



Published in final edited form as:

Arthritis Care Res (Hoboken). 2017 August ; 69(8): 1260–1265. doi:10.1002/acr.23103.

STANDARDS OF COMPARISON AND DISCORDANCE IN RHEUMATOID ARTHRITIS GLOBAL ASSESSMENTS BETWEEN PATIENTS AND CLINICIANS

Michael M. Ward, MD, MPH, Lori C. Guthrie, BSN, and Maria I. Alba, MD

Intramural Research Program, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD.

Abstract

Objective—Patient-physician discordance in health status ratings may arise because patients use temporal comparisons (comparing their current status with their previous status), while clinicians use social comparisons (comparing this patient’s status to that of other patients, or to the full range of disease severity possible) to guide their assessments. We compared discordance between patients with rheumatoid arthritis (RA) and clinicians using either the conventional patient global assessment (PGA) or a rating scale with five anchors describing different health states. We hypothesized that discordance would be smaller with the rating scale because clinicians likely used similar social comparisons when making global assessments.

Methods—We prospectively studied 206 patients with active RA, and assessed the PGA (0 – 100), rating scale (0 – 100), and evaluator global assessment (EGA; 0 – 100) on each of two visits (total 401 visits). We compared the PGA/EGA discordance and the rating scale/EGA discordance at each visit.

Results—The mean (\pm standard deviation) PGA/EGA discordance was 8.5 ± 22.4 , and the mean rating scale/EGA discordance was 2.3 ± 24.0 . The intraclass correlation, measuring agreement, was higher between the rating scale and EGA than between the PGA and EGA (0.39 versus 0.31). Agreement was larger at low levels of RA activity on both pairs of measures.

Conclusion—Discordance between patients’ global assessments and evaluators’ global assessments was smaller when patients used a social standard of comparison than when they marked the PGA, suggesting that differences in standards of comparison contribute to patient-clinician discordance when the PGA is used.

Keywords

patient global assessment; rheumatoid arthritis; rating scale; discordance

Discordance between patients and clinicians in their appraisals of rheumatoid arthritis (RA) activity has been widely noted [1–8]. Most often, patients rate their RA as more active than

Address for correspondence: Dr. Michael Ward, NIAMS/NIH, 10 CRC, 4-1339, 10 Center Drive, MSC 1468, Bethesda, MD 20892-1468. Telephone (301) 496-7263; Facsimile (301) 480-2714; wardm1@mail.nih.gov.

The authors have no financial conflicts of interest related to this work.

clinicians. Differences of more than 20 points (on a 100 point global assessment scale) are not uncommon. Discordance of this degree raises questions about the validity of these measures, and about whose assessment should be used to guide treatment decisions [2,8].

Investigations of the reasons for discordance have focused on identifying factors that independently influence the patient global assessment (PGA) and the evaluator global assessment (EGA) [3–8]. The PGA is most strongly associated with subjective pain severity, but also with depression or anxiety [3–7]. These psychological influences have been implicated as important sources of discordance, because they can affect how patients perceive their symptoms in ways that may not readily be appreciated by others [4,8]. The EGA is most influenced by joint counts and acute phase reactants, as would be expected given the clinician's vantage as an external observer [5–7]. Discordance in global assessments may also reflect errors in the measurement process, and true differences in appraisals [2].

According to Rapkin and Schwartz's model of quality of life self-assessment, the process of self-appraisal involves four cognitive steps: selection of frames of reference; sampling of experiences within the frame of reference; judgment of these experiences against a standard of comparison; and integration into a single appraisal [9]. Frames of reference refer to the domains of quality of life that patients may consider in response to a global question, such as symptoms, functioning, and emotional well-being. When sampling experiences within these frames of reference, patients may recall their typical days, or may focus on their recent experiences or on the most vivid high or low points. Standards of comparison may be others with the same illness, people of the same age, hypothetical peers conjured from feedback from family members or clinicians, or comparisons with one's previous state of health [10,11]. Variations in any of these processes may lead to differences in self-appraisals among patients, and between patients and clinicians, despite similar levels of health.

