Construct Validity of the Modified Numeric Rating Scale of Patient Global Assessment in Psoriatic Arthritis

YING-YING LEUNG, KWOK-WAH HO, TRACY-YANNER ZHU, LAI-SHAN TAM, EMILY WAILIN KUN, and EDMUND KWOK-MING LI

ABSTRACT. ***Objective.*** The construct validity of the patient global health assessment (PGA) in psoriatic arthritis (PsA) has not been analyzed, despite its common use. We evaluated the construct validity of a numeric rating scale (NRS) of the PGA in PsA.

Methods. Patients with PsA who fulfilled the ClASsification for Psoriatic ARthritis (CASPAR) criteria were recruited at a tertiary referral center. Demographic data were collected and PGA data were determined from administration of an 11-point NRS (0 to 10 points representing best to worst status). Convergent and discriminant validity were evaluated by correlation between PGA and clinical variables. Patients were grouped as having severe disease based on Disease Activity Score 28-joint count (DAS28) > 5.1, Health Assessment Questionnaire (HAQ) > 1.0, walking with aids, and social welfare-dependent. Patients were grouped as being in remission by DAS28 < 2.6 and the Minimal Disease Activity Criteria. Known-group validity of PGA was evaluated.

Results. A total of 125 patients (52% men) were studied. Convergent validity revealed strong correlations of PGA with pain score, HAQ, and DAS28; and weak correlations with skin severity score, physician's global assessment and morning stiffness. In multivariate analysis, PGA was associated with pain, physical function, mental function, and skin severity score. PGA distinguished different levels of severity well, as determined by comparison with different known groups with large effect sizes.

Conclusion. Judged on an NRS, the PGA had good construct validity and satisfactorily distinguished all levels of severity in PsA. (First Release March 15 2012; J Rheumatol 2012;39:844–8; doi:10.3899/jrheum.110919)

Key Indexing Terms:
PSORIATIC ARTHRITIS OUTCOME MEASURE PAIN
QUALITY OF LIFE NUMERIC RATING SCALE

Psoriatic arthritis (PsA) has deleterious effects on joint and skin, causing joint deformities, impaired physical function, and diminished health-related quality of life (HRQOL). As identified by the Group for Research and Assessment of Psoriasis and Psoriatic Arthritis (GRAPPA) for the Outcome Measures in Rheumatology Clinical Trials (OMERACT) working group, physical function, pain, patient's global health assessment (PGA), and quality of life were among the core set of domains that should be included in clinical trials and longitudinal observational studies[1].

The PGA is also commonly used in daily clinical practice, clinical trials, and composite indices in PsA to evaluate disease activity and response to treatment. PGA is usually measured as a self-report rating of global health status using a 0–100 mm visual analog scale (VAS). Another method is the use of an 11-point numeric rating scale (NRS), with numbers anchored at 2 extreme ends. Zero point represents best status, and 10 points the worst status. NRS of the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI), Bath Ankylosing Spondylitis Functional Index (BASFI), and Dougados Functional Index were found to be similar to the VAS in ankylosing spondylitis (AS)[2]. Construct validity of the NRS and VAS in PGA was found to be similar in rheumatoid arthritis (RA)[3]. Although it is widely used, the construct validity of PGA in PsA has not been examined. We evaluated the construct validity of an NRS of PGA in PsA.

MATERIALS AND METHODS

Data collection. We recruited consecutive patients with PsA according to the ClASsification for Psoriatic ARthritis (CASPAR) criteria[4] who attended an outpatient specialist clinic in a single tertiary rheumatology center from January 2008 to December 2008. All patients were Han Chinese who read traditional Chinese characters. These patients were assessed according to a standard protocol. Clinical data were collected including number of tender and swollen joints, damaged joint counts, dactylitis, and enthesitis using the Maastricht Ankylosing Spondylitis Entheses Score[5] and the Psoriasis Area and Severity Index (PASI)[6]. Functional status and HRQOL were assessed by

From the Department of Medicine and Geriatrics, Tai Po Hospital, Hong Kong; Department of Statistics, Chinese University of Hong Kong, Hong Kong; and Department of Medicine and Therapeutics, Chinese University of Hong Kong, Hong Kong, China.

