. Comparison of the Lupus Foundation of America-Rapid Evaluation of Activity in Lupus to More Complex Disease Activity Instruments As Evaluated by Clinical Investigators or Real-World Clinicians. Arthritis Care Res. 2018;70(7):1058–63.

29. Parodis I, Sjöwall C, Jönsen A, Ramsköld D, Zickert A, Frodlund M, et al. Smoking and pre-existing organ damage reduce the efficacy of belimumab in systemic lupus erythematosus. Autoimmun Rev. 2017;16(4):343–51

30. Devilliers H, Bonithon-Kopp C, Jolly M. The lupus impact tracker is responsive to changes in clinical activity measured by the systemic lupus erythematosus responder index. Lupus. 2017;26(4):396–402

31. Askanase A, Li X, Pong A, Shum K, Kamp S, Carthen F, et al. Preliminary test of the LFA rapid evaluation of activity in lupus (LFA-REAL): an efficient outcome measure correlates with validated instruments. Lupus Sci Med. 2015;2(1):e000075.

32. Touma Z, Urowitz MB, Taghavi-Zadeh S, Ibañez D, Gladman DD. Systemic lupus erythematosus disease activity Index 2000 Responder Index 50: sensitivity to response at 6 and 12 months. Rheumatology (Oxford). 2012;51(10):1814–9.

33. Mina R, Klein-Gitelman MS, Ravelli A, Beresford MW, Avcin T, Espada G, et al. Inactive disease and remission in childhood-onset systemic lupus erythematosus. Arthritis Care Res. 2012;64(5):683–93.

34. Brunner HI, Mina R, Pilkington C, Beresford MW, Reiff A, Levy DM, et al. Preliminary criteria for global flares in childhood-onset systemic lupus erythematosus. Arthritis Care Res. 2011;63(9):1213–23.

35. Navarra SV, Guzmán RM, Gallacher AE, Hall S, Levy RA, Jimenez RE, et al. Efficacy and safety of belimumab in patients with active systemic lupus erythematosus: a randomised, placebo-controlled, phase 3 trial. Lancet 2011;377(9767):721–31.

36. Vilá LM, Alarcón GS, McGwin G, Bastian HM, Fessler BJ, Reveille JD, et al. Systemic lupus erythematosus in a multiethnic US cohort, XXXVII: association of lymphopenia with clinical manifestations, serologic abnormalities, disease activity, and damage accrual. Arthritis Rheum. 2006;55(5):799–806

37. Ward MM, Marx AS, Barry NN. Psychological distress and changes in the activity of systemic lupus erythematosus. Rheumatology (Oxford). 2002;41(2):184–8.

38. Ward MM, Marx AS, Barry NN. The rating scale preference measure as an evaluative measure in systemic lupus erythematosus. Lupus. 2000;9(9):696–701.

39. Ward MM, Marx AS, Barry NN. Comparison of the validity and sensitivity to change of 5 activity indices in systemic lupus erythematosus. J Rheumatol. 2000;27(3):664–70.

40. Petri M, Howard D, Repke J. Frequency of lupus flare in pregnancy: The hopkins lupus pregnancy center experience. Arthritis Rheum. 1991;34(12):1538–45

41. Touma Z, Gladman DD, Ibañez D, Urowitz MB. Development and initial validation of the systemic lupus erythematosus disease activity index 2000 responder index 50. J Rheumatol. 2011;38(2):275–84.

42. Gladman DD, Goldsmith CH, Urowitz MB, Bacon P, Bombardier C, Isenberg D, et al. Sensitivity to change of 3 Systemic Lupus Erythematosus Disease Activity Indices: international validation. J Rheumatol. 1994;21(8):1468–71.

43. Ho A, Barr SG, Magder LS, Petri M. A decrease in complement is associated with increased renal and hematologic activity in patients with systemic lupus erythematosus. Arthritis Rheum. 2001;44(10):2350–7.

