CrossMark

Adapting summary scores for the PROMIS-29 v2.0 for use among older adults with multiple chronic conditions

Wenjing Huang[1] · Adam J. Rose[2,3] · Elizabeth Bayliss[4,5] · Lesley Baseman[6] · Emily Butcher[1] · Rosa-Elena Garcia[1] · Maria Orlando Edelen[2]

Abstract

Purpose The patient-reported outcomes measurement information system 29-item profile (PROMIS-29 v2.0) is a widely used health-related quality of life (HRQoL) measure. Summary scores for physical and mental HRQoL have recently been developed for the PROMIS-29 using a general population. Our purpose was to adapt these summary scores to a population of older adults with multiple chronic conditions.

Methods We collected the PROMIS-29 v2.0 for 1359 primary care patients age 65+ with at least 2 of 13 chronic conditions. PROMIS-29 has 7 domains, plus a single-item pain intensity scale. We used exploratory factor analysis (EFA), followed by confirmatory factor analysis (CFA), to examine the number of factors that best captured these eight scores. We used previous results from a recent study by Hays et al. (Qual Life Res 27:1885–1891, 2018) to standardize scoring coefficients, normed to the general population.

Results The mean age was 80.7, and 67% of participants were age 80 or older. Our results indicated a 2-factor solution, with these factors representing physical and mental HRQoL, respectively. We call these factors the physical health score (PHS) and the mental health score (MHS). We normed these summary scores to the general US population. The mean MHS for our population of was 50.1, similar to the US population, while the mean PHS was 42.2, almost a full standard deviation below the US population.

Conclusions We describe the adaptation of physical and mental health summary scores of the PROMIS-29 for use with a population of older adults with multiple chronic conditions.

Keywords Physical health · Mental health · Quality of life · PROMIS · PROMIS-29 · Comorbidity · Geriatrics

Introduction

Health-related quality of life (HRQoL) is generally understood as the ways in which a medical condition and/or therapy affect a patient's well-being [1, 2]. HRQoL is both subjective and multidimensional, taking into account multiple aspects of well-being, including physical and occupational function, psychological state, social interaction, and somatic sensation. When compared with more objective measures such as survival, HRQoL captures a different, and potentially complementary, dimension of the impact of illness and therapy on a patient's life. This is especially true with older adults with a considerable illness burden and limited life expectancy.

Patient-reported outcome measures (PROs), which are often based upon HRQoL, may improve clinicians' ability to monitor symptoms and treatment effectiveness, and may

✉ Wenjing Huang
  whuang@rand.org

1 RAND Corporation, 1776 Main St., Santa Monica,
   CA 90407, USA

2 RAND Corporation, Boston, MA, USA

3 Section of General Internal Medicine, Boston University
   School of Medicine, Boston, MA, USA

4 Institute for Health Research, Kaiser Permanente Colorado,
   Denver, CO, USA

5 Department of Family Medicine, University of Colorado
   School of Medicine, Aurora, CO, USA

6 RAND Corporation, Arlington, VA, USA

engage patients in their care through better patient–physician communication [3, 4]. The most widely used HRQoL instruments to date have included the Veterans RAND 36-Item Survey Instrument or VR-36 [5] which was developed in the early 1990s and has an established application of two summary scores. The patient-reported outcomes measurement information system 29-Item profile measure (PROMIS-29 V2.0) is a relatively new HRQoL instrument (http://www.healthmeasures.net/explore-measurement-systems/promis), which was developed using modern measurement theory and was calibrated and scored based on contemporary samples [6].

PROMIS profile instruments are a collection of short forms containing a fixed number of items from the same 7 PROMIS domains (physical function, anxiety, depression, fatigue, sleep disturbance, ability to participate in social roles and activities, and pain interference) plus a single item on pain intensity. Depending on how many questions are asked in each domain, there are three different lengths of PROMIS Profile forms: PROMIS-29, PROMIS-43, and PROMIS-57. The PROMIS-29 is the shortest, with 4 questions asked for each of 7 domains, plus the single pain intensity item. Each short form is scored separately, yielding a total of 7 domain scores. There are two versions of PROMIS-29, and the one we used is the second version, PROMIS-29 v2.0. The PROMIS Profile instruments are intended for adults (ages 18+), and are designed to be focused on general HRQoL, as opposed to disease-specific measures intended to be used with a population who all have a certain medical condition (e.g., heart failure). All PROMIS instruments assess HRQoL over the past 7 days, with the exception of two domains (physical function and ability to participate in social roles and activities), which do not specify a timeframe.

