



Published in final edited form as:

Qual Life Res. 2016 November ; 25(11): 2811–2823. doi:10.1007/s11136-016-1325-3.

Reliability and validity of PROMIS measures administered by telephone interview in a longitudinal localized prostate cancer study

Caroleen W. Quach, MSPH¹, Michelle M. Langer, PhD^{2,*},⁷, Ronald C. Chen, MD, MPH^{2,3,4}, David Thissen, PhD⁵, Deborah S. Usinger, BA^{2,3}, Marc A. Emerson, MPH⁶, and Bryce B. Reeve, PhD^{1,2}

¹Department of Health Policy and Management, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC

²Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC

³Cecil G. Sheps Center for Health Services Research, University of North Carolina at Chapel Hill, Chapel Hill, NC

⁴Department of Radiation Oncology, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC

⁵Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill, Chapel Hill, NC

⁶Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC

⁷American Institutes of Research, Chapel Hill, NC

*Dr. Langer's affiliation when the study was conducted was Lineberger; her current affiliation is American Institutes of Research

Abstract

Purpose: To evaluate the reliability and validity of six PROMIS measures (anxiety, depression, fatigue, pain interference, physical function, sleep disturbance) telephone-administered to a diverse, population-based cohort of localized prostate cancer patients.

CORRESPONDING AUTHOR: Caroleen W. Quach, cquach@unc.edu, Telephone: (919) 962-5434, Fax: (919) 843-6308.

Part of this research was presented at the International Society for Quality of Life 22nd Annual Conference in Vancouver, BC, Canada in October 2015.

Disclosure of Potential Conflicts of Interest

The authors declare that they have no conflicts of interest related to this research.

COMPLIANCE WITH ETHICAL STANDARDS

Research Involving Human Participants and/or Animals

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This article does not contain any studies with animals performed by any of the authors.

Informed Consent

Informed consent was obtained from all individual participants included in the study.

Methods: Newly-diagnosed men were enrolled in the North Carolina Prostate cancer Comparative Effectiveness and Survivorship Study. PROMIS measures were telephone-administered pre-treatment (baseline), and at 3-months and 12-months post-treatment initiation (N=778). Reliability was evaluated using Cronbach's alpha. Dimensionality was examined with bifactor models and explained common variance (ECV). Ordinal logistic regression models were used to detect potential differential item functioning (DIF) for key demographic groups. Convergent and discriminant validity were assessed by correlations with the legacy instruments Memorial Anxiety Scale for Prostate Cancer and SF-12v2. Known-groups validity was examined by age, race/ethnicity, comorbidity, and treatment.

Results: Each PROMIS measure had high Cronbach's alpha values (0.86 to 0.96) and was sufficiently unidimensional. Floor effects were observed for anxiety, depression, and pain interference measures; ceiling effects were observed for physical function. No DIF was detected. Convergent validity was established with moderate to strong correlations between PROMIS and legacy measures (0.41 to 0.77) of similar constructs. Discriminant validity was demonstrated with weak correlations between measures of dissimilar domains (-0.20 to -0.31). PROMIS measures detected differences across age, race/ethnicity, and comorbidity groups; no differences were found by treatment.

Conclusions: This study provides support for the reliability and construct validity of six PROMIS measures in prostate cancer, as well as the utility of telephone administration for assessing HRQoL in low literacy and hard-to-reach populations.

Keywords

Prostate cancer; reliability; validity; psychometric validation; comparative effectiveness research

INTRODUCTION

Prostate cancer is the most common solid tumor malignancy in American men [1]. In 2015, there will be an estimated 220,800 new cases of prostate cancer, and 27,540 men will die of this disease [2]. The disease and its treatments cause significant burden in terms of morbidity, mortality, health-related quality of life (HRQoL), and costs—to patients as well as the U.S. healthcare system. Though several treatment options for localized prostate cancer, such as radical prostatectomy, radiation therapy, and hormone therapy, are available, none has been shown to be clearly superior in terms of survival [3–5]. The disease and its treatments have different effects on HRQoL [5,6]; thus, there is a need to have valid and reliable patient-reported outcome (PRO) measures of symptoms and functioning to facilitate comparative effectiveness research (CER) in prostate cancer [7].

The Patient-Reported Outcomes Measurement Information System® (PROMIS), a National Institutes of Health initiative, has developed an extensive set of self-report questionnaires that measure a variety of physical, mental, and social health domains that are relevant to men with prostate cancer [8,9]. While cancer patients were included in the initial validation of the PROMIS measures [9], no study, to our knowledge, has examined the psychometric properties of the PROMIS domains specifically in prostate cancer patients. Performance of both general HRQoL domains and prostate cancer-specific concerns need to be examined.

This disease-specific validation evidence is needed to support inclusion of PROMIS measures in future prostate cancer CER studies. Our study complements the validation work of other researchers who are evaluating PROMIS measures in multiple cancer populations (including prostate) [10].

In addition, patients with low literacy are often excluded from PRO research [11]. Previous validation studies of PROMIS measures have been limited to participants who are able to read and respond to PROMIS items on a computer, handheld device, or paper. For example, data collection in the initial PROMIS validation study was exclusively electronic. Thus patients with low literacy or who do not have access to a computer were not able to participate in these studies. Use of telephone interviews to collect patient self-report data will allow us to reach low literacy and vulnerable populations that are often underrepresented in clinical research.

The North Carolina Prostate cancer Comparative Effectiveness and Survivorship Study (NC ProCESS) was designed as a prospective cohort study to compare the effectiveness of different treatment options for localized prostate cancer with respect to key patient outcomes, including cancer control and HRQoL [12]. The NC ProCESS cohort is diverse with respect to race (27% African-American), education (34% with a high school degree or less), age (49% of participants are 65 years or older), and residence (50% of participants live in medically underserved areas) [12]. Patient-reported HRQoL data were collected pre-treatment (baseline), and at 3- and 12-months post-treatment initiation. Thus, the diversity of the patient population, the mode of data collection, and longitudinal nature of the data make the NC ProCESS an ideal platform for psychometric evaluation of the PROMIS measures in prostate cancer. Specifically, this study evaluates the reliability and validity of six PROMIS measures (anxiety, depression, fatigue, pain interference, physical function, sleep disturbance) administered via telephone interview in a diverse, population-based cohort of localized prostate cancer patients. A subsequent publication will include an assessment of the psychometric properties of disease-specific measures such as sexual function in the NC ProCESS cohort.

