

UCSF

UC San Francisco Previously Published Works

Title

Effects of Language, Insurance, and Race/Ethnicity on Measurement Properties of the PROMIS Physical Function Short Form 10a in Rheumatoid Arthritis

Permalink

<https://escholarship.org/uc/item/7gr0z73s>

Journal

Arthritis Care & Research, 71(7)

ISSN

2151-464X

Authors

Izadi, Zara

Katz, Patricia P

Schmajuk, Gabriela

et al.

Publication Date

2019-07-01

DOI

10.1002/acr.23723

Peer reviewed



# HHS Public Access

Author manuscript

*Arthritis Care Res (Hoboken)*. Author manuscript; available in PMC 2020 August 26.

Published in final edited form as:

*Arthritis Care Res (Hoboken)*. 2019 July ; 71(7): 925–935. doi:10.1002/acr.23723.

## Effects of language, insurance and race/ethnicity on measurement properties of the PROMIS Physical Function Short Form 10a in rheumatoid arthritis

**Zara Izadi,**

Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA

**Patricia P. Katz,**

Division of Rheumatology, University of California, San Francisco, San Francisco, CA

**Gabriela Schmajuk,**

Division of Rheumatology, San Francisco Veterans Affairs Medical Center, San Francisco, CA

**Julie Gandrup,**

Division of Rheumatology, University of California, San Francisco, San Francisco, CA

**Jing Li,**

Division of Rheumatology, University of California, San Francisco, San Francisco, CA

**Milena Gianfrancesco,**

Division of Rheumatology, University of California, San Francisco, San Francisco, CA

**Jinoos Yazdany**

Division of Rheumatology, University of California, San Francisco, San Francisco, CA

### Abstract

**Objective:** Most studies evaluating patient-reported outcomes such as the PROMIS Physical Function Short Form 10a (PF10a) in rheumatoid arthritis (RA) have been performed in Caucasian and English-speaking populations. We assessed the measurement properties of the PF10a in a racially/ethnically diverse population with RA. We determined the effect of non-English language proficiency, insurance status and race/ethnicity, on the validity and responsiveness of the PF10a.

**Methods:** Data were abstracted from electronic health records for all RA patients seen in a university-based rheumatology clinic between 2013 and 2017. We evaluated the PF10a's use, floor and ceiling effects, and construct validity across categories of language preference, insurance and race/ethnicity. We used standardized response means and linear mixed-effects models to evaluate the responsiveness of the PF10a to longitudinal changes in the Clinical Disease Activity Index (CDAI) across population subgroups.

---

**Corresponding Author:** Zara Izadi, PharmD, Department of Epidemiology and Biostatistics, Mission Hall: Global Health & Clinical Sciences Building, 550 16th Street, 2nd Floor, Box #0560, San Francisco, CA 94158-2549, Tel: (510) 737-9909, zara.izadi@ucsf.edu.

COI statement:

All authors have no commercial or financial conflicts of interest to disclose.

**Results:** We included 595 patients in a cross-sectional analysis of validity and 341 patients in longitudinal responsiveness analyses of the PF10a. The PF10a had acceptable floor and ceiling effects and was successfully implemented. We observed good construct validity and responsiveness to changes in CDAI among whites, English-speakers and privately-insured patients. However, constructs evaluated by the PF10a were less correlated with clinical measures among Chinese-speakers and Hispanics, and less sensitive to clinical improvements among Medicaid patients and Spanish-speakers.

**Conclusion:** While the PF10a has good measurement properties and is both practical and acceptable for implementation in routine clinical practice, we also found important differences across racial/ethnic groups and those with limited English proficiency that warrant further investigation.

---

Assessment of patient-reported physical function is important for monitoring individuals with rheumatoid arthritis (RA) and provision of patient-centered care. Integration of patient-reported physical function into routine clinical care has been shown to be a feasible mechanism for incorporating patient preferences into a treat-to-target approach for managing RA (1). Incorporation of patient-reported measures of physical function in RA is a nationally-endorsed quality measure and recommended in American College of Rheumatology guidelines (2).

The Patient Reported Outcome Measurement Information System (PROMIS®) was developed by the National Institute of Health (NIH) to provide standard metrics for measuring patient-reported outcomes across chronic conditions. To date, a number of researchers have examined the measurement properties of PROMIS measures in rheumatic conditions, with the largest concentration of work on the physical functioning measures in RA (3–11). While PROMIS physical function measures have been evaluated in white and English-speaking populations with RA, no studies have examined their validity or responsiveness in other racial and ethnic groups, non-English speakers or populations with low socioeconomic status.

Previous studies have shown that sociodemographic factors can affect multiple aspects of care in RA including mortality and disability, disease activity, prevalence of comorbidities, patient-reported outcomes, access to treatment/health services, treatment preferences and medication adherence, health literacy, and trust in providers (12–35). A better understanding of the effects of sociodemographic factors on the validity of PROMIS physical function measures will determine their generalizability across diverse communities. This study investigates the effect of language preference, race/ethnicity, and insurance status (as a proxy for low income) on measurement properties of the PROMIS Physical Function Short Form 10a (PF10a) including floor and ceiling effects, construct validity, and responsiveness to improvements and deteriorations in clinical disease activity over time.

## Patients and Methods

### Data sources

Provider information, clinical and demographic data were extracted from the electronic health record (EHR) for all patients seen in the UCSF Rheumatology clinic with at least one

face-to-face encounter with a rheumatologist that was associated with an ICD-9 code for RA between February 1, 2013 and October 31, 2017. The UCSF Committee on Human Research approved this study.

### Study population

To assess cross-sectional validity, we included patients who had at least one encounter with complete data, including a pain score, PF10a score, Clinical Disease Activity Index (CDAI) score, serum C-reactive protein (CRP) and erythrocyte sedimentation rate (ESR). For the longitudinal analysis of responsiveness, we restricted analysis to the cohort of patients with at least two encounters with complete data. For inclusion in the analysis, encounters were required to be spaced between 1-12 months apart to capture separate episodes of care. Baseline was defined as the first encounter with complete data.

### Measures

Physical function was measured using the PF10a for all patients. The PF10a is a 10-item questionnaire assessing current self-reported physical function. Raw scores range from 10 to 50 and can be translated into T-scores, with a mean of 50 and a standard deviation of 10, for comparison with the U.S. general population mean; for this study, all reported PF10a scores are T-scores. A higher PROMIS-PF10a T-score represents better physical function. Chinese and Spanish PF10a forms were obtained from [www.nih.promis.gov](http://www.nih.promis.gov) and were utilized for patients who preferred these languages. All forms were scored and entered by clinic staff prior to the clinic visit.

RA disease activity was measured using the Clinical Disease Activity Index (CDAI) (19), a composite measure of Patient Global Assessment (visual analog scale from 0-100 mm), Evaluator Global Assessment (visual analog scale from 0-100 mm), 28-tender joint counts (TJC) and 28-swollen joint counts (SJC). Scores range from 0-76 with higher values reflecting more severe disease. All patients completed a visual analog scale (0-100 mm) for pain at each encounter, and CRP or ESR was measured at least every 3 months.