44. Askanase AD, Nguyen SC, Costenbader K, Lim SS, Kamen D, Aranow C, et al. Comparison of the Lupus Foundation of America-Rapid Evaluation of Activity in Lupus to More Complex Disease Activity Instruments As Evaluated by Clinical Investigators or Real-World Clinicians. Arthritis Care Res. 2018;70(7):1058–63.

45. Zen M, Iaccarino L, Gatto M, Saccon F, Larosa M, Ghirardello A, et al. Lupus low disease activity state is associated with a decrease in damage progression in Caucasian patients with SLE, but overlaps with remission. Ann Rheum Dis. 2018;77(1):104–10

46. Golder V, Kandane-Rathnayake R, Hoi AY-B, Huq M, Louthrenoo W, An Y, et al. Association of the lupus low disease activity state (LLDAS) with health-related quality of life in a multinational prospective study. Arthritis Res Ther. 2017;19(1):62

47. Schneider M, Mosca M, Pego-Reigosa J-M, Gunnarsson I, Maurel F, Garofano A, et al. Cross-cultural validation of Lupus Impact Tracker in five European clinical practice settings. Rheumatology (Oxford). 2017;56(5):818–28.

48. Mankee A, Petri M, Magder LS. Lupus anticoagulant, disease activity and low complement in the first trimester are predictive of pregnancy loss. Lupus Sci Med. 2015;2(1):e000095.

49. Furie R, Petri MA, Strand V, Gladman DD, Zhong ZJ, Freimuth WW, et al. Clinical, laboratory and health-related quality of life correlates of Systemic Lupus Erythematosus Responder Index response: a post hoc analysis of the phase 3 belimumab trials. Lupus Sci Med. 2014;1(1):e000031

50. Thanou A, Chakravarty E, James JA, Merrill JT. How should lupus flares be measured? Deconstruction of the safety of estrogen in lupus erythematosus national assessment-systemic lupus erythematosus disease activity index flare index. Rheumatology (Oxford). 2014;53(12):2175–81.

51. Isenberg DA, Allen E, Farewell V, D'Cruz D, Alarcón GS, Aranow C, et al. An assessment of disease flare in patients with systemic lupus erythematosus: a comparison of BILAG 2004 and the flare version of SELENA. Ann Rheum Dis. 2011;70(1):54–9.

52. Strand V, Petri M, Kalunian K, Gordon C, Wallace DJ, Hobbs K, et al. Epratuzumab for patients with moderate to severe flaring SLE: health-related quality of life outcomes and corticosteroid use in the randomized controlled ALLEVIATE trials and extension study SL0006. Rheumatology (Oxford). 2014;53(3):502–11.

53. Fatemi A, Raeisi A, Sayedbonakdar Z, Smiley A. Sensitivity analyses of four systemic lupus erythematosus disease activity indices in predicting the treatment changes in consecutive visits: a longitudinal study. Clin Rheumatol. 2018 Apr;37(4):955–62.

54. Apostolopoulos D, Kandane-Rathnayake R, Louthrenoo W, Luo SF, Wu YJ, Lateef A, et al. Factors associated with damage accrual in patients with systemic lupus erythematosus with no clinical or serological disease activity: a multicentre cohort study. Lancet Rheumatol 2020; 2: e24–30.

55. Pisetsky DS, Clowse MEB, Criscione-Schreiber LG, Rogers JL. A Novel System to Categorize the Symptoms of Systemic Lupus Erythematosus. Arthritis Care Res (Hoboken). 2019 Jun;71(6):735-741.

56. Arnaud L, Gavand, PE, Voll R, Schwarting A, Maurier F, Blaison G, et al. Predictors of fatigue and severe fatigue in a large international cohort of patients with systemic lupus erythematosus and a systematic review of the literature. Rheumatology (Oxford). 2019, 58, 987–996.

57. Mertz P, Schlencker A, Schneider M, Gavand PE, Martin T, Arnaud L. Towards a practical management of fatigue in systemic lupus erythematosus. Lupus Sci. Med. 2020, 7, e000441.

58. Arnaud L, Mertz P, Amoura Z, Voll RE, Schwarting A, Maurier F, et al. Patterns of fatigue and association with disease activity and clinical manifestations in systemic lupus erythematosus. Rheumatology (Oxford). 2021;60(6):2672-2677.

