Published in final edited form as:

Qual Life Res. 2018 November ; 27(11): 2935–2944. doi:10.1007/s11136-018-1958-5.

Evaluating the PROMIS-29 v2.0 for Use Among Older Adults with Multiple Chronic Conditions

Adam J. Rose, MD MSc^{1,2}, Elizabeth Bayliss, MD^{3,4}, Wenjing Huang, PhD⁵, Lesley Baseman⁶, Emily Butcher¹, Rosa-Elena García, MPH⁵, and Maria Orlando Edelen, PhD¹

¹RAND Corporation, Boston, MA

²Section of General Internal Medicine, Boston University School of Medicine, Boston, MA

³Institute for Health Research, Kaiser Permanente Colorado, Aurora, CO

⁴Department of Family Medicine, University of Colorado School of Medicine, Aurora, CO

⁵RAND Corporation, Santa Monica, CA

⁶RAND Corporation, Arlington, VA

Abstract

Purpose: The Patient-Reported Outcomes Measurement Information System 29-item profile (PROMIS-29 v2.0), which measures health-related quality of life (HRQoL), has had limited evaluation among older adults (age 65+) with multiple chronic conditions. Our purpose was to establish convergent validity for PROMIS-29 in this population.

Methods: We collected the PROMIS-29 v2.0 and the Veterans RAND 36 (VR-36) for 1,359 primary care patients age 65+ with at least 2 of 13 chronic conditions, oversampling those age 80+. We conducted multiple analyses to examine score differences across subgroups, differential item functioning (DIF), and comparisons of PROMIS-29 v2.0 and VR-36 scores.

Results: The mean age was 80.7, and all patients had at least 2 of 13 chronic conditions. Older age, female sex, Hispanic ethnicity, and more chronic conditions were associated with worse physical health scores (PHS) and mental health scores (MHS) on the PROMIS-29 v2.0 – findings which are in the expected direction. None of the 700 pairs of items met criteria for DIF. PHS and MHS were highly intercorrelated ($r = 0.74$, $p < 0.001$ for this and all other findings). PHS was more highly correlated with the VR-36 Physical Component Score (PCS) than the Mental Component Score (MCS) ($r = 0.85$ and 0.32 , respectively), while MHS was highly correlated with both ($r = 0.70$ and 0.64 , respectively).

Address for Correspondence: Adam J. Rose, MD MSc, RAND Corporation, 20 Park Plaza, Suite 920, Boston, MA 02116, Phone: 617-338-2059 X 8672, Fax: 617-357-7470, arose@rand.org.

Disclosures: The authors have no conflicts of interest to report.

Conflict of Interest: The authors declare that they have no relevant conflicts of interest.

Compliance with ethical standards: The authors declare that this study was conducted in accordance with appropriate ethical standards for research, including the Declaration of Helsinki. The study was approved by the RAND Human Subjects Research Protection Committee (Study #2015-0956-AM05), and by the Kaiser Permanente Colorado IRB (IRB Number CO-15-2199).

Informed Consent: Participants provided informed consent, with a waiver of documentation of informed consent.

Conclusions: PROMIS-29 v2.0 demonstrates expected bivariate relationships with key person-level characteristics and does not show DIF. PROMIS-29 v2.0 scores are highly correlated with VR-36 scores. These results provide support for the validity of PROMIS-29 v2.0 as a measure of HRQoL among older adults with multiple chronic conditions.

Keywords

quality of life; PROMIS; geriatrics; comorbidity; chronic disease; elderly

INTRODUCTION

Accurately measuring health-related quality of life (HRQoL) is important for many applications, including research, quality improvement, performance measurement, and informing clinical care. The Patient-Reported Outcomes Measurement Information System 29-Item Profile Measure (PROMIS-29 v2.0) is a relatively new HRQoL instrument, which was developed using modern measurement theory and calibrated and scored based on contemporary samples [1, 2]. By design, PROMIS-29 v2.0 contains some content, such as a scale about sleep disturbance, which is not directly measured in many older HRQoL instruments. The PROMIS-29 v2.0 is meant to be an efficient means of assessing a broad range of HRQoL domains, providing a comprehensive assessment of a patient's HRQoL. The instrument includes four items from each of the seven PROMIS domains (anxiety, depression, fatigue, pain interference, physical functioning, sleep disturbance, and ability to participate in social roles), as well as an additional pain intensity item [3].

The PROMIS-29 v2.0 has been tested in a variety of patient populations. The instrument has been used to assess HRQoL in chiropractic patients [4] and in people with neuroendocrine tumors [5, 6], systemic sclerosis [3], irritable bowel syndrome [7], rheumatoid arthritis and osteoarthritis [8], and systemic lupus erythematosus [9], and HIV [10]. While the PROMIS-29 v2.0 has been used and evaluated in these varied populations, one particularly important group for whom it has not been explicitly evaluated is older patients with multiple chronic conditions (MCCs), a population that often has intensive health care needs requiring complex care coordination. The measurement of HRQoL may be particularly relevant for this population, since other ways of measuring quality of care or patient outcomes (including certain processes of care or long-term survival) may not be as relevant to patients who are older and who have an increasing burden of chronic illness. The importance of HRQoL does not wane with age, with limited life expectancy, or in the presence of MCCs.

