

Health Literacy Predicts Discrepancies Between Traditional Written Patient Assessments and Verbally Administered Assessments in Rheumatoid Arthritis

Joel M. Hirsh, Lisa A. Davis, Itziar Quinzanos, Angela Keniston, and Liron Caplan

ABSTRACT. Objective. Patient assessments of disease activity (PtGA) and general health (GH) measured by visual analog scale (VAS) are widely used in rheumatoid arthritis (RA) clinical practice and research. These require comprehension of the question's wording and translation of disease activity onto a written VAS, which is problematic for patients with limited health literacy (HL) or difficulty completing forms. This study's objective was to validate verbally administered versions of patient assessments and identify factors that might explain discrepancies between verbal and written measures.

Methods. We enrolled patients with RA at the Denver Health rheumatology clinic ($n = 300$). Subjects were randomized to complete the traditional written PtGA and GH and one of the verbal assessments. Subjects provided a verbal numeric response after reading the question, having the question read to them in person, or hearing the question over the phone. Spearman and Lin correlations comparing written and verbal assessments were determined. Multivariate logistic regression was performed to explain any discrepancies.

Results. The instruments administered verbally in-person showed good, but not excellent, correlation with traditional written VAS forms (Spearman coefficients 0.59 to 0.70; $p < 0.001$ for all correlations). Twenty-three percent of subjects were unable to complete 1 of the written VAS assessments without assistance. HL predicted missing written data and discrepancies between verbal and written assessments ($p < 0.05$ for all correlations).

Conclusion. Providers should use verbal versions of PtGA and GH with caution while caring for patients unable to complete traditional written version. Limited HL is widely prevalent and a barrier to obtaining patient-oriented data. (First Release Dec 15 2013; J Rheumatol 2014;41:256–64; doi:10.3899/jrheum.130664)

Key Indexing Terms:

RHEUMATOID ARTHRITIS

QUESTIONNAIRES

OUTCOME ASSESSMENT

From the Division of Rheumatology, Department of Medicine, Denver Health Medical Center, in conjunction with the Division of Rheumatology, Department of Medicine, University of Colorado Denver School of Medicine (SOM), and the Department of Medicine, Denver Veterans Affairs (VA) Medical Center, Denver, Colorado, USA.

Supported by a UCB pharmaceuticals Investigator Initiated Studies grant to Denver Health. Dr. Caplan is supported by VA Health Services Research and Development Career Development Award 07-221. Dr. Davis is supported by the Rheumatology Research Foundation Scientist Development Award. The views expressed do not represent the views of University of Colorado SOM, the Department of Veterans Affairs, or the Denver Health and Hospital Authority.

J.M. Hirsh, MD; L.A. Davis, MD, MSCS, Division of Rheumatology, Department of Medicine, Denver Health Medical Center, and Division of Rheumatology, Department of Medicine, University of Colorado Denver SOM; I. Quinzanos, MD, Division of Rheumatology, Department of Medicine, Denver Health Medical Center, Division of Rheumatology, Department of Medicine, University of Colorado Denver SOM, and Department of Medicine, Denver VA Medical Center; A. Keniston, MSPH, Division of Rheumatology, Department of Medicine, Denver Health Medical Center; L. Caplan, MD, PhD, Division of Rheumatology, Department of Medicine, University of Colorado Denver SOM, and Department of Medicine, Denver VA Medical Center.

Address correspondence to Dr. J. Hirsh, Denver Health and Hospital Authority, 777 Bannock St., Mail Code 4000, Denver, Colorado 80204, USA. E-mail: joel.hirshmd@dhha.org

Accepted for publication October 25, 2013.

Patient global assessments of disease activity (PtGA) in rheumatoid arthritis (RA) and general health (GH) as assessed by self-administered visual analog scales (VAS) are widely used in RA clinical practice and research trials¹. These patient-reported outcomes have assumed critical importance in recent years owing to the prevailing movement to guide RA therapy based on disease activity (or “treat-to-target”) as measured by composite indices, all of which include either the GH-VAS or the PtGA-VAS². The profile of these measures has increased even further with remission emerging as a realistic goal for RA treatment and the recent publication of remission criteria that mandate low PtGA-VAS or GH-VAS levels³.

The PtGA-VAS and GH-VAS have received additional scrutiny because of evidence showing discrepancies with evaluator global assessment of disease activity (EGA) and other disease activity indices^{4,5,6}. Recent research also demonstrates that these patient assessments alone may prevent patients from being classified as in remission^{7,8}. This emphasis on the PtGA-VAS and GH-VAS is likely to continue with the potential adoption of patient-based

remission criteria⁹. It will also intensify if RA medication efficacy trials increasingly align with the treat-to-target model, rather than the current use of the American College of Rheumatology (ACR) 20, 50, and 70 criteria, which allow for an improvement without a mandated low level of the PtGA-VAS or GH-VAS^{10,11}.

Despite this recent emphasis on the PtGA-VAS and GH-VAS, completion of PtGA in the VAS format might be problematic for many patients. This includes patients with poor vision, limited literacy, limited numeracy, patients unable to write because of hand deformity, neurologic disorders or muscular diseases, and patients unable to keep in-person appointments. These conditions and circumstances are common: one-third of patients in the United States have basic or below basic health literacy^{5,12} and about 3% of elderly adults have functional visual impairment¹³. There is no published literature regarding how assessments of PtGA and GH are gathered in these patients who are unable to complete the standard written form, but in all likelihood, the information is gathered verbally.