### Other variables

Other time-varying variables included body mass index (BMI) and smoking status which were recorded at each encounter. Baseline variables included demographics, number of comorbidities (Charlson comorbidity score) and medication use. Demographics included date of birth, sex, self-reported race/ethnicity, preferred language, and insurance category (Private, Medicare, Medicaid). For medication use at baseline, physician medication orders for all oral or intravenous drugs including biologic disease modifying anti-rheumatic drugs (DMARDs), targeted small-molecule DMARDs, non-biologic DMARDs, and glucocorticoids, associated with a rheumatology encounter within 12-months before baseline, were retrieved from the EHR.

### Statistical analysis

Pearson's chi-squared, one-way ANOVA or Kruskal-Wallis test were selected for descriptive statistics based on the type and sample distribution of the variable being analyzed.

**Floor and ceiling effects**—The proportion of individuals with floor (defined as worst score; 14.1) and ceiling (defined as best score; 61.7) effects for PF10a was calculated across different categories of language, insurance, and race/ethnicity.

**Validity**—Construct validity, the extent to which a test measures the concept or construct that it is intended to measure, was assessed by looking at convergent, discriminant, and known-group validity. Convergent and discriminant validity explain how a measure conforms to a similar or different measure and were assessed by comparing correlation of PF10a to that of patient global RA assessments, pain, swollen- and tender- joint count, ESR and CRP with Spearman's correlation coefficient, as not all scores were normally distributed. We hypothesized that the PF10a would correlate strongly ( $r < -0.60$ ) with other patient-reported measures (Patient Global Assessment VAS, Pain VAS), and moderately ( $-0.30 < r < -0.60$ ) with clinical outcome measures (28 tender and swollen joint counts) (36). Spearman's correlation coefficient was compared across different categories of language, insurance, and race/ethnicity.

Known-groups validity explains how the measure discriminates between groups that are known to be different. This was investigated by evaluating differences in mean PF10a scores among predefined groups by age or disease severity. PF10a was hypothesized to show lower scores in older patients (age  $\geq 65$  compared to age  $< 50$ ), and those with higher disease activity (CDAI score  $> 22$  (severe) compared to CDAI score  $\leq 10$  (remission or mild)). T-tests were used to compare mean group differences in each category by language, insurance, and race/ethnicity, and Cohen's  $d$  effect size (the difference in mean scores divided by the pooled standard deviation) was calculated. Effect size values for dichotomous variables were categorized as small ( $< 0.5$ ), medium (0.5-0.8), or large ( $> 0.8$ ) (37).

**Responsiveness**—Responsiveness was determined by analyzing changes in PF10a scores in relation to changes in disease activity (CDAI). Previous studies have defined a minimally important difference (MID) in CDAI as a 12-point change (38). To estimate the standardized response mean (SRM), patients with PF10a scores recorded on two encounters 1-12 month apart were divided into three groups: those with a 12-point decrease in CDAI (clinical improvement), those with a 12-point increase in CDAI (clinical deterioration), and those with a  $< 12$ -point change in CDAI (no change). We investigated the association between language preference, insurance status and race/ethnicity, and mean score changes of the PF10a using a test for trend (39) and calculated the ratio of the mean score change to the standard deviation of that change (SRM) across subgroups. Values were categorized as small ( $< 0.5$ ), medium (0.5-0.8), and large ( $> 0.8$ ) (40) and compared across different categories of language, insurance, and race/ethnicity.

Finally, we used multi-level mixed effects linear regression to assess the responsiveness of PF10a to changes over the follow up period by modeling the relationship between changes in PF10a and changes in CDAI among all patients with at least two clinical encounters. We used a random effects model (Model 1), allowing each subject to have his/her own starting intercept and disease trajectory. Also, since there may have been systematic differences in how providers rate swollen and tender joints in the CDAI, we accounted for clustering by provider. Because most patients saw the same provider across all visits, a nested random

effects model was used. The association between change in CDAI score and change in PF10a might be influenced by the magnitude of the initial PF10a score; we therefore adjusted for the initial PF10a score. Since different patients had follow-up visits at different times, we also incorporated time as a linear predictor in the model. To assess differences in PF10a responsiveness across population groups, we fitted three additional models, with an interaction term between change in CDAI score (since the previous encounter) and either language (Model 2), insurance (Model 3), or race/ethnicity (Model 4) in each model. To assess responsiveness to both improvements and deteriorations in CDIA, we fitted splines with a single knot at CDAI of 0. In our fully-adjusted analyses, in addition to the above terms, we also included baseline covariates (age, sex, smoking status, Charlson comorbidity score (41), and medications), and time-dependent covariates (CRP, and BMI). Analyses were performed using Stata (StataCorp. 2014. Stata Statistical Software: Release 14. College Station, TX: StataCorp LP).

## Results

Data from 846 RA patients and 5,834 encounters (mean (SD) encounters per patient of 10 (5), range 1-31) were extracted from the EHR. PF10a scores were recorded for 833 (98%) patients at 5,174 (89%) encounters. The final dataset for cross-sectional analysis included 595 patients. Of these, 341 patients had complete data on at least two encounters spaced 1-12 months apart (mean (SD) encounters per patient of 6 (3), range 2-15) and were included in the longitudinal analysis of responsiveness (see Figure 1 for study populations). A total of 32 providers contributed data for analysis with a mean (SD) RA encounters per provider of 23 (11), range 2-45.

Baseline clinical characteristics of patients included in the cross-sectional sample were representative of RA populations previously described (9, 42) and similar to patients included in the longitudinal cohort (Table 1). The majority of patients were female (83%) with mean (SD) age of 56 (15) years. The group was racially/ethnically diverse (50% non-white) and 14% preferred a language other than English. Most patients had Medicare (47%) or private (40%) insurance. Mean (SD) PF10a score was 40 (11), nearly a standard deviation lower than the overall US population mean; about half had moderate or severe disease activity scores at baseline and the majority had received at least one non-biologic DMARD (52%) or a biologic (43%) at baseline.

Preferred language, insurance, and racial/ethnic groups differed by age, disease activity and number of comorbidities at baseline. Chinese-speakers were on average older than Spanish or English-speakers (68, 59 and 55 years;  $p=0.022$ ). Significantly more patients with Medicaid coverage had moderate-severe disease activity (CDAI  $\geq 10$ ), than Medicare or privately insured patients (73%, 53%, 41%;  $p<0.001$ ). Medicaid patients and African Americans had a statistically significantly higher median Charlson comorbidity score at baseline (2), than other insurance groups (all 1;  $p=0.015$ ) and other race/ethnicities (all 1;  $p<0.001$ ). Baseline PF10a scores were lower among non-English speakers, Medicaid patients and African Americans (Table 1, Supplementary material).

## Floor and ceiling effects

Clustering at the floor was low, ranging from 0-7.5% across all language, insurance and racial/ethnic groups (Table 2, Supplementary material). Similarly, ceiling effects ranged from nil among Chinese-speakers to 11% among privately insured patients.