We therefore measured the PROMIS-29 v2.0, together with another widely used HRQoL instrument (the Veterans RAND 36-Item Survey Instrument, or VR-36 [11]) in a population of adults age 65 and older who had at least 2 of 13 pre-specified chronic health conditions. Our objective was to establish construct validity for the PROMIS-29 v2.0 in this population of older adults with MCCs.

METHODS

Setting

Participants were recruited from Kaiser Permanente Colorado (KPCO), a not-for-profit integrated delivery system that directly provides primary and specialty care, including both ambulatory and hospital-based care. The Institute for Health Research at KPCO maintains a virtual data warehouse (VDW), which includes data from the electronic health records, pharmacy fills, claims, demographics, and the membership and administrative systems. We obtained approvals from the institutional review boards of both KPCO (IRB Number CO-15–2199) and the RAND Corporation (Protocol #2015–0956-AM05). The work described in this manuscript was part of a broader program of work, which has been described in full elsewhere [12].

Eligibility and Identification of Participants

KPCO members were eligible to participate if they were age 65 or older, were assigned to a primary care provider at a KPCO ambulatory clinic, had been seen for clinical care at least once in the past 12 months, had a valid email address, and had at least 2 of 13 specific chronic conditions (the conditions are found in Table 1). Further details regarding eligibility can be found in Huang et al. (under review).

Survey Design/Contents

The survey consisted of the PROMIS-29 v2.0 and the VR-36. In addition to the eight scores for the PROMIS-29 v2.0 (anxiety, depression, fatigue, pain intensity, pain interference, physical function, sleep disturbance, social roles), we also used two summary scores for the PROMIS-29 v2.0, the PROMIS-29 v2.0 Physical Health Summary Score (PHS) and the PROMIS-29 v2.0 Mental Health Summary Score (MHS), which were intended to measure the two major domains of HRQoL. The process for developing the PHS and MHS has been described by Hays, et al. in their recent manuscript [12]. The process for adapting them for use in this sample is described Huang et al.

The VR-36 assesses quality of life across eight domains [13]. It also has two summary scores, both of which draw on all eight domains to some extent: the Physical Component Score, or PCS, and the Mental Component Score, or MCS [11, 14]. The VR-36 is a modification of the 36-item short-form (SF-36), which was developed in response to findings from the Medical Outcomes Study [11, 14]. The response scale for the two role functioning items (Role Physical and Role Emotional) was changed to a five-point scale instead of a dichotomized scale, in order to reduce ceiling and floor effects and to increase the explanatory power of the items [13]; a similar change was also made in a later version of the SF-36. The VR-36 has been widely evaluated in populations with diverse conditions, including posttraumatic stress disorder [15], coronary disease [16], and multiple sclerosis [17]. It has also been included as part of the Healthcare Effectiveness Data and Information Set (HEDIS) performance measures since 2006 [18].

Data Collection

Survey data collection procedures are described in detail in Huang et al. (under review). Briefly, a total of 4,991 patients were deemed to be eligible based on the criteria described above. Of these, 283 patients opted out, 677 had an invalid mail or email address, and 282 had died, leaving a sample of 3,749 patients. Because one purpose of our study was to investigate survey mode effects in older patients, we randomly assigned 2,764 of these to a web survey, 376 to a mail survey, and 372 to a phone survey. All survey activities were conducted by the RAND Survey Research Group. Non-responders to the web survey were sent weekly email reminders for up to 3 weeks. Non-responders to the mail survey were sent additional packets at four and six weeks after the initial mailing. Those assigned to the telephone survey were contacted an average of five times. While the results are not shown here, it is noteworthy that we did not find mode effects between the web, mail, and telephone surveys, nor did we find important differences between responders and non-responders; details can also be found online at https://www.rand.org/pubs/research_reports/RR2176.html.

In addition to collecting survey data, we queried the KPCO VDW. This enabled us to supplement self-report information from the survey with information that was both a) more detailed than would have been possible for many respondents to report and b) collected without adding to respondent burden. We collected data on age at baseline, sex, race/ethnicity, area socioeconomic status (SES, measured as described below), chronic health conditions, number of home health visits, primary and specialty care visits, and hospitalizations. Decisions regarding how to divide variables into categories were based on their univariate distribution, with a goal of creating roughly equal sized groups for analysis whenever possible.

Because a majority of our sample was White and non-Hispanic, and several other categories were relatively uncommon, we created four merged categories that would each contain sufficient numbers to support analyses: White non-Hispanic, non-White and non-Hispanic, Hispanic (any race), and missing race non-Hispanic. We assessed area SES using the Census tract of residence, linked with the percentage of households below the federal poverty level (FPL) from the American Community Survey 2010–2015 estimates [19].

Statistical Analyses

The process of establishing construct validity for the PROMIS-29 v2.0 in this sample of older adults with MCC had three main parts. First, we used bivariate analyses to examine the relationship between PROMIS-29 v2.0 scores and important stratifying patient-level variables. Second, we looked for evidence of differential item functioning (DIF) for the items of the PROMIS-29 v2.0, based on key stratifying variables. Third, we examined correlations between PROMIS-29 v2.0 scores and scores of another HRQoL instrument, the VR-36. Details regarding these three analyses follow.

