Follow us @BoneJointOpen

Freely available online OPEN ACCESS

BJO

GENERAL ORTHOPAEDICS

Can PROMIS measures be used to create subgroups for patients seeking orthopaedic care?

S. Z. George,
X. Yan,
S. Luo,
S. A. Olson,
E. K. Reinke,
M. P. Bolognesi,
M. E. Horn

From Duke University, Durham, North Carolina, USA

Aims

Patient-reported outcome measures have become an important part of routine care. The aim of this study was to determine if Patient-Reported Outcomes Measurement Information System (PROMIS) measures can be used to create patient subgroups for individuals seeking orthopaedic care.

Methods

This was a cross-sectional study of patients from Duke University Department of Orthopaedic Surgery clinics (14 ambulatory and four hospital-based). There were two separate cohorts recruited by convenience sampling (i.e. patients were included in the analysis only if they completed PROMIS measures during a new patient visit). Cohort #1 (n = 12,141; December 2017 to December 2018,) included PROMIS short forms for eight domains (Physical Function, Pain Interference, Pain Intensity, Depression, Anxiety, Sleep Quality, Participation in Social Roles, and Fatigue) and Cohort #2 (n = 4,638; January 2019 to August 2019) included PROMIS Computer Adaptive Testing instruments for four domains (Physical Function, Pain Interference, Depression, and Sleep Quality). Cluster analysis (K-means method) empirically derived subgroups and subgroup differences in clinical and sociodemographic factors were identified with one-way analysis of variance.

Results

Cluster analysis yielded four subgroups with similar clinical characteristics in Cohort #1 and #2. The subgroups were: 1) *Normal Function:* within normal limits in Physical Function, Pain Interference, Depression, and Sleep Quality; 2) *Mild Impairment:* mild deficits in Physical Function, Pain Interference, and Sleep Quality but with Depression within normal limits; 3) *Impaired Function, Not Distressed:* moderate deficits in Physical Function and Pain Interference, but within normal limits for Depression and Sleep Quality; and 4) *Impaired Function, Distressed:* moderate (Physical Function, Pain Interference, and Sleep Quality) and mild (Depression) deficits.

Conclusion

These findings suggest orthopaedic patient subgroups differing in physical function, pain, and psychosocial distress can be created from as few as four different PROMIS measures. Longitudinal research is necessary to determine whether these subgroups have prognostic validity.

Cite this article: *Bone Jt Open* 2021;2-7:493–502.

Keywords: Phenotypes, Patient-reported outcomes, Prognostic, Physical function, Pain interference

Introduction

Patient-reported outcome measures (PROMs) have become an important part of routine care for almost all medical specialties, including orthopaedics.[1,2] PROMs capture a patient's perspective on their medical condition by quantifying information related to symptoms, functional level, quality of life, social roles, and satisfaction with care.[1] Traditionally, PROMs have been used to determine care outcomes, but they can also be used to establish prognosis. For example, PROMs can be used to estimate risk of having prolonged disability from low back pain,[3,4] and improve

Correspondence should be sent to Steven Z. George; email: steven.george@duke.edu

doi: 10.1302/2633-1462.27.BJO-2021-0045.R1

Bone Jt Open 2021;2-7:493–502.

the prediction of 12-month pain, disability, and satisfaction outcomes following spine surgery.[5]

A set of standardized outcome measures, the Patient Reported Outcomes Measurement Information System (PROMIS), has been increasingly used as an outcome measure for patients seeking orthopaedic care due to its flexibility for use across heterogenous clinical populations. PROMIS has been validated in orthopaedic populations and domains typically reported in the medical literature include but are not limited to physical function,[6] pain interference,[7] pain intensity,[8] anxiety,[9] sleep disturbance,[10] fatigue,[11] depression,[12] and participation in social roles.[13] Moreover, an advantage of PROMIS measures is that they are not region-specific and as such can be used to inform outcomes and prognosis across different orthopaedic patient populations.[14-17] Furthermore, there has been a noticeable uptake in use of PROMIS measures in orthopaedics as evidenced by the large increase of PROMIS measures reported within the past three years.[18] This increased uptake supports the exploration of whether PROMIS measures can create prognostic profiles for orthopaedic patient populations.