We hypothesized that use of different standards of comparison by patients and clinicians contributes to discordant global arthritis ratings. Patients may tend to rate their current status relative to normal, on one hand, and to the most active arthritis they have experienced in the past, on the other hand [12]. Conversely, clinicians may more likely judge patients against the most active RA possible or the most severely affected patient they have seen in their careers [5–7]. Consequently, discordance may occur because patients may use a temporal comparison against their past experience, while clinicians may use a social comparison, encompassing the entire range of RA activity [9].

To examine this hypothesis, we asked patients with RA to rate their global status not only using the common PGA scale, but also using a rating scale with anchors denoting health states. In rating themselves against these fixed health states, patients were making social comparisons against hypothetical "other" patients. We hypothesized that if standards of comparison contributed to patient-clinician discordance, then the rating scale/EGA discordance would be smaller than the PGA/EGA discordance. If standards of comparison did not contribute to patient-clinician discordance, the PGA/EGA and rating scale/EGA discordance would be similar. We also hypothesized that discordance would be lower in patients with low RA activity.

METHODS

Patients

We recruited patients with RA to participate in a prospective longitudinal study of measures of RA activity [13]. Inclusion criteria were fulfillment of the revised American College of Rheumatology criteria for the classification of RA, age ≥ 18 years, and presence of active RA and escalation of anti-rheumatic treatment at the baseline visit. The study protocol was approved by the institutional review board. All patients provided written informed consent.

Of 262 participants, 206 participants enrolled in a sub-study of preference measures, which included the rating scale. Data from these 206 participants were included in this analysis. Participants in the sub-study were older than non-participants (53.0 ± 13.5 years versus 44.8 ± 11.9 years) and more often men (25% versus 11%), but had similar baseline Disease Activity Score-28 (DAS28) values (5.3 ± 1.0 versus 5.5 ± 1.0).

Study Design

This study involved an outpatient visit at baseline and a follow-up visit to assess responses either one month or four months later (depending on the specific treatment prescribed) [13]. Identical procedures were done at each visit, and included questionnaires, a computer-administered assessment that included the rating scale, and brief histories, joint counts, and EGAs (0 – 100, with higher scores indicating more active RA) by one of the four study rheumatologists, in this sequence. The rheumatologists were unaware of the participants' responses when marking the EGA.

Study Measures

The PGA was measured using a visual analog scale with anchors of 0= "Very Well" and 100= "Very Poor". The prompt for the PGA was: "Considering all the ways that rheumatoid arthritis affects you, rate how you are doing on the following scale by placing a mark on the line."

The rating scale was a vertical visual analog scale with a top anchor of "Perfect Health" and bottom anchor of "Worst Imaginable Health," which were scored as 100 and 0, respectively [14]. "Perfect Health" was described as "health that was ideal or perfect in every way." In addition to the endpoint anchors, participants were asked to rate three other health states describing mild, moderate, and severe RA activity, based on the McMaster Utility Measurement Questionnaire (Appendix, supplemental material). Then they were asked to rate their current status. Only the participant's current rating was used in the analysis. Participants did the rating scale after completing the PGA and several other questionnaires.

We used the three-variable DAS28 as the measure of RA activity, excluding the PGA to avoid overestimating its associations with the PGA.

Statistical Analysis

We defined discordance as the difference between the patient-reported score (rating scale or PGA) and the physician-reported score (EGA). The direction of scoring of the rating scale

was reversed (0 = best; 100 = worst) for all analyses to make it analogous to the PGA. Positive values for discordance indicated higher patient-reported scores. We pooled results from the baseline and follow-up visits for analysis. We computed the rating scale/EGA discordance and the PGA/EGA discordance separately for each visit, and compared the magnitude of discordance using a paired t test. While this comparison provides a measure of discordance averaged over all subjects, it can mask agreement on the individual level. To measure agreement, we used intraclass correlations (ICC) and Bland-Altman plots. We computed ICCs between the rating scale and EGA, and between the PGA and EGA, using a linear mixed model that accounted for the repeated observations of patients and for differences in measure variances on the pre-treatment and post-treatment visits. We computed 95% confidence intervals for the ICCs using 2000 bootstrapped samples. Higher ICCs indicate better agreement, but standards for weak, moderate, or strong agreement are not well-established. Therefore we made only relative comparisons. Bland-Altman plots provide a visual representation of agreement, as well as any systematic differences, in a graph of the difference between a pair of measures plotted against the mean of the pair. Measures with high agreement have differences that are tightly clustered around the mean. To determine if discordance varied with arthritis activity, we examined mean discordance and intraclass correlations by categories of DAS28.