Profile measures such as the SF-36 provide a wealth of information; for example, the SF-36 has 8 scales. However, it has also been important to summarize HRQoL with fewer scores [7]. In the case of SF-36 and related instruments, this has been done with the summary scores, called the physical component score (PCS) and the mental component score (MCS), which represent two underlying factors based on factor analyses of SF-36 scale scores [8, 9]. While all 8 scales of the SF-36 contribute to both PCS and MCS to some extent, some of the 8 scales are more strongly represented within one or the other. With PCS, physical health is defined mostly by measures of physical functioning, pain, and role limitations due to physical health problems. With MCS, mental health is primarily measured by emotional well-being and role limitations caused by emotional problems [8, 9]. Other scales, such as general health perceptions, vitality, and social functioning, contribute relatively strongly to both physical and mental health. One important feature for PCS and MCS is that in their construction, they were forced to have zero

correlation with each other, which may conflict with the reality of how most people experience HRQoL. This, in turn, may contribute to inconsistency between SF-36 scales scores and summary scores (see more in-depth discussion in [8]). In a recent study, Hays and colleagues used a sample that represented the general US population to combine the 7 PROMIS-29 v2.0 domain scores and the single pain intensity item into two composite scores, physical and mental health summary scores [10]. The new summary scores were constructed allowing physical and mental health to be inter-correlated, rather than forcing a zero correlation between them. Here, we adapted these scores to a population much older and with a much higher illness burden than the general US population. Our sample was collected prospectively, and consisted of patients age 65 + with at least 2 of 13 specified chronic conditions.

Our objective was to assess the utility of these adapted scores as a tool to characterize HRQoL in this population, as an extension of their original derivation across the entire population of US adults. We address this task through two companion manuscripts. In this paper, we demonstrate how we created similar summary scores with the PROMIS-29 v2.0 for the population of older adults with multiple chronic conditions (MCCs). In a companion paper [11], we establish construct validity for these summary scores by comparing them against summary scores of the VR-36, a modification of SF-36 which was developed in response to findings from the Medical Outcomes Study [5, 12].

Methods

Participants

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. 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 (which demonstrates that they were not permanently residing at a skilled nursing facility), 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 13 conditions include arthritis, cancer, chronic lung disease, congestive heart failure, depression, diabetes, hypertension, inflammatory bowel disease, ischemic heart disease, osteoporosis, other heart problems, sciatica, and stroke. These conditions were selected because they are assessed as part of the Medicare Health Outcomes Survey (http://www.hosonline.org/). Each condition was defined using a list of International Classification of Diseases – Clinical Modification, Version 10 (ICD-10) codes; these codes and more information on how we chose

the 13 conditions can be found in "Appendix 1". We over-sampled participants age 80 and older to ensure an adequate representation of the oldest-old. We obtained approvals from the institutional review boards of both KPCO and the RAND Corporation. The work described in this manuscript was part of a broader program of work, which has been described in full elsewhere [13].

Data collection

The survey included the PROMIS-29 v2.0 questionnaire. Survey data were collected between 11/11/2016 and 3/6/2017. A total of 4991 patients were deemed to be eligible based on the criteria described above. Participants were given 2 weeks to opt out, after which time they were presumed to be eligible for initial contact. KPCO also performed one final check to remove any deceased respondents from the list. A total of 283 eligible patients opted out, 677 were excluded due to an invalid mailing address or email address, and 282 had died. The final sample file therefore contained 3749 potential participants who had valid addresses, were not deceased, and had not opted out.

We randomly assigned participants to different survey modes, with the same survey instrument administered across all modes. Initially, 2764 participants were assigned to the web survey, 376 to the mail survey, and 372 to the phone survey. The number assigned to the mail and phone surveys was based on previous experience by the RAND Survey Research Group with similar surveys, as well as the final number of responses we wanted to gather. We assigned a much larger number to the web survey, based on conservative assumption about the response rate.

Those assigned to the web survey were sent an email invitation, followed by weekly email reminders for 3 weeks if they had not completed the survey. Participants who only partially completed the web-based survey were sent reminders asking them to finish the survey they had started. Those invited to complete the survey by mail received an initial packet that included a personalized invitation letter, information sheet, mail survey, and prepaid reply envelope. Materials identical to the first packet were sent to non-responders 4 and 6 weeks after the first survey mailing. For those invited to complete the survey by phone, telephone center staff contacted respondents an average of five times. We anticipated that some of the participants would require help from a proxy, such as a family member, to complete the survey, and therefore proxy responses were allowed for all modes.

Analytic approach

Scoring the PROMIS-29

We scored the 7 domains of the PROMIS-29 using an online scoring tool intended for this purpose (https://www.assessmentcenter.net/ac_scoringservice). More extensive information about scoring is available at http://www.healthmeasures.net. These scores were calculated based on the principles of item response theory (IRT) [14], and hence they are called IRT scale scores. Two summary scores were developed to measure the physical and mental domains of HRQoL based on the 7 IRT scales scores and the single pain intensity item. The process for developing PROMIS-29 summary scores for the general population has been described by Hays et al [10]; below, we describe how we adapted it for use in this population.

Initial exploration of our sample

We first estimated the reliability of the 4-item short forms for each of the 7 domains by calculating Cronbach's Alpha [15]. Together with the single pain intensity item, we examined the Pearson product-moment correlations among these 8 scores.