METHODS

Participants

The NC ProCESS is a population-based cohort study that recruited newly-diagnosed prostate cancer patients from all 100 counties in North Carolina through the Rapid Case Ascertainment (RCA) mechanism of the North Carolina Central Cancer Registry. New cases are typically reported to RCA within 1-2 weeks of diagnosis. A letter was mailed to each newly-diagnosed patient's physician to explain NC ProCESS. Following a 2-week physician opt-out period, a letter and brochure were mailed to the patient describing the study. Of 2,473 eligible men, a total of 1,419 men (57%) enrolled from January 2011 to June 2013. Study participants speak English and receive their cancer care in North Carolina. All baseline data were collected prior to treatment, and participants continue to be prospectively followed. Median time from diagnosis to the baseline survey was 5 weeks. Additional details on NC ProCESS are described elsewhere [12,13]. In September 2012, we received additional funding from the National Cancer Institute (NCI) to add PROMIS measures to the

NC ProCESS and perform a psychometric evaluation of the PROMIS measures. Because the PROMIS measures were added approximately 21 months after the NC ProCESS started, smaller numbers of participants completed PROMIS measures at baseline (n=333), 3-month (n=411), and 12-month (N=778) time points than in the larger NC ProCESS: n=1,456 at baseline, n=1,163 at 3-months, n=1,079 at 12-months.

This study (#10-1483) was approved by the University of North Carolina Institutional Review Board.

Data and Measures

The NC ProCESS collects self-reported socio-demographic characteristics, treatment type, and comorbid conditions. Responses to PROMIS measures were collected prospectively via telephone interview at baseline (pretreatment) and at 3- and 12-months after initiating treatment. This study evaluated six PROMIS measures, including anxiety (5 items), depression (5 items), fatigue (5 items), pain interference (5 items), physical function (6 items), and sleep disturbance (4 items). These items appear on the PROMIS short forms available through the Assessment Center (<http://www.assessmentcenter.net/>); these are version 1 of the short forms that were available at the time we started the study (2011) [14]. Each measure includes the items in the 4-item short forms and an additional 1-2 items from the domain bank (except for sleep disturbance). The additional items were selected to examine item-level performance as well as scale-level performance. PROMIS item banks have undergone rigorous evaluation and were calibrated with item response theory (IRT) models, which place each item on a common metric [9,15,16]. This common metric is a key advantage of the PROMIS measures as it allows for development of fixed, short forms and application of computerized adaptive testing (CAT). Short forms were selected over CAT as our study measures were embedded in a parent study (NC ProCESS) so integration of a CAT method was not feasible. Regardless of the type of measure, scores can be compared and combined [8]. Scores for each PROMIS domain are reported on a T-score metric with a mean of 50 and standard deviation (SD) of 10 in the U.S. general population. The T-score metric is a linear transformation from the IRT theta scale: $T\text{-score} = 10 * \theta + 50$. Higher PROMIS scores reflect a higher level of the construct measured. For example, higher anxiety scores indicate more anxiety, and higher physical function scores indicate better physical function. PROMIS items use a 5-point ordered categorical (or ordinal) response scale.

The NC ProCESS included two existing or legacy PRO instruments: the Medical Outcomes Study Short Form-12 version 2 (SF-12v2; administered at all three assessment points) and the Memorial Anxiety Scale for Prostate Cancer (MAX-PC; administered at the 12-month follow-up). The SF-12v2 includes summary scores for Physical Component Summary (PCS) and Mental Component Summary (MCS), and 8 sub-domain scores for Physical Function (PF), Role-Physical (RP), Bodily Pain (BP), General Health (GH), Social Function (SF), Role-Emotional (RE), Mental Health (MH), and Vitality (VT) [17]. Higher SF-12v2 scores reflect better HRQoL. However, the standard scoring algorithm for the component summary scores, which are based on an orthogonal (uncorrelated) rotation model, may yield inconsistent results compared to the subscale scores [18,19]. The MAX-PC, which was

specifically developed for prostate cancer, includes 18 items that are aggregated into an overall score, where higher scores reflect more anxiety about prostate cancer [20,21].

Analysis

Descriptive Statistics—Item-level descriptive statistics included frequencies, percent missing, mean, and standard deviation (SD). Floor and ceiling effects were calculated as the proportion of men who had the minimum and maximum summed score, respectively; this was examined for each measure and by assessment point.

Reliability—Each PROMIS measure was evaluated by Cronbach's alpha, a traditional measure of reliability [22]. Alpha values of 0.70 or greater are an acceptable minimum for group-level assessment [23,24].

Validity—Structural validity was evaluated with a factor analytic approach using IRT to confirm unidimensionality. For each PROMIS measure, a unidimensional confirmatory factor analysis (CFA) with the graded response IRT model [25,26] was conducted using IRTPRO (Scientific Software International, Lincolnwood, IL). Overall model fit was evaluated based on the root mean square error of approximation (RMSEA) [27], where a smaller RMSEA value indicates a closer fit. An RMSEA value ≤ 0.06 is considered to reflect good fit, values ≤ 0.08 are fair, and values above 0.10 generally reflect poor fit [28]. The $S-X^2$ statistic [29–31] was used to assess item-level fit, for which a nonsignificant result ($p > 0.05$, adjusted for multiple comparisons) was an indicator of adequate model fit. Although there is less power to detect item misfit with shorter tests, we are also assuming item fit to the IRT graded response model would not be significantly different from the initial validity analyses with larger PROMIS item pools. The standardized local dependence (LD) X^2 statistic [32] was used to identify items that were excessively related after controlling for the underlying domain; values larger than 10 indicated substantial LD. An additional check on the dimensionality of the data was performed by estimating a bifactor graded response IRT model [33] with each identified LD pair or set of items as a second order factor. Any LD violations were deemed negligible if the explained common variance (ECV) was at least 0.90 [34–36]. ECV represents the variance explained by the general factor in the bifactor model.

After confirming the unidimensionality (structural validity) of each PROMIS measure, potential differential item functioning (DIF) was examined to detect whether items behave differently across age (<65 vs. ≥ 65 years), education (high school or less vs. more than high school), and ethnicity/race (non-Hispanic whites [NHW] vs. non-Hispanic blacks) groups. For each item within a domain, an ordinal logistic regression (OLR) model was used to examine whether item responses were significantly associated with group membership after controlling for participants' summed score on the measure. Uniform DIF was detected by a likelihood ratio test comparing an OLR model with one predictor, summed score, to an OLR model with an additional predictor, group membership, representing a shift in the use of the response options due to group membership (e.g., after controlling for overall level of symptoms, one group has a higher endorsement rate of item response options reflecting greater symptom severity). Non-uniform DIF was detected by a likelihood ratio test

comparing the OLR model with two predictors, summed score and group membership, to an OLR model with an additional interaction term, representing a difference in how strongly the item is related to the underlying construct due to group membership (e.g., the item provides better measurement of functional status for one group versus another). With each paired-group analysis, an initial OLR model was fitted to identify an anchor group of items without DIF. For each sequential OLR model, any items previously identified as having DIF were removed from the summed score computation. The final OLR model used a summed score computed with only the DIF-free anchor items to test for DIF. The Benjamini-Hochberg procedure was used to make inferential decisions in the context of the multiple comparisons [37,38]. In addition to examining the significance ($p < 0.05$), magnitude of DIF was further evaluated by examining the expected item scores and estimating the effect sizes ($R^2 > 0.02$ indicative of salient DIF) [39].