Some existing literature has established the correlation of verbal rating scales (VRS) for pain (rather than PtGA and GH) with VAS measures in patients with inflammatory arthritis¹⁴. These verbal measures perform suboptimally in patients with less education or limited literacy^{15,16}. Interestingly, patients with less educational attainment often prefer the Pain-VRS and find it easier to understand than the traditional written Pain-VAS¹⁷. One pilot study gathered verbally administered GH in patients with RA¹⁸, but there are no published data regarding the validity of verbally attained PtGA or GH.

To investigate the validity of verbally administered PtGA and GH in English-speaking and Spanish-speaking patients with RA, we performed a cross-sectional observational study at an academic, safety-net rheumatology clinic. We hypothesized that verbally obtained responses would strongly correlate with the traditional VAS versions. In addition, we anticipated that sociodemographic factors such as education may be associated with discrepancies between verbally obtained and traditional written patient VAS assessments. We also hypothesized that sociodemographic factors may be associated with missing and noninterpretable written VAS patient assessments. Finally, to explore the potential use of verbally obtained information by telephone in phase IV clinical studies and between-visit care, we examined various methods of administering a verbally obtained patient assessment (in person, by telephone) to determine whether these methods affected the correlation of verbal and written patient assessments.

MATERIALS AND METHODS

Patients who met the ACR 1987¹⁹ criteria for RA were eligible for recruitment. We excluded subjects less than 18 years of age, prisoners, persons with uncontrolled psychiatric illness, and patients with vision worse than 20/100, as measured by a Snellen eye chart. Patients who

received an intramuscular or intraarticular corticosteroid injection during their visit were also excluded, because the injections could potentially confound the comparisons of self-assessments performed at the time of the study visit. Subjects received \$10 to participate in the study. The subjects for this study were recruited on the day of a regularly scheduled patient visit from the Denver Health rheumatology clinic between November of 2011 and January of 2013. Denver Health and Hospital Authority is an urban safety-net system that serves 150,000 patients annually, of which 78% are minorities and 50% are uninsured²⁰. A bilingual research assistant identified eligible subjects by reviewing the daily clinic schedule and medical charts for patients who self-identified either English or Spanish as their primary language. We recruited subjects at the end of their rheumatology clinic visits by first providing patients with a verbal description of the study prior to the use of any written materials. The recruitment process was conducted in a confidential workroom with sensitivity and respect to ensure the dignity of persons with limited literacy²¹ and to mitigate against response bias. Subjects were randomly assigned to complete written assessments at either the beginning or end of the study visit. Subjects completed all written assessments in their primary language without assistance from the research assistant. The complete study packet was 11 pages. The written VAS assessments were placed either first or last in the study packet to minimize the risk of them being skipped. Figure 1 is a flowchart of the study visit.

We used the following definitions, measures, and assessments.

Written patient global assessments. Subjects were asked to complete written assessments based on terminology for the PtGA from the Disease Activity Score 28-joint count (DAS28)²² and GH from the Multi-dimensional Health Assessment Questionnaire (MDHAQ)²³.

For the PtGA-VAS, subjects provided a written global assessment of their disease severity by recording their responses as marks on a horizontal visual analog scale, 100 mm in length, in response to the question, "Please mark an 'X' on the line below to show how active has your rheumatoid arthritis been during the past seven days." The left anchor was "no disease activity" and the right anchor was "high disease activity"²².

For the GH-VAS, subjects provided a written global health assessment on a horizontal VAS 100 mm in length based on the standard terminology used in MDHAQ: "Considering all the ways in which illness and health conditions may affect you at this time, please make an 'X' on the line below to show how you are doing." The left anchor was "very well" and the right anchor was "very poorly"²³.

Verbal patient global assessments. Subjects were randomly assigned to complete a verbal numeric response of between 0 and 100 for both versions of the patient assessments at either the beginning or end of the study visit. Subjects were randomized to provide 1 of 4 different forms of the verbally obtained (VO) patient assessments.

VO-PtGA1 and VO-GH1 were obtained by the patient reading the global assessments in their primary language and providing a verbal numeric response to the research assistant in person.

VO-PtGA2 and VO-GH2 were obtained by the research assistant reading the global assessments to the patient in their primary language and the patient providing a verbal numeric response to the research assistant in person.

VO-PtGA3 and VO-GH3 were obtained by the research assistant reading the global assessments to the patient in their primary language and the patient providing a verbal numeric response by telephone in a research visit room at the start of the study visit.

VO-PtGA4 and VO-GH4 were obtained by the research assistant reading the global assessments to the patient in their primary language and the patient providing a verbal numeric response by telephone within 3 h after their clinic visit on their mobile or home telephones.

Twenty-eight-joint tender joint count (28TJC) and swollen joint count (28SJC), evaluator global assessment of disease activity (EGA). The provider who performed a history and physical examination during the clinical visit recorded the results of a 28TJC and 28SJC and EGA using a