We then proceeded to exploratory factor analysis (EFA), using the PROMAX oblique rotation and maximum likelihood (ML) estimation method on these 8 scores. Eigenvalues [16], scree plot [17], and proportion of variance explained were used to guide decisions about the number of factors to extract from the 8 scores. We also tested alternatives of combining some of the scores, including combining the two pain scores (pain intensity and pain interference), and combining depression and anxiety into a single score that represents emotional distress. Based on the rotated factor loading pattern, we examined the interpretability of EFA results. When the criteria of the number of factors extracted were satisfied and the results were interpretable, we proceeded to confirmatory factor analysis (CFA) to examine model fit. The CFA model was based on the same loading pattern that we extracted from EFA results.

Adapting the summary scores to our sample

The PROMIS-29 summary scores were originally developed by Dr. Hays and colleagues using a nationally representative sample recruited from the Opinions4Good (Op4G) panel (http://op4g.com/our-panel/). Op4G participants all completed the PROMIS-29 profile measure, v. 2.0, on the internet, and the sample represents the US adult population in terms of gender, region, race/ethnicity, and education. The average age of Op4G sample is 46 with a range of 18–88 years old.

We compared results from our MCC sample with those from Hays and colleagues' Op4G sample to ensure consistent factor loading patters across the two samples. The standardized scoring coefficients were derived from the Op4G sample and were applied to our MCC sample to score our participants. This process provides the two

summary scores, the PROMIS-29 physical health summary score (PHS) and mental health summary score (MHS). We used SAS, version 9.4 [18] for EFA and correlation analyses. All CFA were conducted using Mplus version 8 [19].

Results

Survey response

Out of a sample of 3749 people who received a survey, a total of 1359 participants responded, for a total response rate of 36%. Response rates varied by survey modality, and were 25% for the web mode, 51% for the mail mode, and 48% for the phone mode. The web mode was noteworthy in that 37% of those who began to respond to the web-based survey ultimately abandoned the effort before completing enough of the survey to support including them in the HRQoL analyses, which we defined by having enough responses to calculate at least one PROMIS-29 scale score. With the phone mode, the interviewer made sure the respondent completed all items; with the mail mode, it was also uncommon for the respondent to return a survey without enough information to support analysis.

Characteristics of the study participants are shown in Table 1. The mean age of the MCC sample was 81 (SD 6.9), ranging from 66 to 89. A majority of participants (56%) were age 80 or older, and a majority (89%) were White/non-Hispanic. By definition, all had at least 2 of 13 chronic conditions; 35% had exactly 2, 31% 3, 18% 4, and 16% 5 or more. Prevalence of some representative chronic conditions was relatively high; for example, 38% had chronic lung disease, and 31% had diabetes.

"Appendix 2" shows a comparison of responders with non-responders based on data derived from the KPCO electronic medical record, as well as a regression analysis to predict response based on these characteristics. Responders and non-responders were generally comparable in terms of age, burden of chronic conditions, and healthcare utilization. We did not find mode effects between the web, mail, and telephone surveys, nor did we find important differences between responders and non-responders in terms of age, burden of chronic conditions, and healthcare utilization. Women were somewhat less likely to respond compared to men (adjusted odds ratio (AOR) 0.77, $p = 0.002$), and compared to White non-Hispanics, Hispanics and non-Whites were less likely to respond (AOR 0.47 and 0.58, respectively; $p < 0.001$). Details can also be found online at https://www.rand.org/pubs/research_reports/RR2176.html.

Table 1 Respondent demographic and clinical characteristics

	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
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
Other heart problems	35
Sciatica	5
Stroke	6

Percentages given except where otherwise noted. ($n = 1,359$)

PROMIS-29 scores

The means, standard deviations, minimum, and maximum of the 8 scores are presented in Table 2. The IRT scale scores we obtained from the online scoring service follow the same PROMIS T-score metric with a mean of 50 and standard deviation of 10. The single pain intensity item is scaled from 0 to 10, with 0 being no pain and 10 being worst imaginable pain.

We measured internal consistency reliabilities of the 7 PROMIS-29 health domain scales that are based on the

Table 2 Items used for PROMIS-29 summary score calculations

Domains	PROMIS-29 items	Mean	SD	Min	Max
1. Physical function	B1–B4: physical function	41.24	9.02	22.90	56.90
2. Anxiety	C1–C4: anxiety	49.71	8.60	40.30	81.40
3. Depression	D1–D4: depression	49.44	8.35	41.00	79.30
4. Fatigue	E1–E4: fatigue	50.85	9.04	33.70	75.80
5. Sleep	F1–F4: sleep disturbance	47.29	8.84	32.00	73.30
6. Social	G1–G4: ability to participate in social roles & activities	50.19	9.30	27.50	64.20
7. Pain interference	H1–H4: pain interference	53.33	9.24	41.60	75.60
8. Pain intensity	H5: PROMIS-29 Pain Intensity Score	3.11	2.40	0.00	10.00

$N = 1296$

At least one item in each short form has response. None is missing the intensity item. The first 7 domains are represented by IRT scale scores calculated from response pattern scoring at Assessment Center and are on a T-score metric with means of 50 and standard deviations of 10. The last item (pain intensity) is a single item on a scale of 0–10