Relevance to current literature. PROMIS measures could be used to identify patients seeking orthopaedic care who have similar multidomain profiles. Classifying orthopaedic patients into homogenous subgroups that inform health status beyond the medical diagnosis is a critical first step towards promoting individualized care recommendations. The addition of a subgroup profile (i.e. phenotype) could be used to establish a more accurate prognosis or identify additional treatment needs. Therefore, the aim of this study was to use PROMIS measures to empirically derive patient subgroups from individuals seeking orthopaedic care at a large, academic tertiary medical centre. This study answered the following research questions: 1) Can PROMIS measures be used to create patient subgroups for individuals seeking orthopaedic care?; and 2) What are the clinical and sociodemographic features of these patient subgroups?

Methods

Study design and setting. This was a cross-sectional study of patients seeking care from Duke University Department of Orthopaedic Surgery clinics (14 ambulatory and four hospital-based). These clinics included services from eight specialties including Joint Reconstruction, Spine, Neurosurgery, Sports Medicine, Trauma, Oncology, Foot and Ankle, and Hand.

Participants. Participants were recruited via convenience sampling. Patients included in this analysis were 1) seeking care at Duke Health in the aforementioned Department of Orthopaedic Surgery clinics; 2) completed PROMIS measures associated with their initial clinic visit or "new patient" visit to an orthopaedic surgeon or accompanying alternate physical provider (e.g. physician assistant); and 3) were aged 18 years or older. We excluded patients who completed PROMIS measures associated with their appointment but then either cancelled or did not attend their scheduled appointment. Patients who did not complete PROMIS measures associated with their appointment were not included in this analysis. Informed consent was not obtained from patients as PROMIS measures were collected as part of routine care episodes. Data were extracted retrospectively from the electronic health record (Epic Systems) and this study was granted exempt status by the Duke University Institutional Review Board.

Description of data collection. PROMIS measures were incorporated into the Department's routine clinical care in two separate phases, resulting in two different patient cohorts. Cohort #1 (December 2017 to December 2018) included PROMIS short forms to collect information on eight different domains. Cohort #2 (January 2019 to August 2019) included PROMIS Computer Adaptive Testing (CAT) to collect information on four different domains. More details on the PROMIS measures are provided in the next section (PROMIS Measures). In both cohorts, data were extracted from the electronic health record in the same way. In addition to PROMIS measures, the data extracted included corresponding demographic information, encounter information, and provider speciality.

Variables. PROMIS measures assess physical, mental, and social health.[14-17] PROMIS measures have been validated in the general population, individuals with chronic health conditions, and multiple orthopaedic populations.[17,19,20] In the development cohort we collected PROMIS short form instruments for physical function, pain interference, pain intensity, depression, anxiety, fatigue, sleep disturbance, and ability to participate in social roles. In the validation cohort, a reduced set of PROMIS domains was collected via the CAT instrument for physical function or physical function upper limb, pain interference, depression, and sleep disturbance. The transition to CAT measures was made as part of an implementation strategy that sought to decrease patient burden, improve the precision of domain estimates, and reduce floor and ceiling effects of the PROMIS measures. The timing of the transition to CAT measures was driven by the availability of the CAT instrument within the EHR platform.

PROMIS domains are scored separately on a T-score metric, with standard population values for mean scores (50) and standard deviation (SD) (10). Thus for all domains a score of 60 is one SD above the population mean for a given domain. A higher PROMIS score for a domain means an increase in what is being measured. For example a score of 60 for physical function domain means higher functioning than a score of 45 on the same domain. Higher PROMIS scores could be associated with a desirable or undesirable characteristic, depending upon the domain. For example, higher physical function and social role participation scores are likely to be desirable, while higher pain interference and

Table I. Descriptive summary of sociodemographic information.