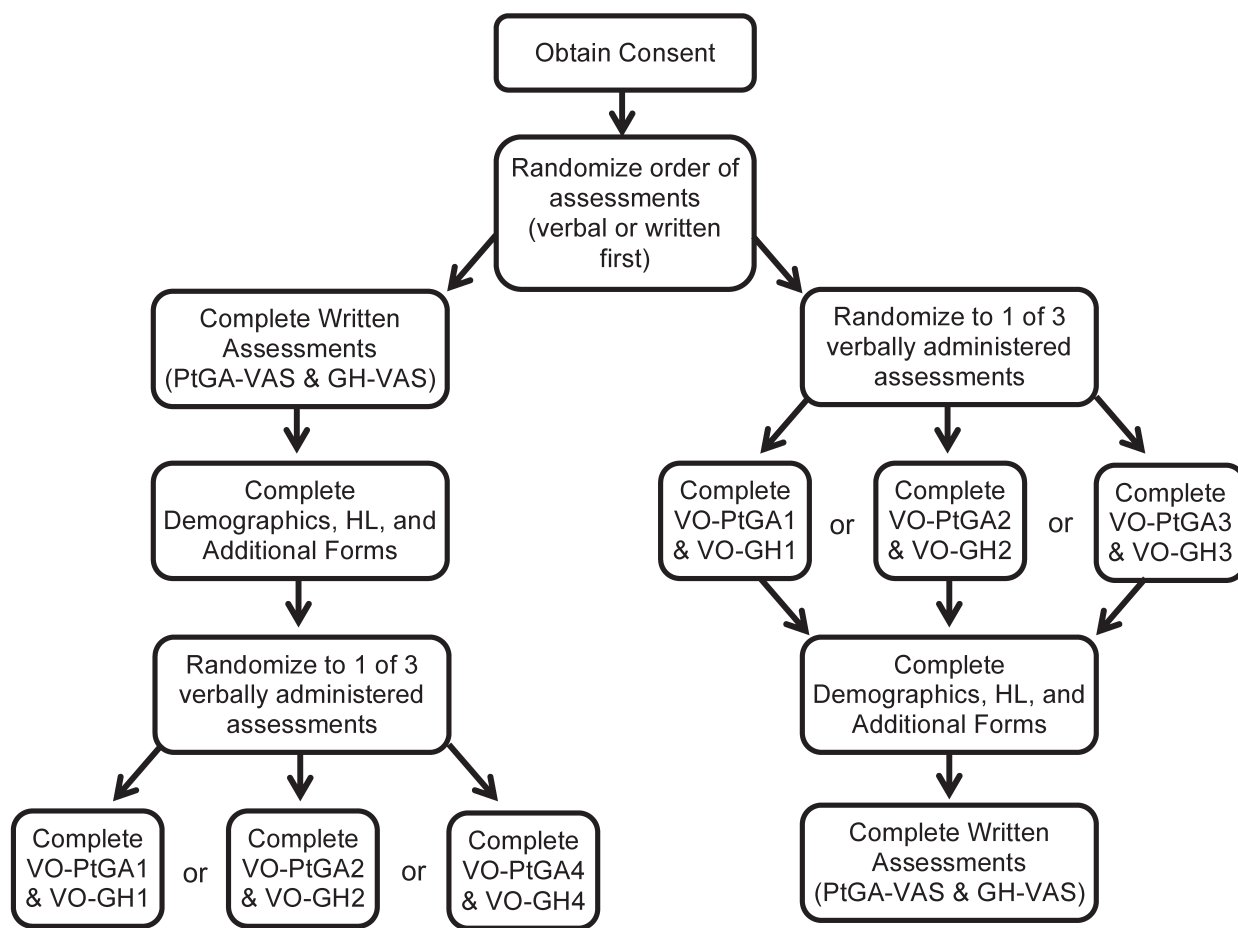


Figure 1. Participant study visit flow. HL: healthy literacy; PtGA: patient global assessment from the Disease Activity Score with 28-joint count; VAS: visual analog scale; GH: global health assessment from the Multidimensional Assessment Questionnaire; VO: verbally obtained; VO-PtGA1/VO-GH1: patient reads the global assessment questions and provides verbal numeric responses to the research assistant in person; VO-PtGA2/VO-GH2: research assistant reads the global assessment questions to the patient and the patient provides verbal numeric responses to the research assistant in person; VO-PtGA3/VO-GH3: research assistant reads the global assessment questions to the patient and the patient provides verbal numeric responses by telephone in a quiet room; VO-PtGA4/VO-GH4: research assistant reads the global assessment questions to the patient and the patient provides verbal numeric responses by telephone after the visit.

horizontal 100 mm VAS. A score of 100 signified “very poorly” and a score of 0 signified “very well”²³. The provider was blinded to all patient global assessments.

Demographic information. We obtained by written self-report from the subjects the following: patient’s primary language, age, sex, ethnicity, race, marital status, highest level of education completed, tobacco use history, annual household income, and employment status.

Tests of functional health literacy (HL). We used the modified Short Test of Functional Health Literacy in Adults (S-TOFHLA), a validated English and Spanish language instrument to assess functional HL²⁴. The large-print version of the modified S-TOFHLA is a 12-min, 14-point font cloze procedure that contains 2 reading comprehension sections about medical subjects and 4 questions assessing numeracy²⁴. A cloze procedure is an evaluation in which words are deleted from written passages regarding medical subject matter, and respondents attempt to select the correct missing words. Inadequate HL was defined as a total score of < or equal to 53. Marginal HL was defined as scores between 54–66.

Laboratory studies. We performed chart review to determine results of prior anticyclic citrullinated peptide antibody (anti-CCP) tests. Erythrocyte

sedimentation rate (ESR, mm/h) and C-reactive protein (mg/dl) levels were obtained as part of clinical care following the completion of the study paperwork. All laboratory analyses were performed at the Denver Health Medical Center clinical laboratory.

Prior patient experience. We performed chart review to determine whether subjects had completed the PtGA-VAS previously, and their history of biologic medication use and current prednisone use/dose.

MDHAQ-pain. The subjects circled a number between 0 and 10 along a horizontal scale in response to the question “How much pain have you had because of your condition over the past week?”²³.