Table 3 Pearson product-moment correlations among the PROMIS-29 domain scores

	Physical function	Anxiety	Depression	Fatigue	Sleep	Social	Pain interference	Pain intensity
Physical function	1							
Anxiety	−0.36	1						
Depression	−0.45	0.74	1					
Fatigue	−0.58	0.53	0.58	1				
Sleep	−0.27	0.38	0.37	0.41	1			
Social	0.72	−0.48	−0.57	−0.70	−0.36	1		
Pain interference	−0.55	0.42	0.43	0.56	0.35	−0.57	1	
Pain intensity	−0.46	0.34	0.35	0.47	0.36	−0.45	0.79	1

4-item short forms. Cronbach's Alphas by domain were as follows: physical functioning (0.91), anxiety (0.87), depressive symptoms (0.89), fatigue (0.92), sleep disturbance (0.87), ability to participate in social roles and activities (0.93), and pain interference (0.95). Product-moment correlations among these 7 domain scales and the single pain intensity item are presented in Table 3. The size of these correlations ranges from −0.70 (fatigue and ability to participate in social roles) to 0.79 (pain interference and pain intensity). All correlations are statistically significant with p values < 0.001.

Factor analyses

EFA results indicated a two-factor solution based on an overall evaluation of eigenvalues and scree plot. Kaiser's criterion [16] considers factors with an eigenvalue greater than one as common factors. As is shown in the scree plot (see Fig. 1), two eigenvalues are above one, suggesting two factors be kept. The scree plot suggested solutions with one or three factors. Although previous research [20] suggests a two-factor solution, we reserved the possibility of one-factor
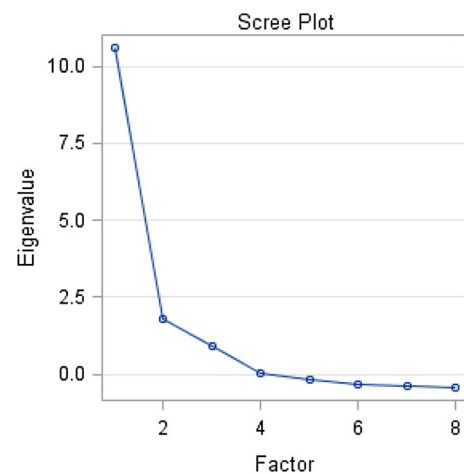

Fig. 1 Scree plot and variance explained

or three-factor solutions to be further explored beyond EFA stage.

We then proceeded to CFA to compare 1-, 2-, or 3-factor solutions. We also explored the possibility of

combining some items given their higher correlations (see Table 3) at the CFA stage. Specifically, we combined the pain intensity and pain interference items (to create a single item "pain," $r = 0.79$), and also the anxiety and depression items ("emotional distress," $r = 0.74$). Our CFA results showed that the 2-factor solution based on these 6 variables (fatigue, sleep disturbance, social ability, physical function, plus the two combined variables—pain and emotional distress) fit the data best. Specifically, the root mean square error of approximation (RMSEA) value for this model was 0.056, with Comparative Fit Index (CFI) of 0.99 and Tucker-Lewis Index (TLI, also called the non-normed fit index or NNFI) of 0.98; the accepted standard is that a RMSEA value smaller than 0.06, and CFI and TLI values above 0.95, are indicators of a relatively good fit between the hypothesized model and the observed data [21]. The reduced 6-indicator set includes four indicators representing original PROMIS domain scores (i.e., fatigue, sleep disturbance, social ability, and physical function), and two new indicators emotional distress, created by averaging the anxiety and depression domain scores, and pain which was the average of the pain intensity scale and the pain interference domain score. All items were standardized to account for differences in their original scales. This result is both consistent with previous research [20] and what was found in Hays et al.'s study [10] from which we derived the scoring coefficients we used here.

We had therefore developed a two-factor model; the two factors were correlated ($r = 0.59$) in our final CFA model. Consistent with the findings in Hays et al [10], we considered the first factor to be physical HRQoL, which we call the physical health score (PHS). The PHS was defined by physical functioning, fatigue, ability to participate in social roles and activities, and the combined pain measure. We considered the second factor to be mental HRQoL and it was defined by all indicators except physical functioning. We call this second factor the mental health score (MHS).

This two-factor solution allowed cross-loadings of fatigue, social ability, and pain on both factors, suggesting that they contribute to both factors (see Table 4 for the factor loading pattern, which is very close to what Hays et al [10] found). Previous research evaluating summary scores from the SF-36 [20] found that fatigue and sleep disturbance contributed equally to both physical and mental quality of life. In our case, however, fatigue contributed more to mental HRQoL (-0.76) than to physical HRQoL (-0.14), while sleep disturbance contributed only to mental HRQoL (-0.51). Two-factor solutions of the SF-36 also showed that ability to participate in social roles and activities was only related to physical health [20]. Our findings showed that it

Table 4 Factor loadings and standardized scoring coefficients from CFA model with 6 indicators