Variable	Development cohort (n = 12,141)	Validation cohort (n = 4,638)
Mean age, yrs (SD)	55.4 (15.7)	55.0 (15.0)
Sex, n (frequency)		
Not reported/declined	2 (0)	0 (0)
Female	7,427 (0.61)	2,949 (0.64)
Male	4,712 (0.39)	1,689 (0.36)
Race, n (frequency)		
Two or more races	161 (0.01)	59 (0.01)
American Indian or Alaskan Native	40 (0)	12 (0)
Asian	268 (0.02)	107 (0.02)
Black or African American	1,761 (0.15)	606 (0.13)
Caucasian/White	9,528 (0.78)	3,674 (0.79)
Native Hawaiian or Other Pacific Islander	9 (0)	3 (0)
Not reported/Declined	269 (0.02)	139 (0.03)
Other	105 (0.01)	38 (0.01)
Ethnicity, n (frequency)		
Hispanic	245 (0.02)	100 (0.02)
Not Hispanic/Latino	11,438 (0.94)	4,305 (0.93)
Not reported/Declined	458 (0.04)	233 (0.05)
Marital status, n (frequency)		
Divorced	937 (0.08)	339 (0.07)
Legally separated	136 (0.01)	41 (0.01)
Life partner	53 (0)	25 (0.01)
Married	7,727 (0.64)	2,867 (0.62)
Single	2,347 (0.19)	973 (0.21)
Unknown	391 (0.03)	187 (0.04)
Widowed	550 (0.05)	206 (0.04)
Rural/urban, n (frequency)		
Others	836 (0.07)	384 (0.08)
Rural NC	1,035 (0.09)	403 (0.09)
Urban NC	10,270 (0.85)	3,851 (0.83)
Primary insurance, n (frequency)		
Medicare	2,826 (0.23)	1,065 (0.23)
Medicaid	149 (0.01)	63 (0.01)
Workers Compensation	101 (0.01)	45 (0.01)
Private	9,065 (0.75)	3,465 (0.75)

NC, North Carolina; SD, standard deviation.

depression scores are likely to be undesirable. For reporting PROMIS measures for the derived subgroups we 1) calculated the mean and SD of PROMIS domain scores using the aforementioned T-score metric and 2) determined whether the mean PROMIS domain score corresponded to "within normal limits", "mild", "moderate", or "severe" categories.[20-22]

Description of study participants. In Cohort #1, PROMIS short forms were collected in 12,141 patients (mean age 55.4 years (SD 15.7), 61% female (n = 7,427)), and in Cohort #2 PROMIS CAT was collected in 4,638 patients (mean age 55.0 years (SD 15.9), 64% female (n = 2,949)). Additional sociodemographic information from the two cohorts is presented in Table I.

Statistical analysis. First, we calculated descriptive statistics to characterize the cohorts. Means and SDs were reported for continuous variables, and percentages were reported for categorical variables. Second, we performed data analysis using R Statistical Software version R 4.0.2 (Austria). Patient subgroups were derived using k-means cluster analysis from eight PROMIS domains (Physical Function, Pain Interference, Pain Intensity, Depression, Anxiety, Sleep Quality, Participation in Social Roles, and Fatigue) for the development cohort and four PROMIS domains (Physical Function, Pain Interference, Depression, and Sleep Quality) for the validation cohort. Gap statistics and partitioning plots were used to determine the number of clusters that would create PROMIS-based subgroups. Third, the accuracy of subgroup membership was investigated by iteratively removing each PROMIS domain and determining how much error (i.e. misclassification of subgroup membership) occurred for predicting subgroup membership without a given domain. This meant that domains most important for determining subgroup membership would have higher misclassification rates. Finally, one-way analysis of variance (ANOVA) was used to identify overall between-group differences across the empirically derived subgroups. Post-hoc testing for individual

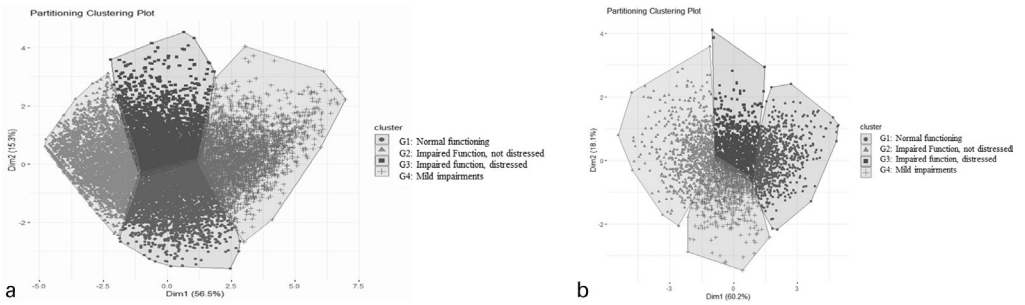

Fig. 1

a) This graph shows the partitioning cluster plot for Cohort #1 (Patient-Reported Outcomes Measurement Information System (PROMIS) Short Forms, eight Domains). This graph represents subgroup position relative to each other in two dimensions after the cluster analyses. The x and y axes provide coordinates for this 2D space. These coordinates provide subgroup location through standardized residuals and should not be interpreted as having any absolute value. b) This graph shows the partitioning cluster plot for Cohort #2 (PROMIS Computer Adaptive Testing (CAT), four Domains). This graph represents subgroup position relative to each other in two dimensions after the cluster analyses. The x and y axes provide coordinates for this 2D space. These coordinates provide subgroup location through standardized residuals and should not be interpreted as having any absolute value.