Convergent and discriminant validity were evaluated using Spearman's rank correlations of PROMIS anxiety, depression, and physical function scores with SF-12v2 MCS and PCS and MAX-PC scores [17,20,21]. We expected moderate to strong correlations between common measures of physical health (PROMIS physical function with SF-12v2 PCS and PF scales), mental health (PROMIS anxiety and depression with MAX-PC, SF-12v2 MCS and MH scales), fatigue (PROMIS fatigue with SF-12v2 MCS and VT scales), and pain interference (PROMIS pain interference with SF-12v2 BP scales). We expected weak correlations between measures of dissimilar constructs, such as physical function with mental health. Correlation coefficients (r) were interpreted using Dancey and Reidy's classifications [40]: $r = 0$ corresponds to no correlation, $0 < r < 0.4$ is a weak correlation, $0.4 < r < 0.7$ is a moderate correlation, $0.7 < r < 1$ is a strong correlation, and 1 is perfect correlation.

Known-groups validity was examined with t -tests of mean PROMIS T-scores for age (<65 vs. ≥65 years), race/ethnicity (NHW vs. Black), mental health comorbidities (no mental health comorbidity vs. ≥1 mental health comorbidities); limiting physical health/other comorbidities (no comorbidity that limits usual or daily activities vs. ≥1 comorbidities that limit usual or daily activities), and treatment groups. The purpose was to assess the extent to which PROMIS measures could discriminate between groups that should, in theory, differ—based on evidence from the published literature. Given the review of the literature summarized in the Discussion section, we hypothesized that younger men (aged <65 years) would have worse HRQoL versus men aged ≥65 years (except for physical functioning), Black men would have worse HRQoL compared to NHW men, and men with ≥1 [limiting] comorbidities would report worse HRQoL versus men with no [limiting] comorbidities. For the known-groups validity analysis by treatment at 3-months, men were classified into one of three groups (active surveillance, prostatectomy, radiation) according to treatment received by the 3-month follow-up. For the known-groups validity analysis by treatment at 12-months, men were classified into one of four groups (active surveillance, prostatectomy, radiation, hormone therapy) according to treatment received by the 12-month survey. All combinations of self-reported treatments received by the 12-month assessment are presented in Table 1. Men who received multiple treatments were excluded from these known-groups analyses by treatment. The radiation group includes men who received external beam radiation therapy, brachytherapy, or proton therapy. Hormone therapy was not included as a treatment group in the 3-month known-groups analyses because there were few men who

received hormone therapy alone (n=14) relative to the other three treatment groups. We compared means across all treatment groups using a one-way analysis of variance (ANOVA). Pairwise *t*-test comparisons will be performed if the overall F-test is significant.

For evaluation of convergent, discriminant, and known-groups validity, statistical significance was defined at the 0.05 alpha level; these analyses were conducted using SAS 9.3 (SAS Institute, Cary, NC).

RESULTS

Descriptive Statistics

This population-based cohort of 778 men diagnosed with localized prostate cancer is socio-demographically and clinically diverse (Table 1). NHW and Black men comprised 71% and 26% of the sample, respectively. Mean age at baseline was 65 (SD 7.6) years old. Approximately 32% of participants had a high school education or less.

Men in this sample had better HRQoL compared to the U.S. general population in each of the 6 PROMIS domains at the 12-month assessment, and in 5 out of 6 domains at the baseline and 3-month assessments (Table 2). Future work will include estimation of minimally important differences (MIDs) of PROMIS measures in prostate cancer. For now, we compare scores using recommended T-score MID ranges for six PROMIS-Cancer scales in advanced-stage cancer patients: 17-item Fatigue (2.5–4.5), 7-item Fatigue (3.0–5.0), 10-item Pain Interference (4.0– 6.0), 10-item Physical Functioning (4.0–6.0), 9-item Emotional Distress-Anxiety (3.0–4.5), and 10-item Emotional Distress-Depression (3.0–4.5) [41]. Based on those published MID estimates (the most relevant estimates to our sample and measures), differences in anxiety at 12-months, depression at 3-months and 12-months, and fatigue at baseline and 12-months may be clinically meaningful.

Mean±SD T-scores at 12-months ranged from 45.5±8.5 (anxiety) to 49.7±6.2 (sleep disturbance). Missing data was <1%. On the 12-month assessment, substantial floor effects occurred for PROMIS anxiety, depression, and pain interference measures (range 45–58%). Ceiling effects were consistently low across symptom domains (<2%), but high for the physical function measure (44%). Similar patterns of floor and ceiling effects were observed at the baseline and 3-month assessments.

Reliability

Reliability of each PROMIS measure at the 12-month assessment is presented in Table 3. Each PROMIS measure with 5 or more items had high values of Cronbach's alpha (each alpha >0.90). Four-item versions of each PROMIS short form (which can be downloaded from the PROMIS Assessment Center) also yielded good reliability estimates between 0.86 and 0.96 (data not presented). The addition of 1–2 items to the 4-item short forms resulted in either no change in alpha values, or an increase of 0.01 to 0.02. Similar magnitudes were observed for the baseline and 3-month assessments. Small increases in alpha values were expected as longer scales are less vulnerable to measurement error and tend to be more reliable.

We present the 12-month sample data for the reliability and validity analyses due to the larger sample size. Similar results were found for the baseline and 3-month assessments.

Validity

Structural Validity—The structural validity (i.e., unidimensionality) of each PROMIS measure was supported. Although some item-level misfit was found for the anxiety, depression, fatigue, and sleep disturbance measures, the overall model fit was acceptable. The PROMIS pain interference measure had a high RMSEA value (RMSEA=0.33); however, no items were flagged for poor fit. Additionally, all CFAs fit to the measures produced high factor loadings (>0.70) (Table 3). Potential LD was found for some item pairs on all measures except physical functioning. However, bifactor models fit to account for residual variance in these flagged item pairs resulted in ECV values greater than 0.90 for all measures. Therefore, these PROMIS measures can be considered unidimensional for this population of men with prostate cancer.

Differential Item Functioning—Significant DIF was flagged for a few items across PROMIS measures. However, effect sizes were all small ($R^2 < 0.02$), and therefore, no items were considered to function differently by age, education, or ethnicity/race. Thus, these PROMIS items can be considered unbiased for this localized prostate cancer population, measuring the domains of anxiety, depression, fatigue, pain interference, physical function and sleep disturbance based only on the respective underlying trait, and not conditional on group membership.

Convergent and Discriminant Validity—Convergent validity was established with moderate correlations of PROMIS anxiety and depression scores with the MAX-PC, SF-12v2 MCS and MH scores (range 0.41 to -0.64) (Table 3). PROMIS physical function was strongly correlated with the SF-12v2 PCS (0.73) and PF subscale (0.77). PROMIS fatigue and pain interference measures were moderately to strongly correlated with legacy measures of similar constructs (range -0.50 to -0.66).

Discriminant validity was demonstrated with weak correlations of PROMIS physical function with the MAX-PC and SF-12v2 MCS (-0.31 and 0.21, respectively) and weak correlations of SF-12v2 PCS with PROMIS anxiety and depression scales (-0.20 and -0.22, respectively).