We used SAS version 9.3 (SAS Institute, Cary, NC) programs for analysis.

RESULTS

Participant characteristics

Participants were predominantly middle-aged white women with moderately or highly active RA at study entry (Table 1). Ninety-two percent of participants reported that RA was their main health problem. Eleven participants completed the necessary evaluations on only one study visit, while 195 participants completed them on both visits, for a total of 401 visits. Median time needed to complete the rating scale was 6 minutes (range 3 – 14 minutes).

Among all visits, the mean (\pm standard deviation) PGA was 45.4 ± 25.2 , mean rating scale was 39.2 ± 24.4 , and mean EGA was 36.9 ± 19.8 . Correlations between the rating scale and DAS28 (Spearman $r = 0.42$; $p < .0001$) and between the PGA and DAS28 (Spearman $r = 0.49$; $P < .0001$) were similar.

PGA/EGA and Rating scale/EGA Discordance

The mean PGA/EGA discordance was 8.5 ± 22.4 (Table 2). The median [25th, 75th percentile] discordance was 8 [-7, 25], indicating that on 25% of visits, the PGA was more than 25 points higher than the EGA. The mean rating scale/EGA discordance was 2.3 ± 24.0 and the median discordance was 2 [-14, 20]. The rating scale had fewer extreme positive values relative to the EGA than did the PGA. However, rating scale responses were more likely to be lower than the EGA. Comparing values at the same visit, the rating scale/EGA discordance was significantly smaller than the PGA/EGA discordance.

The ICC between the rating scale and EGA was larger, indicating greater agreement, than the ICC between the PGA and EGA among all visits (Table 2). The Bland-Altman plots

showed similar patterns for each pair of measures. There was less agreement between measures as the mean increased (i.e. with ratings of more active RA) (Figure 1). A notable difference between the plots was tighter clustering around the mean, indicating closer agreement, on the rating scale/EGA plot compared to the PGA/EGA plot for means less than 30. The PGA-rating scale ICC was 0.18 (95% Confidence Interval 0.09, 0.27).

Associations with RA Activity

Rating scale/EGA discordance was smaller than PGA/EGA discordance in each category of RA activity (Table 2). ICCs for both pairs of measures were largest at low disease activity or remission. Agreement was poorer among visits with higher RA activity. ICCs for the rating scale/EGA pairing were larger than ICCs for the PGA/EGA pairing at each level of RA activity. The difference in ICCs was greatest at high RA activity, and the difference was small among visits with moderate RA activity. Confidence intervals for the ICCs at low disease activity were wide due to the small number of visits in this category.

A plot of discordance by DAS28 also showed the rating scale/EGA discordance was smaller than the PGA/EGA discordance at low DAS28 values (Supplemental figure 1).

DISCUSSION

That the rating scale/EGA discordance was lower than the PGA/EGA discordance is consistent with the hypothesis that differences in standards of comparison contribute to patient-clinician discordance in RA global assessments. When asked to mark the PGA, patients are given few, if any, cues to the intended standard of comparison. In most clinical contexts, and particularly in clinical trials, patients likely infer that the intended comparison is relative to their previous visit [11,14]. Patients are conditioned to use personal temporal comparisons because changes in their health are the central focus of clinical encounters. In contrast, when patients marked the rating scale, which had cues from the marker states to consider how they compare to other hypothetical patients with RA, they rated themselves somewhat differently. In general, patients rated their RA as less active using the rating scale than the PGA. This difference indicates that the standard of comparison influences how patients rate their RA.

Patients' rating scale assessments were closer to the evaluators' ratings, indicating that clinicians were likely using a similar standard of comparison. This is perhaps not surprising, as clinicians are trained to assess health and disease on absolute scales. Their experience with other patients informs how a given patient's presentation relates to the range of possibilities of RA activity, making these assessments implicitly social comparisons. Placing a given patient on this spectrum helps inform the clinician's prognosis and treatment recommendation. Use by clinicians of an absolute standard is supported by studies demonstrating that the strongest correlates of the EGA are joint counts and acute phase reactants [5–7]. Our results suggest that patient-clinician discordance in global RA ratings arises in part because patients typically use a personal relative standard of comparison while clinicians use a social standard of comparison.

