



Published in final edited form as:

Arthritis Care Res (Hoboken). 2022 August ; 74(8): 1269–1276. doi:10.1002/acr.24574.

Physician Global Assessment as a Disease Activity Measure for Relapsing Polychondritis

Emily Rose, BS¹, Marcela A. Ferrada, MD¹, Kaitlin A. Quinn, MD¹, Wendy Goodspeed, RN¹, Laurent Arnaud, MD, PhD², Aman Sharma, MD³, Hajime Yoshifuji, MD, PhD⁴, Jeff Kim, MD⁵, Clint Allen, MD⁶, Arlene Sirajuddin, MD⁷, Marcus Chen, MD⁷, Peter C. Grayson, MD, MSc¹

¹Systemic Autoimmunity Branch, National Institutes of Health, NIAMS, Bethesda, MD, USA

²Department of Rheumatology, National Reference Center for Autoimmune Diseases (RESO), Hôpitaux Universitaires de Strasbourg, Strasbourg, France

³Clinical Immunology and Rheumatology Wing, Department of Internal Medicine, Postgraduate Institute of Medical Education and Research, Sector-12, Chandigarh-160012, India

⁴Department of Rheumatology and Clinical Immunology, Graduate School of Medicine, Kyoto University

⁵Office of Clinical Director, National Institutes on Deafness and Other Communication Disorder, National Institutes of Health, Bethesda, Maryland, USA

⁶Otolaryngology, National Institute on Deafness and Other Communication Disorders, Bethesda, Maryland, USA

⁷National Heart Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland, USA.

Abstract

Objective: Relapsing polychondritis (RP) is a systemic inflammatory disorder of cartilage that lacks validated disease activity measures. Physician global assessment (PhGA), a measure of disease activity commonly used in rheumatologic diseases, has not been tested in a cohort of patients with RP.

Methods: Adult patients in an observational cohort of RP underwent standardized, comprehensive evaluation at approximately 6-month intervals. PhGA was scored by three physicians from the evaluating institution on a scale of 0 to 10 for each visit. A random subset of twenty visits was scored by three, independent physicians not affiliated with the evaluating institution. Treatment change between consecutive visits was categorized as increased, decreased or unchanged.

Results: 78 patients were evaluated over 164 visits. The interclass correlation coefficient (ICC) (2, 1) for the three raters from the evaluating institution was excellent (0.79, 95% CI: 0.73–0.84) but was poor in the subset of cases scored by the additional raters (ICC (2,1) = 0.27, 95% CI: –0.01–0.53). Median PhGA was 3 (range 0–7). PhGA weakly correlated with CRP ($r_s = 0.30$, $p <$

0.01). In response to increased treatment, median PhGA decreased from 3 (IQR: 2–4) to 2 (IQR: 2–3) ($p < 0.01$) but rarely went to 0.

Conclusion: Within a single-center, PhGA can be used to quantify disease activity and monitor disease response in RP. Persistent disease activity despite treatment, rather than a relapsing-remitting pattern, is observed for most patients with RP. Reliability of PhGA may not generalize across different institutions. A validated disease-specific activity index is needed in RP.

Keywords

relapsing polychondritis; outcome measures; physician global assessment; disease activity

Relapsing polychondritis (RP) is a rare systemic inflammatory disorder that affects multiple organs with a predilection for cartilaginous structures such as the ear, nose, airway, and joints. (1) RP can also affect the eyes, central nervous system, vasculature, inner ear, and skin. (2) (1) Given the rarity of the disease, clinical assessment has not been standardized, and the disease lacks validated measures of disease activity.(3)

Clinical assessment tools are useful to measure disease activity and treatment response in clinical trials and in daily practice. The Relapsing Polychondritis Disease Activity Index (RPDAI) is a proposed tool to quantify disease activity in RP. (4) The RPDAI was developed using clinical vignettes rather than patient data and has not been validated in independent cohorts of patients with RP. When deriving weights for items within the RPDAI, the Physician Global Assessment (PhGA) score was used as the gold standard to quantify disease activity, yet the PhGA has also never been systematically studied or validated for use in RP.

The PhGA is frequently used to measure disease activity and track response to treatment in a variety of rheumatologic diseases. (5) (6) (7, 8) (9) (10) (11, 12) (13) (14) Because performance characteristics of the PhGA may vary by disease, the PhGA should be tested specifically in a cohort of patients with RP.

Therefore, this study sought to characterize PhGA to measure disease activity using data from a prospective, observational cohort of patients with RP.

Patients and Methods

Study Population

Patients 18 years who fulfilled existing diagnostic criteria for RP were recruited into a prospective, observational cohort at the National Institutes of Health (NIH) Clinical Center from August 2016 to October 2019 (NCT02257866). (15–17) Patients were recruited at any point in the disease course from different countries with no geographic or demographic restrictions. Consecutive patients with clinical visits to the NIH in the time period stated were included. Patients were evaluated by an investigative team with expertise in RP at approximately 6-month intervals. Written informed consent was obtained from all patients, and the study was approved by local ethics review at the NIH.