Known-groups Validity—The known-groups validity analyses at the baseline, 3-month, and 12-month assessments are presented in Table 4 (by age and race) and Table 5 (by comorbidities).

As hypothesized *a priori*, PROMIS measures detected differences by age (Table 4). At baseline, men aged <65 years had greater anxiety, depression, fatigue, and sleep disturbance than men aged ≥ 65 years. At the 3-month follow-up, younger men (<65) reported significantly worse HRQoL on five out of six measures compared to older men (≥ 65). By 12-months post-treatment initiation, younger men still reported greater anxiety, depression, and sleep disturbance than older men.

Similarly, PROMIS measures demonstrated known-groups validity by race/ethnicity (Table 4). Black men reported greater fatigue, pain interference, and worse physical function than NHW men at baseline (each $p < 0.05$). At 3-months, Black men experienced greater anxiety, depression, pain interference, and worse physical function than NHW men. At 12-months, Black men continued to report greater pain interference and worse physical function compared to NHW men.

PROMIS measures were also able to detect differences across mental health comorbidity groups (0 vs. 1 condition) and physical health/other comorbidity groups (0 vs. 1 limiting condition) for all six domains at each of the three time points (Table 5).

However, PROMIS measures did not detect differences by treatment. At 3-months, a total of 375 men were analyzed across three treatment groups: active surveillance ($n=133$), prostatectomy ($n=146$), and radiation therapy ($n=96$). The overall F-test from the one-way ANOVA across treatment groups revealed no statistically significant differences. For the 12-month known-groups validity analysis, 735 men were analyzed across four treatment groups: active surveillance ($n=263$), prostatectomy ($n=264$), radiation ($n=167$), and hormone therapy ($n=41$). No significant differences were found across treatments.

DISCUSSION

The goal of this study was to evaluate the psychometric properties of six PROMIS measures in a diverse cohort of men with localized prostate cancer who completed assessments via telephone interview. This study provided evidence for the reliability and validity of each PROMIS measure of anxiety, depression, fatigue, pain interference, physical function, and sleep disturbance embedded in a longitudinal CER study.

Notably, we observed substantial floor and ceiling effects in most of the PROMIS measures. This may be in part attributed to our shorter measures comprising 4 to 6 items each. Shorter scales may result in more narrowly defined constructs and affect our ability to reliably assess individual change [42–44]. Still, shorter scales have the advantages of convenience and ease of administration, and are important for minimizing respondent burden, particularly in a clinical setting where time is typically limited. Implications for use of PROMIS measures that exhibit floor and/or ceiling effects in future prostate cancer studies will depend on the study objectives. As an example, if distinguishing among men with low levels of depression is of interest in a study to assess treatment effects on depression, then the PROMIS depression measure may not be ideal. In contrast, those floor effects may be negligible if the presence of depression has already been documented in the study population. Thus use of longer PROMIS forms and/or CAT may be another alternative.

Overall our findings for the known-groups validity evaluation are consistent with previous studies. For the analyses by age, Hampson et al. [45] and Penson et al. [46] also found that younger men had worse mental health (as measured by the SF-36) compared to older men pre-treatment. Similarly, Eton, Lepore, and Helgeson observed better mental health among older men at approximately 7-weeks post treatment [47]. However, though we expected younger men to have better physical functioning [45,47,48], we did not observe any

differences by age at any of the three time points. Regarding the known-groups analysis by race, Jayadevappa et al. (2007) reported similar results that African-American men (versus White men) had significantly higher levels on all SF-36 subscales except bodily pain pretreatment [49]. At 12-months, African-American men had lower role-emotional and bodily pain scores. Jayadevappa et al. (2009) observed that unadjusted baseline SF-36 subscale scores were significantly lower for African-American men [50]. Additionally, it took longer for physical function, role-physical, role-emotional, and general health scores to return to baseline values for African-American men; however, race was no longer a predictor of HRQoL after adjusting for demographic and clinical variables. Brassell et al. [51] reported significant differences between African-American and Caucasian men at baseline for every SF-36 domain but no significant racial group differences in change over a 24-month period for any SF-36 domain. For the comorbidity analyses, we observed that men with one or more comorbidities had significantly higher levels of all measured symptoms and poorer physical functioning compared to men with no comorbidities, which persisted from pre-treatment to 12 months post-treatment. Our findings were supported in a pretreatment localized prostate cancer sample (Penson et al. [46] found that men with a greater number of comorbid conditions had lower HRQoL scores in all eight SF-36 domains), and in a large PROMIS cancer sample (Rothrock et al. [52] observed that having a comorbid condition was associated with greater anxiety, depression, fatigue, pain interference, and worse physical function compared to those with no comorbidity). Lastly, we did not observe HRQoL differences by treatment. Lack of support for our hypotheses may be in part due to our focus on general HRQoL measures as opposed to disease-specific [53], and may also be related to our shorter measures.

This study had several limitations. The study was limited to those who spoke English and received their cancer care in North Carolina. Still, the population-based NC ProCESS sample is more socio-demographically diverse than one typically found in a clinical trial or single-institution study, which may improve generalizability to the broader target prostate cancer population. The average duration of each telephone-based interview was long: baseline=45 minutes, 3-months=35 minutes, and 12-months=44 minutes. It is unknown if participants' responses were affected by interview length. Further, it is unknown what impact the telephone-based PRO assessments had on participants' responses, though measurement equivalence of telephone versus computer administration method will be evaluated in a follow-up study. Notably, a prior study reported no significant differences in score levels, reliability, or validity of PROMIS scales by the method of administration [54]. Although interviewer-administered assessments helped to reach more individuals and minimize missing data, low literacy and language and cultural differences remain barriers to PRO measurement. Finally, a full psychometric review of a PRO measure should include evaluation of responsiveness, or the extent to which a measure can detect change over time. Responsiveness of PROMIS measures in this population will be examined in a follow-up study.

This study had many notable strengths. Unlike clinical trials which rarely reflect the diversity of patients, our sample included 26% African Americans, 53% aged 65 years, and 32% of participants with a high school education or less. Ninety-six percent of participants had one or more comorbid conditions, which allowed us to examine the validity of PROMIS

measures relative to conditions that affect more emotional distress domains (e.g., depression, anxiety) or more physical health domains (e.g., arthritis, back pain, chronic obstructive pulmonary disease). This study provided psychometric evidence to support the use of PROMIS measures for men with prostate cancer via phone-based interviews. Telephone-based administration is critical for reaching individuals that historically have been excluded from PRO and clinical research, including those with low literacy rates. In sum, this study provides psychometric evidence for the reliability and construct validity of six PROMIS measures in a prostate cancer population, thus supporting their use in CER studies and oncology trials.

Acknowledgments

Funding

This research was supported by grants from the Agency for Healthcare Research and Quality (HHSA29020050040ITO6) and the National Cancer Institute (R01CA174453).