Bivariate Analyses—We began by examining the relationship between PROMIS-29 v2.0 scores and important stratifying patient-level variables. We performed these analyses for ten PROMIS-29 v2.0 scores, including the eight usual PROMIS-29 v2.0 scores (anxiety,

depression, fatigue, pain intensity, pain interference, physical function, sleep disturbance, social roles) as well as the two summary scores (PHS and MHS). Stratifying patient variables included sociodemographic and clinical variables. We performed bivariate linear regressions to examine the association between each stratifying (independent) variable and the ten PROMIS-29 v2.0 scores (continuous dependent variables). In these regressions, independent variables were modeled as class variables. The goal was to ensure that PROMIS-29 v2.0 scores behave as expected across these categories, based on what is known from previous research.

Testing for Differential Item Function—The second way we assessed the construct validity of the PROMIS-29 v2.0 in this sample was to look for DIF. DIF refers to a systematic tendency for respondents of a certain type to score higher or lower on a particular survey item than expected based on their overall scale score. For example, women might systematically score higher than men on an item related to wrist pain, beyond what would be predicted from their levels of the symptom as estimated by the overall pain score of which the wrist pain item was part. The presence of DIF would imply a threat to the construct validity of PROMIS-29 v2.0 in this sample, whereas its absence would help build a case for such validity.

We tested for DIF using the widely accepted method described by Gelin and Zumbo, which uses ordinal logistic regression, an approach that can accommodate items with non-binary response options [[20]. We corrected for multiple comparisons using a Benjamini-Hochberg correction [21]. We tested items from the PROMIS-29 v2.0 for DIF according to every respondent-level characteristic including survey mode by estimating a series of nested models. These models predicted each item response as the outcome. The first model specified a main effect for the case mix adjusted score on the PROMIS-29 v2.0 scale score that the item is from (Model 1; item = score). Case mix adjustment was performed using a model containing all of the variables in Table 2. The next model added a main effect for the grouping variable indicator (Model 2; item = score group). The last model added an interaction effect (score*group; Model 3; item= score group score*group).

For each set of analyses, the stratifying variable was omitted from the case mix model used to generate the case mix adjusted score (e.g., we did not adjust for sex in the case mix model when testing for the presence of DIF by sex). When a respondent-level characteristic had more than two levels, we created a binary version of the variable based on the univariate distribution of the data, with the exception of survey mode in which we tested all levels of the variable against one another. Binary variables were constructed with a goal of roughly equal size groups, whenever possible.

We tested 28 of the 29 items in the PROMIS-29 v2.0. Because the pain intensity item stands alone and is not part of a scale, we did not test it for DIF, as DIF applies only to items that are included in multi-item scales. We therefore tested 28 items from the PROMIS-29 v2.0 according to 25 binary variables, for a total of 700 analyses to evaluate DIF. The Benjamini-Hochberg correction for multiple testing, using 0.1 as the base significance level, was applied once for each of the 25 binary variables we tested; thus it was applied for groups of 28 analyses. Statistical significance was determined based on a comparison of three nested

models, namely Model 1 (case mix adjusted scale score only), Model 2 (scale score and group indicator), and Model 3 (scale score, group indicator, and the interaction between these two predictors). In addition to statistical significance, we considered the magnitude of the effect size for DIF, in part because the ability to demonstrate a statistically significant effect is highly related to sample size and not necessarily to the importance of a given effect size to the patient. Per the recommendations of Gelin and Zumbo, an effect size can be considered negligible in magnitude if the increase in R^2 between two nested models is less than 0.035 [20].

Comparisons with VR-36 Scores—The third way we assessed the construct validity of the PROMIS-29 v2.0 as an HRQoL instrument in this population was to examine Pearson correlations of the ten PROMIS-29 v2.0 scores with the two summary scores of the VR-36, the PCS and the MCS. Finding a high correlation between summary scores on the two HRQoL instruments (particularly a correlation near or above 0.6) would be generally supportive of the idea that they are measuring the same underlying construct, namely HRQoL. All statistical analyses were conducted using SAS, version 9.4 [22].

RESULTS

Bivariate Analyses: Impact of Patient Characteristics on PROMIS-29 Scores

Out of a sample of 3,749 people who received a survey, a total of 1,359 participants responded, for a total response rate of 36%. Further details regarding response rate can be found from Huang et al. paper. Characteristics of responders are shown in Table 1. The mean age was 80.7 years, and 56% of participants were age 80 or older. A majority of participants (89%) were White non-Hispanic. Relatively few (9%) lived in high-poverty areas (defined as 20% or more of households in the Census tract below the federal poverty level). Per the study inclusion criteria, all had at least 2 of 13 chronic conditions; 35% had exactly 2, 31% had 3, 18% had 4, and 16% had 5 or more. Prevalence of many chronic conditions was relatively high compared to the U.S. population; for example, 38% had chronic lung disease, and 31% had diabetes. More than half had at least one home health encounter in the past 12 months. Because eligibility for home health services in the United States usually requires a recent hospitalization or inability to self-care, this is a marker for a relatively severe decrement in function, at least temporarily.

The mean PROMIS-29 v2.0 Mental Health Score (MHS) was 50.1, similar to the U.S. population, while the mean PROMIS-29 v2.0 Physical Health Score (PHS) was 42.2, almost one standard deviation below the U.S. population.