59. Gabba A, Piga M, Vacca A, Porru G, Garau P, Cauli A, et al. Joint and tendon involvement in systemic lupus erythematosus: an ultrasound study of hands and wrists in 108 patients. Rheumatology (Oxford). 2012;51(12):2278-85.

60. Liang MH, Socher SA, Roberts WN, Esdaile JM. Measurement of systemic lupus erythematosus activity in clinical research. Arthritis Rheum. 1988;31(7):817–25.

61. Buyon JP, Petri MA, Kim MY, Kalunian KC, Grossman J, Hahn BH, et al. The effect of combined estrogen and progesterone hormone replacement therapy on disease activity in systemic lupus erythematosus: a randomized trial. Ann Intern Med. 2005;142:953–62

62. Wallace DJ, Kalunian K, Petri MA, Strand V, Houssiau FA, Pike M, et al. Efficacy and safety of epratuzumab in patients with moderate/severe active systemic lupus erythematosus: results from EMBLEM, a phase IIb, randomized, double-blind, placebo-controlled, multicentre study. Ann Rheum Dis. 2014;73:183–90.

63. Franklyn K, Lau CS, Navarra SV, Louthrenoo W, Lateef A, Hamijoyo L, et al. Definition and initial validation of a Lupus Low Disease Activity State (LLDAS). Ann Rheum Dis. 2016;75(9):1615–21.

64. van Vollenhoven R, Voskuyl A, Bertsias G, Aranow C, Aringer M, Arnaud L, et al. A framework for remission in SLE: consensus findings from a large international task force on definitions of remission in SLE (DORIS). Ann Rheum Dis. 2017;76(3):554–61.

65. van Vollenhoven R, Bertsias G, Doria A , et al. 2021 DORIS definition of remission in SLE: final recommendations from an international task force. Lupus Sci Med. 2021;8:e000538.

66. Liang MH, Socher SA, Larson MG, Schur PH. Reliability and validity of six systems for the clinical assessment of disease activity in systemic lupus erythematosus. Arthritis Rheum. 1989;32(9):1107–18

67. FitzGerald JD, Grossman JM. Validity and reliability of retrospective assessment of disease activity and flare in observational cohorts of lupus patients. Lupus. 1999;8(8):638–44.

68. Touma Z, Gladman DD, Su J, Anderson N, Urowitz MB. A novel lupus activity index accounting for glucocorticoids: SLEDAI-2K glucocorticoid index. Rheumatology. 2018;57(8):1370–6.

69. Touma Z, Urowitz MB, Fortin PR, Landolt C, Toloza SM, Riddell C, et al. Systemic lupus erythematosus disease activity index 2000 responder index-50: a reliable index for measuring improvement in disease activity. J Rheumatol. 2011;38(5):868–73

70. Gladman DD, Goldsmith CH, Urowitz MB, Bacon P, Bombardier C, Isenberg D, et al. Sensitivity to change of 3 Systemic Lupus Erythematosus Disease Activity Indices: international validation. J Rheumatol. 1994;21(8):1468–71.

71. Bergier H, Duron L, Sordet C, Kawka L, Schlencker A, Chasset F, et al. Digital health, big data and smart technologies for the care of patients with systemic autoimmune diseases: Where do we stand? Autoimmun Rev. 2021;20(8):102864.

72. Dietz B, Katz P, Dall'Era M, Murphy LB, Lanata C, Trupin L, et sl. Major Depression and Adverse Patient-Reported Outcomes in Systemic Lupus Erythematosus: Results From a Prospective Longitudinal Cohort. Arthritis Care Res (Hoboken). 2021;73(1):48-54.

73. Nowicka-Sauer K, Hajduk A, Kujawska-Danecka H, Banaszkiewicz D, Smoleńska Ż, Czuszyńska Z, Siebert J. Illness perception is significantly determined by depression and anxiety in systemic lupus erythematosus. Lupus. 2018;27(3):454-460.