Data Collection

At each visit, patients underwent a standardized comprehensive evaluation that included a clinical rheumatology and otolaryngology evaluation with direct laryngoscopy, laboratory studies, audiology, echocardiogram, chest imaging with a dynamic computed tomography (CT) scan and any additional clinically appropriate evaluations such as ophthalmology or pulmonology. Clinical signs and symptoms related to common organ involvement due to RP were recorded as either present or absent within the past month. Blood was collected on the day of the study visit for laboratory assessments including ESR, CRP, and complement levels. All laboratory testing was performed in the NIH Department of Laboratory Medicine. Elevated acute phase reactants were defined based on levels above the laboratory normal range which included the following: CRP 0–4.99 mg/L, ESR 0–42 mm/hr for females and ESR 0–25 mm/hr for males.

Physician global assessment (PhGA) scoring

From the data collected at each patient visit, de-identified clinical vignettes were created that summarized current symptoms, physical examination findings, laboratory results, imaging findings, and other diagnostics (e.g. audiogram) at each visit. Using these clinical vignettes derived from patient data, three practicing rheumatologists from the evaluating institution (MF, KQ, PG) with expertise caring for patients with RP scored the PhGAs for each study visit blinded to each other. The amount of direct clinical observation of patients in the cohort differed between the raters, and one rater only directly evaluated <10% of the patients in the cohort. PhGA was scored on a scale of 0 defined as absence of disease activity to 10 defined as maximum disease activity, based on an overall assessment of disease activity in the past month. Raters were instructed to consider only symptoms directly attributable to disease activity rather than damage. The raters agreed that a score of 0 would define clinical remission. Ratings were analyzed for discordance, defined as a difference of 3 or more points on the PhGA between any 2 raters. (18) For visits with concordant ratings, a final PhGA was assigned based on averaging the three raters and rounding to the nearest whole number. Discordant PhGAs were adjudicated by group discussion among the raters. To compare frequencies of organ involvement across levels of PhGA, PhGAs were categorized in the following manner: PhGA=0, PhGA=1 or 2, PhGA=3 or 4 and PhGA = 5. To determine whether acute phase reactants impacted rating of PhGA, a subset of 60 randomly selected study visits were scored a second time, except ESR and CRP values were withheld. Scores when ESR and CRP were withheld were compared to scores when this information was available. To describe the cumulative number of symptoms, a disease activity summary score was created. The disease activity summary score was defined as the total number of active disease manifestations on a scale of 0–9 that a patient experienced in the previous month of the following 9 items: auricular chondritis, nasal chondritis, chondritis of the chest wall, cardiovascular involvement, respiratory chondritis, eye involvement, hearing loss, vestibular dysfunction, and joint involvement. Each item equaled one point.

To assess interrater reliability of PhGA among physicians who practice outside of the evaluating institution, clinical vignettes from 20 unique patients (25% of the cohort) were randomly selected and independently scored by three physicians (LA, AS, and HY) from different regions of the world (France, India and Japan, respectively). Each of these raters

has > 10 years of experience caring for patients with RP, and none of these raters directly evaluated any of the patients in this study.

Response to Treatment

Response to treatment was analyzed in patients who provided data from at least two study visits. A detailed history of immunosuppressive therapy including glucocorticoids, synthetic DMARDs [including targeted synthetic agents (JAK inhibitors)] and biologic DMARDs was recorded at each visit. Treatment decisions were made by the referring physicians rather than the investigative team. The PhGA raters were blinded to treatment data. Treatment change between consecutive visits was categorized as increased, decreased or unchanged. A change in treatment was defined as any of the following: a daily prednisone dose change of >5 mg, the addition or removal of a DMARD therapy, a decrease in dose of DMARD therapy 50% or an increase in dose of DMARD therapy 50% at time of the follow-up visit relative to previous visit. (13)

Statistical Analysis

Demographic and clinical variables were expressed as median and interquartile range (IQR), or frequencies, according to data type. The intra-class correlation coefficient (ICC) (2,1) and 95% confidence intervals (95%CI) were calculated using the *icc* function of the R package *irr* to determine inter-rater reliability between the PhGA raters. This model was chosen to assess inter-rater reliability because of its feasibility with more than two raters. (18) ICC results were interpreted as follows: ICC<0.4 was poor, ICC between 0.40 and 0.59 was fair, ICC between 0.60 and 0.74 was good and ICC between 0.75 and 1.0 was excellent. (19) Spearman's rank correlation analysis was used to evaluate the strength of association between PhGA and continuous variables and the correlation between different raters. Organ-specific signs and symptoms of RP were compared between the proportion of visits with PhGA 2 and the proportion of visits with PhGA 3 using Fisher's exact test. Wilcoxon signed rank test was used to compare changes in PhGA and CRP in association with treatment status and with addition/increase or subtraction/decrease of specific categories of medications. Chi-Square was used to compare the proportion of patients that increased treatment in the first visit interval compared to other visit intervals. *P* values <0.05 were considered significant.

All Statistical analysis was performed using JMP 14, PRISM 8 or RStudio 1.2.5001.