Y-Y. Leung, MBChB, Specialist Resident, Department of Medicine and Geriatrics, Tai Po Hospital; K-W. Ho, PhD, Instructor, Department of Statistics, Chinese University of Hong Kong; T-Y. Zhu, PhD, Research Coordinator; L-S. Tam, MD, Professor, Department of Medicine and Therapeutics, Chinese University of Hong Kong; E.W. Kun, FRCP, Consultant, Department of Medicine and Geriatrics, Tai Po Hospital; E.K-M. Li, MD, FRCP, Professor, Department of Medicine and Therapeutics, Chinese University of Hong Kong.

Address correspondence to Dr. Y-Y. Leung, Department of Rheumatology and Immunology, Singapore General Hospital, Outram Road, Singapore 169608. E-mail: katyccc@hotmail.com

Accepted for publication December 2, 2011.

the following patient-reported outcomes: Health Assessment Questionnaire (HAQ) and the Short-Form 36 Survey (SF-36) Chinese (Hong Kong) version[7]. The norm-based Physical Component Summary (PCS) and Mental Component Summary (MCS) were computed. PGA and pain were rated on an 11-point NRS with 0 representing the best status and 10 points the worst status. Patients were required to choose the exact number to represent their health status and not to choose between scores. A research assistant was responsible for checking the completeness of patient-reported outcomes and clarified with patients if they meant 4 or 5, if they chose between 4 and 5, for example. The translated text of the pain and PGA assessment form is shown in Figure 1.

Construct validity and statistical analysis. Construct validity was assessed by convergent and discriminant validity and by assessment of known-group validity. To assess convergent validity, we determined the Spearman's rho between PGA and various clinical variables. We expect strong correlations between instruments or scales that are conceptually related (convergent validity) and weak correlations in scales that are conceptually unrelated (discriminant validity). A Spearman's rho > 0.5 was considered strong, while 0.3–0.49 was moderate and < 0.3 was weak[8]. Known-group validity was tested by comparing scale scores across groups known to be different[9,10]. Patients were grouped as having severe disease based on Disease Activity Score 28-joint count (DAS28) > 5.1, HAQ > 1.0, walking with aids, or dependency on social welfare. Patients were grouped as having minimal disease based on DAS28 < 2.6 or fulfillment of the Minimal Disease Activity (MDA) criteria[11]. The ability of the instrument to differentiate between known groups was calculated as the statistically significant difference by Mann-Whitney U test between groups. The effect size was calculated as the standardized mean difference described by Cohen[12], that is, the difference in mean scores divided by the pooled standard deviation. The effect size was categorized as small (0.2–0.5), moderate (0.5–0.8), or large (> 0.8). Data were analyzed using SPSS version 11.0. A value of $p < 0.05$ was considered significant.

RESULTS

A total of 125 patients with PsA (65 men, 60 women) were studied. The mean (± SD) age was 47.5 (± 12.4) years and mean duration of illness 8.2 (± 6.8) years. All patients were Han Chinese ethnicity; their demographic and disease characteristics are summarized in Table 1. The PsA cohort had moderate pain score and disability. HRQOL judged by SF-36 scores was much lower than for the normal population.

Convergent and discriminant validity. Analysis of convergent and discriminant validity of the PGA revealed correlations in the expected directions. There were strong correlations of PGA with pain score, DAS28, and HAQ scores. The correlations between PGA and PASI, physician's global assessment, erythrocyte sedimentation rate, and duration of morning stiffness were weak. Essentially, no correlation existed between PGA and age or duration of illness (Table 2). In multivariate regression analysis, PGA was associated with pain score, the PCS and MCS of the SF-36, and the PASI (Table 3). This 4-variable multivariate model explained 47.7% of the variance in PGA.