References

1. National Cancer Institute (2015). SEER Cancer Statistics Factsheets: Prostate Cancer. <http://seer.cancer.gov/statfacts/html/prost.html>.
2. American Cancer Society (2015). Prostate Cancer Overview. <http://www.cancer.org/cancer/prostatecancer/detailedguide/prostate-cancer-what-is-prostate-cancer>.
3. Wilt TJ, MacDonald R, Rutks I, Shamliyan TA, Taylor BC, & Kane RL (2008). Systematic review: comparative effectiveness and harms of treatments for clinically localized prostate cancer. *Ann Intern Med*, 148(6), 435–448, doi:0000605-200803180-00209 [pii]. [PubMed: 18252677]
4. Xiong T, Turner RM, Wei Y, Neal DE, Lyratzopoulos G, & Higgins JP (2014). Comparative efficacy and safety of treatments for localised prostate cancer: an application of network meta-analysis. *BMJ Open*, 4(5), e004285, doi:10.1136/bmjopen-2013-004285.
5. Sun F, Oyesanmi O, Fontanarosa J, Reston J, Guzzo T, & Schoelles K (12 2014). Therapies for Clinically Localized Prostate Cancer: Update of a 2008 Systematic Review. Comparative Effectiveness Review No. 146. (Prepared by the ECRI Institute–Penn Medicine Evidence-based Practice Center under Contract No. 290-2007-10063.) AHRQ Publication No. 15-EHC004-EF. . Rockville, MD: Agency for Healthcare Research and Quality.
6. Hoffman RM, Penson DF, Zietman AL, & Barry MJ (2013). Comparative effectiveness research in localized prostate cancer treatment. *J Comp Eff Res*, 2(6), 583–593, doi:10.2217/ce.13.66. . [PubMed: 24236797]
7. Aaronson N, Alonso J, Burnam A, Lohr KN, Patrick DL, Perrin E, et al. (2002). Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res*, 11(3), 193–205. [PubMed: 12074258]
8. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, et al. (2007). The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. *Med Care*, 45(5 Suppl 1), S3–s11, doi:10.1097/01.mlr.0000258615.42478.55.
9. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, et al. (2010). 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*, 63(11), 1179–1194, doi: 10.1016/j.jclinepi.2010.04.011. [PubMed: 20685078]
10. Jensen RE, Potosky AL, Reeve BB, Hahn E, Cella D, Fries J, et al. (2015). Validation of the PROMIS physical function measures in a diverse US population-based cohort of cancer patients. *Qual Life Res*, doi: 10.1007/s11136-015-0992-9.

11. Hahn EA, Cella D, Dobrez D, Shiimoto G, Marcus E, Taylor SG, et al. (2004). The talking touchscreen: a new approach to outcomes assessment in low literacy. *Psycho- Oncology*, 13(2), 86–95. [PubMed: 14872527]
12. Chen RC, Carpenter WR, Kim M, Hendrix LH, Agans RP, Meyer AM, et al. (2015). Design of the North Carolina Prostate Cancer Comparative Effectiveness and Survivorship Study (NC ProCESS). *J Comp Eff Res*, 4(1), 3–9, doi: 10.2217/cer.14.67. [PubMed: 25565065]
13. Greenberg CC, Wind JK, Chang GJ, Chen RC, & Schrag D (2013). Stakeholder engagement for comparative effectiveness research in cancer care: experience of the DEcIDE Cancer Consortium. *J Comp Eff Res*, 2(2), 117–125, doi: 10.2217/cer.12.80. [PubMed: 24236554]
14. Gershon RC, Rothrock N, Hanrahan R, Bass M, & Cella D (2010). The Use of PROMIS and Assessment Center to Deliver Patient-Reported Outcome Measures in Clinical Research. *J Appl Meas*, 11(3), 304–314. [PubMed: 20847477]
15. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, et al. (2007). Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care*, 45(5 Suppl 1), S22–31, doi: 10.1097/01.mlr.0000250483.85507.04. [PubMed: 17443115]
16. Riley WT, Rothrock N, Brace B, Christodolou C, Cook K, Hahn EA, et al. (2010). Patient-reported outcomes measurement information system (PROMIS) domain names and definitions revisions: further evaluation of content validity in IRT-derived item banks. *Qual Life Res*, 19(9), 1311–1321, doi: 10.1007/s11136-010-9694-5. [PubMed: 20593306]
17. Ware JE, Kosinski M, Turner-Bowker D, & Gandek B (2002). User's Manual for the SF-12v2® Health Survey (With a Supplement Documenting SF-12® Health Survey). Lincoln, RI: QualityMetric Incorporated.
18. Ware JE, Kosinski M, Turner-Bowker DM, & Gandek B (2002). How to score version 2 of the SF-12 health survey (with a supplement documenting version 1). Lincoln, RI: QualityMetric Incorporated.
19. Farivar SS, Cunningham WE, & Hays RD (2007). Correlated physical and mental health summary scores for the SF-36 and SF-12 Health Survey, V.I. *Health Qual Life Outcomes*, 5, 54, doi: 10.1186/1477-7525-5-54. [PubMed: 17825096]
20. Roth AJ, Rosenfeld B, Kornblith AB, Gibson C, Scher HI, Curley-Smart T, et al. (2003). The memorial anxiety scale for prostate cancer: validation of a new scale to measure anxiety in men with with prostate cancer. *Cancer*, 97(11), 2910–2918, doi: 10.1002/cncr.11386. [PubMed: 12767107]
21. Roth A, Nelson CJ, Rosenfeld B, Warshowski A, O'Shea N, Scher H, et al. (2006). Assessing anxiety in men with prostate cancer: further data on the reliability and validity of the Memorial Anxiety Scale for Prostate Cancer (MAX-PC). *Psychosomatics*, 47(4), 340–347, doi:10.1176/appi.psy.47.4.340. [PubMed: 16844894]
22. Cronbach LJ (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334.
23. Streiner DL, & Norman GR (1995). *Health measurement scales : a practical guide to their development and use*. Oxford; New York: Oxford University Press.
24. Nunnally J (1978). *Psychometric methods*. New York: McGraw-Hill.
25. Samejima F (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika monograph supplement*.
26. Samejima F (1997). Graded response model In *Handbook of modern item response theory* (pp. 85–100): Springer.
27. Steiger JH, & Lind J (1980). Statistically-based tests for the number of common factors. Paper presented at the Paper presented at the Annual Spring Meeting of the Psychometric Society, Iowa City, IA
28. Browne MW, Cudeck R, Bollen KA, & Long JS (1993). *Alternative ways of assessing model fit* (Vol. 154). Newbury Park, CA: Sage Publications.
29. Orlando M, & Thissen D (2000). Likelihood-based item-fit indices for dichotomous item response theory models. *Applied Psychological Measurement*, 24(1), 50–64.