## Validity

In examining convergent and discriminant validity, PF10a scores were strongly correlated ( $r = -0.60$ ) with patient global assessment of RA activity in all groups except Chinese-speakers ( $r = -0.52$ ) and Medicaid patients ( $r = -0.53$ ) (Table 2). PF10a scores also had strong correlations with pain in all groups except Chinese-speakers ( $r = -0.52$ ) and African Americans ( $r = -0.53$ ). Correlations between PF10a and clinical outcomes (SJC and TJC) were moderate ( $-0.3 < r < -0.6$ ) in most groups; Chinese-speakers and Hispanics had weak correlations ( $r > -0.3$ ) with both outcomes, while Spanish-speakers and African Americans had moderate correlations with TJC but weak correlations with SJC. Correlations between PF10a and inflammatory markers were weak or negligible among groups except English-speakers, whites and privately insured patients who had moderate correlations with CRP.

In examining known group validity, patients who had more active disease (CDAI $>22$ ), had significantly lower mean PF10a scores, as hypothesized (Table 3). In the group dichotomized by disease activity (CDAI $>22$  vs. CDAI  $\leq 10$ ), effect size (Cohen's  $d$ ) was large ( $>0.8$ ) and statistically significant with respect to all sociodemographic variables except for Hispanic race which had a small and non-significant effect size. Older patients also had lower mean PF10a scores compared to younger patients; however, differences were not clinically or statistically significant in most groups. As expected, younger Medicare patients had significantly worse physical functioning.

## Responsiveness

Of the 341 patients with at least two clinical encounters, median (IQR) interval between visits was 126 (97-202) days. Patients with two encounters were divided into 3 subgroups based on whether they had a 12-point change in CDAI (clinical improvement, no change, and clinical deterioration). Mean PF10a scores decreased with clinical deteriorations, remained constant with no clinical change and increased with clinical improvements over time (Table 3, Supplementary material). Mean score changes differed significantly between groups ( $p < 0.05$ ) in all groups of language, insurance and race/ethnicity except Chinese-speakers, African Americans and Asians. In the improvement group, the SRM was large ( $>0.8$ ) in English and Chinese-speakers, those with private insurance, whites and Hispanics; small ( $<0.5$ ) in Spanish-speakers and African Americans; and medium (0.5-0.8) in all other groups. In the deterioration group, SRM was large or medium in all language, insurance, and race/ethnicity groups with sufficient numbers for analysis.

Linear mixed effects regression showed that both clinical improvements and deteriorations were associated with changes in PF10a scores over time ( $p < 0.001$ ), suggesting that PF10a is responsive to changes in clinical disease activity. In a model without interaction terms (Model 1), a 12-point increase in CDAI was associated with a 2.93 (95% CI 2.06, 3.80) point decrease in PF10a, and a 12-point decrease in CDAI was associated with a 2.70 (2.00,



3.41) point increase in PF10a scores. We constructed three additional models (Model 2-4), incorporating an interaction term between CDAI and either language, insurance or race/ethnicity. Model 2 showed that PF10a is more responsive to clinical deteriorations among Chinese-speakers than English-speakers (a 12-point increase in CDAI was associated with a 5.96-point decrease in PF10a among Chinese-speakers and a 2.92-point decrease in PF10a among English-speakers;  $p=0.036$ ) (Table 4, and Figure 2a). The PF10a was less responsive to clinical improvements among Spanish-speakers than English-speakers (a 12-point decrease in CDAI was associated with a 0.67-point increase in PF10a among Spanish-speakers and a 3.08-point increase in PF10a among English-speakers;  $p=0.029$ ). Among Chinese-speakers, PF10a appeared to be more responsive to clinical deteriorations, than clinical improvements, although this shift in responsiveness was not statistically significant ( $p=0.065$ ). Model 3 showed highest sensitivity to changes in disease activity among individuals with private insurance. The responsiveness of the PROMIS PF10a to clinical deteriorations was lower among Medicare ( $-1.62$ ;  $p<0.001$ ) and Medicaid ( $-1.79$ ;  $p=0.005$ ) patients, than privately insured patients ( $-5.33$ ) (Figure 2b). Responsiveness to clinical improvements was also lower among Medicaid patients than privately insured patients ( $0.70$ ,  $3.63$ ;  $p=0.005$ ). Model 4 showed highest responsiveness to clinical deteriorations among Asians ( $-3.49$ ) and highest responsiveness to clinical improvements among whites ( $3.41$ ). Differences in responsiveness to clinical improvements or deteriorations between whites and non-whites did not reach statistical significance.

In all models, consistent results were obtained after adjusting for age, sex, BMI, smoking, baseline medications, baseline total Charlson comorbidities score and CRP (data not shown).

## Discussion

Consistent with prior research, PF10a has strong measurement properties and is responsive to longitudinal changes in disease activity among whites, English-speakers and privately insured patients. However, our study highlights important differences across racial/ethnic groups and those with limited English proficiency.

Impressively, PF10a scores were recorded for 98% of eligible patients at 89% of encounters, even among those with non-English language proficiency. This finding demonstrates that PF10a can be collected efficiently and consistently over a prolonged period in a busy clinic providing care to a diverse community. Fewer ceiling effects were noted in this clinical sample than in some research samples (22). Given that floor and ceiling effects were below the commonly accepted criteria of 15% (43) across all categories of language, insurance and race/ethnicity, PF10a seems both practical and acceptable for use in a general practice setting.

In our evaluations of convergent and discriminant validity, PF10a scores generally correlated strongly with other patient-reported outcomes, moderately with clinical measures and weakly with laboratory measures. Although findings are consistent with prior research (8, 9) and reflective of the instrument's convergent and discriminant validity, we observed some deviations among language, insurance and race/ethnicity groups. Most notably, non-English speakers, Hispanics and African Americans had weaker correlations between PF10a scores



and clinical outcomes like the tender and swollen joint counts. While some of these correlations may have been limited by small samples, unravelling the contributions of other factors that may contribute to these findings is important. Some literature suggests that a higher prevalence of depression and chronic pain in some populations may hinder correlations between patient-reported and physician reported outcomes in RA (22, 27, 44). Further, we found that Medicaid patients had weaker correlations between PF10a and patient global assessment of RA disease activity. Median baseline comorbidity scores were significantly higher among Medicaid patients than patients with private insurance. Since PF10a items are generic and do not specifically address RA-related impairments in physical functioning (45), weaker correlations between PF10a and patient global assessment may be attributed to non-RA comorbidities that were more prevalent among Medicaid patients.

Known group differences by disease activity largely performed as hypothesized. A smaller effect size observed among Hispanics is likely the result of weak correlations between PF10a and clinical measures among this group coupled with small sample sizes. Known group differences by age did not perform as hypothesized and PF10a score differences among those <50 years compared to those ≥ 65 years were mostly small or statistically non-significant. We observed better than expected PF10a scores among older individuals (data not shown) which may indicate a response-shift, reflecting how patients adapt to and report their level of physical functioning over time. Response-shifts occur as patients recalibrate as they learn to live with RA. For instance, some RA patients have reported that when they record a score of “0” on a questionnaire, this does not necessarily represent the absence of a symptom, but instead reflects a new baseline of “what is normal for me” (46).