Results

Characteristics of Cohort

Seventy-eight patients were evaluated over 164 study visits over a median follow up period of 6 months (IQR 0–16). Forty-seven patients (60.3%) had at least one follow up visit with a median follow-up interval of 6 months (IQR 5–9). The cohort was predominantly Caucasian (82.0%) and female (83.3%). The median age at the baseline visit was 45 years (IQR 36–55), and the median disease duration from symptom onset was 8 years (IQR 4–16). The median (IQR) C-reactive protein was 2.2 (0.7–6.9) mg/L at baseline visit and the erythrocyte sedimentation rate (ESR) was 10.5 (5–20.8) mm/hr. Most patients (78.2%) were

on treatment at their first visit. Demographics and clinical characteristics at the baseline visit are shown in Table 1.

Physician Global Assessment (PhGA) Ratings

Out of 164 study visits, the median (IQR) for PhGA ratings by each of the three raters at the evaluating institution was 3 (2–4), 2 (1–3) and 2.5 (2–3) respectively. For 6 out of 164 visits (3.7%) the ratings by the 3 raters were discordant. After adjudicating the discordant ratings, the median (IQR) assigned PhGA was 3 (2–3), ranging from 0–7. The distribution of the final assigned PhGA score is shown in Figure 1A. Patients had PhGA scores of 0 at 4 visits (2.4%), had PhGA score of 1–2 in 76 (46.3%) of the visits, had a PhGA score of 3–4 at 62 visits (37.8%), and had a PhGA score 5 at 22 visits (13.4%).

Inter-Rater Reliability of PhGA Ratings

When comparing ratings across all study visits by the three physicians from the evaluating institution, the ICC (2,1) was excellent at 0.79 (95% CI: 0.73 to 0.84). For the 20 visits randomly selected for additional review, the ICC (2,1) was excellent at 0.85 (95% CI: 0.72 to 0.93) for the three physicians at the evaluating institution but was poor at 0.27 (95% CI: –0.01 to 0.53) for the additional raters from outside institutions. The ICC (2,1) for all six raters in the subset of 20 visits was poor at 0.21 (95% CI: 0.06–0.44). A scatter plot of the ratings for all six raters is shown in Figure 2A. There was a significant positive correlation ($p < 0.01$) between each of the three raters from the evaluating institution. No correlation was observed between the outside raters with each other or with the raters from the evaluating institution. (Figure 2B)

Laboratory Variables and PhGA

ESR was elevated above the laboratory normal range at 22 out of 162 visits (13.6%), and CRP was elevated at 44 visits out of 163 visits (27.0%). In each case, elevated acute phase reactants were believed to be associated with disease activity and not related to a secondary process. PhGA score was weakly correlated with CRP ($r_s = 0.30$, $p < 0.01$) but not with ESR ($r_s = 0.13$, $p = 0.10$). Similar results were found when only analyzing data from the baseline visits (data not shown).

To investigate the contribution of acute phase reactants (ESR and CRP) to PhGA scores, 60 clinical vignettes were rescored without providing values for acute phase reactants. For visits when at least one of the acute phase reactants was elevated, the median (IQR) PhGA ratings was 3 (2–4) when the acute phase reactant information was provided; however, repeat PhGA ratings were significantly lower with a median (IQR) of 2 (1–3) when information about acute phase reactants was omitted, $p = 0.02$. (Figure 1B) For visits with normal values for the acute phase reactants, there was no difference in PhGA ratings when information about these tests was available (2(2–3)) compared to when the information was withheld (2 (1–3)) ($p = 0.32$). (Figure 1B) While CRP positively correlated with PhGA ratings when this information was provided ($r_s = 0.25$, $p = 0.05$) (Figure S1A), CRP did not correlate with PhGA score when information about acute phase reactants was withheld ($r_s = -0.013$, $p = 0.92$). (Figure S1B)

Association between Clinical Symptoms and PhGA

To determine signs and symptoms that contribute to PhGA ratings, symptom frequencies were assessed. PhGA scores did not significantly differ based on presence or absence of most symptoms, except significantly greater PhGA scores were observed in association with arthralgia/stiffness and wheezing. (Table 2) There was no difference in symptom frequencies assessed between visits with PhGA 2 and with PhGA 3 (Table S1).

The disease activity summary score ranged from 0 to 5 with a median of 2 and an IQR of 1–3. Across all visits, the disease activity summary score was not correlated to PhGA score, $r_s = 0.087$, $p = 0.27$. However, for the 60 visits that were rescored without acute phase reactant information provided to the raters, PhGA ratings were significantly correlated with the disease activity summary score, $r_s = 0.51$, $p < 0.01$. (Figure S1 C and D)

Change in PhGA Over Time

Out of the 86 follow-up visits from 47 patients, the interval changes in PhGA scores ranged from –5 to +3. Between consecutive visits, the median (IQR) change was 0 (–1 to 1). PhGA scores increased by 2 or more between 12 visit intervals (14.0%), decreased by 2 or more between 16 visit intervals (18.6%), and remained unchanged for 30 visit intervals. (Figure 3A) The median (IQR) for the first four study visits were the following: 3 (2–5), 2 (1–3), 2 (2–3), and 2 (1–3) respectively. There was a significant decrease between the baseline visit compared to each of the next three follow up visits ($p < 0.01$). There were no significant differences in PhGA scores between any of the follow up visits. (Figure 3B)