74. Piga M, Congia M, Gabba A, Figus F, Floris A, Mathieu A, et al. Musculoskeletal manifestations as determinants of quality of life impairment in patients with systemic lupus erythematosus. Lupus. 2018 Feb;27(2):190-198.

75. Gladman DD, Urowitz MB, Keystone EC. Serologically active clinically quiescent systemic lupus erythematosus: a discordance between clinical and serologic features. Am J Med 1979; 66: 210–5.

76. Steiman AJ, Gladman DD, Ibañez D, Urowitz MB. Outcomes in patients with systemic lupus erythematosus with and without a prolonged serologically active clinically quiescent period. Arthritis Care Res (Hoboken). 2012 Apr;64(4):511-8.

77. Steiman AJ, Gladman DD, Ibañez D, Urowitz MB. Prolonged serologically active clinically quiescent systemic lupus erythematosus: frequency and outcome. J Rheumatol. 2010;37(9):1822-7.

Table 1. List of statements reaching consensus during the three rounds of the Delphi process.

PICOS statements	LoA (mean ± SD)	% of votes ≥9/10
1. Overarching principles for scoring PGA		
1.1 The PGA is an outcome instrument for measuring disease activity in SLE	9.7 ± 0.9	94.9%
1.2 The PGA is a visual analogue scale for measuring disease activity in SLE.	9.8 ± 0.4	97.5%
1.3 The PGA should be used to rate overall disease activity in SLE	9.7 ± 0.8	96.2%
1.4 The PGA rating should reflect the severity of active manifestations	9.6 ± 1.4	93.7%
1.5 The PGA should be used to rate disease activity only (NOT organ damage)	9.8 ± 0.9	96.2%
1.6 The PGA reflects the clinician's judgment about disease activity in SLE	9.9 ± 0.5	96.2%
1.7 The PGA should be scored independently of pre-specified thresholds used in other scores	9.6 ± 0.9	93.7%
1.8 The PGA should assess disease activity according to the physician's perspective	9.8 ± 0.5	97.5%
1.9 The PGA should be rated only by a medical doctor	9.6 ± 1.0	94.9%
1.10 Only physicians with expertise in SLE can rate disease activity using the PGA	9.5 ± 0.9	91.1%
1.11 The PGA should not take into account subjective findings (e.g. headache, arthralgias, fatigue.) if those are not clearly related to disease activity	9.5 ± 1.4	93.7%
1.12 The PGA should be scored taking into account common clinical laboratory parameters such as urinalysis, serum creatinine level and blood cell count	9.6 ± 1.1	93.7%
2. PGA scale design		
2.3 The PGA is rated on a 0-3 scale	9.7 ± 0.5	94.9%
2.4 The PGA should incorporate inner markers for 1 and 2 (with a graph showing a PGA and 0, 1, 2, 3) without any additional comment regarding whether this is mild, moderate, etc.	9.5 ± 1.0	93.6%
3. Practical considerations for PGA scoring		
3.1 Preliminary training on a set of "training cases" is mandatory before a physician can rate the PGA	9.2 ± 1.8	89.8%
3.2 For a given patient, the PGA should be preferably scored by the same rater at each visit	9.6 ± 1.1	94.9%
3.3 The PGA should preferably be scored during the consultation and, if needed, amended as soon as all elements for PGA rating are available	9.7 ± 0.8	92.4%
3.4 The timeframe to consider when rating the PGA should be explicitly stated	9.5 ± 1.3	91.1%
3.5 The timeframe to consider when rating the PGA is the last month	9.5 ± 1.4	89.8%
3.6 The PGA should be rated by putting a vertical tick (recommended in trials by EMA and FDA)	9.9 ± 0.3	100%
3.7 The PGA can be scored either on a printed sheet or using a web/app-based scale	9.9 ± 0.3	100%
3.8 The PGA score should be expressed using a continuous measure with one decimal (e.g. 2.3 on a 0-3 scale)	9.5 ± 1.4	93.7%
4. Relationship between PGA values and severity of disease activity		
4.1 On the PGA scale, the anchor "0" refers to "No disease activity"	9.9 ± 0.2	100%
4.2 On the PGA scale , the anchor "3" refers to "Most severe disease activity"	9.9 ± 0.2	100%
4.3 A PGA score ≥ 0.5 and <= 1 reflects MILD disease activity	9.7 ± 1.0	93.7%
4.4 A PGA Score >1 and <=2 reflects MODERATE disease activity	9.8 ± 0.5	96.2%
4.5 A PGA score >2 and <=3 reflects SEVERE disease activity	9.8 ± 0.5	96.2%