Standardized response means, obtained in our evaluations of responsiveness, show PF10a captured expected change and stability in scores across language, insurance, and race/ethnicity groups with sufficient numbers for analysis. While another approach to evaluating responsiveness relies on patient self-reported change anchors obtained at a fixed time point, this was not possible in our retrospective analysis of clinical data. Mixed effects modelling has been used previously to assess longitudinal responsiveness of a measure (47) and was used to model the relationship between changes in CDAI and changes in PF10a score. Among our entire eligible clinic population, we found that a 12-point increase in CDAI was associated with a 2.93-point decrease in PF10a, and a 12-point decrease in CDAI was associated with a 2.70-point increase in PF10a. Importantly, these findings are quantitatively consistent with prior evaluations of the responsiveness of PROMIS physical function measures anchored by deteriorations in clinical disease activity (9) or using patient-reported change anchors (7). However, we found that PF10a responsiveness to clinical improvements and deteriorations varied among population subgroups and was most notably influenced by insurance type and language preference. Patients with Medicaid coverage had worse baseline RA disease activity. Worse general health states among non-English speakers and low socioeconomic groups have been described previously (48). One possible explanation for PF10a’s poor responsiveness to clinical improvements among Medicaid patients and Spanish-speakers may be average time spent in ill-states (11). It is possible that patients under-report their physical function during periods of clinical improvement because they reference their usual state rather than their improved state. Future research should examine responsiveness of PF10a among non-English speakers and low socioeconomic groups to

patient-reported change using validated change anchors. Responsiveness may also be dependent upon patients' physical functioning at baseline; Spanish-speakers, Medicaid patients and African Americans had small PF10a score changes in response to CDAI worsening because their baseline PF10a scores were already poor and could not deteriorate much more (22).

Strengths of our study include representation of RA patients across the spectrum of RA disease activity, and inclusion of data from a large, real-world cohort. However, our study has some limitations. First, we were not able to examine individual item characteristics of the PF10a, which might inform internal consistency of items across language, insurance and race/ethnicity groups. Second, the use of CDAI as an anchor for changes in clinical disease activity may have introduced incorporation bias due to strong correlations between the PF10a score and patient global assessment, which is a component of CDAI. Incorporation bias occurs when a reference standard is used that incorporates some of the test that is the subject of investigation. The result is a bias toward stronger associations between PF10a and CDAI among subgroups in whom PF10a is strongly correlated with patient global assessment. However, correlations between PF10a and patient global assessment varied only slightly among our population subgroups, making incorporation bias less likely in this study. Finally, sample size was modest among non-English speakers and non-whites, and we were underpowered to examine these subgroups in some analyses.

Reliable, precise, and accurate measurement of symptoms and functional status across the continuum of disease activity has never been more important to optimize RA treatment given that remission or low disease activity is the current target for management (2, 49). While ongoing efforts are in place to investigate the cross-cultural validity of PROMIS measures, our study is the first to evaluate the validity and responsiveness of a PROMIS measure across different languages, races/ethnicities and insurance groups in a real-world clinic population and serves as an important step in the ongoing evaluation of the PROMIS Physical Function item bank. Our study demonstrated constructs evaluated by PF10a were less correlated with clinical measures among Chinese-speakers and Hispanics, and less sensitive to clinical improvements among Medicaid patients and Spanish-speakers.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## References

1. Bacalao EJ, Greene GJ, Beaumont JL, Eisenstein A, Muftic A, Mandelin AM, et al. Standardizing and personalizing the treat to target (T2T) approach for rheumatoid arthritis using the Patient-Reported Outcomes Measurement Information System (PROMIS): baseline findings on patient-centered treatment priorities. *Clinical rheumatology*. 2017;36(8):1729–36. [PubMed: 28653263]
2. Singh JA, Saag KG, Bridges SL Jr, Akl EA, Bannuru RR, Sullivan MC, et al. 2015 American College of Rheumatology Guideline for the Treatment of Rheumatoid Arthritis. *Arthritis care & research*. 2016;68(1):1–25. [PubMed: 26545825]
3. Fries JF, Cella D, Rose M, Krishnan E, Bruce B. Progress in assessing physical function in arthritis: PROMIS short forms and computerized adaptive testing. *The Journal of rheumatology*. 2009;36(9):2061–6. [PubMed: 19738214]