Change of PhGA in Response to Treatment

Out of the 86 follow-up visits from 47 patients, there was an increase in treatment over 55 visits, a decrease in treatment over 14 visits, and no change in treatment over 17 visits. The proportion of visits with increased treatment was greater in the first visit interval (74.5%) compared to all other intervals ($p = 0.03$). Patients who had an increase in treatment between visits showed a decrease in PhGA score from 3 (2–4) to 2 (2–3) ($p < 0.01$). (Figure 3C) PhGA did not change between visits in patients who decreased treatment or who had no change in treatment. For visit intervals with no change in treatment, PhGA ratings stayed the same 2 (1–3) ($p = 0.77$) and for visits that decreased in treatment, PhGA scores went from 2 (2–3) to 2 (1–3) ($p = 0.49$). Patients who increased treatment had greater median (IQR) PhGA scores at the visit before the treatment change compared to patients who did not change treatment: 3 (2–4) vs 2 (1–3), $p = 0.02$. There were no other significant differences in PhGA scores between the pre or post treatment groups. Similarly, there was a significant decrease in median (IQR) CRP following an increase in treatment from 2.3 (0.6–4.4) mg/L to 0.9 (0.3–4) mg/L ($p = 0.04$) but no change in CRP with either decreasing or not changing treatment. There were no changes in ESR between visits with increasing, decreasing, or not changing treatment. Adding either a synthetic (including targeted synthetic) DMARD or a biologic DMARD was associated with a significant decrease in PhGA ratings between visits from 3 (2–3.25) to 2 (1–3) and from 3.5 (3–5) to 3 (2–3) ($p = 0.04$ and $p < 0.01$, respectively). (Figure 3D) Withdrawal of a synthetic or biologic DMARD was not associated with a change in PhGA score. Prednisone dose was not correlated with PhGA score, $r_s = -0.087$, $p = 0.27$.

Discussion

This study was undertaken to characterize PhGA for RP, a rare disease that lacks validated outcome measures. PhGA was not a reliable measure of disease activity amongst raters from different institutions, highlighting a need to develop and validate disease-specific activity indices in RP. Despite this limitation, PhGA was reliable and useful to monitor disease activity in response to treatment within a single-center observational cohort study of patients with RP. Persistent disease activity was common in RP despite treatment, and most patients with RP did not experience a relapsing-remitting pattern of disease activity.

The factors that influence PhGA ratings in RP are complex and can lead to differences in clinical assessment of disease activity. For example, accurate determination of airway inflammation based off clinical symptoms and direct observation of only the upper airway can be challenging. In this study, very few individual signs or symptoms were independently associated with PhGA score, highlighting that disease activity assessment in RP is likely driven by multifaceted factors. Issues of study design may contribute to the lack of strong associations between PhGA score and individual clinical symptoms. Patients were often evaluated later into the disease course while receiving treatment, which may have blunted associations between individual symptoms and disease activity assessment. Additionally, clinical symptoms were categorized as present or absent. Degree of severity of individual symptoms may have been more strongly associated with PhGA score.

Multiple measures point to a moderate influence of ESR and CRP, particularly when elevated, on PhGA ratings. CRP weakly correlated with PhGA score. However, at most visits, patients had normal acute reactants despite ongoing symptoms attributable to RP. Without the acute phase reactant information, symptom-based variables appear to carry more weight in the PhGA rating. Since it can be difficult in RP to differentiate between disease activity and damage, knowledge of the acute phase reactant levels influences to what extent the rater considers clinical symptoms as representative of active disease. These findings are consistent with findings in other rheumatologic diseases, such as systemic lupus erythematosus (SLE), where laboratory data has been shown to significantly impact PhGA assessment. (20) Together, these data show that laboratory markers interplay with other factors in the assessment of disease activity in RP.

The distribution of PhGA scores provides insight into disease activity profiles in patients with RP. In this study, most patients had PhGA scores between 1 and 3. There were few patients with PhGA scores greater than 3 and only a small number of patients with achieved a PhGA score of 0. The small number of patients with a PhGA score of 0 points to the difficulty of completely eliminating disease activity, even with treatment. Selection bias may affect the distribution of PhGA scores, as patients on either extreme on the PhGA scale may have been less likely to enroll and travel to the NIH to participate in this study. In some cases, when clinical symptoms were minor, it may have been difficult to distinguish between other etiologies rather than true disease activity, which is affirmed by the rating differences of PhGA when information about elevated acute phase reactants is provided versus withheld. Differentiating activity from damage remains an important challenge in the assessment and care of patients with RP.

Examination of PhGA scores over time provides insight into the natural history of RP and treatment response. Based on change in PhGA ratings over consecutive visits, RP worsens in only a few patients on treatment, improves slightly in many patients with treatment, and stays the same only in those patients who already had low disease activity. Contrary to the disease name—relapsing polychondritis—patients with multiple visits did not show a relapsing-remitting pattern. Rather, there was a significant decrease in disease activity between the first and second visits, and most patients remained relatively stable over subsequent visits with ongoing persistent mild disease activity despite treatment. Treatment escalation most often occurred after the initial study visit. Increasing treatment with either synthetic, targeted synthetic, or biologic DMARDs was associated with a lower PhGA score at subsequent follow-up visits, but PhGA score was not correlated with glucocorticoid dose. Improvement in PhGA following increased treatment suggests that aggressive pharmacologic measures for patients with active disease does help, as suggested by others. (21) Alternatively, improvement in PhGA score only during the initial observation interval could represent regression to the mean or the natural history of disease.