	Factor loadings		Standardized scoring coefficients[b]	
	Factor 1 physical QoL	Factor 2 mental QoL	Factor 1 physical QoL	Factor 2 mental QoL
Physical functioning	1[a]		0.872	− 0.015
Emotional distress		− 0.67	0.003	− 0.257
Fatigue	− 0.14	− 0.76	− 0.009	− 0.351
Sleep		− 0.51	0.002	− 0.139
Social	0.40	0.55	0.113	0.252
Pain	− 0.23	− 0.45	− 0.094	− 0.154

Results representing PROMIS-29 physical and mental QoL sum scores

Take physical QoL summary score as an example. We first combine all the 6 indicators:

PHS_$z = 0.872 \times$ (physical functioning z score) $+ 0.003 \times$ (combined emotional distress z score) $+ (-0.009) \times$ (fatigue z score) $+ 0.002 \times$ (sleep z score) $+ 0.113 \times$ (social z score) $+ (-0.094) \times$ (combined pain z score). PHS is then put on the PROMIS T-score metric as PHS_$t = 50 + 10 \times$ PHS_z

[a]Factor loading of physical functioning on the first factor is set to 1. We used the same constraint as in Hays et al. [10]. Estimated correlation between the two factors is 0.59

[b]All indictors are standardized on a z score metric with means of 0 and SD of 1 before being combined into summary scores with the use of standardized scoring coefficients

was related to both physical and mental health, and in fact that it contributed more strongly to mental HRQoL (0.55) than to physical HRQoL (0.40).

Scoring PHS and MHS

We proceeded to calculate the PHS and MHS scores in our MCC sample, using the standardized factor scoring coefficients that Hays had developed using the Op4G sample. The last two columns in Table 4, and a note below the table, explain how to calculate these summary scores. These scores were set on the PROMIS T-score metric (i.e., normed to the general US population) with a mean of 50 and standard deviation of 10. The mean MHS for the MCC sample was 50.1, similar to the US population, while the mean PHS was 42.2, almost a standard deviation below the US population. The correlation between the two measures of PROMIS-29 HRQoL in our sample was 0.74, implying that the two concepts of physical and mental HRQoL may overlap fairly strongly among the older and sicker population we studied here.

Discussion

Despite the ever-increasing use of PROMIS-29 v2.0, this instrument has not been explicitly evaluated among older patients with multiple chronic conditions, a population that often requires intensive health care and complex care coordination. The need for measuring HRQoL, however, is particularly relevant for this population. This is because given the limited life expectancy of most older adults with multiple chronic conditions, other ways of measuring quality of care (e.g., certain processes of care) or patient outcomes (e.g., long-term survival) may be less relevant.

The utility of PROMIS-29 is enhanced by the use of two summary scores. These two scores serve to distill the information from the seven domains of the PROMIS profile, and the single pain intensity item, into two scores, which represent the physical and mental aspects of HRQoL. Here, we adapted these two scores, which we call PHS and MHS, for use in a population of community-dwelling older adults (age 65+) who had at least two of thirteen pre-specified chronic health conditions. This extends their use beyond the general US population in which they were originally derived [6]. This paper also provides guidance for those who wish to adapt these summary scores for use in other populations, and to calibrate them to allow a comparison between these populations and the general US population, as we did.

Previous efforts relying upon the PROMIS-29 measures have used the seven parent scales, plus the pain intensity item, to compare populations [22]. A reliance on 8 scales to track changes in HRQoL over time, or to compare populations with each other, necessitates multiple comparisons and is not easily or intuitively grasped. In contrast, reliance on the physical and mental aspects of HRQoL concords with long-held understandings of how HRQoL works, and allows the PROMIS-29 to be used much as other older HRQoL instruments have been used. For example, the SF-36 has long featured summary scores, called the physical component score (PCS) and the mental component score (MCS). In one study, the HRQoL burden of low back pain was estimated at -5.5 points on PCS and -2.8 points on MCS [23].

Hays and colleagues' study [10], using a nationally representative sample, showed strong support for some of the psychometric properties of the PROMIS-29 v2.0 summary scores. Their results provide the factor scoring coefficients (see Table 4) for estimating the summary scores that should apply to the general US population. The present study helps open up a similar possibility for the use of the PHS and MHS, and also points the way toward how these scores can be adapted for use in a particular population. In addition to

future validation work that is needed to assess their scoring system in different samples, algorithms have been developed in the PROMIS project to estimate the EQ-5D-3L [24] and the HUI-3 [25] from the PROMIS-29 scales. Preference-based scoring functions can also be estimated directly from the PROMIS-29 [26–28].

We acknowledge several limitations. First, older KPCO members with MCC may be somewhat different from analogous older patients with MCC drawn from the general US population in several ways—not least of which is that older KPCO members are mostly enrolled in a medicare advantage benefit plan. Another limitation of the study is that our sample had limited diversity in terms of race/ethnicity, and relatively few participants lived in a high-poverty zip code.

Despite these limitations, this study has adapted the PHS and MHS, summary scores for the PROMIS-29 v2.0, for use in a population of older adults with a considerable illness burden. It is particularly important to have HRQoL summary scores for this older and sicker population with more chronic conditions, because this is a population that often has intensive health care needs requiring complex care coordination. Future research should use the PHS and MHS summary scores to quantify the impact on HRQoL from specific chronic conditions, as measured using the PROMIS-29.