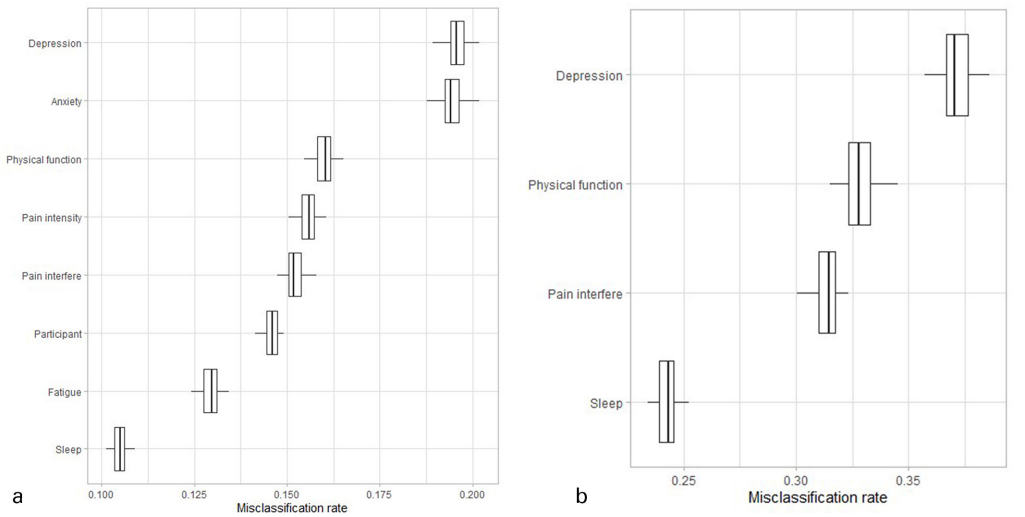

Fig. 2

a) This graph shows the subgroup misclassification rates for Cohort #1 (Patient-Reported Outcomes Measurement Information System (PROMIS) Short Forms, eight Domains). The misclassification rate provides an estimate of how much the accuracy of determining subgroup membership is changed with the iterative removal of one PROMIS domain and then remaining domains were used to determine subgroup membership. One way to interpret this figure is that the PROMIS domains with higher misclassification rates were more important for determining subgroup membership, relative to other domains. In a), Depression was the domain with the highest misclassification rate (most important for subgroup membership) while Sleep Quality had the highest misclassification rate (least important for subgroup membership). b) This graph shows the subgroup misclassification rates for Cohort #2 (PROMIS Computer Adaptive Testing (CAT), four Domains). The misclassification rate provides an estimate of how much the accuracy of determining subgroup membership is changed with the iterative removal of one PROMIS domain and then remaining domains were used to determine subgroup membership. One way to interpret this is that the PROMIS domains with higher misclassification rates were more important for determining subgroup membership, relative to other domains. In b), Depression was the domain with the highest misclassification rate (most important for subgroup membership) while Sleep Quality had the highest misclassification rate (least important for subgroup membership).

group comparisons was completed with Bonferroni correction. Finally in a descriptive analysis (i.e. no inferential statistics), we reported frequencies of the newly derived subgroups by high-volume clinical specialties.

Results

Cohort #1. Cluster solutions were derived from the eight PROMIS domains collected via short forms in the development cohort. A four-cluster solution is presented after review of gap statistic, partitioning plot (Figure 1a), and cluster sizes. The PROMIS domains that contributed most to accuracy of subgroup membership were Depression and Anxiety. Domain importance was determined by misclassification rate for subgroup membership without each domain, and this rate approached 20% for Depression and Anxiety (Figure 2a). The Physical Function, Pain Intensity, Pain Interference, Social Role Participation, and Fatigue domains all had moderate influence on subgroup membership accuracy, with respective misclassification rates ranging from 12.5% to 16.5% when these domains were not used to determine subgroup membership. The Sleep

Table II. Clinical characteristics of empirically derived Patient-Reported Outcomes Measurement Information System (PROMIS) subgroups. Data are presented as mean (SD).