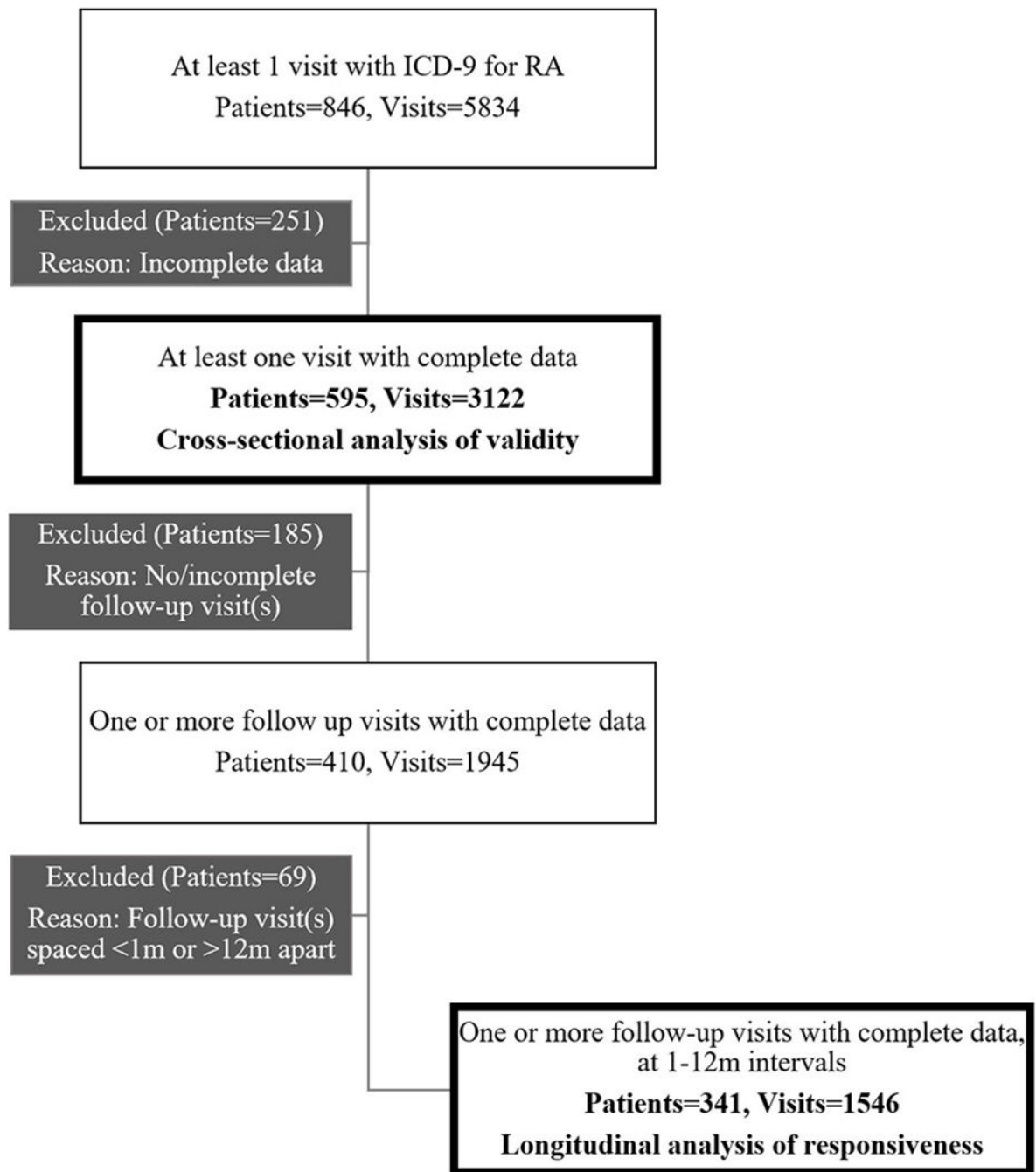
4. Fries JF, Krishnan E, Rose M, Lingala B, Bruce B. Improved responsiveness and reduced sample size requirements of PROMIS physical function scales with item response theory. *Arthritis research & therapy*. 2011;13(5):R147. [PubMed: 21914216]
5. Bjorner JB, Rose M, Gandek B, Stone AA, Junghaenel DU, Ware JE Jr. Method of administration of PROMIS scales did not significantly impact score level, reliability, or validity. *Journal of clinical epidemiology*. 2014;67(1):108–13. [PubMed: 24262772]
6. Bartlett SJ, Orbai AM, Duncan T, DeLeon E, Ruffing V, Clegg-Smith K, et al. Reliability and Validity of Selected PROMIS Measures in People with Rheumatoid Arthritis. *PloS one*. 2015;10(9):e0138543. [PubMed: 26379233]
7. Hays RD, Spritzer KL, Fries JF, Krishnan E. Responsiveness and minimally important difference for the patient-reported outcomes measurement information system (PROMIS) 20-item physical functioning short form in a prospective observational study of rheumatoid arthritis. *Annals of the rheumatic diseases*. 2015;74(1):104–7. [PubMed: 24095937]
8. Oude Voshaar MA, Ten Klooster PM, Glas CA, Vonkeman HE, Taal E, Krishnan E, et al. Validity and measurement precision of the PROMIS physical function item bank and a content validity-driven 20-item short form in rheumatoid arthritis compared with traditional measures. *Rheumatology (Oxford, England)*. 2015;54(12):2221–9.
9. Wahl E, Gross A, Chernitskiy V, Trupin L, Gensler L, Chaganti K, et al. Validity and Responsiveness of a 10-Item Patient-Reported Measure of Physical Function in a Rheumatoid Arthritis Clinic Population. *Arthritis care & research*. 2017;69(3):338–46. [PubMed: 27332620]
10. Cook KF, Jensen SE, Schalet BD, Beaumont JL, Amtmann D, Czajkowski S, et al. PROMIS measures of pain, fatigue, negative affect, physical function, and social function demonstrated clinical validity across a range of chronic conditions. *Journal of clinical epidemiology*. 2016;73:89–102. [PubMed: 26952842]
11. Schalet BD, Hays RD, Jensen SE, Beaumont JL, Fries JF, Cella D. Validity of PROMIS physical function measured in diverse clinical samples. *Journal of clinical epidemiology*. 2016;73:112–8. [PubMed: 26970039]
12. Baldassari AR, Cleveland RJ, Luong MN, Jonas BL, Conn DL, Moreland LW, et al. Socioeconomic factors and self-reported health outcomes in African Americans with rheumatoid arthritis from the Southeastern United States: the contribution of childhood socioeconomic status. *BMC musculoskeletal disorders*. 2016;17:10. [PubMed: 26754747]
13. Barton JL, Schmajuk G, Trupin L, Graf J, Imboden J, Yelin EH, et al. Poor knowledge of methotrexate associated with older age and limited English-language proficiency in a diverse rheumatoid arthritis cohort. *Arthritis research & therapy*. 2013;15(5):R157. [PubMed: 24432366]
14. Barton JL, Trupin L, Tonner C, Imboden J, Katz P, Schillinger D, et al. English language proficiency, health literacy, and trust in physician are associated with shared decision making in rheumatoid arthritis. *The Journal of rheumatology*. 2014;41(7):1290–7. [PubMed: 24931952]
15. Chen CH, Huang KY, Wang JY, Huang HB, Chou P, Lee CC. Combined effect of individual and neighbourhood socioeconomic status on mortality of rheumatoid arthritis patients under universal health care coverage system. *Family practice*. 2015;32(1):41–8. [PubMed: 25304308]
16. Constantinescu F, Goucher S, Weinstein A, Fraenkel L. Racial disparities in treatment preferences for rheumatoid arthritis. *Medical care*. 2009;47(3):350–5. [PubMed: 19165120]
17. Constantinescu F, Goucher S, Weinstein A, Smith W, Fraenkel L. Understanding why rheumatoid arthritis patient treatment preferences differ by race. *Arthritis and rheumatism*. 2009;61(4):413–8. [PubMed: 19333986]
18. Greenberg JD, Spruill TM, Shan Y, Reed G, Kremer JM, Potter J, et al. Racial and ethnic disparities in disease activity in patients with rheumatoid arthritis. *The American journal of medicine*. 2013;126(12):1089–98. [PubMed: 24262723]
19. Hughes LB, Beasley TM, Patel H, Tiwari HK, Morgan SL, Baggott JE, et al. Racial or ethnic differences in allele frequencies of single-nucleotide polymorphisms in the methylenetetrahydrofolate reductase gene and their influence on response to methotrexate in rheumatoid arthritis. *Annals of the rheumatic diseases*. 2006;65(9):1213–8. [PubMed: 16439441]