This study has some limitations. This was not an inception cohort; therefore, patients were enrolled at various stages of disease and many had RP for years. This limitation is somewhat countered by a relatively large sample size for a rare disease, prospective data collection, and a standardized assessment protocol. Studying patients later in the disease course may have restricted the range of disease activity observed in this study. Cumulative glucocorticoid dosing was not examined which could explain the lack of association of PhGA score and glucocorticoid dose. While this study begins to explore the association of laboratory and symptom-based factors with PhGA ratings in RP, the composite drivers of PhGA scores are yet to be determined.

Validation of a disease specific activity index is an unmet need in RP. While PhGA is simple to implement, it is a generic measure of disease activity that lacks the nuance of a disease-specific index. A disease-specific activity index may standardize activity assessment in RP and improve agreement among investigators from different institutions. The RPDAI is a proposed disease specific tool to quantify disease activity in RP, and efforts are currently underway to modify and validate this index. (4)

Assessment of the PhGA in a cohort of patients with RP is a necessary step towards the goal of outcome measure development and validation in RP. PhGA is useful to monitor disease activity in RP but may not be reliable across different institutions. The complexity of factors influencing PhGA ratings highlights the challenges of clinical assessment in RP. Development of disease-specific outcome measures are a prerequisite to the successful conduct of much-needed randomized clinical trials in RP.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Financial supports and conflicts disclosure:

This study was supported by the Intramural Research Program at the National Institute of Arthritis and Musculoskeletal and Skin Diseases. This research was made possible through the NIH Medical Research Scholars Program, a public-private partnership supported jointly by the NIH and contributions to the Foundation for the NIH from the Doris Duke Charitable Foundation, Genentech, the American Association for Dental Research, the Colgate-Palmolive Company, and other private donors. The authors have no conflicts of interest to disclose.

References

1. Mathian A, Miyara M, Cohen-Aubart F, Haroche J, Hie M, Pha M, et al. Relapsing polychondritis: A 2016 update on clinical features, diagnostic tools, treatment and biological drug use. *Best Practice & Research Clinical Rheumatology*. 2016;30(2):316–33. [PubMed: 27886803]
2. Puechal X, Terrier B, Mouthon L, Costedoat-Chalumeau N, Guillemin L, Le Jeunne C. Relapsing polychondritis. *Joint Bone Spine*. 2014;81(2):118–24. [PubMed: 24556284]
3. Rednic S, Damian L, Talarico R, Scire CA, Tobias A, Costedoat-Chalumeau N, et al. Relapsing polychondritis: state of the art on clinical practice guidelines. *RMD Open*. 2018;4(Suppl 1):e000788. [PubMed: 30402273]
4. Arnaud L, Devilliers H, Peng SL, Mathian A, Costedoat-Chalumeau N, Buckner J, et al. The Relapsing Polychondritis Disease Activity Index: development of a disease activity score for relapsing polychondritis. *Autoimmun Rev*. 2012;12(2):204–9. [PubMed: 22771427]
5. Mukhtyar C, Lee R, Brown D, Carruthers D, Dasgupta B, Dubey S, et al. Modification and validation of the Birmingham Vasculitis Activity Score (version 3). *Annals of the Rheumatic Diseases*. 2009;68(12):1827–32. [PubMed: 19054820]
6. Griffiths B, Mosca M, Gordon C. Assessment of patients with systemic lupus erythematosus and the use of lupus disease activity indices. *Best Practice & Research Clinical Rheumatology*. 2005;19(5):685–708. [PubMed: 16150398]
7. ROHEKAR G, POPE J. Test-Retest Reliability of Patient Global Assessment and Physician Global Assessment in Rheumatoid Arthritis. *The Journal of Rheumatology*. 2009;36(10):2178–82. [PubMed: 19755617]
8. Cauli A, Gladman DD, Mathieu A, Olivieri I, Porru G, Tak PP, et al. Physician's Global Assessment in Psoriatic Arthritis: A Multicenter GRAPPA Study. *The Journal of Rheumatology*. 2018;45(9):1256–62. [PubMed: 29907666]
9. Hinojosa-Azaola A, Jimenez-Gonzalez A, Alcocer-Castillejos N. Patient and physician perspectives on the impact of health-related quality of life in Mexican patients with ANCA-associated vasculitis. *Rheumatol Int*. 2018;38(4):631–40. [PubMed: 29247347]
10. Rimland CA, Quinn KA, Rosenblum JS, Schwartz MN, Gribbons KB, Novakovich E, et al. Outcome Measures in Large-Vessel Vasculitis: Relationship Between Patient, Physician, Imaging, and Laboratory-Based Assessments. *Arthritis Care & Research*. n/a(n/a).
11. Tago M, Sawada T, Nishiyama S, Tahara K, Kato E, Hayashi H, et al. Influence of large joint involvement on patient–physician discordance in global assessment of rheumatoid arthritis disease activity analyzed by a novel joint index. *International Journal of Rheumatic Diseases*. 2018;21(6):1237–45. [PubMed: 29611628]
12. Chessa E PM, Floris A, Devilliers H, Cauli A, Arnaud L. Use of Physician Global Assessment (PGA) in Systemic lupus erythematosus: a systematic review of its psychometric properties. *Rheumatology (Oxford)*. 2020;In press.
13. Banerjee S, Quinn KA, Gribbons KB, Rosenblum JS, Civelek AC, Novakovich E, et al. Effect of Treatment on Imaging, Clinical, and Serologic Assessments of Disease Activity in Large-vessel Vasculitis. *The Journal of Rheumatology*. 2020;47(1):99–107. [PubMed: 30877209]
14. Kaneko Y, Takeuchi T, Cai Z, Sato M, Awakura K, Gaich C, et al. Determinants of Patient's Global Assessment of Disease Activity and Physician's Global Assessment of Disease Activity in patients with rheumatoid arthritis: A post hoc analysis of overall and Japanese results from phase 3 clinical trials. *Modern Rheumatology*. 2018;28(6):960–7. [PubMed: 29278339]
15. McAdam LP, O'Hanlan MA, Bluestone R, Pearson CM. Relapsing polychondritis: prospective study of 23 patients and a review of the literature. *Medicine (Baltimore)*. 1976;55(3):193–215. [PubMed: 775252]