Table 2 shows bivariate analyses regarding how PROMIS-29 v2.0 scores vary according to patient-level characteristics. Older age was associated with poorer HRQoL; respondents of age 85+ had PHS scores that were 6.60 points lower than those age 65–69 ($p < 0.001$), and MHS scores that were 1.82 points lower ($p < 0.05$). The effect of age on physical HRQoL was much more pronounced than on mental HRQoL, as has been noted in previous studies [23]. Hispanics and females had consistently poorer HRQoL scores in this sample, compared to other groups. For example, females scored 3.51 and 2.50 points lower on PHS and MHS, respectively, than males ($p < 0.001$ for both), whereas Hispanics scored 2.93 and

3.94 points lower, respectively, than White non-Hispanics ($p < 0.001$ for both). As might be expected, HRQoL was lower with an increasing number of comorbid conditions, with a more pronounced effect on physical than mental HRQoL.

The HRQoL effects of specific chronic conditions generally concurred with what would be expected. Congestive heart failure was associated with considerably poorer physical HRQoL (PHS was 4.66 lower for patients with congestive heart failure than for those without, the largest difference of any single condition). However, congestive heart failure was not associated with more depression or more fatigue, was not associated with more pain, and the overall difference in MHS was relatively modest (2.94 points lower). In contrast, patients with major depression had the largest difference on MHS for any single condition (4.96 points lower), but also had a meaningfully lower PHS than patients without depression (3.37 points lower). Patients with depression also had the second-largest difference in pain intensity for any single condition (1.01 points higher), second only to sciatica (1.37 points higher).

Testing for DIF

Complete results for all 700 DIF analyses are shown in Appendix A. Of the 700 pairs of items that we tested for DIF, 37 demonstrated statistical significance at the Benjamini-Hochberg-corrected 0.01 level. However, they all had negligible effect sizes, ranging from a change in the R^2 of 0.007 (the smallest) to 0.033 (the largest, but still within the “negligible” range). Thus, after an extensive search for DIF, we found that only a small number of pairs of items demonstrated statistically significant DIF, and all had negligible effect sizes.

Relationship of PROMIS-29 and VR-36 Scores

We assessed the Pearson correlation between eight PROMIS-29 v2.0 scores and summary scores from the VR-36 (PCS and MCS), which were measured concurrently (Table 3). The direction of the associations shown in this table are not informative, as the polarity of many PROMIS-29 v2.0 scores is negative. Without exception, all correlations were positive in the sense that better HRQoL on one score correlated with better HRQoL on the other score. Four PROMIS-29 v2.0 scores showed large associations with PCS, as measured by a correlation of 0.5 or higher [24]: physical function, social roles, pain interference, and pain intensity. Scores on one PROMIS-29 v2.0 domain, fatigue, were showed large associations with both PCS and MCS. Two other scores, anxiety and depression, showed large associations with MCS scores. Finally, scores on the PROMIS-29 v2.0 sleep disturbance score were only moderately correlated with the score of both the PCS and the MCS.

We also compared the two summary scores of the PROMIS-29 v2.0 (PHS and MHS) with the VR-36 PCS and MCS, using Pearson correlation (Table 3). Because they are constructed to be orthogonal to each other, the correlation between PCS and MCS scores is low [25]; here, it was 0.13 ($p < 0.001$ for all correlations discussed). The PHS and MHS were much more highly correlated with each other ($r = 0.74$), and both were highly correlated with the PCS (0.85 for PHS and 0.70 for MHS). MHS was strongly correlated with the VR-36 MCS (0.64), whereas the correlation between the PHS and the MCS was considerably weaker

(0.32). Interestingly, the MHS was more strongly correlated with PCS (0.70) than MCS (0.64).

DISCUSSION

We collected primary data in a population of older adults with MCCs, to establish the construct validity of the PROMIS-29 v2.0 when used in this population. Construct validity was established in three ways: 1) PROMIS-29 v2.0 scores varied with respondent characteristics in ways that were both expected and concordant with the previous literature for other HRQoL instruments; 2) PROMIS-29 v2.0 items did not exhibit significant DIF in this population; and finally, 3) PROMIS-29 v2.0 scores exhibited acceptable correlations with scores from another well-used HRQoL instrument, the VR-36.

The findings of our bivariate analyses, such as showing that increasing age was more strongly associated with poorer physical than mental HRQoL, support the validity of PROMIS-29 v2.0 for use in this population, but in many cases are not novel findings. The lack of DIF for the PROMIS-29 v2.0 in this population, while also reassuring regarding validity, does not provide novel insights in and of itself. However, it is important to note that we did not observe any DIF according to survey mode, thus providing a basis for survey administration of the PROMIS-29 v2.0 via phone, mail, or web-based surveys.

Perhaps the most interesting results, in terms of what they tell us beyond simply establishing validity, are the relationships between PROMIS-29 v2.0 and VR-36 scores in this population. One finding that clearly emerges from this report is that the two summary scores of the PROMIS-29 v2.0 (PHS and MHS) are highly correlated ($r = 0.74$), while the two summary scores of the VR-36 are only weakly correlated ($r = 0.13$ in our sample). These findings generally are consistent with what is known about the relationship between the PCS and MCS and among scores on the eight scores of the PROMIS-29 v2.0. All of the PROMIS-29 v2.0 scores were positively correlated with all of the VR-36 scores, taking into account the negative polarity of several PROMIS-29 v2.0 items. In general, these results are supportive of the construct validity of PROMIS-29 v2.0 as a valid measurement of HRQoL in this population. It is also worth noting that both the PHS and the MHS measure a construct more closely related to the PCS than the MCS. The implication of this finding is that the PROMIS-29 v2.0 summary scores are not the same as the PCS and MCS, and should not be interpreted as if they are. This may reflect differences in how the scores are formulated; the PCS and MCS are designed to be orthogonal to each other and therefore have a low correlation with each other, which is not true of the PHS and MHS. In fact, this may represent a weakness in the construction of the VR-36 and related measures, because it conflicts with the reality of how most people experience HRQoL. This, in turn, may contribute to inconsistency between VR-36 scale scores and summary scores [26–28]. Thus, it has been argued that the PROMIS-29, by not forcing physical and mental HRQoL to be orthogonal, represents an important advance in measurement [12]. Future studies should explore the extent to which the summary scores of the PROMIS-29 v2.0 may measure a somewhat different aspect of HRQoL than the VR-12 and related measures.