MDHAQ-fatigue. The subjects provided an assessment of fatigue by circling a number between 0 and 10 using a horizontal scale in response to the question “How much of a problem has unusual fatigue or tiredness been for you over the past week?”²³.

Depression. Subjects completed the 2-question patient health questionnaire (PHQ-2) depression screen²⁵.

Outcome measures. The correlation coefficient of the VO-PtGA and VO-GH with PtGA-VAS and GH-VAS were the primary outcome measures. Written assessments were classified as noninterpretable if the

subject attempted to complete the PtGA or GH but provided a response that could not be scored. Examples of such responses include writing along the VAS line or circling an anchor. Written assessments were classified as missing if the PtGA or GH were left blank by the subject on the form after (1) the research assistant explained the need to complete the forms, (2) the subject was given as much time as required to complete the form, and (3) the subject returned the form to the research assistant. Verbal responses were considered missing if the patient failed to return telephone calls.

Statistical methods. Prior to the study, we determined that a sample size of 100 (representing each examined VO and written VAS pair) would achieve 87% power to detect a difference in correlation of 0.3. To determine whether we could aggregate verbal self-assessment scores determined by telephone in a quiet room and scores determined by a phone after the visit, we used a t-test to compare mean scores. To determine whether the order of written assessments was associated with missing data, we performed the chi-square test and compared the presence of missing data between subjects who completed the written assessments at the end of the protocol to those who did so at the onset of the study.

We addressed missing data by imputing missing data based on a linear regression that included all clinical and demographic characteristics from Table 1. Correlation of the VO-PtGA and VO-GH were determined by Spearman coefficients, Bland-Altman 95% limits of agreement,

Bland-Altman plots, and Lin's Concordance correlation coefficients. Comparisons were made to written VAS assessments and the EGA. We modeled all global assessments as continuous variables. HL, estimated by the S-TOFHLA, was also modeled as a continuous score, as has been done previously²⁶. We defined disagreement/discrepancy between verbal and written assessment as those with a difference in scores of at least 2 SD of the written versions or EGA. We performed an initial univariate logistic regression to evaluate which variables predicted disagreement. We then performed a multivariate logistic regression including all terms with a p-value of 0.5. In the final model, p-values < 0.05 were considered significant. Regression analyses to determine predictors of missing or noninterpretable patient assessments followed identical procedures. All analyses were performed using Stata software version 12 (StataCorp).

The Colorado Multiple Institutional Review Board approved our study.

RESULTS

Three hundred thirty-seven patients were screened for the study, which represents about 90% of the eligible subjects in our clinic. Fourteen persons declined to participate in the study for unspecified reasons. Fifteen patients were excluded because of complete functional illiteracy. Six

Table 1. Demographics of study cohort.

Variable	Mean, %	SD	Min	Max
Age, yrs, mean	53.0	13.0	18.4	79.8
Male	23.0			
Married	33.0			
Spanish speaker	28.0			
Race/ethnicity				
White	46.7			
Black or African American	17.0			
Hispanic or Latino	57.8			
Smoker, ever	54.4			
Attended college	30.3			
Annual income at or above \$15K (US)	27.2			
Depression screen-positive	24.0			
Currently employed	23.3			
Rheumatoid arthritis-related features				
MDHAQ score (0–3)	1.0	0.7	0	3
MDHAQ fatigue score (0–10)	5.3	3.3	0	10
MDHAQ pain score (0–10)	5.2	3.1	0	10
Tender joint count (0–28)	5.1	5.8	0	28
Swollen joint count (0–28)	3.8	4.7	0	22
Sedimentation rate (mm/h)	22.3	19.5	3	112
C-reactive protein (mg/dl)	10.5	17.1	0	130
DAS28 (4 variable w/sedimentation rate)	4.0	1.5	1	8
Anti-CCP-positive	82.0			
Recent prednisone use	51.7			
Recent biologic use	30.0			
Patient global health from MDHAQ (0–100)	41.1	29.3	0	100
PtGA from DAS (0–100)	48.4	32.4	0	100
Previously completed DAS28 form	62.5	0.0	0	0
EGA (0–100)	34.6	27.5	0	100
Health literacy				
S-TOFHLA comprehension score (0–72)	54.7	19.9	0	72
S-TOFHLA numeracy score (0–28)	22.2	6.9	0	28
S-TOFHLA total score (0–100)	77.6	23.5	8	100

Min: minimum; Max: maximum; MDHAQ: Multidimensional Health Assessment Questionnaire; anti-CCP: anticyclic citrullinated peptide antibody; DAS28: 28-joint Disease Activity Score; S-TOFHLA: Short Test of Functional Health Literacy in Adults; EGA: evaluator global assessment; PtGA: patient global assessment.

patients were excluded owing to uncontrolled psychiatric disease. Two patients were excluded because of severe visual, hearing, or speech impairments, producing a study cohort of 300 subjects.

The subjects were predominantly female, unmarried, English-speaking, and currently unemployed. Sixty-three percent had completed the PtGA-VAS and none had completed the GH-VAS previously as part of their clinical care at Denver Health. This cohort represents a diverse population with about half being nonwhite and 58% Hispanic. Only a third had attended college. Eighty-two percent were anti-CCP-positive and the mean DAS28 was 4.0. Twenty-eight percent of patients had either inadequate or marginal functional HL. Table 1 summarizes the demographic characteristics of the study population.