Compliance with ethical standards

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

Ethical approval Approved by the Human Subjects Research Protection Committee of the RAND Corporation and the Institutional Review Board of Kaiser Permanente Colorado. The authors declare that this study was conducted in accordance with appropriate ethical standards for research, including the Declaration of Helsinki.

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

Appendix

Appendix 1: Conditions used in this study

See Table 5.

Table 5 Layman's terms and ICD-10 codes for chronic conditions

Brief title for chronic condition	Full title—Layman's terms	ICD-10 codes
Arthritis	Arthritis of the hip, knee, hand, or wrist	M16.x, M17.x, M18.x, M19.03X, M19.04x, M19.13x, M19.14x, M19.23x, M19.24x
Cancer	Cancer of the colon, rectum, lung, breast, or prostate	C50.x, C61.x, C34.x, C18x-C20x
Chronic Lung Disease	Emphysema, asthma, or COPD (chronic obstructive pulmonary disease)	I27.8, I27.9, J40.x–J47.x, J60.x–J67.x, J68.4, J70.1, J70.3
Congestive heart failure	Congestive heart failure	I09.81, I11.0, I13.0, I13.2, I25.5, I42.0, I42.5–I42.9, I43.x, I50.x
Depression	Depression	F32.x, F33.x
Diabetes	Diabetes, high blood sugar, or sugar in the urine	E08x-E13x
Hypertension	High blood pressure	I10.x, I11.x–I13.x, I15.x
Inflammatory bowel disease	Crohn's disease, ulcerative colitis, or inflammatory bowel disease	K50.x-K51.x
Ischemic heart disease	Angina pectoris, coronary artery disease, myocardial infarction, or heart attack	I20.x-I25.x
Osteoporosis	Osteoporosis, sometimes called thin or brittle bones	M80.x, M81.x
Other heart problems	Other heart conditions, such as problems with heart valves or the rhythm of your heartbeat	I44.1–I44.3, I45.6, I45.9, I47.x–I49.x, R00.0, R00.1, R00.8, T82.1, Z45.0, Z95.0, A52.0, I05.x–I08.x, I09.1, I09.8, I34.x–I39.x, Q23.0–Q23.3, Z95.2–Z95.4
Sciatica	Sciatica (pain or numbness that travels down your leg to below your knee)	M54.3x, M54.4x
Stroke	Stroke	I60.x, I61.x, I62.x, I63.x, I69.x

Development of code list

To develop an appropriate list of codes for the present study, we started with a set of codes developed by Quan et al., which were based on those originally established by Elixhauser for risk adjustment of chronic conditions [29]. We further modified the Quan set of codes in consultation with clinically trained members of our team and other practicing clinicians at RAND.

We modified the Quan set of codes in two main ways. First, we added codes to capture some conditions that were not included by Quan. The purpose of risk adjustment models such as the Elixhauser/Quan model is to detect chronic conditions that drive hospitalizations, costs, or mortality. Our purpose here, in contrast, was to detect conditions that would have an important effect on HRQoL. We therefore added codes to the Quan set to capture certain conditions that can impact HRQoL but are not a major cause of hospitalization or death, such as osteoarthritis, osteoporosis, and sciatica.

Second, we added codes to extend the spectrum of disease for some conditions that were already part of the Quan model, because the spectrum of disease that we needed to detect was different. Quan was interested in detecting only the most severe manifestations of disease, because these tend to drive morbidity and mortality. For our study, we were also interested in capturing less severe manifestations of disease, because those can impact HRQoL as well. For ischemic heart disease, for example, we added the codes for angina pectoris to the codes that Quan had used to capture more severe manifestations of disease, such as myocardial infarction. We reasoned that, although angina pectoris is less severe, it could still meaningfully impact HRQoL.

Appendix 2: Comparison of survey responders with non-responders based on data derived from the KPCO electronic medical record

See Tables 6 and 7.

Table 6 Bivariate comparisons of survey responders and non-responders based on measured characteristics

	Responders (n = 1359) (%)	Non-responders (n = 2390) (%)	p value
Age groups			0.08
65–69	10	10	
70–74	13	12	
75–79	10	8	
80–84	38	38	
85+	28	32	
Sex			<0.001
Male	48	40	
Female	52	60	
Race/ethnicity			<0.001
White/non-Hispanic	89	83	
Hispanic	4	8	
Non-White/non-Hispanic	4	7	
Missing race/non-Hispanic	3	2	
Percent below poverty in census tract			0.004
0–9.99%	59	56	
10–19.99%	32	30	
20%+	9	13	
Total number of chronic conditions			0.20
2	35	34	
3	31	30	
4	18	18	
5+	16	18	
Presence of a specific chronic condition			
Arthritis	25	21	0.02
Cancer	9	7	0.02
Chronic lung disease	38	35	0.056
Congestive heart failure	16	20	0.001
Depression	23	27	0.008
Diabetes	31	34	0.02
Hypertension	82	85	0.06
Inflammatory bowel disease	1	1	0.80
Ischemic heart disease	29	29	0.54
Osteoporosis	23	28	<0.001
Other Heart Problems	35	33	0.32
Sciatica	5	5	0.90
Stroke	6	8	0.09
Any home health encounters in past 12 months?			0.56
No	54	53	
Yes	46	47	
Number of primary care visits in past 12 months			0.01
0–3	43	48	
4–6	35	30	
7–9	12	12	
10+	10	10	
Number of specialty care visits in past 12 months			<0.001
0–3	56	63	
4–6	22	20	
7–9	11	9	
10+	11	8	