We also note that we settled on a two-factor solution for the summary scores of the PROMIS-29, which concords with the findings of Hays, et al. [12]. However, we agree with Hays et al. that there could be other possibilities for future research, including preference-based scoring functions [29–31], that could result in different solutions. Thus, the present study should not be taken as the final word on summary scores for PROMIS-29.

This study has important strengths. It represents a painstaking primary data collection effort, combined with clinical and sociodemographic data, and represents a singular effort to measure the HRQoL of a large population of older adults with MCCs in a primary care setting. The results clearly establish the validity of the PROMIS-29 v2.0 for use in such a population. However, some limitations should also be noted. First, like any survey effort, ours was subject to non-response, which did vary by survey mode. Biased response, when present, can lead to samples that are less than fully representative of the population being studied [32,33]. Nevertheless, it should be noted that we only found minimal differences between responders and non-responders, none of which seemed to represent a serious threat to study validity. Another limitation that should be kept in mind is that KPCO members may be somewhat different from the general U.S. population, and largely reflect enrollment in a Medicare Advantage benefit plan. This is important, because seniors in the US who choose a Medicare Advantage plan as opposed to fee-for-service Medicare generally have different expectations about their future health utilization, and are often healthier individuals than those who choose fee-for-service. In addition, our sample had limited diversity in terms of race/ethnicity and residence in high-poverty ZIP codes. It would be useful to extend these results to other, more diverse populations.

Despite these limitations, the results of this study support the validity of PROMIS-29 v2.0 for use in populations of older adults with MCCs, including those over age 80. As HRQoL measurement continues to be used for diverse purposes, and as the PROMIS-29 v2.0 instrument grows in stature as one of the most widely-used instruments to measure it, we can be confident that HRQoL measurement using the PROMIS-29 v2.0 instrument among the oldest and sickest ambulatory patients will also be valid and reliable.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

none

Funding: Funded by the National Institute on Aging (contract #HHSN271201500064C NIH NIA, PI: Edelen). The funder had no role in data collection, data analysis, interpretation, manuscript drafting, manuscript revision, or decision to submit for publication.

References

1. *HealthMeasures*. PROMIS.
2. Cella D, et al., The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *J Clin Epidemiol*, 2010 63(11): p. 1179–94. [PubMed: 20685078]

3. Hinchcliff M, et al., Validity of two new patient-reported outcome measures in systemic sclerosis: Patient-Reported Outcomes Measurement Information System 29-item Health Profile and Functional Assessment of Chronic Illness Therapy-Dyspnea short form. *Arthritis Care Res (Hoboken)*, 2011 63(11): p. 1620–8. [PubMed: 22034123]
4. Alcantara J, Ohm J, and Alcantara J, The use of PROMIS and the RAND VSQ9 in chiropractic patients receiving care with the Webster Technique. *Complement Ther Clin Pract*, 2016 23: p. 110–6. [PubMed: 25990687]
5. Beaumont JL, et al., Comparison of health-related quality of life in patients with neuroendocrine tumors with quality of life in the general US population. *Pancreas*, 2012 41(3): p. 461–6. [PubMed: 22422138]
6. Pearman TP, et al., Health-related quality of life in patients with neuroendocrine tumors: an investigation of treatment type, disease status, and symptom burden. *Support Care Cancer*, 2016 24(9): p. 3695–703. [PubMed: 27029477]
7. IsHak WW, et al., Patient-Reported Outcomes of Quality of Life, Functioning, and GI/Psychiatric Symptom Severity in Patients with Inflammatory Bowel Disease (IBD). *Inflamm Bowel Dis*, 2017 23(5): p. 798–803. [PubMed: 28301432]
8. Katz P, Pedro S, and Michaud K, Performance of the PROMIS 29-Item Profile in Rheumatoid Arthritis, Osteoarthritis, Fibromyalgia, and Systemic Lupus Erythematosus. *Arthritis Care Res (Hoboken)*, 2016.
9. Lai JS, et al., An evaluation of health-related quality of life in patients with systemic lupus erythematosus using PROMIS and Neuro-QoL. *Clin Rheumatol*, 2017 36(3): p. 555–562. [PubMed: 27848056]
10. Schnall R, et al., *A Health-Related Quality-of-Life Measure for Use in Patients with HIV: A Validation Study*. *AIDS Patient Care STDS*, 2017 31(2): p. 43–48. [PubMed: 28051875]
11. Ware JE, Jr. and Sherbourne CD, The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*, 1992 30(6): p. 473–483. [PubMed: 1593914]
12. Hays RD, et al., PROMIS-29 v.2.0 Profile physical and mental health summary scores. *Quality of Life Research*. Epub ahead of print. 10.1007/s11136-018-1842-3.
13. Kazis LE, et al., Improving the response choices on the veterans SF-36 health survey role functioning scales: results from the Veterans Health Study. *J Ambul Care Manage*, 2004 27(3): p. 263–80. [PubMed: 15287216]
14. Tarlov AR, et al., The medical outcomes study: An application of methods for monitoring the results of medical care. *JAMA*, 1989 262(7): p. 925–930. [PubMed: 2754793]
15. Goldberg J, et al., The association of PTSD with physical and mental health functioning and disability (VA Cooperative Study #569: the course and consequences of posttraumatic stress disorder in Vietnam-era veteran twins). *Qual Life Res*, 2014 23(5): p. 1579–91. [PubMed: 24318083]
16. Bishawi M, et al., Changes in health-related quality of life in off-pump versus on-pump cardiac surgery: Veterans Affairs Randomized On/Off Bypass trial. *Ann Thorac Surg*, 2013 95(6): p. 1946–51. [PubMed: 23453761]
17. Turner AP, Kivlahan DR, and Haselkorn JK, Exercise and quality of life among people with multiple sclerosis: looking beyond physical functioning to mental health and participation in life. *Arch Phys Med Rehabil*, 2009 90(3): p. 420–8. [PubMed: 19254606]
18. NCQA HEDIS 2006 *Specifications for the Medicare Health Outcomes Survey. Volume 6*. 2006, National Committee for Quality Assurance.
19. Krieger N, et al., Painting a truer picture of US socioeconomic and racial/ethnic health inequalities: the Public Health Disparities Geocoding Project. *Am J Public Health*, 2005 95(2): p. 312–23. [PubMed: 15671470]
20. Gelin MN and Zumbo BD, Differential item functioning results may change depending on how an item is scored: An illustration with the Center for Epidemiologic Studies Depression Scale. *Educational and Psychological Measurement*, 2003 63(1): p. 65–74.