16. Michet CJ Jr., McKenna CH, Luthra HS, O'Fallon WM. Relapsing polychondritis. Survival and predictive role of early disease manifestations. *Ann Intern Med.* 1986;104(1):74–8. [PubMed: 3484422]
17. Damiani JM, Levine HL. Relapsing polychondritis--report of ten cases. *Laryngoscope.* 1979;89(6 Pt 1):929–46. [PubMed: 449538]
18. Trevethan R Intraclass correlation coefficients: clearing the air, extending some cautions, and making some requests. *Health Services and Outcomes Research Methodology.* 2017;17(2):127–43.
19. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. US: American Psychological Association; 1994. p. 284–90.
20. 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. *Annals of the Rheumatic Diseases.* 2020:annrhumdis-2019–216753.
21. Moulis G, Pugno G, Costedoat-Chalumeau N, Mathian A, Leroux G, Boutemy J, et al. Efficacy and safety of biologics in relapsing polychondritis: a French national multicentre study. *Ann Rheum Dis.* 2018;77(8):1172–8. [PubMed: 29535124]

Significance and Innovations

- Physician global assessment (PhGA) can be used to monitor disease activity in relapsing polychondritis (RP)
- Acute phase reactants are not elevated in most patients with RP and weakly correlate with PhGA
- Persistently active disease is more common than a relapsing-remitting pattern of disease activity in RP
- Despite treatment, few patients with RP achieve a PhGA score of 0, indicating absence of disease activity

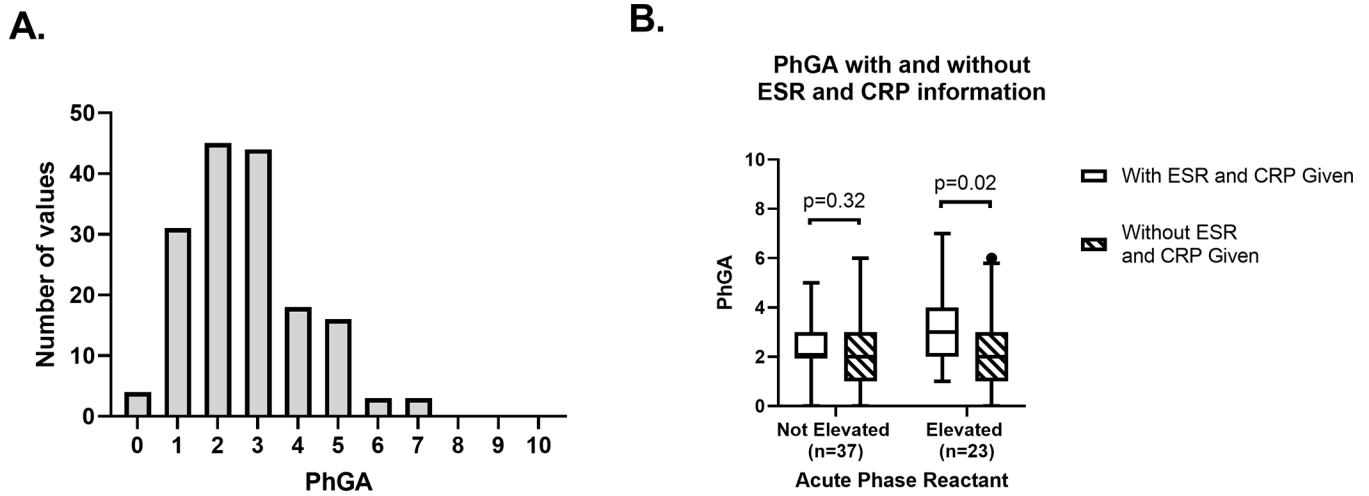
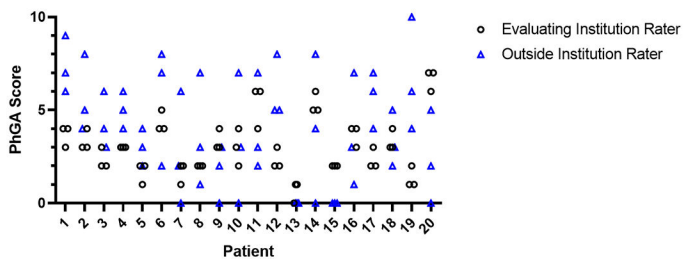


Figure 1. Distribution of physician global assessment (PhGA) scores in relapsing polychondritis. Histogram shows the distribution of PhGA scores for 164 visits; median (interquartile range) was 3 (2–3) (A). Box and whisker plot of PhGA scores when information about ESR and CRP was provided to (no stripes) or withheld from (stripes) the raters. Whiskers show 5th to 95th percentile, 60 visits included (B).