Few studies have examined the cognitive processes that patients with RA use to arrive at their PGA, including what standards of comparison they typically use. In a qualitative study of 10 patients, Hooper reported that all patients used temporal comparisons, while 7 patients also used social comparisons [12]. Affleck and Tennen reported that 59% of patients used temporal comparisons to describe the severity of their RA, and only 14% compared themselves to other patients [15]. Studies in other diseases have also indicated that temporal comparisons were more commonly used, and were more strongly associated with self-appraisals of health, than were social comparisons [10,11,14]. To our knowledge, previous studies have not examined differences in standards of comparison as a potential source of patient-clinician discordance in RA global assessments.

Patient-clinician agreement using either measure was inversely related to RA activity, possibly in part due to floor effects. The rating scale/EGA agreement was particularly high among patients in remission or with low RA activity. The rating scale may therefore be a more valid global measure of RA activity in remission. Although agreement was lower at higher RA activity, the relative difference in agreement between the rating scale and PGA was more pronounced. With the rating scale, patients and clinicians more closely agreed on how well the patient was doing, likely because the health states provided well-defined anchors to which patients could easily compare themselves.

Our study is limited in including only four rheumatologists as evaluators, which may have reduced the degree of discordance and affect generalizability. The rating scale measured RA-related health, rather than RA activity specifically, although studies have indicated that in the context of clinical assessment these measures are interchangeable [16]. Because most patients reported that RA was their major health problem, their health rating likely reflected their rating of RA. Direction of scores differed between the PGA and rating scale, which may have confused patients, but differences in scale orientation, labels, and modes of administration likely minimized carry-over effects. The standards of comparison used by patients and clinicians were implicit. Future research should examine these standards explicitly. Although we used a computer assessment for the rating scale, a written assessment may be more feasible.

Our findings suggest that understanding patient-clinician discordance may benefit from investigation of factors beyond identification of correlates of the PGA and EGA. Insights from cognitive psychology on how individuals form self-assessments may be useful in understanding variations in the PGA in RA and other rheumatic diseases. Discordance may be decreased by providing cues to patients on the standards of comparison to use so that these are similar to the standards used by clinicians. These ratings will then likely be social comparisons relative to others with RA, rather than a temporal comparison with their prior health state. However, temporal comparisons are likely more relevant for clinical decision-making, creating a tension between minimizing discordance and using the more appropriate standard of comparison for decision-making. If the PGA is intended to provide temporal comparisons, clinicians should also use temporal comparisons if minimizing discordance is deemed important.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the Intramural Research Program, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health and US PHS grant RO1-AR45177.