21. Benjamini Y and Hochberg Y, Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 1995: p. 289–300.
22. Bindman AB, Keane D, and Lurie N, Measuring health changes among severely ill patients. The floor phenomenon. *Med Care*, 1990 28(12): p. 1142–52. [PubMed: 2250498]
23. Selim AJ, et al., The health status of elderly veteran enrollees in the Veterans Health Administration. *J Am Geriatr Soc*, 2004 52(8): p. 1271–6. [PubMed: 15271113]
24. Cohen J, *Statistical Power Analysis for the Behavioral Sciences*. 2 ed. 1988, Hillsdale: Lawrence Erlbaum Associates.
25. Rogers WH, et al., Comparing the Health Status of VA and Non-VA Ambulatory Patients: The Veterans' Health and Medical Outcomes Studies. *The Journal of Ambulatory Care Management*, 2004 27(3): p. 249–262. [PubMed: 15287215]
26. Farivar SS, Cunningham WE, & Hays RD Correlated physical and mental health summary scores for the SF-36 and SF-12 health survey, V. 1. *Health and Quality of Life Outcomes*, 2007 5:54 18. [PubMed: 17428340]
27. Taft C, Karlsson J, & Sullivan M Do SF-36 summary component scores accurately summarize subscale scores? *Qual Life Res*, 2001 10(5):p. 395–404. [PubMed: 11763202]
28. Simon GE, et al. SF-36 summary scores: Are physical and mental health truly distinct? *Med Care*, 1998 36(4):p. 567–572. [PubMed: 9544596]
29. Craig BM, et al. US valuation of health outcomes measured using the PROMIS-29. *Value in Health*, 2014 17(8): p. 846–853. [PubMed: 25498780]
30. Hanmer J, et al. Selection of key health domains from PROMIS for a generic preference-based scoring system. *Qual Life Res*, 2017 26(12):p. 3377–3385. [PubMed: 28825177]
31. Hanmer J, et al. The PROMIS of QALYs. *Health and Quality of Life Outcomes*, 2015 13:122. [PubMed: 26260491]
32. Krosnick JA, *SURVEY RESEARCH*. *Annual Review of Psychology*, 1999 50(1): p. 537–567.
33. Rogelberg SG and Stanton JM, Introduction. *Organizational Research Methods*, 2007 10(2): p. 195–209.

Table 1:

Respondent demographic and clinical characteristics, and summary health-related quality of life scores.
Percentages given except where otherwise noted. (n = 1,359)

	Percentage (except where noted)
Mean Age, Years (SD)	80.7 (6.8)
Age Groups	
65–69	10%
70–74	13%
75–79	10%
80–84	38%
85+	28%
Sex	
Male	48%
Female	52%
Race/Ethnicity	
White/Non-Hispanic	89%
Hispanic	4%
Non-White/Non-Hispanic	4%
Missing Race/Non-Hispanic	3%
Percent Below Poverty in Census Tract	
0–9.99%	59%
10–19.99%	32%
20%+	9%
Total Number of 13 Chronic Conditions	
2	35%
3	31%
4	18%
5+	16%
Presence of a Specific Chronic Condition	
Arthritis	25%
Cancer	9%
Chronic Lung Disease	38%
Congestive Heart Failure	16%
Depression	23%
Diabetes	31%
Hypertension	82%
Inflammatory Bowel Disease	1%
Ischemic Heart Disease	29%
Osteoporosis	23%