A



B

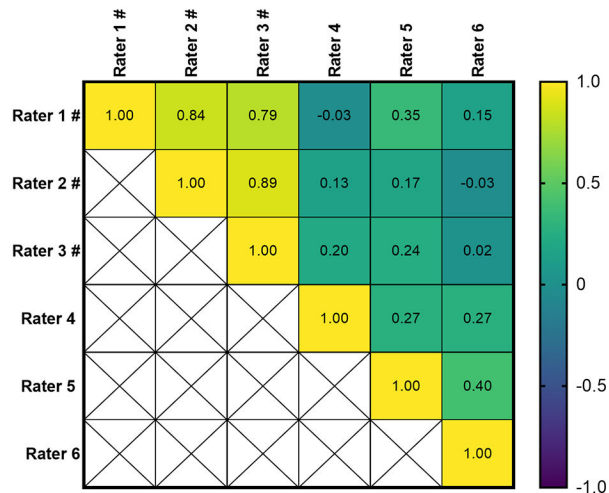
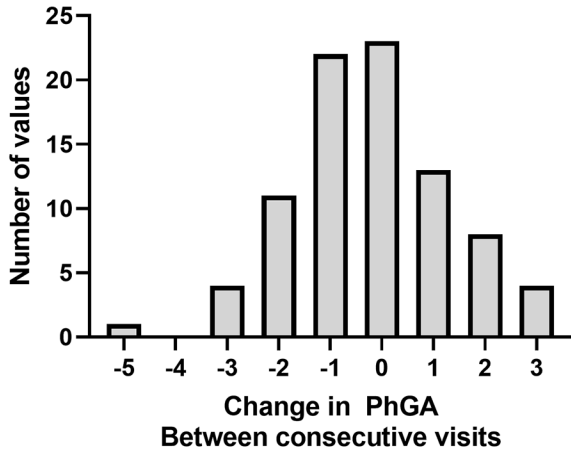


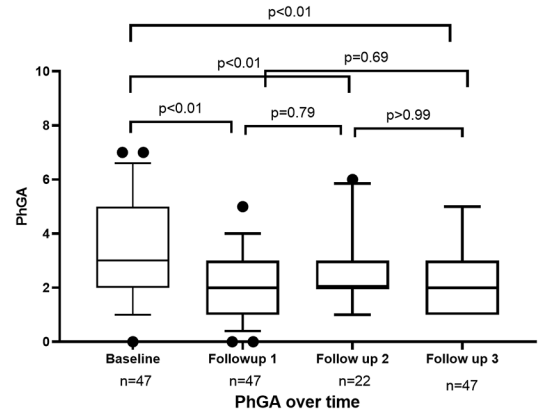
Figure 2. Physician global assessment (PhGA) scores of clinical vignettes by six independent raters.

Scatter plot shows the PhGA scores for 20 patients by 3 raters from the evaluating institution (black circles) and 3 additional outside raters (blue triangles) (A). Correlation matrix shows the spearman’s correlation between each the 6 raters for the 20 patient visits. # denotes a rater from the evaluating institution. Scores were significantly correlated ($p < 0.01$) only between the 3 raters from the evaluating institution (B).

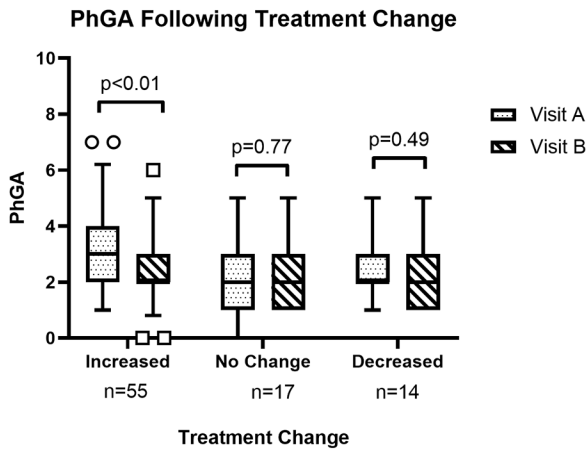
A.



B.



C.



D.