20. Jacobi CE, Mol GD, Boshuizen HC, Rupp I, Dinant HJ, Van Den Bos GA. Impact of socioeconomic status on the course of rheumatoid arthritis and on related use of health care services. *Arthritis and rheumatism*. 2003;49(4):567–73. [PubMed: 12910565]
21. Johnson BK, Bahce-Altuntas A. Too Little Too Late: Effect of Poor Access to Biologics for Patients with Rheumatoid Arthritis. *The Journal of rheumatology*. 2017;44(12):1765–6. [PubMed: 29196544]
22. Katz PP, Barton J, Trupin L, Schmajuk G, Yazdany J, Ruiz PJ, et al. Poverty, Depression, or Lost in Translation? Ethnic and Language Variation in Patient-Reported Outcomes in Rheumatoid Arthritis. *Arthritis care & research*. 2016;68(5):621–8. [PubMed: 26414775]
23. Kerr GS, Swearingen C, Mikuls TR, Yazici Y. Use of Biologic Therapy in Racial Minorities With Rheumatoid Arthritis From 2 US Health Care Systems. *Journal of clinical rheumatology : practical reports on rheumatic & musculoskeletal diseases*. 2017;23(1):12–8. [PubMed: 28002151]
24. Lapcevic M, Vukovic M, Gvozdenovic BS, Mijoljevic V, Marjanovic S. Socioeconomic and therapy factor influence on self-reported fatigue, anxiety and depression in rheumatoid arthritis patients. *Rev Bras Reumatol Engl Ed*. 2017;57(6):545–56. [PubMed: 29173692]
25. Lee SJ, Kremer J, Kavanaugh A. Treatment disparity related to race/ethnicity and education in rheumatoid arthritis patients: comment on the article by Constantinescu et al. *Arthritis and rheumatism*. 2009;61(8):1141–2. [PubMed: 19644891]
26. Li X, Sundquist J, Sundquist K. Socioeconomic and occupational risk factors for rheumatoid arthritis: a nationwide study based on hospitalizations in Sweden. *The Journal of rheumatology*. 2008;35(6):986–91. [PubMed: 18464310]
27. Margaretten M, Barton J, Julian L, Katz P, Trupin L, Tonner C, et al. Socioeconomic determinants of disability and depression in patients with rheumatoid arthritis. *Arthritis care & research*. 2011;63(2):240–6. [PubMed: 20824800]
28. McBurney CA, Vina ER. Racial and ethnic disparities in rheumatoid arthritis. *Current rheumatology reports*. 2012;14(5):463–71. [PubMed: 22773376]
29. Mikuls TR, Kazi S, Cipher D, Hooker R, Kerr GS, Richards JS, et al. The association of race and ethnicity with disease expression in male US veterans with rheumatoid arthritis. *The Journal of rheumatology*. 2007;34(7):1480–4. [PubMed: 17552044]
30. Molina E, Del Rincon I, Restrepo JF, Battafarano DF, Escalante A. Association of socioeconomic status with treatment delays, disease activity, joint damage, and disability in rheumatoid arthritis. *Arthritis care & research*. 2015;67(7):940–6. [PubMed: 25581770]
31. Parks CG, D'Aloisio AA, DeRoo LA, Huiber K, Rider LG, Miller FW, et al. Childhood socioeconomic factors and perinatal characteristics influence development of rheumatoid arthritis in adulthood. *Annals of the rheumatic diseases*. 2013;72(3):350–6. [PubMed: 22586176]
32. Solomon A, Christian BF, Dessein PH. Potential determinants of poor disease outcome in socioeconomically disadvantaged patients with rheumatoid arthritis. *The Journal of rheumatology*. 2008;35(9):1895–6; author reply 6. [PubMed: 18785299]
33. Solomon DH, Ayanian JZ, Yelin E, Shaykevich T, Brookhart MA, Katz JN. Use of disease-modifying medications for rheumatoid arthritis by race and ethnicity in the National Ambulatory Medical Care Survey. *Arthritis care & research*. 2012;64(2):184–9. [PubMed: 22012868]
34. Yang G, Bykerk VP, Boire G, Hitchon CA, Thorne JC, Tin D, et al. Does socioeconomic status affect outcomes in early inflammatory arthritis? Data from a canadian multisite suspected rheumatoid arthritis inception cohort. *The Journal of rheumatology*. 2015;42(1):46–54. [PubMed: 25399388]
35. Yazici Y, Kautiainen H, Sokka T. Differences in clinical status measures in different ethnic/racial groups with early rheumatoid arthritis: implications for interpretation of clinical trial data. *The Journal of rheumatology*. 2007;34(2):311–5. [PubMed: 17304656]
36. Oude Voshaar MA, ten Klooster PM, Taal E, van de Laar MA. Measurement properties of physical function scales validated for use in patients with rheumatoid arthritis: a systematic review of the literature. *Health and quality of life outcomes*. 2011;9:99. [PubMed: 22059801]
37. Arat S, Lenaerts JL, De Langhe E, Verschueren P, Moons P, Vandenberghe J, et al. Illness representations of systemic lupus erythematosus and systemic sclerosis: a comparison of patients,

- their rheumatologists and their general practitioners. *Lupus science & medicine*. 2017;4(1):e000232. [PubMed: 29177061]
38. Ward MM, Guthrie LC, Alba MI. Clinically important changes in individual and composite measures of rheumatoid arthritis activity: thresholds applicable in clinical trials. *Annals of the rheumatic diseases*. 2015;74(9):1691–6. [PubMed: 24794149]
  39. Vittinghoff E. *Regression methods in biostatistics: linear, logistic, survival, and repeated measures models*. 2nd ed New York: Springer; 2012:82.
  40. Guyatt GH, Deyo RA, Charlson M, Levine MN, Mitchell A. Responsiveness and validity in health status measurement: a clarification. *Journal of clinical epidemiology*. 1989;42(5):403–8. [PubMed: 2659745]
  41. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of chronic diseases*. 1987;40(5):373–83. [PubMed: 3558716]
  42. Katz P, Pedro S, Michaud K. Performance of the Patient-Reported Outcomes Measurement Information System 29-Item Profile in Rheumatoid Arthritis, Osteoarthritis, Fibromyalgia, and Systemic Lupus Erythematosus. *Arthritis care & research*. 2017;69(9):1312–21. [PubMed: 28029753]
  43. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of clinical epidemiology*. 2007;60(1):34–42. [PubMed: 17161752]
  44. Karpouzas GA, Dolatabadi S, Moran R, Li N, Nicassio PM, Weisman MH. Correlates and predictors of disability in vulnerable US Hispanics with rheumatoid arthritis. *Arthritis care & research*. 2012;64(9):1274–81. [PubMed: 22489073]
  45. Lynch AD, Dodds NE, Yu L, Pilkonis PA, Irrgang JJ. Individuals with knee impairments identify items in need of clarification in the Patient Reported Outcomes Measurement Information System (PROMIS(R)) pain interference and physical function item banks - a qualitative study. *Health and quality of life outcomes*. 2016;14:77. [PubMed: 27169858]
  46. Orbai AM, Smith KC, Bartlett SJ, De Leon E, Bingham CO 3rd “Stiffness has different meanings, I think, to everyone”: examining stiffness from the perspective of people living with rheumatoid arthritis. *Arthritis care & research*. 2014;66(11):1662–72. [PubMed: 24891304]
  47. Koster N, Knol DL, Uitdehaag BM, Scheltens P, Sikkes SA. The sensitivity to change over time of the Amsterdam IADL Questionnaire((c)). *Alzheimer’s & dementia : the journal of the Alzheimer’s Association*. 2015;11(10):1231–40.
  48. Barton JL, Trupin L, Schillinger D, Gansky SA, Tonner C, Margaretten M, et al. Racial and ethnic disparities in disease activity and function among persons with rheumatoid arthritis from university-affiliated clinics. *Arthritis care & research*. 2011;63(9):1238–46. [PubMed: 21671414]
  49. Smolen JS, Landewe R, Bijlsma J, Burmester G, Chatzidionysiou K, Dougados M, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2016 update. *Annals of the rheumatic diseases*. 2017;76(6):960–77. [PubMed: 28264816]