Cohort #1				
PROMIS domain	**Normal Function (n = 3,331)**	**Mild Impairments (n = 2,736)**	**Impaired Function, Not Distressed (n = 3,259)**	**Impaired Function, Distressed (n = 2,815)**
Physical Function	44.77 (6.07)	40.54 (6.38)	32.05 (6.45)	28.39 (6.11)
Pain Interference	49.79 (6.61)	55.39 (5.82)	62.52 (4.38)	66.09 (3.52)
Pain Intensity	2.44 (1.84)	3.37 (1.75)	5.97 (1.82)	6.79 (1.76)
Depression	40.57 (4.08)	50.67 (5.93)	42.81 (5.33)	57.83 (7.15)
Anxiety	42.3 (6.09)	54.38 (5.52)	45.61 (7.0)	60.61 (6.82)
Sleep Quality	42.65 (7.76)	51.1 (7.37)	49.77 (8.47)	59.32 (8.29)
Participation in Social Roles	59.88 (6.28)	50.76 (6.16)	45.77 (6.99)	37.62 (6.45)
Fatigue	40.7 (6.71)	51.83 (6.29)	50.29 (7.89)	62.44 (7.24)
Cohort #2				
PROMIS domain	**Normal Function (n = 1,078)**	**Mild Impairments (n = 1,212)**	**Impaired Function, Not Distressed (n = 1,143)**	**Impaired Function, Distressed (n = 1,205)**
Physical Function	51.19 (6.68)	44.44 (5.09)	37.65 (4.87)	33.19 (5.27)
Pain Interference	51.53 (5.62)	57.74 (4.45)	63.07 (4.46)	68.68 (4.46)
Depression	42.42 (6.74)	53.16 (5.35)	44.22 (6.33)	58.88 (6.74)
Sleep Quality	44.94 (7.53)	54.65 (6.01)	52.12 (7.28)	62.06 (7.26)

Colour coding corresponds to PROMIS score interpretation based on population comparisons (Green = Within Normal Limits, Yellow = Mild Deficit, Orange = Moderate Deficit, Red = Severe Deficit).
All p-values < 0.001, calculated using analysis of variance for any between-group difference.
SD, standard deviation.

domain contributed the least to subgroup membership accuracy, having a misclassification rate of only 10.5%.

Cohort #2. Cluster solutions were independently derived from the four PROMIS domains collected via CAT in the validation cohort. A four-cluster solution is presented after review of gap statistic, partitioning plot (Figure 1b), and subgroup size. The PROMIS domain that contributed most to accuracy of subgroup membership was Depression. When Depression was removed from the cluster analysis the misclassification rate approached 37.5% indicated much lower accuracy of subgroup determination (Figure 2b). The Physical Function and Pain Interference domains had misclassification rates ranging just above and below 32.5% respectively, suggesting again that there would be lower accuracy of subgroup determination without these domains. Finally, the Sleep domain had a misclassification rate of just below 25%. Relative to the other domains this was the least important factor for determining subgroup membership. However, a misclassification rate approaching 25% suggests that Sleep is still an important domain to consider for improving accuracy of subgroup membership.

Clinical and sociodemographic features of subgroups. As expected, subgroup differences existed for all PROMIS domains in the development cohort (Table II). Statistical differences in PROMIS domains and using the expected population based values (i.e. "within normal limits", "moderate deficit") for PROMIS measures guided subgroup labels:

1. *Normal Function* (n = 3,331): characterized by within normal limits in all PROMIS domains except for moderate deficit in Social Role Participation.
2. *Mild Impairment* (n = 2,736): characterized by mild deficits in Physical Function, Pain Interference, and Pain Intensity but within normal limits for other PROMIS domains (e.g. Depression, Anxiety, Sleep Quality, Social Role Participation, and Fatigue).
3. *Impaired Function, Not Distressed* (n = 3259): characterized by moderate deficits in Physical Function, Pain Interference, and Pain Intensity but within normal limits for other PROMIS domains (e.g. Depression, Anxiety, Sleep Quality, Social Role Participation, and Fatigue).
4. *Impaired Function, Distressed* (n = 2,815) characterized by severe (Physical Function), moderate (Pain Interference, Pain Intensity, Anxiety, Social Role Participation, and Fatigue) and mild deficits (Depression and Sleep Quality).