	Percentage (except where noted)
Other Heart Problems	35%
Sciatica	5%
Stroke	6%
Any Home Health Encounters in Past 12 Months?	
No	46%
Yes	54%
Number of Primary Care Visits in Past 12 Months	
0-3	43%
4-6	35%
7-9	12%
10+	10%
Number of Specialty Care Visits in Past 12 Months	
0-3	56%
4-6	22%
7-9	11%
10+	11%
Number of Hospitalizations in Past 12 Months	
0	87%
1	10%
2+	3%
Summary PROMIS-29 v2.0 Scores (Mean, SD)	
PHS	42.2 (9.2)
MHS	50.1 (8.0)
Summary VR-36 Scores (Mean, SD)	
PCS	38.1 (11.5)
MCS	55.4 (9.8)

PHS: PROMIS-29 v2.0 Physical Health Score; MHS: PROMIS-29 v2.0 Mental Health Score; PCS: Physical Component Score; MCS: Mental Component Score.

Table 2: Bivariate analyses for associations between PROMIS-29 v2.0 scores and measured respondent characteristics.

	PHS	MHS	Anxiety	Depression	Fatigue	Pain	Physical Function	Sleep Disturbance	Social Roles	Pain Intensity
Age Groups										
65–69	REF	REF	REF	REF	REF	REF	REF	REF	REF	REF
70–74	1.01	1.26	–0.92	–0.94	–1.50	–0.33	0.98	–1.14	0.40	–0.11
75–79	–1.08	0.89	–1.13	–0.74	–1.23	0.06	–1.59	–1.26	–0.71	0.08
80–84	–3.06 [†]	0.12	–0.37	0.06	–0.02	–0.27	–3.52 [†]	–1.64	–1.67	–0.15
85+	–6.60 [†]	–1.82 [*]	0.62	2.44 [*]	1.48	0.94	–7.46 [†]	–0.93	–5.23 [†]	0.02
Sex										
Male	REF	REF	REF	REF	REF	REF	REF	REF	REF	REF
Female	–3.51 [†]	–2.50 [†]	1.83 [†]	1.50 [*]	2.37 [†]	2.94 [†]	–3.22 [†]	1.59 [†]	–2.53 [†]	0.90 [†]
Race/Ethnicity										
White/Non-Hispanic	REF	REF	REF	REF	REF	REF	REF	REF	REF	REF
Hispanic	–2.93 [*]	–3.94 [†]	2.53 [*]	3.49 [*]	3.31 [*]	4.79 [†]	–2.36	3.92 [*]	–3.03 [*]	1.02 [*]
Non-White/Non-Hispanic	0.02	–0.07	1.55	0.61	–1.58	1.41	0.80	0.69	0.84	0.50
Missing Race/Non-Hispanic	2.20	0.69	–0.91	0.56	–1.22	–0.35	2.17	0.35	–0.43	0.11
Percent Below Poverty in Census Tract										
0–9.99%	REF	REF	REF	REF	REF	REF	REF	REF	REF	REF
10–19.99%	–1.64 [*]	–0.98 [*]	0.74	0.91	1.07 [*]	0.72	–1.59 [*]	–0.05	–1.28 [*]	0.08
20%+	–1.06	–1.08	0.99	1.49	1.14	–0.74	–1.13	0.63	–0.65	0.07
Total Number of Chronic Conditions										
2	REF	REF	REF	REF	REF	REF	REF	REF	REF	REF
3	–2.62 [†]	–2.00 [†]	0.93	1.05	1.97 [†]	2.24 [†]	–2.49 [†]	0.78	–2.26 [†]	0.31