### Significance and Innovations

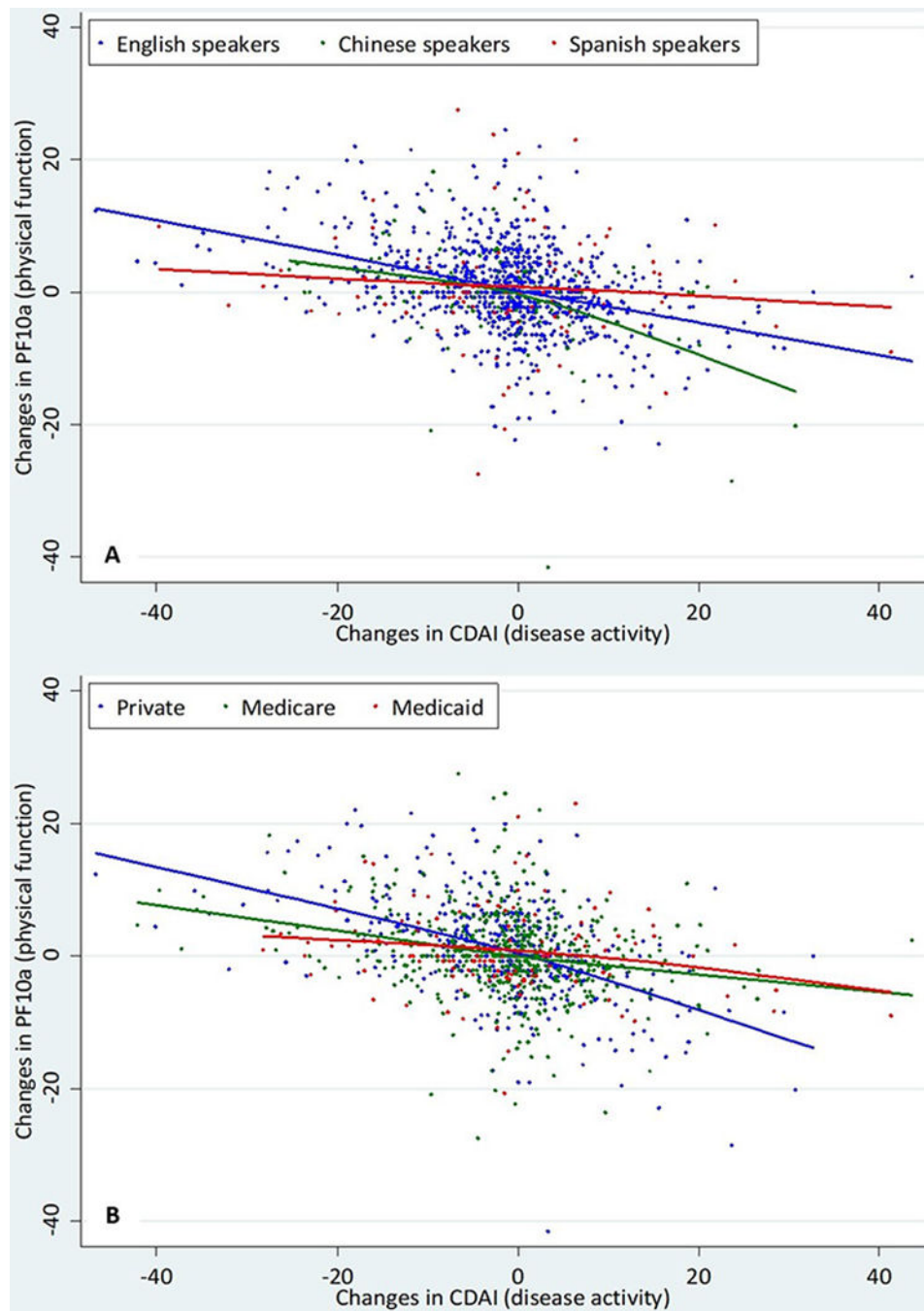
- There is growing interest nationally in using patient-reported outcomes (PROs) in routine clinical care to engage patients, monitor outcomes and inform treatment decisions. However, most studies evaluating PROs have been performed in white, English-speaking populations.
- We studied the validity and responsiveness of a PROMIS measure in RA across different languages, races/ethnicities and insurance groups in a real-world clinic population
- We found that constructs evaluated by PROMIS-PF10a are less correlated with clinical outcomes among Chinese-speakers and Hispanics and that PF10a has less sensitivity to clinical improvements among Medicaid patients and Spanish-speakers.



**Figure 1.**

Data set for cross-sectional analysis of validity and longitudinal analysis of responsiveness. ICD -9 = International Classification of Diseases, Ninth Revision; RA = rheumatoid arthritis.





**Figure 2.** Linear mixed-effects regression models showing the effect of language and insurance status on the Physical Function Short Form 10a (PF 10a) responsiveness to clinical improvements and deteriorations over time. CDAI = Clinical Disease Activity Index.

**Table 1:**

Characteristics of the RA clinic population at baseline

	Cross-sectional sample N = 595	Longitudinal cohort N = 341
Age in years, Mean $\pm$ SD	56.5 (15.3)	55.8 (15.4)
Female, n (%)	493 (83)	282 (83)
BMI, mean $\pm$ SD	26.7 (6.4)	26.2 (5.9)
Race/Ethnicity, n (%)		
White	297 (50)	163 (48)
African American	38 (6)	19 (6)
Hispanic	27 (5)	14 (4)
Asian	115 (19)	73 (21)
Other	118 (20)	72 (21)
Preferred language, n (%)		
English	512 (86)	288 (85)
Spanish	43 (8)	28 (8)
Chinese	40 (6)	25 (7)
Insurance type, n (%)		
Private	236 (40)	142 (42)
Medicare	281 (47)	155 (45)
Medicaid	78 (13)	44 (13)
Smoking, n (%)	137 (23)	79 (23)
Total Charlson score, median (IQR)	1 (1-2)	1 (1-2)
Medication, n (%)		
DMARD & Biologic Naïve	26 (5)	11 (3)
DMARD only	311 (52)	181 (53)
Biologic with or without DMARD	258 (43)	149 (44)
<b>Clinical Parameters</b>		
RA disease activity, n (%)		
CDAI remission	85 (14)	49 (14)
CDAI low	207 (35)	111 (33)
CDAI moderate	168 (28)	104 (30)
CDAI high	135 (23)	77 (23)
Physician GA VAS, median (IQR)	23 (10-44)	24 (10-44)
Patient GA VAS, median (IQR)	40 (15-64)	35 (15-62)
PF10a, mean $\pm$ SD	40.1 (10.7)	40.8 (10.4)
Pain VAS, median (IQR)	40 (15-68)	33 (12-65)
Tender 28-Joint Count, median (IQR)	2 (0-6)	2 (0-6)
Swollen 28-Joint Count, median (IQR)	2 (0-5)	2 (0-6)
CRP mg/dl, median (IQR)	4 (2-9.2)	4 (2-8.6)
ESR mm/hr, median (IQR)	19 (9-36)	20 (10-35)