Results of the criterion validity testing for the verbally obtained global assessments appear in Table 2. Any scoreable response was included in the criterion validity analyses. Sixty-nine subjects (23% of the cohort) had at least 1 missing or noninterpretable VAS response that was imputed. Such responses were received on 25 of both the GH-VAS and PtGA-VAS. Eleven of 50 patients randomized to perform the VO-PtGA4 and VO-GH4 at the conclusion of the study were unreachable by telephone. Patients were 22% less likely to respond to calls after the visit, and there was a statistical trend toward a difference in scores ($p = 0.07$) between VO-PtGA3 (self-assessment by telephone in a quiet room) and VO-PtGA4 (response by telephone after visit). Further, a visual inspection of score distributions for VO-PtGA3 and VO-PtGA4 suggested that the non-responders likely constituted those subjects who would have provided high self-assessment scores. This finding was corroborated by imputation of the missing scores, which predicted high self-assessment scores for subjects who

failed to respond to calls after the visit, whether the patient assessment adopted the DAS or MDHAQ formatted questions. For these reasons, we did not combine VO-PtGA3 and VO-PtGA4 scores in the analysis.

Results of additional assessments using Lin's concordance correlation, Bland-Altman 95% Limits of Agreement, and Bland-Altman plots (supplementary material available from the author on request) were consistent with the findings of Table 2. The correlation of the in-person verbally administered patient assessments with the traditional written/VAS patient assessments ranged from Spearman coefficients of 0.59 to 0.70, which is generally considered moderate to good correlation. The results were nonsignificant ($p > 0.05$) for verbal assessments obtained by phone at the end of the study visit.

About one-third of patients provided VO-PtGA responses that were discrepant by more than 2 SD with their PtGA-VAS. Similar findings were noted for the VO-GH responses (Figure 2). These discrepancies affected the classification of subjects' disease activity: 17% of subjects' disease activity classification changed if a verbally obtained PtGA replaced the PtGA-VAS in the DAS28-ESR calculation (Supplemental Table 2 available from the author on request).

With the exception of the VO-PtGA4 and VO-GH4, correlations of the verbally obtained globally assessments with the EGA demonstrated coefficients that ranged from 0.38 to 0.54, which is considered fair to moderate correlation. Similar results were noted, however, between the written patient global assessments and EGA. The correlations all demonstrated statistically significant relationships ($p < 0.001$ for all correlations).

We performed univariate and multivariate logistic regression analysis to determine which clinical and

Table 2. Correlations of various verbally administered patient assessments, standard written assessments, and evaluator global assessments.

Patient Assessment	Instrument Origin	Who Reads Patient Assessment?	Format of Patient's Response	n	Correlation with EGA	n	Correlation with Written VAS
Standard GH-VAS	MDHAQ	Patient	Written	300	0.400	—	—
VO-GH1	MDHAQ	Patient	Verbal	100	0.419	100	0.590
VO-GH2	MDHAQ	Staff, in person	Verbal	100	0.421	100	0.658
VO-GH3	MDHAQ	Staff, by phone in room	Verbal	50	0.381	50	0.593
VO-GH4	MDHAQ	Staff, by phone after visit	Verbal	50	0.353	50	0.376
Standard PtGA-VAS	DAS28	Patient	Written	300	0.452	—	—
VO-PtGA1	DAS28	Patient	Verbal	100	0.538	100	0.704
VO-PtGA2	DAS28	Staff, in person	Verbal	100	0.406	100	0.657
VO-PtGA3	DAS28	Staff, by phone in room	Verbal	50	0.546	50	0.445
VO-PtGA4	DAS28	Staff, by phone after visit	Verbal	50	0.264	50	0.261

All correlations are significant to $p < 0.01$, except VO-PtGA4 ($p > 0.05$ for correlations and Limits of Agreement). EGA: evaluator global assessment; VAS: visual analog scale; GH: global health; VO: verbally obtained; MDHAQ: Multidimensional Health Assessment Questionnaire; VO-PtGA1/VO-GH1: patient reads the global assessment questions and provides verbal numeric responses to the research assistant in person; VO-PtGA2/VO-GH2: research assistant reads the global assessment questions to the patient and the patient provides verbal numeric responses to the research assistant in person; VO-PtGA3/VO-GH3: research assistant reads the global assessment questions to the patient and the patient provides verbal numeric responses by telephone in a quiet room; VO-PtGA4/VO-GH4: research assistant reads the global assessment questions to the patient and the patient provides verbal numeric responses by telephone after the visit; DAS28: Disease Activity Score with 28-joint count.