Known-group validity. Table 4 summarizes the known-group validity of PGA. The expected tendency was found. Patients with severe disease had higher PGA, while patients with mild disease had lower PGA. The differences were statistically significant. The effect sizes for severe disease were large, including HAQ > 1.0, DAS28 > 5.1, and walking with aids. The effect sizes for remission criteria by DAS28 < 2.6 and MDA criteria were also large.

DISCUSSION

PGA is an important core domain for the assessment of PsA. Ours is the first study, to our knowledge, to describe the construct validity of PGA on an 11-point NRS for PsA. In the convergent validity analysis, moderate correlations of PGA with indexes of active disease were observed, whereas correlations with unrelated variables like age and duration of illness were weak. This showed that PGA identifies the activity status in PsA well. Interestingly, PGA was found to be correlated with HAQ, which may indicate that physical disability also affects how patients perceive their overall health status.

From the known-group validity analysis, PGA differentiated between groups of patients with health status that differed according to patient characteristics. PGA differentiated patients with poor functional status and high disease activity by high HAQ scores, walking disability, dependency on wel-

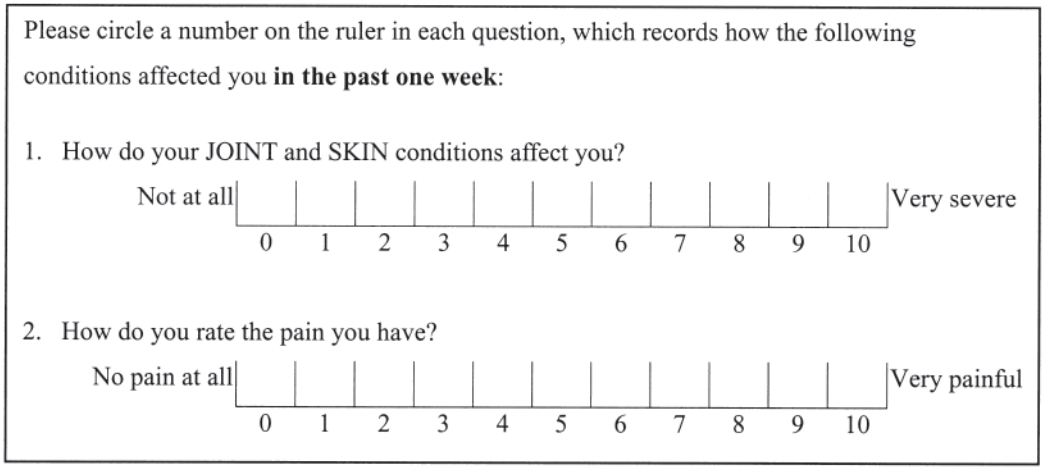

Figure 1. The pain and patient global health assessment form.

Table 1. Demographic and disease characteristics of the cohort with psoriatic arthritis.

Characteristic	Mean (± SD)
Age, yrs	47.5 (± 12.4)
Sex, % female	48
Duration of illness, yrs	8.2 (± 6.8)
Tender joint count, 0–68	3.98 (± 5.22)
Swollen joint count, 0–66	1.84 (± 2.67)
Damaged joint count, 0–68	3.07 (± 4.49)
Dactylitis, 0–20	0.33 (± 0.95)
MASES, 0–13	1.37 (± 2.30)
ESR, mm/h	31.7 (± 28.1)
DAS28	3.8 (± 1.5)
Pain, 0–10	4.78 (± 2.57)
PGA, 0–10	4.56 (± 2.32)
MDGA, 0–10	2.06 (± 1.98)
HAQ, 0–3	0.62 (± 0.60)
PASI, 0–72	5.48 (± 7.33)
SF-36 PCS	39.89 (± 9.43)
SF-36 MCS	43.03 (± 11.49)

MASES: Maastricht Ankylosing Spondylitis Entheses Score; ESR: erythrocyte sedimentation rate; PGA: patient's global assessment of disease activity; MDGA: physician's global assessment of disease activity; HAQ: Health Assessment Questionnaire; PASI: Psoriasis Area and Severity Index; SF-36 PCS: physical component summary score of Medical Outcomes Study Short Form-36; SF-36 MCS: mental component summary score of SF-36.