Table 6 (continued)

	Responders (n=1359) (%)	Non-responders (n=2390) (%)	p value
Number of hospitalizations in past 12 Months			<0.001
0	87	81	
1	10	14	
2+	3	5	

Table 7 Odds ratios for responding to the initial survey (n=3749; 1359 responders and 2390 non-responders)

Variable	Adjusted odds ratio	Confidence interval	p value
Survey Mode			<0.001
Mail	Ref	–	
Phone	0.70	0.54–0.92	
Web	0.22	0.18–0.28	
Web-Mail	0.58	0.43–0.78	
Web-Phone	0.60	0.44–0.80	
Age Groups			0.24
65–69	Ref	–	
70–74	1.12	0.83–1.52	
75–79	1.14	0.82–1.58	
80–84	0.96	0.75–1.25	
85+	0.88	0.67–1.16	
Sex			0.002
Male	Ref	–	
Female	0.77	0.66–0.91	
Race/Ethnicity			<0.001
White/Non-Hispanic	Ref	–	
Hispanic	0.47	0.33–0.66	
Non-White/Non-Hispanic	0.58	0.42–0.80	
Missing Race/Non-Hispanic	1.22	0.77–1.96	
Percent Below Poverty in Census Tract			0.04
0–9.99%	Ref	–	
10–19.99%	1.10	0.94–1.30	
20%+	0.80	0.63–1.02	
Total number of chronic conditions			0.76
2	Ref	–	
3	1.11	0.85–1.46	
4	1.26	0.79–2.00	
5+	1.52	0.70–3.30	
Presence of a specific chronic condition			
Arthritis	1.05	0.80–1.37	0.73
Cancer	1.06	0.76–1.47	0.73
Chronic lung disease	1.03	0.80–1.32	0.84
Congestive heart failure	0.68	0.51–0.90	0.007
Depression	0.72	0.55–0.93	0.01
Diabetes	0.81	0.62–1.04	0.098
Hypertension	0.84	0.63–1.11	0.22
Inflammatory bowel disease	0.78	0.41–1.47	0.44
Ischemic heart disease	0.84	0.65–1.09	0.19
Osteoporosis	0.78	0.59–1.02	0.07

Table 7 (continued)

Variable	Adjusted odds ratio	Confidence interval	p value
Other heart problems	1.01	0.78–1.32	0.93
Sciatica	0.90	0.62–1.30	0.57
Stroke	0.89	0.62–1.26	0.50
Any home health encounters in past 12 months?			0.81
No	Ref	–	
Yes	0.98	0.84–1.15	
Number of primary care visits in past 12 months			0.02
0–3	Ref	–	
4–6	1.30	1.1–1.53	
7–9	1.12	0.88–1.43	
10+	1.21	0.93–1.58	
Number of specialty care visits in past 12 months			0.008
0–3	Ref	–	
4–6	1.20	1.01–1.45	
7–9	1.41	1.11–1.81	
10+	1.35	1.05–1.74	
Number of hospitalizations in past 12 months			0.003
0	Ref	–	
1	0.70	0.55–0.89	
2+	0.63	0.42–0.94	

References

1. Cella, D. F. (1995). Measuring quality of life in palliative care. *Seminars in Oncology, 22*(2 Suppl 3), 73–81. https://www.ncbi.nlm.nih.gov/pubmed/7537908.

2. Schipper, H., Clinch, J. J., & Olweny, C. L. M. (1996). Quality of life studies: Definitions and conceptual issues. In B. Spilker (Ed.), *Quality of life and pharmacoeconomics in clinical trials* (pp. 11–23). Philadelphia: Lippincott-Raven Publishers.

3. Bevans, M., Ross, A., & Cella, D. (2014). Patient-reported outcomes measurement information system (PROMIS): Efficient, standardized tools to measure self-reported health and quality of life. *Nursing Outlook, 62*(5), 339–345. https://doi.org/10.1016/j.outlook.2014.05.009.

4. Chen, J., Ou, L., & Hollis, S. J. (2013). A systematic review of the impact of routine collection of patient reported outcome measures on patients, providers and health organizations in an oncologic setting. *BMC Health Services Research, 13*, 211. https://doi.org/10.1186/1472-6963-13-211.

5. Ware, J. E. Jr., & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care, 30*(6), 473–483.

6. 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. *Journal of Clinical Epidemiology, 63*(11), 1179–1194. https://doi.org/10.1016/j.jclinepi.2010.04.011.