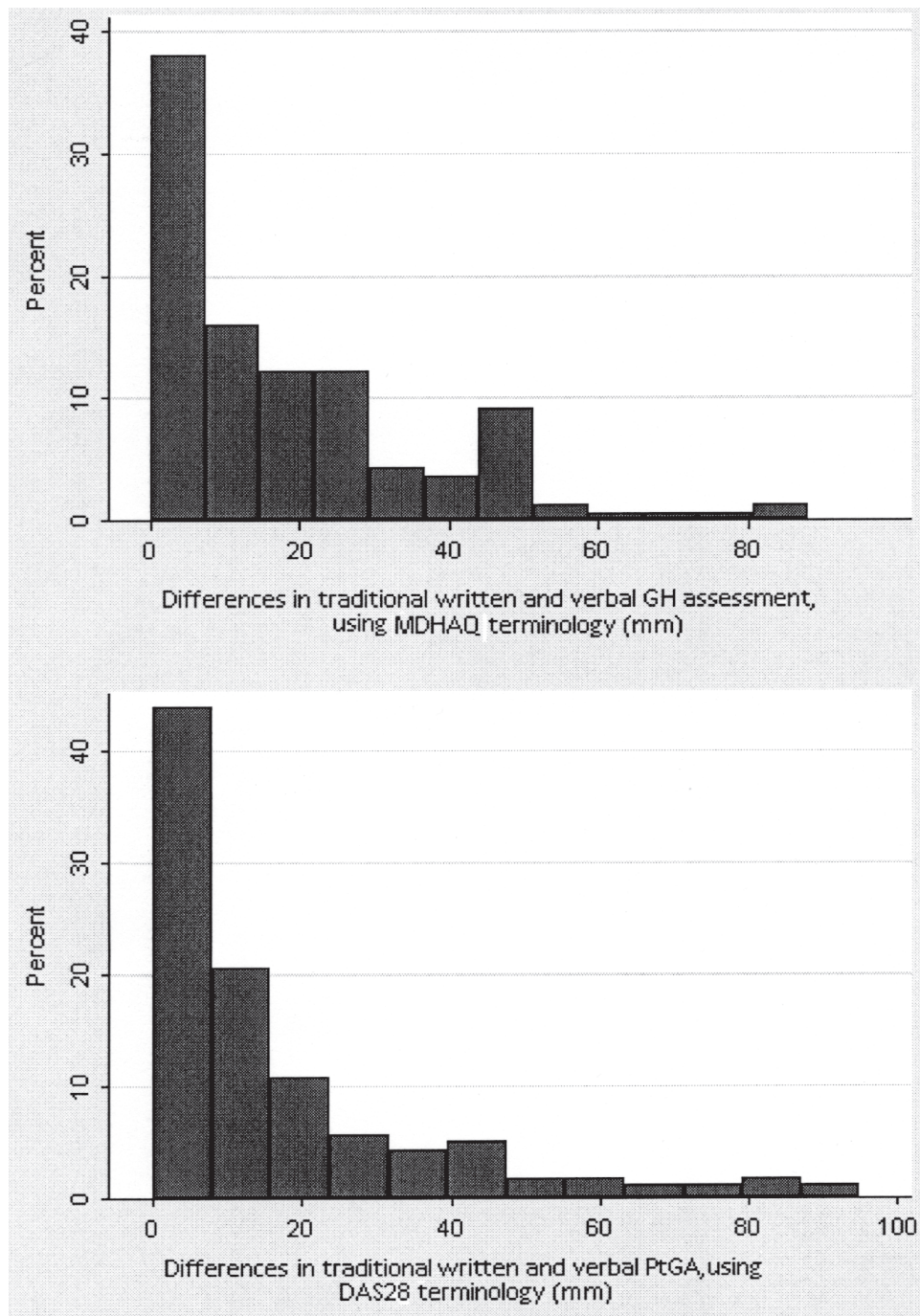


Figure 2. Distributions of the absolute differences in traditional written patient assessments and verbally administered patient assessments for the study cohort. GH: global health assessment; MDHAQ: Multidimensional Health Assessment Questionnaire; PtGA: patient global assessment; DAS28: Disease Activity Score with 28-joint count.

demographic features predicted a greater than 2 SD difference between the verbally administered patient assessments and traditional VAS assessments (Table 3; univariate results available from the author on request). In multivariate analysis, discrepancy between the VO-PtGA and PtGA-VAS was only consistently predicted by the S-TOFHLA compre-

hension score (greater comprehension decreased the odds of discrepancy). The S-TOFHLA numeracy score was the sole statistically significant predictor of discrepancy between the VO-GH and GH-VAS.

We also performed univariate and multivariate logistic regression analysis to determine the odds of missing written

Table 3. Final multivariate logistic regression models predicting odds of a discrepancy between verbally administered and written patient assessment and odds of a missing patient assessment score.

Variable	OR	Std. Err.	p	95% CI
Odds of a verbally administered PtGA and traditional written PtGA-VAS discrepancy				
S-TOFHLA comprehension score	0.959	0.012	0.001	0.936–0.983
Odds of a verbally administered GH and traditional written GH-VAS discrepancy				
S-TOFHLA numeracy score	0.940	0.028	0.042	0.887–0.998
Odds of a missing PtGA value				
S-TOFHLA comprehension score	0.899	0.021	0.000	0.859–0.941
Age, yrs	0.952	0.018	0.011	0.917–0.989
Odds of a missing GH value				
S-TOFHLA comprehension score	0.965	0.010	0.001	0.946–0.985

Std. Err.: standard error; PtGA: patient global assessment, as used in Disease Activity Score with 28-joint count (DAS28); VAS: visual analog scale; S-TOFHLA: Short Test of Functional Health Literacy in Adults; GH: global health assessment, as used in Multidimensional Health Assessment Questionnaire.

PtGA or GH values and noninterpretable VAS responses (univariate results are available from the author on request). In multivariate regression, missing and noninterpretable PtGA data was only predicted by the S-TOFHLA comprehension score and younger age (Table 3). Numeracy as defined by the S-TOFHLA numeracy score was the only variable that predicted missing and noninterpretable GH data (greater numeracy associated with fewer missing values).

DISCUSSION

We were able to demonstrate a moderate correlation between in person verbally obtained PtGA and GH and traditionally obtained written VAS versions of these measures. This correlation was not as strong as we had hypothesized. We were also able to confirm that socio-demographic factors are associated with the presence of a discrepancy between verbally and traditionally obtained patient assessments. Specifically, we demonstrated that limited HL predicts the presence of a clinically relevant discrepancy. An additional important finding of our study was that a large number of subjects with limited HL did not complete the traditional written VAS versions of the PtGA and GH without assistance. The relationship between limited HL and missing VAS data suggests that some patients are facing significant challenges completing this instrument. Given the association of limited HL and missing data, our estimates for the association of HL with discrepant assessments likely represents an underestimate for the true effect size (compared with complete data).