Table 2. Spearman's correlation of patient global assessment of disease activity with clinical variables. Bold type indicates strong associations.

Characteristic	Spearman's rho
Age	0.03
Duration of illness	–0.008
Tender joint count	0.34**
Swollen joint count	0.15
Damaged joint count	0.10
Dactylitis	0.03
MASES	0.19*
ESR	0.23**
Duration of morning stiffness	0.26**
DAS28	**0.50****
Pain	**0.54****
MDGA	0.29**
HAQ	**0.54****
PASI	0.24**
SF-36 PCS	–0.49**
SF-36 MCS	–0.47**

* $p < 0.05$; ** $p < 0.01$. SF-36 PCS: norm-based physical component summary score of Medical Outcomes Study Short Form-36; SF-36 MCS: norm-based mental component summary score of SF-36. Other definitions as in Table 1.

fare, and high disease activity scores. On the other hand, PGA could also differentiate the patient group with low disease activity by the DAS28 remission criteria and the MDA criteria.

We found that PGA was influenced by pain, physical function, and mental health as well as skin condition. However, these 4 factors jointly explained only 47.7% of the variance of PGA, indicating that PGA is a distinct construct on its own.

There has been controversy about whether patients' global assessment for arthritis and skin should be separated[13]. PGA as a single measure has been used in the development of other composite measures such as the Psoriatic Arthritis Joint Activity Index[14,15] and the Disease Activity index for Psoriatic Arthritis[16]. These composite measures have been validated in large samples of PsA subjects undergoing anti-tumor necrosis factor therapies in randomized controlled trials. A recent multicenter GRAPPA and OMERACT study, mainly in rheumatology clinics, revealed that assessment of joint and skin conditions together is reliable, and joint activity has a major influence on PGA. The authors also illustrated that the skin and joint activities do not correlate with each other[17]. There are certain circumstances under which joint or skin activity should be assessed separately, for example when a drug therapy may adequately improve one of these domains but not the other. In our study, a single PGA score represented a patient's perception of both joint and skin disease. Although the contribution of skin score to change in PGA was small, it remained significant in the multivariate analysis.

Our patients with PsA were recruited from a tertiary rheumatology center. There was a predominance of joint symptoms over skin symptoms. The mean PASI score was relatively low in our cohort (5.48). As a result, PGA had stronger correlation with tender joint count, DAS28, and pain score than with PASI. This situation also occurred in the GRAPPA study on PGA, which recruited patients from mainly rheumatology centers[17]. However, the majority of our patients were also followed by dermatologists, while patients with arthritis not being followed by a rheumatologist were very rare. We believe this represented the general population with PsA. It is important to note, however, that PASI still emerged as an important variable associated with PGA in multivariate analysis, which may indicate that even if skin involvement is not severe on objective assessment, patients still perceive that it has an important influence on their overall health status.

In the field of pain measurement, the NRS and VAS are highly correlated with one another[18]. In clinical trials of RA, the NRS was demonstrated to be more reliable than the VAS, especially with less educated patients[19]. Van Tubergen, *et al* observed high agreement between the VAS and NRS of the BASDAI, BASFI, and Dougados Functional Index in AS[2]. For the PGA in RA, Lati, *et al* also demonstrated that the VAS and NRS had comparable construct validity[3]. Some might think that the VAS has an infinite number of responses and thus would be more sensitive to change. For example, patients could choose a score of 42 on the 100-mm VAS, but cannot choose between 4 and 5 in an 11-point NRS. The VAS and NRS in pain scores were demonstrated to be slightly different in the degree of responsiveness, but both were capable of detecting changes in response after treatment[20]. The presump-

Table 3. Multivariate regression analysis of variables associated with patient global assessment of disease activity.