N: Number; SD: Standard deviation; BMI: Body mass index; IQR: interquartile range; DAMRD: disease modifying antirheumatic drug; CDAI: clinical disease activity index; GA: Global assessment; VAS: Visual analogue score; PF10a: PROMIS Physical Function Short Form 10a; CRP: C-reactive protein; ESR: Erythrocyte sedimentation rate.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2:**

Construct validity analysis, showing Spearman's correlation coefficients between PROMIS PF10a scores and patient reported outcomes, physician assessed outcomes, and inflammatory markers among different subgroups

	Spearman's correlation coefficients with PF10a ( <i>r</i> )					
	Patient reported outcomes		Physician assessed outcomes		Inflammatory markers	
	PGA (VAS)	Pain (VAS)	TJC	SJC	CRP	ESR
Language						
English, N=512	-0.71 *	-0.68 *	-0.45 *	-0.34 *	-0.31 *	-0.29 *
Spanish, N=43	-0.60 *	-0.61 *	-0.35 *	-0.24	-0.02	-0.12
Chinese, N=40	-0.52 *	-0.52 *	-0.28	-0.19	-0.13	-0.08
Insurance						
Private, N=236	-0.72 *	-0.66 *	-0.61 *	-0.49 *	-0.33 *	-0.24 *
Medicare, 281	-0.62 *	-0.60 *	-0.29 *	-0.17 *	-0.23 *	-0.24 *
Medicaid, 78	-0.53 *	-0.61 *	-0.42 *	-0.31 *	-0.14	-0.14
Race/Ethnicity						
White, N=297	-0.69 *	-0.67 *	-0.44 *	-0.33 *	-0.34 *	-0.28 *
African American, N=38	-0.62 *	-0.53 *	-0.45 *	-0.12	-0.04	-0.32
Hispanic, N=38	-0.62 *	-0.64 *	-0.25	-0.05	0.05	0.08
Asian, N=115	-0.64 *	-0.65 *	-0.40 *	-0.33 *	-0.26 *	-0.13
Other, N=118	-0.69 *	-0.65 *	-0.48 *	-0.41 *	-0.14	-0.31 *

Data source: Cross-sectional sample (N=595)

PGA: Patient global assessment; VAS: Visual analog scale; TJC: Tender joint count; SJC: Swollen joint count; CRP: C-reactive protein; ESR: Erythrocyte sedimentation rate; PF10a: Physical function short-form 10a

\* P<0.05

r<-0.6: Strong correlation; -0.3> r >-0.6: Moderate correlation; r>-0.3: Weak correlation

**Table 3:**

Known-group validity for PF10a using Cohen's d effect size

	Age				Clinical disease activity			
	<50 years		65 years		CDAI $\leq 10$		CDAI >22	
	N	N	Difference in mean PF10a	Cohen's d	N	N	Difference in mean PF10a	Cohen's d
Language								
English	176	155	2.5*	0.24*	263	113	13.3*	1.44*
Spanish	9	18	6.6	0.69	18	13	11.6*	1.21*
Chinese	3	26	4.2	0.44	11	9	10.9*	0.95*
Insurance								
Private	131	12	6.1*	0.64*	140	50	16.5*	2.15*
Medicare	22	184	-6.0*	-0.59*	131	55	9.9*	1.04*
Medicaid	35	3	-8.6	-1.00	21	30	9.6*	1.06*
Race/Ethnicity								
White	94	94	2.3	0.23	161	54	13.3*	1.49*
African American	6	16	0.8	0.07	15	10	9.9*	0.96*
Hispanic	12	3	0.9	0.09	13	6	5.3	0.48
Asian	28	50	6.6*	0.64*	47	28	13.6*	1.37*
Other	48	36	3.8	0.36	56	37	13.7*	1.50*

Data source: Cross-sectional sample (N=595)

CDAI: Clinical disease activity index; PF10a: Physical function short-form 10a

Difference in mean PF10a score is the difference between mean PF10a among those <50 years compared to those 65 years or among those with CDAI  $\leq 10$  compared to those with CDAI >22; Cohen's d is the difference in mean scores divided by the pooled standard deviation; Effect size (Cohen's d) values are categorized as small (<0.5), medium (0.5-0.8), or large (>0.8).

\* P<0.05 using Student's t-test; for Cohen's d, 95% confidence intervals do not cross 0

**Table 4:**

Effect of language preference, insurance status, and race/ethnicity on responsiveness of PF10a to changes in clinical disease activity

		Clinical Improvement (A 12-point decrease in CDAI)			Clinical Deterioration (A 12-point increase in CDAI)		
		Change in PF10a score ( $\beta$ )	95% CI	P*	Change in PF10a score ( $\beta$ )	95% CI	P*
Language							
<b>Model 2</b>	English (Ref)	3.08	2.31, 3.86	NA	-2.92	-3.93, -1.92	NA
	Spanish	0.67	-1.37, 2.70	0.029	-0.83	-3.10, 1.44	0.098
	Chinese	1.84	-0.73, 4.42	0.363	-5.96	-8.61, -3.31	0.036
Insurance							
<b>Model 3</b>	Private (Ref)	3.63	2.59, 4.66	NA	-5.33	-6.74, -3.92	NA
	Medicare	2.21	1.17, 3.25	0.055	-1.62	-2.91, -0.34	<0.001
	Medicaid	0.70	-1.27, 2.67	0.005	-1.79	-3.80, 0.22	0.005
Race/Ethnicity							
<b>Model 4</b>	White (Ref)	3.41	2.37, 4.44	NA	-2.96	-4.41, -1.51	NA
	African American	2.47	-0.17, 5.12	0.520	-0.77	-4.36, 2.82	0.268
	Hispanic/Latino	2.79	-0.24, 5.82	0.708	-2.81	-5.99, 0.37	0.934
	Asian	1.50	-0.12, 3.14	0.052	-3.49	-5.15, -1.82	0.636
	Other	2.36	0.95, 3.78	0.241	-2.92	-4.74, -1.09	0.974

Data source: Longitudinal cohort (N=341, Encounters=1546)

CDAI: Clinical disease activity index; CI: Confidence interval; Ref: Reference category

Results are from the linear mixed effects regression and adjusted for baseline PF10a, and time; Model 2 incorporates an interaction term between changes in CDAI score (from previous visit) and preferred language; Model 3 incorporates an interaction term between changes in CDAI score and insurance status; Model 4 incorporates an interaction term between changes in CDAI score and race/ethnicity;  $\beta$  represents magnitude of change in PF10a score; P\* is the p-value for statistical significance of effect modification by non-English language, non-Private insurance or non-white race.