The correlation coefficients were similar between written VAS forms and verbally obtained assessments, regardless of whether the patients read the question themselves or the research assistant read the question to the patient in person. This finding supports the future use of in-person verbally administered patient assessments for population-based studies and randomized controlled trials when subjects are unable to complete the written forms. The modest corre-

lation of verbally administered patient assessments with traditional written assessment, however, suggests that clinical decision making for the individual patient should not be made based on verbal patient assessments alone. In instances in which providers are unable to use the traditional written instruments, additional weight should be given to other indicators (laboratory tests, etc.).

Our results expand the science of patient-reported outcomes in 3 key areas. First, our study is the first, to our knowledge, to investigate the validity of verbally administered versions of the PtGA and GH. Second, we have enriched the understanding of the literacy burdens of the instruments used to collect patient-reported outcomes. And third, we established the relationship between limited HL and incomplete or inaccurately completed patient assessments in subjects with RA. We are unaware of previous research documenting how frequently patients need assistance with the GH-VAS or PtGA-VAS. Our finding that more than 1 in 5 patients in our clinic provided blank or “un-score-able” VAS responses raises concern that critical patient report outcomes are functioning as a hidden barrier to assessing disease activity in research trials and clinical care. The link between limited HL and incomplete or inaccurately completed PtGA-VAS or GH-VAS responses has not been previously reported by other groups. We are performing ongoing research to determine the role of patient and system factors that are hindering proper completion of the GH-VAS and PtGA-VAS.

Previous research has investigated techniques to facilitate the capture and measurement of the PtGA and GH. Pincus, *et al* studied modifications to the GH-VAS and other VAS metrics including the use of 21 circles at 5 mm intervals²⁶. Studies have also attempted to capture the GH or PtGA electronically with Web-based portals, computer touch screens, and personal digital assistant-based formats; these investigations, in general, demonstrate modest associations with traditional VAS techniques^{27,28,29}. Verbal rating scales of pain have been studied in patients with inflam-

matory arthritis¹⁵, but our study is the first, to our knowledge, to systematically assess the criterion validity of verbally obtained PtGA and GH. This is important because these measures are obtained frequently in research trials and increasingly in clinical practice, given the numerous conditions and circumstances in which the traditional written assessments are not available.

This study has several limitations. Its cross-sectional design does not permit us to determine whether limited HL was causally associated with the missing data or discrepancy between verbal and written assessments. It is possible that limited HL is a proxy for a factor inadequately controlled for in our regression analysis, such as education and income level, which we only captured as categorical variables. We did, however, examine a large number of variables that may influence HL, including race, ethnicity, age, employment, income, and language. The relatively low number of Spanish-speaking patients may have limited our power to detect the relationship of primary language with our various outcomes. We also did not perform 2 written GH-VAS and PtGA-VAS on each subject, so we are unable to ascertain the baseline test-retest variability of these instruments. Because we did not want to influence patient responses through subtle prompts, we deliberately gave no guidance to patients with regard to completion of the instruments, apart from the actual assessment question itself. This contributed to missing GH-VAS and PtGA-VAS data in about 20% of our patients. Imputation of missing data, however, had no effect on our findings.

It is possible that minor assistance may have improved the performance of the verbal assessments and reduced the rates of missing data; however, this approach would also have the potential to bias our results. Nearly two-thirds of our subjects had completed PtGA-VAS previously, but none had completed a GH-VAS at Denver Health. This was not a likely source of bias, because similar discrepancies were seen with the written and verbal forms of both instruments, and prior completion of a PtGA-VAS did not predict discrepant scores. Subjects who completed the written VAS first were less likely to have missing values for their self-assessments; however, the relationships between HL scores on the S-TOFHLA and odds of a missing PtGA or GH (Table 3) remained after stratifying the analysis by the order in which the written self-assessment was administered.

With regards to strengths, our study included a broad range of patients in terms of sex, age, race, ethnicity, and literacy skills, typical of an urban public health hospital in the United States. The results likely have external validity when applied to other diverse populations, but may not apply to more homogenous subpopulations.

Our findings are relevant to both rheumatology practice and research. Providers and researchers should be cognizant of limited HL and appreciate that some of our

patient-reported outcomes, although laudable in their attempt to honor the patient's perspective, may not truly be reflective of the patient's disease state. Additionally, accurate completion of the PtGA-VAS and GH-VAS is challenging for patients with limited HL and introduces additional complexity to the delivery of care to these patients³⁰. Our group is engaged in efforts to understand these instruments better from the patient perspective, with a focus on which domains (anchors, wording of question, scale, etc.) engender the most confusion. National quality measures for the care of patients with RA need to reflect the challenges that patients with impaired HL face in attempting to complete patient-reported measures. There has been increased awareness regarding the myriad of differently worded versions of the PtGA³¹, but greater focus and standardization needs to be applied to the instructions and explanations given to the patients when they require assistance completing these instruments. Additional research is needed on how best to capture the PtGA and GH in patients with low HL or inability to complete written testing for other reasons.