Variables	Beta	p	Collinearity	
			Tolerance	VIF
Pain	0.348	< 0.001	0.587	1.703
SF-36 MCS	–0.285	< 0.001	0.824	1.213
SF-36 PCS	–0.236	0.005	0.679	1.473
PASI	0.152	0.029	0.977	1.023

SF-36 PCS: physical component summary score of Medical Outcomes Study Short Form-36; SF-36 MCS: mental component summary score of SF-36; PASI: Psoriasis Area and Severity Index; VIF: variance inflation factor.

Table 4. Known-group validity of patient global assessment of disease activity.

Known Groups (no. patients)		PGA, mean (± SD)	Effect Size Cohen's d
Severe disease			
HAQ > 1.0	Yes (35)	6.23 (± 1.72)**	0.82
	No (88)	3.90 (± 2.22)	
DAS28 > 5.1	Yes (31)	5.90 (± 2.09)**	1.17
	No (94)	4.12 (± 2.23)	
Walking with aids	Yes (12)	6.17 (± 1.34)*	0.93
	No (113)	4.39 (± 2.34)	
Social welfare dependence	Yes (27)	5.74 (± 1.87)**	0.72
	No (96)	4.21 (± 2.35)	
Remission criteria			
DAS28 < 2.6	Yes (29)	3.17 (± 2.25)**	–0.82
	No (96)	4.98 (± 2.18)	
MDA	Yes (16)	2.06 (± 2.02)**	–1.39
	No (104)	4.95 (± 2.14)	

* p < 0.05; ** p < 0.01. HAQ: Health Assessment Questionnaire; DAS28: Disease Activity Score 28; MDA: minimal disease activity criteria.

tion that the VAS has an infinite-number response and is more sensitive to change has been disproved. Jensen, *et al*[21] demonstrated that little information was lost when a 101-point NRS was transformed to an 11- or 21-point NRS. A recent systemic review of 44 studies comparing NRS and VAS demonstrated that NRS are applicable for unidimensional assessment of pain in most settings and were noted to have higher compliance rates, better responsiveness and ease of use, and good applicability relative to VAS in 11 out of 19 studies[22]. The NRS can also be administered in written or verbal form, and may be advantageous for future development into computerized administrative systems. The NRS was therefore chosen in our study for its simplicity and ease of use. It may be important to compare PGA using VAS and NRS in future studies.

There are several limitations to our study. First, it was a cross-sectional design that did not address sensitivity to change. Second, the test-retest reliability was not assessed. Third, our study cohort consisted of single-ethnicity patients with PsA (Han Chinese), with long duration of illness, from a tertiary rheumatology center, which limits its generalizability to the entire population with PsA. Moreover, we did not have data to compare the performance of PGA as assessed by NRS and VAS.

PGA on an 11-point numeric rating scale is an instrument for measuring disease activity in PsA that has good construct validity, and it measures a unique construct in PsA.

REFERENCES

1. Gladman DD, Mease PJ, Strand V, Healy P, Helliwell PS, Fitzgerald O, et al. Consensus on a core set of domains for psoriatic arthritis. J Rheumatol 2007;34:1167-70.
2. Van Tubergen A, Debats I, Rvser L, Londono J, Burgos-Vargas R, Cardiel MH, et al. Use of a numerical rating scale as an answer modality in ankylosing spondylitis-specific questionnaires. Arthritis Rheum 2002;47:242-8.
3. Lati C, Guthrie LC, Ward MM. Comparison of the construct validity and sensitivity to change of the visual analog scale and a modified rating scale as measures of patient global assessment in rheumatoid arthritis. J Rheumatol 2010;37:717-22.
4. Taylor W, Gladman D, Helliwell P, Marchesoni A, Mease P, Meilants H, CASPAR Study Group. Classification criteria for psoriatic arthritis: Development of new criteria from a large international study. Arthritis Rheum 2006;54:2665-73.
5. Heuft-Dorenbosch L, Spoorenberg A, van Tubergen A, Landewe R, van ver Tempel H, Mielants H, et al. Assessment of enthesitis in ankylosing spondylitis. Ann Rheum Dis 2003;62:127-32.
6. Fredriksson T, Pettersson U. Severe psoriasis — oral therapy with a new retinoid. Dermatologica 1978;157:238-44.
7. Lam CL, Gandek B, Ren XS, Chan MS. Tests of scaling assumptions and construct validity of the Chinese (HK) version of