In the validation cohort the cluster solution was very similar to the development cohort. That is, subgroup differences existed for all PROMIS domains (Table II) and these statistical differences, along with the expected PROMIS population-based values (i.e. "within normal limits", "moderate deficit") measures guided subgroup labels:

1. *Normal Function* (n = 1,078): characterized by within normal limits in Physical Function, Pain Interference, Depression, and Sleep Quality.

Table III. Demographic characteristics of Patient-Reported Outcomes Measurement Information System (PROMIS) subgroups (Cohort #2).

Group	Normal Function (n = 1,078)	Mild Impairments (n = 1,212)	Impaired Function, Not Distressed (n = 1,143)	Impaired Function, Distressed (n = 1,205)	p-value*
Mean age, yrs (SD)	53.67 (16.86)	54.02 (15.39)	57.72† (16.36)	54.51 (14.96)	< 0.001
Sex, n (%)					< 0.001
Female	580 (53.8)†	775 (63.94)	729 (63.78)	865 (71.78)†	
Male	498 (46.2)†	437 (36.06)	414 (36.22)	340 (28.22)†	
Race, n (%)					< 0.001
2 or more races	15 (1.39)	16 (1.32)	14 (1.22)	14 (1.16)	
American Indian or Alaskan Native	2 (0.19)	3 (0.25)	0 (0)	7 (0.58)	
Asian	38 (3.53)	36 (2.97)	22 (1.92)	11 (0.91)	
Black or African American	126 (11.69)	124 (10.23)	162 (14.17)	194 (16.1)	
Caucasian/White	862 (79.96)	983 (81.11)	898 (78.57)	931 (77.26)	
Native Hawaiian or Other Pacific Islander	0 (0)	2 (0.17)	1 (0.09)	0 (0)	
Not Reported/Declined	28 (2.6)	38 (3.14)	35 (3.06)	38 (3.15)	
Other	7 (0.65)	10 (0.83)	11 (0.96)	10 (0.83)	
Ethnicity, n (%)					0.786
Hispanic	21 (1.95)	22 (1.82)	25 (2.19)	32 (2.66)	
Not Hispanic/Latino	1,007 (93.41)	1,125 (92.82)	1,063 (93)	1,110 (92.12)	
Not reported/declined	50 (4.64)	65 (5.36)	55 (4.81)	63 (5.23)	
Marital status, n (%)					< 0.001
Divorced	55 (5.1)	79 (6.52)	88 (7.7)	117 (9.71)	
Legally separated	6 (0.56)	11 (0.91)	5 (0.44)	19 (1.58)	
Life partner	6 (0.56)	6 (0.5)	6 (0.52)	7 (0.58)	
Married	691 (64.1)	771 (63.61)	725 (63.43)	680 (56.43)	
Single	229 (21.24)	251 (20.71)	224 (19.6)	269 (22.32)	
Unknown	65 (6.03)	45 (3.71)	41 (3.59)	36 (2.99)	
Widowed	26 (2.41)	49 (4.04)	54 (4.72)	77 (6.39)	
Rural/urban, n (%)					< 0.001
Others	68 (6.31)	94 (7.76)	94 (8.22)	128 (10.62)	
Rural NC	68 (6.31)	82 (6.77)	121 (10.59)	132 (10.95)	
Urban NC	942 (87.38)	1,036 (85.48)	928 (81.19)	945 (78.42)	
Primary insurance, n (%)					< 0.001
Medicare	202 (18.74)	224 (18.48)	337 (29.48)	302 (25.06)	
Medicaid	8 (0.74)	8 (0.66)	11 (0.96)	36 (2.99)	
Workers Compensation	7 (0.65)	11 (0.91)	10 (0.87)	17 (1.41)	
Private	861 (79.87)	969 (79.95)	785 (68.68)	850 (70.54)	

*This p-value is for any between group difference; Post-hoc testing was completed only for age and sex due to the number of comparisons and limited cell sizes for other categorical variables.
†Post-hoc difference from other subgroups ($p < 0.01$).
NC, North Carolina; SD, standard deviation.