REFERENCES

1. Khan NA, Spencer HJ, Abda EA, Alten R, Pohl C, Ancuta C, et al. Patient's global assessment of disease activity and patient's assessment of general health for rheumatoid arthritis activity assessment: are they equivalent? *Ann Rheum Dis* 2012;71:1942-9.
2. Anderson J, Caplan L, Yazdany J, Robbins ML, Neogi T, Michaud K, et al. Rheumatoid arthritis disease activity measures: American College of Rheumatology recommendations for use in clinical practice. *Arthritis Care Res* 2012;64:640-7.
3. Felson DT, Smolen JS, Wells G, Zhang B, van Tuyl LH, Funovits J, et al. American College of Rheumatology/European League Against Rheumatism provisional definition of remission in rheumatoid arthritis for clinical trials. *Arthritis Rheum* 2011;63:573-86.
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;62:857-64.
5. Hirsh JM, Boyle DJ, Collier DH, Oxenfeld AJ, Caplan L. Health literacy predicts the discrepancy between patient and provider global assessments of rheumatoid arthritis activity at a public urban rheumatology clinic. *J Rheumatol* 2010;37:961-6.
6. 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;64:206-14.
7. Masri KR, Shaver TS, Shahouri SH, Wang S, Anderson JD, Busch RE, et al. Validity and reliability problems with patient global as a component of the ACR/EULAR remission criteria as used in clinical practice. *J Rheumatol* 2012;39:1139-45.
8. Studenic P, Smolen JS, Aletaha D. Near misses of ACR/EULAR criteria for remission: effects of patient global assessment in Boolean and index-based definitions. *Ann Rheum Dis* 2012;71:1702-5.
9. van Tuyl LH, Boers M. Patient-reported remission in rheumatoid arthritis. *J Rheumatol* 2013;40:350-2.
10. Buch MH, Pavitt S, Parmar M, Emery P. Creative trial design in RA: optimizing patient outcomes. *Nat Rev Rheumatol* 2013;9:183-94.
11. Felson DT, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith C, et al. American College of Rheumatology.

- Preliminary definition of improvement in rheumatoid arthritis. *Arthritis Rheum* 1995;38:727-35.
12. Kutner M, Greenberg E, Jin Y, Boyle B, Hsu Y, Dunleavy E. Literacy in Everyday Life: Results From the 2003 National Assessment of Adult Literacy (NCES 2007-480). [Internet. Accessed Nov. 19, 2013.] Available from: http://nces.ed.gov/Pubs2007/2007480_1.pdf
 13. Tanna AP, Kaye HS. Trends in self-reported visual impairment in the United States: 1984 to 2010. *Ophthalmology* 2012;119:2028-32.
 14. Englbrecht M, Tarner IH, van der Heijde DM, Manger B, Bombardier C, Muller-Ladner U. Measuring pain and efficacy of pain treatment in inflammatory arthritis: a systematic literature review. *J Rheumatol Suppl.* 2012 Sept;90:3-10.
 15. 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.
 16. Mohan H, Ryan J, Whelan B, Wakai A. The end of the line? The Visual Analogue Scale and Verbal Numerical Rating Scale as pain assessment tools in the emergency department. *Emerg Med J* 2010;27:372-5.
 17. Clark P, Lavielle P, Martinez H. Learning from pain scales: patient perspective. *J Rheumatol* 2003;30:1584-8.
 18. Potter T, Wild A, Edwards K, Rai A, Rowe IF. Patients' own ability to assess activity of their rheumatoid arthritis. *Rheumatology* 2006;45:1044.
 19. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988;31:315-24.
 20. Eisert SL, Mehler PS, Gabow PA. Can America's urban safety net systems be a solution to unequal treatment? *J Urban Health* 2008;85:766-78.
 21. Wolf MS, Williams MV, Parker RM, Parikh NS, Nowlan AW, Baker DW. Patients' shame and attitudes toward discussing the results of literacy screening. *J Health Commun* 2007;12:721-32.
 22. Prevoo ML, van 't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, van Riel PL. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis Rheum* 1995;38:44-8.
 23. Pincus T, Sokka T, Kautiainen H. Further development of a physical function scale on a MDHAQ [corrected] for standard care of patients with rheumatic diseases. *J Rheumatol* 2005;32:1432-9.
 24. Baker DW, Williams MV, Parker RM, Gazmararian JA, Nurss J. Development of a brief test to measure functional health literacy. *Patient Educ Couns* 1999;38:33-42.
 25. Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: validity of a two-item depression screener. *Med Care* 2003;41:1284-92.
 26. Pincus T, Bergman M, Sokka T, Roth J, Swearingen C, Yazici Y. Visual analog scales in formats other than a 10 centimeter horizontal line to assess pain and other clinical data. *J Rheumatol* 2008;35:1550-8.
 27. Athale N, Sturley A, Skoczen S, Kavanaugh A, Lenert L. A web-compatible instrument for measuring self-reported disease activity in arthritis. *J Rheumatol* 2004;31:223-8.
 28. Greenwood MC, Hakim AJ, Carson E, Doyle DV. Touch-screen computer systems in the rheumatology clinic offer a reliable and user-friendly means of collecting quality-of-life and outcome data from patients with rheumatoid arthritis. *Rheumatology* 2006;45:66-71.
 29. Heiberg T, Kvien TK, Dale O, Mowinckel P, Aanerud GJ, Songe-Moller AB, et al. Daily health status registration (patient diary) in patients with rheumatoid arthritis: a comparison between personal digital assistant and paper-pencil format. *Arthritis Rheum* 2007;57:454-60.
 30. Paasche-Orlow MK, Wolf MS. The causal pathways linking health literacy to health outcomes. *Am J Health Behav* 2007;31 Suppl 1:S19-26.
 31. French T, Hewlett S, Kirwan J, Sanderson T. Different wording of the patient global visual analogue scale (PG-VAS) affects rheumatoid arthritis patients' scoring and the overall disease activity score (DAS28): a cross-sectional study. *Musculoskeletal Care* 2013 Feb 4 (E-pub ahead of print).