REFERENCES

1. Suarez-Almazor ME, Conner-Spady B, Kendall CJ, Russell AS, Skeith K. Lack of congruence in the ratings of patients' health status by patients and their physicians. *Med Decis Making*. 2001; 21:113–121. [PubMed: 11310944]
2. Hewlett SA. Patients and clinicians have different perspectives on outcomes in arthritis. *J Rheumatol*. 2003; 30:877–879. [PubMed: 12672220]
3. Nicolau G, Yogui MM, Vallochi TL, Gianini RJ, Laurindo IM, Novaes GS. Sources of discrepancy in patient and physician global assessments of rheumatoid arthritis disease activity. *J Rheumatol*. 2004; 31:1293–1296. [PubMed: 15229946]
4. Barton JL, Imboden J, Graf J, Glidden D, Yelin EH, Schillinger D. Patient-physician discordance in assessments of global disease severity in rheumatoid arthritis. *Arthritis Care Res*. 2010; 62:857–864.
5. Khan NA, Spencer HJ, Abda E, Aggarwal A, Alten R, Ancuta C, et al. Determinants of discordance in patients' and physicians' rating of rheumatoid arthritis disease activity. *Arthritis Care Res*. 2012; 64:206–214.
6. Studenic P, Radner H, Smolen JS, Aletaha D. Discrepancies between patients and physicians in their perceptions of rheumatoid arthritis disease activity. *Arthritis Rheum*. 2012; 64:2814–2823. [PubMed: 22810704]
7. Furu M, Hashimoto M, Ito H, Fujii T, Terao C, Yamakawa N, et al. Discordance and accordance between patient's and physician's assessments in rheumatoid arthritis. *Scand J Rheumatol*. 2014; 43:291–295. [PubMed: 24650255]
8. Wolfe F, Michaud K, Busch RE, Katz RS, Rasker JJ, Shahouri SH, et al. Polysymptomatic distress in patients with rheumatoid arthritis: Understanding disproportionate response and its spectrum. *Arthritis Care Res*. 2014; 66:1465–1471.
9. Rapkin BD, Schwartz CE. Toward a theoretical model of quality-of-life appraisal: implications of findings from studies of response shift. *Health Qual Life Outcomes*. 2004; 2:14. [PubMed: 15023229]
10. Suls J, Marco CA, Tobin S. The role of temporal comparison, social comparison, and direct appraisal in the elderly's self-evaluations of health. *J Appl Soc Psychol*. 1991; 21:1125–1144.
11. Idler EL, Hudson SV, Leventhal H. The meanings of self-ratings of health. A qualitative and quantitative approach. *Res Aging*. 1999; 21:458–476.
12. Hooper H, Ryan S, Hassell A. The role of social comparison in coping with rheumatoid arthritis: an interview study. *Musculoskeletal Care*. 2004; 2:195–206. [PubMed: 17041983]
13. Ward MM, Guthrie LC, Alba MI. Clinically important changes in individual and composite measures of rheumatoid arthritis activity. Thresholds applicable in clinical trials. *Ann Rheum Dis*. 2015; 74:1691–1696. [PubMed: 24794149]
14. Bakker C, Rutten M, van Doorslaer E, Bennett K, van der Linden S. Feasibility of utility assessment by rating scale and standard gamble in patients with ankylosing spondylitis or fibromyalgia. *J Rheumatol*. 1994; 21:269–274. [PubMed: 8182636]
14. King KB, Clark PC, Friedman MM. Social comparisons and temporal comparisons after coronary artery surgery. *Heart Lung*. 1999; 28:316–325. [PubMed: 10486448]
15. Affleck, G., Tennen, H. Social comparison and coping with major medical problems. In: Suls, J., Wills, TA., editors. *Social comparison: Contemporary theory and research*. Hillsdale (NJ): Lawrence Erlbaum Associates, Inc; 1991. p. 369-393.

16. Dougados M, Ripert M, Hilliquin P, Fardellone P, Brocq O, Brault Y, et al. The influence of the definition of patient global assessment in assessment of disease activity according to the Disease Activity Score (DAS28) in rheumatoid arthritis. *J Rheumatol.* 2011; 38:2326–2328. [PubMed: 21965642]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

SIGNIFICANCE AND INNOVATION

- This is the first study to examine standards of comparison as a potential source of patient-clinician discordance in global ratings.
- This is one of few studies to examine discordance across different levels of RA activity.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