LoA : level of agreement.

Figure 1. The PGA graph and scoring recommended by the PISCOS experts. In (A) the PGA graph consisting of a 0-3 VAS with anchored values. In (B) the correct way to score the PGA by putting a vertical tick on 0-3 VAS as a continuous measure with one decimal (e.g. 1.4 on a 0-3 scale).
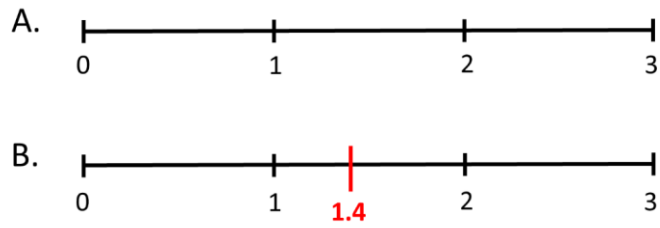
SUPPLEMENTARY DOCUMENTS

Supplementary document #1:

Statements assessed during the PISCOS study

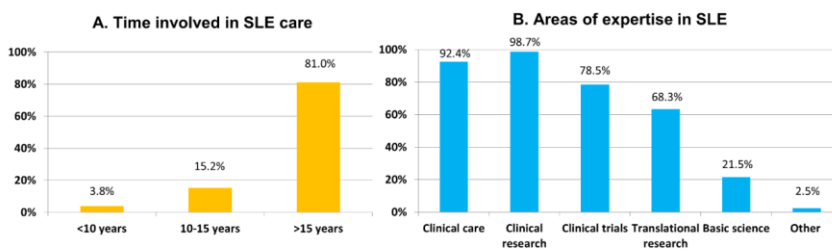| |
|---|
| 1. The PGA is a visual analogue scale for measuring disease activity in SLE |
| 2. The PGA is an outcome instrument for measuring disease activity in SLE |
| 3. The PGA reflects the clinician's judgment about disease activity in SLE |
| 4. The PGA should be used to rate overall disease activity in SLE |
| 5. The PGA should assess disease activity according to the physician's perspective |
| 6. The timeframe to consider when rating the PGA should be explicitly stated (it will be defined later in the study) |
| 7. The PGA should be used to rate disease activity only (NOT organ damage) |
| 8. The PGA should not take into account subjective findings *(e.g. headache, fatigue, arthralgias...)* if those are not clearly related to disease activity |
| 9. The PGA rating should reflect the severity of active manifestations |
| 10. The PGA should be scored independently of pre-specified thresholds used in other scores (e.g., in SELENA-SLEDAI, leucopenia is only rated if <3000/mm3) |
| 11. To avoid any interference with other instruments and threshold effects, the PGA should be rated BEFORE any other instrument for measuring disease activity |
| 12. The PGA should be scored taking into account common laboratory parameters such as urinalysis, serum creatinine level and blood cell count |
| 13. The PGA should be scored taking into account lupus serology (serum complement and anti-dsDNA level) |
| 14. The PGA should be rated only by a medical doctor |
| 15. Any physician involved in the care of SLE patients can rate disease activity using the PGA |
| 16. Only physicians with expertise in SLE can rate disease activity using the PGA |
| 17. Preliminary training on a set of "training cases" is mandatory before a physician can rate the PGA |
| 18. Trainees, interns and residents involved in the care of SLE patients can rate disease activity using the PGA |
| 19. For a given patient, the PGA should be preferably scored by the same rater at each visit |
| 20. When performed at different visits, the PGA rating should take into account the value given at the previous visit. |
| 21. What is the timeframe to consider when rating the PGA: (Multiple choice: last 10 days, last 15 days, last 28 days., last 30 days, other) |
| 22. The PGA should preferably be scored during the consultation and, if needed, amended as soon as all elements for PGA rating are available |
| 23. The PGA should be scored within [X] hours after the visit is completed (Multiple choice 1 hour, 4 hours, 6 hours, 12 hours, 24 hours, other) |
| 24. The PGA is rated on the 0-3 scale |
| 25. The length of the PGA scale (on the scoring sheet) should be 3 cm |
| 26. The length of the PGA scale (on the scoring sheet) should be 10 cm (then converted to a 0-3 score) |
| 27. The PGA should incorporate inner markers for 1 and 2 without any additional comment regarding whether this is mild, moderate, etc. |
| 28. The PGA scale should incorporate inner markers at every 0.5 points |
| 29. The PGA scale should incorporate inner markers every 0.1 points |