30. Orlando M, & Thissen D (2003). Further investigation of the performance of S-X2: An item fit index for use with dichotomous item response theory models. *Applied Psychological Measurement*, 27(4), 289–298.
31. Kang T, & Chen TT (2011). Performance of the generalized S-X2 item fit index for the graded response model. *Asia Pacific Education Review*, 12(1), 89–96, doi:10.1007/s12564-010-9082-4.
32. Chen W-H, & Thissen D (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, 22(3), 265–289.
33. Cai L, Yang JS, & Hansen M (2011). Generalized full-information item bifactor analysis. *Psychological methods*, 16(3), 221–248. [PubMed: 21534682]
34. Ten Berge JM, & So an G (2004). The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika*, 69(4), 613–625.
35. Bentler PM (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika*, 74(1), 137–143. [PubMed: 20161430]
36. Reise SP, Moore TM, & Haviland MG (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of personality assessment*, 92(6), 544–559. [PubMed: 20954056]
37. Benjamini Y, & Hochberg Y (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289–300.
38. Williams VS, Jones LV, & Tukey JW (1999). Controlling error in multiple comparisons, with examples from state-to-state differences in educational achievement. *Journal of Educational and Behavioral Statistics*, 24(1), 42–69.
39. Watt T, Groenvold M, Hegedüs L, Bonnema SJ, Rasmussen ÅK, Feldt-Rasmussen U, et al. (2014). Few items in the thyroid-related quality of life instrument ThyPRO exhibited differential item functioning. *Quality of Life Research*, 23(1), 327–338. [PubMed: 23812962]
40. Dancey C, & Reidy J (2004). *Statistics without Maths for Psychology: Using SPSS for Windows* (3rd edition). Harlow, England: Prentice Hall.
41. Yost KJ, Eton DT, Garcia SF, & Cella D (2011). Minimally important differences were estimated for six Patient-Reported Outcomes Measurement Information System-Cancer scales in advanced-stage cancer patients. *J Clin Epidemiol*, 64(5), 507–516, doi: 10.1016/j.jclinepi.2010.11.018. [PubMed: 21447427]
42. Emons WH, Sijtsma K, & Meijer RR (2007). On the consistency of individual classification using short scales. *Psychological methods*, 12(1), 105–120, doi:10.1037/1082-989x.12.1.105. [PubMed: 17402814]
43. Kruyen PM, Emons WH, & Sijtsma K (2013). Assessing individual change using short tests and questionnaires. *Applied Psychological Measurement*, doi: 10.1177/0146621613510061.
44. Heene M, Bollmann S, & Bühner M (2014). Much ado about nothing, or much to do about something? Effects of scale shortening on criterion validity and mean differences. *Journal of Individual Differences*, 35(4), 245–249, doi:10.1027/1614-0001/a000146.
45. Hampson LA, Cowan JE, Zhao S, Carroll PR, & Cooperberg MR (2015). Impact of Age on Quality-of-life Outcomes After Treatment for Localized Prostate Cancer. *Eur Urol*, doi:10.1016/j.eururo.2015.01.008.
46. Penson DF, Stoddard ML, Pasta DJ, Lubeck DP, Flanders SC, & Litwin MS (2001). The association between socioeconomic status, health insurance coverage, and quality of life in men with prostate cancer. *J Clin Epidemiol*, 54(4), 350–358. [PubMed: 11297885]
47. Eton DT, Lepore SJ, & Helgeson VS (2001). Early quality of life in patients with localized prostate carcinoma: an examination of treatment-related, demographic, and psychosocial factors. *Cancer*, 92(6), 1451–1459. [PubMed: 11745222]
48. Given B, Given C, Azzouz F, & Stommel M (2001). Physical functioning of elderly cancer patients prior to diagnosis and following initial treatment. *Nurs Res*, 50(4), 222–232. [PubMed: 11480531]
49. Jayadevappa R, Johnson JC, Chhatre S, Wein AJ, & Malkowicz SB (2007). Ethnic variation in return to baseline values of patient-reported outcomes in older prostate cancer patients. *Cancer*, 109(11), 2229–2238, doi: 10.1002/cncr.22675. [PubMed: 17443664]

50. Jayadevappa R, Chhatre S, Wein AJ, & Malkowicz SB (2009). Predictors of patient reported outcomes and cost of care in younger men with newly diagnosed prostate cancer. *Prostate*, 69(10), 1067–1076, doi: 10.1002/pros.20955. [PubMed: 19343738]
51. Brassell SA, Elsamamoudi SI, Cullen J, Williams ME, & McLeod DG (2013). Health-related quality of life for men with prostate cancer--an evaluation of outcomes 12–24 months after treatment. *Urol Oncol*, 31(8), 1504–1510, doi:10.1016/j.urolonc.2012.04.008. [PubMed: 22608541]
52. Rothrock NE, Hays RD, Spritzer K, Yount SE, Riley W, & Cella D (2010). Relative to the general US population, chronic diseases are associated with poorer health-related quality of life as measured by the Patient-Reported Outcomes Measurement Information System (PROMIS). *J Clin Epidemiol*, 63(11), 1195–1204, doi:10.1016/j.jclinepi.2010.04.012. [PubMed: 20688471]
53. Litwin MS, Hays RD, Fink A, Ganz PA, Leake B, Leach GE, et al. (1995). Quality-of-life outcomes in men treated for localized prostate cancer. *Jama*, 273(2), 129–135. [PubMed: 7799493]
54. Bjorner JB, Rose M, Gandek B, Stone AA, Junghaenel DU, & Ware JE, Jr. (2014). Method of administration of PROMIS scales did not significantly impact score level, reliability, or validity. *J Clin Epidemiol*, 67(1), 108–113, doi:10.1016/j.jclinepi.2013.07.016. [PubMed: 24262772]
55. PROMIS Health Organization and PROMIS Cooperative Group (2011). PROMIS instrument-level statistics including gender, education level, age bracket, clinical, and levels of self-rated general health subgroups. <http://www.nihpromis.org/science/validitystudies>.
56. Lubeck DP, Kim H, Grossfeld G, Ray P, Penson DF, Flanders SC, et al. (2001). Health related quality of life differences between black and white men with prostate cancer: data from the cancer of the prostate strategic urologic research endeavor. *J Urol*, 166(6), 2281–2285. [PubMed: 11696752]
57. Grandner MA, Martin JL, Patel NP, Jackson NJ, Gehrman PR, Pien G, et al. (2012). Age and sleep disturbances among American men and women: data from the U.S. Behavioral Risk Factor Surveillance System. *Sleep*, 35(3), 395–406, doi:10.5665/sleep.1704. [PubMed: 22379246]
58. Piccolo RS, Yang M, Bliwise DL, Yaggi HK, & Araujo AB (2013). Racial and socioeconomic disparities in sleep and chronic disease: results of a longitudinal investigation. *Ethn Dis*, 23(4), 499–507. [PubMed: 24392615]

Table 1.