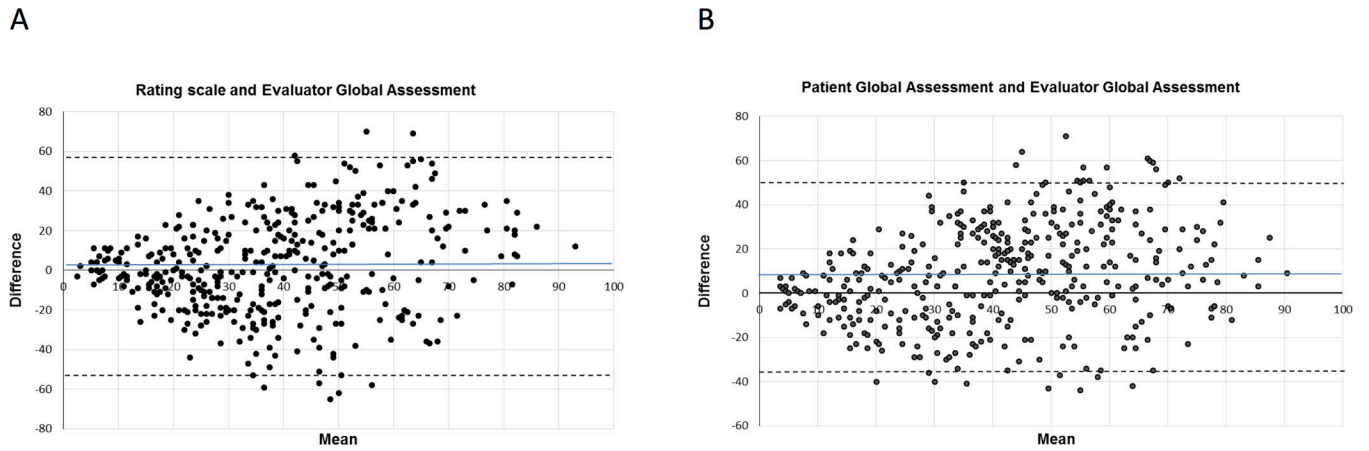


Figure 1. Bland-Altman plots of agreement between the rating scale and evaluator global assessment (A) and the patient global assessment and evaluator global assessment (B). Grey line indicates the mean of all values, and dashed lines indicate 95% levels of agreement.

Table 1

Characteristics of the patients (N =206).

	Baseline (N = 206)	Follow-up (N = 195)
Age, years	53.0 (13.5)	
Women, n (%)	155 (75.2)	
White, n (%)	105 (51.0)	
Black, n (%)	54 (26.2)	
Hispanic, n (%)	28 (13.6)	
Asian, n (%)	17 (8.2)	
Other ethnicity, n (%)	2 (1.0)	
Education, years	13.7 (3.0)	
Duration of RA, years	10.2 (10.1)	
Seropositive, n (%) [*]	140 (75.7)	
Erosive, n (%) [†]	115 (65.0)	
Patient global assessment (0 – 100)	52.8 (25.0)	37.6 (23.1)
Rating scale (0 – 100)	45.0 (25.2)	33.2 (22.0)
Evaluator global assessment (0 – 100)	46.2 (17.5)	27.1 (17.2)
Pain Scale (0 – 100)	58.7 (25.6)	40.7 (26.2)
HAQ (0 – 3)	1.4 (0.7)	1.1 (0.7)
DAS28 using 3 variables	5.3 (1.0)	4.4 (1.1)
DAS28 < 3.2, n (%)	2 (1.0)	39 (20.0)
3.2 DAS28 5.1, n (%)	92 (44.6)	105 (53.8)
DAS28 > 5.1, n (%)	112 (54.4)	51 (26.2)

Values are mean (SD) or n (%).

Values in parentheses after the measures are possible ranges.

RA = rheumatoid arthritis; DAS28 = Disease Activity Score 28; HAQ = Health Assessment Questionnaire.

^{*} Of 185 with results.

[†] Of 177 with results.

Table 2

Mean discordance and intraclass correlations in visits of all patients and in subgroups by level of rheumatoid arthritis activity. EGA = evaluator global assessment. PGA = patient global assessment.

	Number of visits	Mean Discordance (\pm standard deviation)		p*	Intraclass Correlation (95% confidence interval)	
		Rating scale/EGA	PGA/EGA		Rating scale/EGA	PGA/EGA
All visits	401	2.3 \pm 24.0	8.5 \pm 22.4	<.0001	0.39 (0.32, 0.47)	0.31 (0.23, 0.40)
Low disease activity or remission (DAS28 < 3.2)	41	3.8 \pm 14.4	8.6 \pm 20.2	.06	0.58 (0.28, 0.77)	0.49 (0.28, 0.78)
Moderate disease activity (3.2 DAS28 5.1)	197	6.3 \pm 23.8	10.3 \pm 22.4	.007	0.48 (0.40, 0.58)	0.44 (0.36, 0.57)
High disease activity (DAS28 > 5.1)	163	-2.9 \pm 25.2	6.2 \pm 22.8	<.0001	0.36 (0.26, 0.48)	0.16 (0.03, 0.34)

* p for paired differences.