the SF-36 Health Survey. J Clin Epidemiol 1998;51:1139-47.
8. Cohen J. A power primer. Psychol Bull 1992;112:155-9.
9. Marra CA, Woolcott JC, Kopec JA, Shojania K, Offer R, Brazier JE, et al. A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the RAQoL and the HAQ) in rheumatoid arthritis. Soc Sci Med 2005;60:1571-82.
10. Linde L, Sorensen J, Ostergaard M, Horslev-Petersen K, Hetland ML. Health-related quality of life: Validity, reliability, and responsiveness of SF-36, 15D, EQ-5D [corrected] RAQoL, and HAQ in patients with rheumatoid arthritis. J Rheumatol 2008;35:1528-37.
11. Coates LC, Fransen J, Helliwell PS. Defining minimal disease activity in psoriatic arthritis: A proposed objective target for treatment. Ann Rheum Dis 2010;69:48-53.
12. Cohen J. Statistical power for the behavioral sciences. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
13. Gladman DD, Mease PJ, Healy P, Helliwell PS, Fitzgerald O, Cauli A, et al. Outcome measures in psoriatic arthritis. J Rheumatol 2007;34:1159-66.
14. Gladman DD, Tom BD, Mease PJ, Farewell VT. Informing response criteria for psoriatic arthritis. I: Discrimination models based on data from 3 anti-tumor necrosis factor randomized studies. J Rheumatol 2010;37:1892-7.
15. Gladman DD, Tom BD, Mease PJ, Farewell VT. Informing response criteria for psoriatic arthritis (PsA). II: Further considerations and a proposal — The PsA Joint Activity Index. J Rheumatol 2010;37:2559-65.
16. Schoels M, Aletaha D, Funovits J, Kavanaugh A, Baker D, Smolen JS. Application of the DAREA/DAPSA score for assessment of disease activity in psoriatic arthritis. Ann Rheum Dis 2010;69:1441-7.
17. Cauli A, Gladman DD, Mathieu A, Olivieri I, Porru G, Tak PP, et al, for GRAPPA 3PPsA Study Group. Patient global assessment in psoriatic arthritis: A multicenter GRAPPA and OMERACT study. J Rheumatol 2011;38:898-903.
18. Price DD, Bush FM, Long S, Harkins SW. A comparison of pain measurement characteristics of mechanical visual analogue and simple numerical rating scales. Pain 1994;56:217-26.
19. Ferraz MB, Quaresma MR, Aquino LR, Atra E, Tugwell P, Goldsmith CH. Reliability of pain scales in the assessment of literate and illiterate patients with rheumatoid arthritis. J Rheumatol 1990;17:1022-4.
20. Bellamy N, Campbell J, Syrotuik J. Comparative study of self-rating pain scales in rheumatoid arthritis patients. Curr Med Res Opin 1999;15:121-7.
21. Jensen MP, Turner JA, Romano JM. What is the maximum number of levels needed in pain intensity measurement? Pain 1994;58:387-92.
22. Hjermstad MJ, Fayers PM, Haugen DF, Caraceni A, Hanks GW, Loge JH, et al. European Palliative Care Research Collaborative (EPCRC). Studies comparing numerical rating scales, verbal rating scales, and visual analogue scales for assessment of pain intensity in adults: A systematic literature review. J Pain Symptom Manage 2011;41:1073-93.