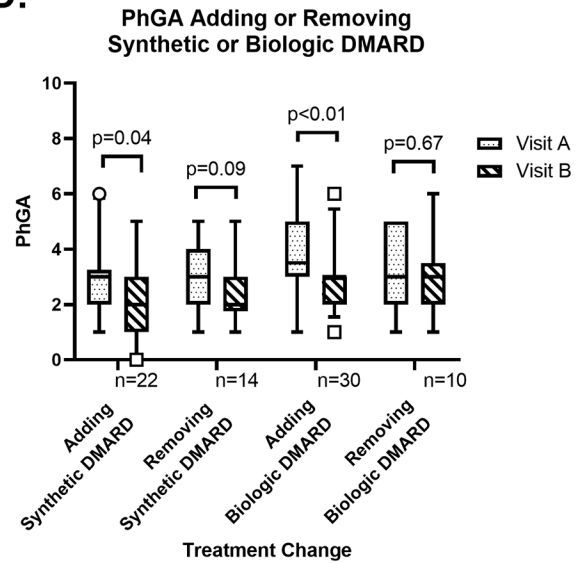


Figure 3. Interval change in physician global assessment (PhGA) score in association with treatment.

Histogram shows change in PhGA scores between consecutive visits over 86 visit intervals (A). Box and whisker plots of PhGA scores from four consecutive study visits in 47 patients with at least two study visits. The median (interquartile range) for the first 4 visits were: 3 (2–5), 2 (1–3), 2 (2–3) and 2 (1–3), respectively (B). Box and whisker plots of PhGA scores from two consecutive study visits categorized by an increase in treatment, no change in treatment, or a decrease in treatment between visits (C). Box and whisker plots of PhGA scores from consecutive study visits in relationship to specific treatment changes between visits (D).

Table 1.

Baseline Study Population Characteristics.

Demographics	n=78
Race (n, % Caucasian)	64 (82.1)
Sex (n, % Female)	65 (83.3)
Age at Visit (years, IQR)	45 (36–55)
Disease Duration (years, IQR)	8 (4–16)
Number of visits Median (IQR)	2 (1–3)
Follow up period (months, IQR)	6 (0–16)
Laboratory	
CRP, mg/L Median (IQR) [#]	2.2 (0.7–6.9)
ESR, mm/hr Median (IQR) ^{##}	10.5 (5–20.8)
Elevated acute phase reactants n (%)	28 (36.8)
Symptom Frequencies at visit attributed to RP in 28 days	
Musculoskeletal	65 (83.3)
Arthralgia or Stiffness	57 (73.1)
Tenosynovitis or Synovitis	25 (33.3)
Nasal Chondritis	34 (43.6)
Auricular Chondritis	19 (24.4)
Chest Wall Chondritis	42 (53.8)
Audi vestibular	2 (2.6)
Vestibular Dysfunction	0
Sensorineural hearing loss	2 (2.6)
Cutaneous	2 (2.6)
Respiratory Chondritis ^{###}	44 (56.4)
Wheezing	3 (3.8)
Voice Changes	17 (21.8)
Dry Cough	24 (30.8)
Dyspnea	29 (37.2)
Cardiovascular involvement	1 (1.3)
Eye involvement	0
Subglottic Inflammation	7 (10.2)
Tracheal wall thickness > 3 mm	18 (24)
Tracheomalacia	31 (42.3)
Bronchomalacia	10 (13.7)
Current Treatment	
No Treatment	17 (21.8)
Prednisone dose (mg)	5 (0–20)
At least one synthetic DMARD	44 (56.4)
At least one biologic DMARD	17 (21.8)
Activity Scores	

Demographics	n=78
PhGA Median (IQR)	3 (2–3)
Disease activity summary score Median (IQR)	3 (2–3)

Values are reported as median and interquartile range or as percentage and n-value.

#. Lab normal range (0–4.99), n=77

##. Lab normal range 0–42 mm/hr for females and 0–25 mm/hr for males, n=76

###. Respiratory Chondritis includes symptoms, not objective findings below

Missing values: Subglottic Inflammation (9), tracheal wall thickness (5), tracheomalacia (5), bronchomalacia (5)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2.

Median PhGA for visits with symptom present compared to visits when symptom not present

Symptom	PhGA Symptom Absent Median (IQR)	PhGA Symptom Present Median (IQR)	P value
Musculoskeletal Involvement	2.5 (1–3)	3 (2–4)	0.57
Arthralgia or Stiffness	2 (1–3)	3 (2–4)	0.024
Tenosynovitis or Synovitis	2.5 (1–4)	3 (2–3)	0.49
Nasal Chondritis	2 (2–3.25)	3 (1.75–3.25)	0.91
Auricular Chondritis	2 (2–3)	3 (2–4)	0.45
Chest Wall Chondritis	2 (2–3)	3 (2–4)	0.42
Audiovestibular	3 (2–3)	2 (1.75–4.25)	0.88
Vestibular Dysfunction	3 (2–3.25)	2 (2–2)	0.45
Sensorineural hearing loss	3 (2–3)	3 (1.4–4.5)	0.94
Cardiovascular involvement	3 (2–3)	3 (3–3)	0.66
Eye involvement	3 (2–3.5)	2 (2–2)	0.36
Respiratory Chondritis	2 (2–3)	3 (2–4)	0.21
Wheezing	2.5 (2–3)	3.5 (2.25–5)	0.045
Voice Changes	3 (2–3)	3 (1.75–4)	0.61
Dry Cough	3 (2–3)	2 (1.75–4)	0.99
Dyspnea	2 (2–3)	3 (2–4)	0.12
Choking sensation	3 (2–3.5)	3 (2–3)	0.87