7. Hays, R. D., Alonso, J., & Coons, S. J. (1998). Possibilities for summarizing health-related quality of life when using a profile instrument. In M. Staquet, R. Hays & P. Fayers (Eds.), *Quality of life assessment in clinical trials: Methods and practice* (153, p. 143). Oxford: Oxford University Press.

8. Farivar, S. S., Cunningham, W. E., & Hays, R. D. (2007). Correlated physical and mental health summary scores for the SF-36 and SF-12 health survey, V. 1. *Health and Quality of Life Outcomes, 5*:54. 18.

9. Hays, R. D., Marshall, G. N., Wang, E. Y. I., & Sherbourne, C. D. (1994). Four-year cross-lagged associations between physical and mental health in the medical outcomes study. *Journal of Consulting and Clinical Psychology, 62*, 441–449.

10. Hays, R. D., Spritzer, K. L., Schalet, B., & Cella, D. (2018). PROMIS®-29 v2.0 physical and mental health summary scores. *Quality of Life Research, 27*(7), 1885–1891. https://doi.org/10.1007/s11136-018-1842-3.

11. Rose, A. J., et al. (2018). Evaluating the PROMIS-29 v2.0 for use among older adults with multiple chronic conditions. *Quality of Life Research*. https://doi.org/10.1007/s11136-018-1958-5

12. Tarlov, A. R., et al. (1989). The medical outcomes study: An application of methods for monitoring the results of medical care. *JAMA, 262*(7), 925–930.

13. Edelen, M. O., Rose, A. J., Bayliss, E., Baseman, L., Butcher, E., Garcia, R. E., Tabano, D., & Stucky, B. D. (2017). *Patient-reported outcome-based performance measures for older adults with multiple chronic conditions*. Santa Monica: RAND Corporation. https://www.rand.org/pubs/research_reports/RR2176.html.

14. Embretson, S. E., & Reise, S. P. (2000). *item response theory for psychologists*. London: Psychology Press.

15. Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334. https://doi.org/10.1007/bf02310555.

16. Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*, 141–151. https://doi.org/10.1177/001316446002000116.

17. Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*, 245–276. https://doi.org/10.1207/s15327906mbr0102_10.

18. SAS (9.4 ed.). Cary, NC: SAS Corporation.

19. Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus User's Guide. Eighth Edition*. Los Angeles: Muthén & Muthén.

20. Hays, R. D., & Stewart, A. L. (1990). The structure of self-reported health in chronic disease patients. *Psychological Assessment, 2*, 22–30. https://doi.org/10.1037/1040-3590.2.1.22.

21. Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118.

22. Beaumont, J. L., Cella, D., Phan, A. T., Choi, S., Liu, Z., & Yao, J. C. (2012). Comparison of health-related quality of life in patients with neuroendocrine tumors with quality of life in the general US population. *Pancreas, 41*(3), 461–466. https://doi.org/10.1097/MPA.0b013e318232804527.

23. Kazis, L. E., Miller, D. R., Skinner, K. M., Lee, A., Ren, X. S., Clark, J. A. el al (2004). Patient-Reported Measures of Health: The Veterans Health Study. *The Journal of ambulatory care management, 27*(1), 70–83. https://www.ncbi.nlm.nih.gov/pubmed/14717468.

24. Revicki, D. A., Kawata, A. K., Harnam, N., Chen, W.-H., Hays, R. D., & Cella, D. (2009). Predicting EuroQol (EQ-5D) scores from the patient-reported outcomes measurement information system (PROMIS) global items and domain item banks in a United States sample. *Quality of Life Research, 18*, 783–791.

25. Hays, R. D., Revicki, D. A., Feeny, D., Fayers, P., Spritzer, K. L., & Cella, D. (2016). Using linear equating to map PROMIS global health items and the PROMIS-29 V. 2 Profile measure to the health utilities index—mark 3. *Pharmacoeconomics, 34*, 1015–1022.

26. Craig, B. M., Reeve, B. B., Brown, P. M., Cella, D., Hays, R. D., Lipscomb, J., Pickard, A. S., & Revicki, D. A. (2014). US valuation of health outcomes measured using the PROMIS-29. *Value in Health, 17*(8), 846–853.

27. Hanmer, J., Cella, D., Feeny, D., Fischhoff, B., Hays, R. D., Hess, R., et al. (2017). Selection of key health domains from PROMIS® for a generic preference-based scoring system. *Quality of Life Research, 26*, 3377–3385.

28. Hanmer, J., Feeny, D., Fischoff, B., Hays, R. D., Hess, R., Pilkonis, P., et al. (2015). The PROMIS of QALYs. *Health and Quality of Life Outcomes, 13*, 122.

29. Quan, H., Sundararajan, V., Halfon, P., Fong, A., Burnand, B., Luthi, J. C., Saunders, L. D., Beck, C. A., Feasby, T. E., & Ghali, W. A. (2005). Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Medical Care, 43*(11), 1130–1139. https://doi.org/10.1097/01.mlr.0000182534.19832.83.