Baseline Patient Characteristics of the 12-month Sample

Characteristic	N=778 Percent (%)
Age at baseline survey	
<65 years	47.3
65 years	52.7
Race	
White	71.9
Black	25.7
Asian or Pacific Islander	0.3
American Indian or Alaskan Native	1.9
Ethnicity	
Hispanic	1.2
Non-Hispanic	98.6
Highest level of education completed	
8th grade or less	3.0
Some high school	7.6
High school graduate	21.5
Some college	28.9
College graduate	39.1
Marital status	
Married	81.2
Divorced	9.1
Widowed	4.1
Never married	3.7
Separated	1.8
Employment status	
Employed full time	33.4
Employed part time	68.7
Unemployed	3.3
Retired	46.9
Disabled and not working	57.6
Smoker status	
Current	11.6
Former	51.5
Never	36.9
Number of comorbid conditions	
0	4.0
1	13.9
2	18.5
3	63.6
Number of mental health comorbid condition(s) that limit(s) usual or daily activities ^a	

Characteristic	N=778 Percent (%)
0	98.7
1	1.3
Number of physical health/other comorbid condition(s) ^b	
0	4.4
1	14.5
2	21.1
3	60.0
Number of physical health/other comorbid condition(s) that limit(s) usual or daily activities ^b	
0	83.4
1	8.2
2	4.8
3	3.6
Treatment(s)	
Radiation ^c + brachytherapy + hormone therapy	0.1
Radiation ^c + hormone therapy	1.5
Radiation ^c + brachytherapy	0.8
Brachytherapy + hormone therapy	0.5
Prostatectomy + radiation	0.3
Prostatectomy + hormone therapy	0.1
Active surveillance	33.8
Prostatectomy	33.9
Radiation ^c	12.1
Brachytherapy	9.4
Hormone therapy	5.3
Other ^d	2.2

Notes:

^aMental health comorbid conditions include Alzheimer's disease, depression, and anxiety.

^bPhysical health/other comorbid conditions include arthritis; back pain; asthma; chronic obstructive pulmonary disease, emphysema, or chronic bronchitis; diabetes; HIV/AIDS; weak or failing kidneys with or without dialysis; osteoporosis, anemia or other blood condition; lupus, rheumatoid arthritis, or other inflammatory condition; another cancer that is not prostate, basal cell skin, or squamous cell skin cancer; a liver condition; pancreatitis; inflammatory bowel disease; stomach ulcer, duodenal ulcer or peptic ulcer; high blood pressure, congestive heart failure; angina; heart attack/acute myocardial infarction; arrhythmia; peripheral artery disease; blood clot in legs or lungs; high cholesterol; and stroke.

^cRadiation includes external beam radiation therapy and proton therapy.

^dOther treatments include cryotherapy, high-intensity focused ultrasound (HIFU), and CyberKnife.

Table 2.

Mean (SD) PROMIS T-Scores by Assessment

Measure	Mean T-score (SD) ^a		
	Baseline (N=333)	3-month (N=411)	12-month (N=778)
Anxiety	47.8 (9.2)	47.1 (8.9)	45.5 (8.5)
Depression	47.1 (8.5)	46.8 (9.0)	46.0 (8.4)
Fatigue	46.2 (10.4)	47.3 (10.1)	46.2 (9.7)
Pain Interference	47.7 (9.1)	47.9 (8.9)	47.7 (8.5)
Sleep Disturbance	49.5 (5.4)	50.1 (6.2)	49.7 (6.2)
Physical Function	50.4 (9.1)	49.4 (8.8)	49.6 (8.5)

Notes: SD=standard deviation.

^aT-scores (mean 50, SD 10) are presented for each PROMIS domain; higher PROMIS T-scores reflect a greater level of the construct measured.

Table 3.

Reliability and Validity of PROMIS Measures at 12-month Assessment (N=778)

Measure ^a	# Items	Reliability	Structural Validity		Convergent Validity ^c	Discriminant Validity ^c
			ECV ^b	Factor Loadings ^b		
Anxiety	5	0.90	0.97	0.90 – 0.96	0.44 (MAX-PC) -0.59 (MCS) -0.60 (MH)	-0.20 (PCS) -- --
Depression	5	0.91	0.98	0.85–0.96	0.41 (MAX-PC) -0.64 (MCS) -0.64 (MH)	-0.22 (PCS) -- --
Fatigue	5	0.94	0.99	0.90–0.96	-0.50 (MCS) -0.60 (VT)	-- --
Pain Interference	5	0.96	0.99	0.95 – 0.98	-0.66 (BP)	--
Sleep Disturbance	4	0.86	0.92	0.77 – 0.92	--	--
Physical Function	6	0.94	-- ^d	0.91 – 0.93	0.73 (PCS) 0.77 (PF)	-0.31 (MAX-PC) 0.21 (MCS)

Notes: ECV=Expected Common Variance; MAX-PC=Memorial Anxiety Scale for Prostate Cancer; MCS=Mental Component Summary from the SF-12v2; PCS=Physical Component Summary from the SF-12v2; MH=Mental Health subscale from the SF-12v2; BP=Bodily Pain subscale from the SF-12v2; VT=Vitality subscale from the SF-12v2; PF=Physical Functioning subscale from the SF-12v2.

-- indicates the analysis was not performed.

^aHigher PROMIS T-scores reflect a greater level of the construct measured. Higher MAX-PC scores indicate greater anxiety in prostate cancer. Higher SF-12v2 scores reflect better health.

^bECV>0.90 and factor loadings >0.70 support unidimensionality.

^cSpearman’s rank correlation coefficients between each PROMIS T-score (mean 50, SD 10) and legacy measure score are statistically significant ($p<0001$).

^dUnidimensionality of the PROMIS physical function measure was assessed using a one-factor confirmatory factor model.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4.