	PHS	MHS	Anxiety	Depression	Fatigue	Pain	Physical Function	Sleep Disturbance	Social Roles	Pain Intensity
4	-4.43 [†]	-3.62 [†]	1.63 [*]	2.23 [†]	3.97 [†]	3.59 [†]	-4.42 [†]	1.27	-4.28 [†]	0.76 [†]
5+	-8.11 [†]	-6.18 [†]	3.48 [†]	4.87 [†]	6.27 [†]	6.14 [†]	-7.96 [†]	2.08 [*]	-7.66 [†]	1.35 [†]
Presence of a Specific Chronic Condition [†]										
Arthritis	-2.41 [†]	-1.43 [*]	0.39	1.04	1.05	3.61 [†]	-2.46 [†]	0.32	-2.20 [†]	0.82 [†]
Cancer	1.18	0.89	-0.42	-0.30	-0.74	-1.11	1.07	-0.33	1.47	-0.42
Chronic Lung Disease	-1.67 [*]	-1.06 [*]	-0.11	0.29	1.28 [*]	1.03	-1.63 [*]	1.07	-1.30 [*]	0.23
Congestive Heart Failure	-4.66 [†]	-2.94 [†]	1.47 [*]	2.30 [†]	3.52 [†]	1.35	-4.61 [†]	1.02	-4.16 [†]	0.32
Depression	-3.37 [†]	-4.96 [†]	5.23 [†]	5.78 [†]	5.11 [†]	4.14 [†]	-3.17 [†]	2.42 [†]	-4.07 [†]	1.01 [†]
Diabetes	-0.82	-0.27	-0.81	-0.03	-0.08	1.22 [*]	-0.83	0.28	-0.86	0.24
Hypertension	-0.70	0.28	0.08	-0.25	0.08	0.88	-1.09	-0.75	-0.33	0.05
Inflammatory Bowel Disease	-0.62	-1.75	0.71	-1.06	3.20	0.33	-0.30	1.35	-0.81	0.16
Ischemic Heart Disease	-1.25 [*]	-0.69	0.06	0.17	0.84	0.97	-1.19 [*]	-0.69	-1.23 [*]	0.11
Osteoporosis	-2.92 [†]	-1.59 [*]	1.22 [*]	1.42 [*]	1.35 [*]	1.54 [*]	-3.09 [†]	0.92	-1.79 [*]	0.47 [*]
Other Heart Problems	-0.91	-1.29 [*]	0.70	0.95 [*]	1.67 [*]	-0.01	-0.97	0.27	-1.58 [*]	-0.05
Sciatica	-1.59	-1.42	0.60	0.05	0.30	4.27 [†]	-1.50	1.88	-1.59	1.37 [†]
Stroke	-4.58 [†]	-3.43 [†]	2.69 [*]	3.42 [†]	3.78 [†]	0.65	-4.62 [†]	-0.73	-5.90 [†]	-0.10
Any Home Health Encounters in Past 12 Months?										
No	REF	REF	REF	REF	REF	REF	REF	REF	REF	REF
Yes	-5.60 [†]	-3.50 [†]	1.77 [†]	2.52 [†]	3.95 [†]	2.75 [†]	-5.48 [†]	0.84	-4.50 [†]	0.52 [†]
Number of Primary Care Visits in Past 12 Months										
0-3	REF	REF	REF	REF	REF	REF	REF	REF	REF	REF
4-6	-0.52	-0.87	0.86	0.40	0.99	1.69 [*]	-0.31	0.43	-0.83	0.30 [*]

	PHS	MHS	Anxiety	Depression	Fatigue	Pain	Physical Function	Sleep Disturbance	Social Roles	Pain Intensity
7-9	-3.52 [†]	-3.05 [†]	1.54 [*]	1.59 [*]	3.40 [†]	4.56 [†]	-3.02 [†]	1.41	-3.07 [†]	1.08 [†]
10+	-5.79 [†]	-4.87 [†]	3.19 [†]	4.06 [†]	5.17 [†]	6.48 [†]	-5.20 [†]	1.96 [*]	-4.83 [†]	1.36 [†]
Number of Specialty Care Visits in Past 12 Months										
0-3	REF	REF	REF	REF	REF	REF	REF	REF	REF	REF
4-6	-0.63	-0.52	-0.52	-0.96	0.72	1.44 [*]	-0.34	0.62	-0.87	0.12
7-9	-1.65 [*]	-1.05	-0.37	-0.61	0.86	2.22 [*]	-1.54	1.37	-0.90	0.59 [*]
10+	-1.65 [*]	-1.44	-0.53	-0.73	2.34 [*]	2.04 [*]	-1.66 [*]	0.25	-1.97 [*]	0.38
Number of Hospitalizations in Past 12 Months										
0	REF	REF	REF	REF	REF	REF	REF	REF	REF	REF
1	-4.03 [†]	-2.96 [†]	1.81 [*]	2.44 [*]	3.71 [†]	2.06 [*]	-4.17 [†]	1.26	-3.64 [†]	0.45 [*]
2+	-7.08 [†]	-4.37 [†]	2.16	3.13 [*]	3.74 [*]	2.83	-6.77 [†]	0.33	-6.95 [†]	0.42

^{*} p < 0.05

[†] p < 0.001

[‡] Compared to the absence of that condition.

PHS: PROMIS-29 v2.0 Physical Health Score; MHS: PROMIS-29 v2.0 Mental Health Score; REF: Reference category.

Table 3:

Pearson correlations between the eight PROMIS-29 v2.0 scores and the two summary scores of the Veterans RAND 36-Item Health Survey (VR-36), measured concurrently. Also, Pearson correlations between the two PROMIS-29 summary scores and the two summary scores of the VR-36, measured concurrently (n = 1198).

Scale or Score	PCS (of the VR36)	MCS (of the VR36)	PHS (of the PROMIS-29)
PROMIS-29 Scales			
Physical Function (P)	0.81	0.28	--
Social Roles (P)	0.72	0.49	--
Pain Interference (N)	-0.70	-0.33	--
Pain Intensity (N)	-0.61	-0.25	--
Fatigue (N)	-0.63	-0.54	--
Anxiety (N)	-0.30	-0.66	--
Depression (N)	-0.40	-0.66	--
Sleep Disturbance (N)	-0.35	-0.38	--
Summary Scores			
PHS (of the PROMIS-29)	0.85	0.32	--
MHS (of the PROMIS-29)	0.70	0.64	0.74
PCS (of the VR-36)	--	0.13	--
MCS (of the VR-36)	--	--	--

(N) means that a higher score indicates worse health-related quality of life, while (P) means that a higher score indicates better health-related quality of life.

Shading indicates a correlation above 0.5. Empty cells are reported elsewhere in the table or would represent a correlation between an item and itself. All correlations statistically significant at the $p < 0.001$ level.

PCS: Physical Component Score; MCS: Mental Component Score; PHS: PROMIS-29 v2.0 Physical Health Score; MHS: PROMIS-29 v2.0 Mental Health Score.