30. The PGA score should be expressed using a continuous measure with one decimal (e.g. 2.3 on a 0-3 scale)
31. The PGA score should be expressed using a continuous measure with 2 decimals (e.g. 2.34 on a 0-3 scale)
32. On the PGA scale, the anchor "0" refers to (Multiple choice: "No disease activity", "Inactive disease", "Absence of disease activity", Other):
33. On the PGA scale , the anchor "3" refers to (Multiple choice: "Most active disease imaginable", "Maximum disease activity", "Most severe disease activity", "Severe disease activity", other):
34. A PGA score ≥ 0.5 and ≤1 reflects MILD disease activity
35. A PGA Score >1 and ≤2 reflects MODERATE disease activity
36. A PGA score >2 and ≤3 reflects SEVERE disease activity
37. A PGA score >2.5 and ≤3 reflects ORGAN-THREATENING disease
38. The PGA should be rated by putting (Multiple choice: figure A: an X on the scale; figure B: a diagonal tick; figure C: a vertical tick; other)
39. The PGA can only be scored on a printed sheet
40. The PGA can be scored either on a printed sheet or using a web/app-based scale
41. If no printed/app/web PGA scale is available, the PGA can be scored by directly providing a numerical value between 0 and 3 instead of using the scale (instead of ticking and then measuring the length up to the tick)

Supplementary document #2

Full description of expert panels

A panel of 79 lupus experts consisting a multidisciplinary investigator group from the five continents was involved in the Delphi exercise. Overall, 66 (83.5%) lupus experts reported Rheumatology as their primary specialty, 11 (13.9%) internal medicine, 1 (1.3%) dermatology, and 1 (1.3%) clinical immunology. The geographic distribution of the panel included 39 (49.3%) experts from Europe, 15 (19.0%) from North America, 13 (16.5%) from Asia-Pacific, 11 (13.9%) from Latin America and 1 (1.3%) from Africa. The vast majority of participants had more than 15 years of involvement in SLE (Supplementary figure 1A) and the most represented areas of expertise (Supplementary Figure 1B) were clinical research (n = 78; 98.7%) and clinical care (n = 73; 92.4%).

Supplementary Figure 1. The panel (A) presents the years on involvement in SLE and the panel (B) the areas of expertisein SLE among the participants in the Delphi survey.