Known-groups Validity by Age and Race

PROMIS Measure	Mean (SD) T-scores ^a							
	Age (Years)				Race			
	Baseline Assessment (N=333)							
	<65 (n=169)	65 (n=164)	t ^b	Hypothesis Source ^c	NHW (n=233)	Black (n=87)	Hypothesis Source ^c	t ^b
Anxiety	50.0 (9.8)	45.4 (8.0)	4.74	[45,46,55]	47.3 (8.2)	48.7 (11.0)	[50,51,56]	1.08
Depression	48.9 (9.0)	45.2 (7.6)	4.04	[45,46,55]	46.5 (7.5)	48.1 (10.2)	[50,51,56]	1.29
Fatigue	47.8 (10.8)	44.6 (9.8)	2.83	[55]	45.2 (9.9)	48.7 (11.2)	[50,51,56]	2.65
Pain Interference	48.5 (9.7)	46.9 (8.4)	1.62	[55]	46.5 (8.2)	50.2 (10.5)	[50,51,56]	2.95
Sleep Disturbance	50.2 (5.2)	48.8 (5.5)	2.44	[57]	49.8 (5.3)	49.0 (5.7)	[58]	1.20
Physical Function	49.9 (9.7)	50.8 (8.5)	0.93	[55,48]	51.1 (8.2)	48.4(10.8)	[50,51,56]	2.12
3-month Assessment (N=411)								
	<65 (n=199)	65 (n=212)	t ^b	Hypothesis Source ^c	NHW (n=290)	Black (n=107)	Hypothesis Source ^c	t ^b
Anxiety	48.5 (9.6)	45.9 (8.0)	3.00	[46,47,55]	46.1 (7.9)	50.1 (10.7)	[50,51,56]	3.58
Depression	48.7 (9.6)	44.9 (8.0)	4.31	[46,47,55]	45.9 (8.3)	49.0 (10.0)	[50,51,56]	2.91
Fatigue	48.9 (10.5)	45.9 (9.6)	3.07	[55]	46.8 (9.7)	48.8 (10.6)	[51,56]	1.75
Pain Interference	49.3 (9.7)	46.6 (7.9)	3.06	[55]	47.0 (8.2)	50.3 (10.1)	[51,56]	2.99
Sleep Disturbance	51.4 (5.7)	48.8 (6.5)	4.28	[57]	50.1 (6.1)	50.2 (6.7)	[58]	0.06
Physical Function	49.1 (9.5)	49.6 (8.2)	0.58	[55]	50.2 (8.1)	47.4 (10.1)	[51,56]	2.60
12-month Assessment (N=778)								
	<65 (n=368)	65 (n=410)	t ^b	Hypothesis Source ^c	NHW (n=553)	Black (n=198)	Hypothesis Source ^c	t ^b
Anxiety	46.7 (9.2)	44.5 (7.7)	3.62	[55]	45.2 (7.9)	46.5 (9.8)	[49,51,56]	1.72
Depression	47.2 (9.0)	44.9 (7.7)	3.88	[55]	45.7 (7.8)	46.7 (9.6)	[49,51,56]	1.33
Fatigue	46.9 (10.5)	45.6 (8.9)	1.91	[55]	46.0 (9.4)	46.6 (10.2)	[49,51,56]	0.81
Pain Interference	48.1 (9.2)	47.2 (7.9)	1.54	[55]	47.1 (7.9)	48.8 (9.5)	[49,51,56]	2.26
Sleep Disturbance	50.3 (6.4)	49.3 (6.0)	2.22	[57]	49.8 (6.0)	49.4 (6.5)	[58]	0.69
Physical Function	49.3 (8.9)	49.8 (8.1)	0.68	[55]	50.5 (8.0)	47.1 (9.1)	[49,51,56]	4.73

Notes: SD=standard deviation; NHW=non-Hispanic White.

^aT-scores (mean 50, SD 10) are presented for each PROMIS measure; higher T-scores reflect a greater level of the construct measured.

^bAbsolute value of t-statistic: |t|>2 indicates statistically significant difference between groups at the 0.05 level.

^cBased on a review of the literature summarized in the Discussion section, we hypothesized that younger men (<65 years) would have worse HRQoL versus men ≥ 65 years (except for physical function), and Black men would have worse HRQoL compared to NHW men.

Table 5.

Known-groups Validity by Comorbidities

PROMIS Measure	Mean (SD) T-scores ^a							
	# of Mental Health Comorbidities ^b				# of Limiting Physical Health/Other Comorbidities ^b			
	Baseline Assessment (N=333)							
	0 (n=271)	1 (n=62)	t ^c	Hypothesis Source ^d	0 (n=280)	1 (n=53)	Hypothesis Source ^d	t ^c
Anxiety	46.7 (8.2)	52.3 (11.6)	3.59	[46,55]	46.5 (8.4)	54.2 (10.7)	[46,55]	4.93
Depression	45.8 (7.5)	52.4 (10.4)	4.66	[46,55]	45.7 (7.6)	54.1 (9.4)	[46,55]	6.13
Fatigue	44.7 (9.5)	52.6 (11.7)	4.92	[46,55]	44.7 (9.7)	54.3 (10.2)	[46,55]	6.57
Pain Interference	46.6 (8.3)	52.3 (11.0)	3.84	[46,55]	45.9 (7.3)	57.3 (11.5)	[46,55]	6.99
Sleep Disturbance	49.2 (5.6)	51.0 (4.7)	2.34	[57]	49.1 (5.4)	51.6 (5.0)	[57]	3.12
Physical Function	51.4 (8.6)	45.6 (9.8)	4.70	[46,55]	52.3 (7.5)	40.0 (9.8)	[46,55]	8.68
3-month Assessment (N=411)								
	0 (n=343)	1 (n=68)	t ^c	Hypothesis Source ^d	0 (n=343)	1 (n=68)	Hypothesis Source ^d	t ^c
Anxiety	45.9 (7.8)	53.5 (11.2)	5.39	[46,55]	45.9 (8.1)	53.6 (9.8)	[46,55]	6.09
Depression	45.5 (7.8)	53.3 (11.3)	5.47	[46,55]	45.6 (8.3)	52.7 (9.9)	[46,55]	5.56
Fatigue	46.1 (9.5)	53.4 (10.9)	5.59	[46,55]	45.9 (9.3)	54.5 (11.0)	[46,55]	6.71
Pain Interference	47.1 (8.2)	52.2 (10.9)	3.66	[46,55]	46.1 (7.4)	57.0 (10.1)	[46,55]	8.45
Sleep Disturbance	49.7 (6.0)	52.1 (7.0)	2.89	[57]	49.8 (6.1)	51.5 (6.5)	[57]	2.04
Physical Function	50.3 (8.5)	44.7 (9.3)	4.93	[46,55]	51.3 (7.6)	39.5 (8.1)	[46,55]	11.63
12-month Assessment (N=778)								
	0 (n=648)	1 (n=130)	t ^c	Hypothesis Source ^d	0 (n=649)	1 (n=129)	Hypothesis Source ^d	t ^c
Anxiety	44.4 (7.6)	50.8 (10.6)	6.53	[46,55]	44.5 (7.6)	50.6 (10.9)	[46,55]	6.14
Depression	45.0 (7.6)	50.7 (10.2)	6.02	[46,55]	44.9 (7.5)	51.1 (10.5)	[46,55]	6.36
Fatigue	45.1 (9.1)	51.7 (10.8)	6.53	[46,55]	45.0 (9.1)	52.1 (10.7)	[46,55]	7.08
Pain Interference	46.9 (8.0)	51.2 (10.3)	4.47	[46,55]	46.1 (7.2)	55.2 (10.4)	[46,55]	9.40
Sleep Disturbance	49.4 (6.2)	51.3 (6.3)	3.24	[55]	49.5 (6.2)	50.8 (6.2)	[57]	2.18
Physical Function	50.2 (8.0)	46.3 (9.8)	4.29	[46,57]	51.3 (7.3)	41.0 (9.0)	[46,55]	12.19

Notes: SD=standard deviation.

^aT-scores (mean 50, SD 10) are presented for each PROMIS measure; higher T-scores reflect a greater level of the construct measured.

^bMental health and physical health/other comorbid conditions are listed under Table 1.

^cAbsolute value of *t*-statistic: |*t*|>2 indicates statistically significant difference between groups at the 0.05 level.

^dBased on a review of the literature summarized in the Discussion section, we hypothesized that men with 1 [limiting] comorbidity would have worse HRQoL versus men with no [limiting] comorbidities.