Supplementary document #3

Statements reaching ≥ 75% agreement during round 1

#	Statement	Agreement (%)/ Mean value (SD)
1	The PGA is **a visual analogue scale** for measuring disease activity in SLE.	Agreement level = 89.74% mean (SD) = 9.1 ± 1.8
2	The PGA is **an outcome instrument** for measuring disease activity in SLE	Agreement level = 78.21% mean (SD) = 7.8 ± 2.7
3	The PGA **reflects the clinician's judgment** about disease activity in SLE	Agreement level = 96.15% mean (SD) = 9.4 ± 1.2
4	The PGA should be used to **rate overall disease activity** in SLE	Agreement level = 89.74% mean (SD) = 8.7 ± 2.2
5	The PGA should assess disease activity **according to the physician's perspective**	Agreement level = 94.87% mean (SD) = 9.1 ± 1.3
6	The **timeframe** to consider when rating the PGA **should be explicitly stated** (it will be defined later in the study)	Agreement level = 97.44% mean (SD) = 9.5 ± 1.1
7	The PGA should be used **to rate disease activity only** (NOT organ damage)	Agreement level = 93.59% mean (SD) = 9.2 ± 2.0
9	The PGA rating should reflect the **severity of active manifestations**	Agreement level = 83.33% mean (SD) = 8.4 ± 2.3
10	The PGA should be scored **independently of pre-specified thresholds used in other scores** (e.g., in SELENA-SLEDAI, leucopenia is only rated if <3000/mm3)	Agreement level = 84.62% mean (SD) = 8.3 ± 2.3
14	The PGA should be rated **only by a medical doctor**	Agreement level = 87.18% mean (SD) = 8.6 ± 2.3
19	For a given patient, **the PGA should be preferably scored by the same rater at each visit**	Agreement level = 85.90% mean (SD) = 8.2 ± 2.1
33	A PGA score between ≥ 0.5 and ≤1 reflects MILD disease activity	Agreement level = 89.61% mean (SD) = 8.4 ± 1.8
34	A PGA Score >1 and ≤ 2 reflects MODERATE disease activity	Agreement level = 90.91% mean (SD) = 8.5 ± 1.6
35	A PGA score >2 and ≤ 3 reflects SEVERE disease activity	Agreement level = 93.51% mean (SD) = 8.7 ± 1.5
37	The PGA should be rated by putting (Multiple choice: figure A: an X on the scale; figure B: a diagonal tick; figure C: a vertical tick; other) Answer: a vertical tick	a vertical tick n= 61; 77,21%
39	The PGA can be scored either on a printed sheet or using a web/app-based scale	Agreement level = 92.11% mean (SD) = 8.8 ± 2.0

Statements reaching ≥ 75% agreement during round 2

#	Statement	2nd round Agreement (%) / Mean value (SD)
1	The PGA should not take into account *subjective* findings (e.g. headache, arthralgias, fatigue) if those are not clearly related to disease activity	Agreement level = 81.00% mean (SD) = 7.4 ± 2.6
3	The PGA should be scored taking into account common laboratory parameters such as urinalysis, serum creatinine level and blood cell count	Agreement level = 89.87% mean (SD) = 8.5 ± 1.8
5	Only physicians with expertise in SLE can rate disease activity using the PGA	Agreement level = 77.22% mean (SD) = 7.8 ± 2.1
6	Preliminary training on a set of "training cases" is mandatory before a physician can rate the PGA	Agreement level = 84.81% mean (SD) = 8.3 ± 1.8
8	What is the timeframe to consider when rating the PGA? Answer: Last month	Last month n=60 75.95%
9	The PGA should preferably be scored during the consultation and, if needed, amended as soon as all elements for PGA rating are available	Agreement level = 88.61% mean (SD) = 8.2 ± 2.0
10	The PGA is rated on the 0-3 scale	Agreement level = 88.61% mean (SD) = 8.7 ± 1.9
12	On the PGA scale, the anchor "0" refers to (please specify the most appropriate wording)? Answer: No disease activity	No disease activity n=67; 86.10%
14	The PGA should incorporate inner markers for 1 and 2 (with a graph showing a PGA and 0, 1, 2, 3) without any additional comment regarding whether this is mild, moderate, etc.	Agreement level = 87.34% mean (SD) = 8.4 ± 2.3

Statements reaching ≥ 75% agreement during round 3

#	Statement	3rd round Agreement (%)/ Mean value (SD)
1	On the PGA scale , the anchor "3" refers to (please specify the most appropriate wording): Answer: Most severe disease activity	Most severe disease activity n=60; 75.95%
3	The PGA score should be expressed using a continuous measure with one decimal (e.g. 2.3 on a 0-3 scale)	Agreement level = 75.95% mean (SD) = 7.4 ± 2.9