2. *Mild Impairment* (n = 1,212): characterized by mild deficits in Physical Function, Pain Interference, and Sleep Quality. Depression was within normal limits when compared to population values.
3. *Impaired Function, Not Distressed* (n = 1,143): characterized by moderate deficits in Physical Function and Pain Interference, but within normal limits for Depression and Sleep Quality.
4. *Impaired Function, Distressed* (n = 1,205): characterized by moderate (Physical Function, Pain Interference, and Sleep Quality) and mild (Depression) deficits.

Sociodemographic differences in the four PROMIS subgroups were further explored in the validation cohort (Table III). Most notably, the *Impaired Function, Not Distressed* subgroup was older than the other subgroups, but there were no other age-related differences between the other subgroups. For differences in sex, the lowest percentage of females corresponded with *Normal Function* subgroup, while the highest percentage of females corresponded with the *Impaired Function, Distressed* subgroup. The *Impaired Function, Not Distressed* and *Mild Impairments* subgroups had similar female/male frequencies. There were also differences in race (*Mild Impairment* highest percentage of Caucasian/White), marital status (*Impaired Function, Distressed* lowest percentage Married), whether living in a rural or urban area (*Impaired Function, Distressed* lowest percentage Urban), and primary insurance type (*Impaired Function, Not Distressed*

Table IV. Frequencies of Patient-Reported Outcomes Measurement Information System (PROMIS) subgroups care seeking by clinical specialties (Cohort #2).

Group	Normal Function	Mild Impairments	Impaired Function, Not Distressed	Impaired Function, Distressed
Clinical Service for Care Seeking*				
Foot-ankle, n (%)	110 (30.9)	100 (28.2)	97 (27.3)	48 (13.5)†
Hand, n (%)	141 (36.8)‡	102 (26.6)	81 (21.1)	59 (15.4)
Spine, n (%)	97 (17.1)	140 (24.6)	137 (24.1)	194 (34.2)‡
Sports med, n (%)	317 (27.3)	347 (29.9)	266 (22.9)	231 (19.9)
Joint reconstruction, n (%)	132 (19.3)	142 (20.8)	194 (28.4)	215 (31.5)‡

*Only specialities with greater than 250 total PROMIS intake measures were included in this table.
†Indicates that a speciality had a lower than 15% frequency.
‡Indicates that a clinical speciality had a greater than 30% frequency.

highest percentage Medicare, *Normal Function* and *Mild Impairment* highest percentage of Private Pay).

Frequencies for the four PROMIS subgroups were reported descriptively for clinical specialties with high clinical volume (Table IV). The Hand clinical speciality had 36.8% in the *Normal* subgroup. The Spine and Joint Reconstruction specialties had 34.2% and 31.5% in the *Impaired Function, Distressed* subgroup, respectively. In contrast, the Foot and Ankle speciality had 13.5% for the *Impaired Function, Distressed* subgroup.

Discussion

Creating patient subgroups is an innovative application of PROMs. PROMs are typically used to assess the patient's current status of a specific domain before and after treatment. However, a single point of measurement can be used to understand the patient's status when initially seeking treatment. In this study we assessed PROMIS measures for physical function, pain, as well as measures of general and mental health. This approach contrasts with use of region-specific measures, which would be generalizable only to given patient populations (e.g. Disability of Arm, Shoulder, and Hand questionnaire used for upper extremity conditions only). Region-specific measures are often used in orthopaedic practice and research, but their specificity is a notable limitation for creating patient subgroups with broad, heterogenous application. There is growing evidence that use of PROMIS measures is rapidly expanding in orthopaedics,[17,23,24] so it was timely to explore how these measures classify patients. The findings from this study are encouraging because they suggest patients with orthopaedic conditions can be broadly classified into four different subgroups. These subgroups are useful because they would allow for direct comparisons (i.e. using the same measure) to be made across different orthopaedic patient populations for physical function, pain interference, and psychosocial distress. In contrast, region-specific measures could not be used to derive these subgroups and make comparisons on a common metric.

The primary limitation of this study is its cross-sectional design and accordingly these data do not support these subgroups being predictive of patient outcomes. Differences in PROMIS domains suggests there is clinical relevance for these subgroups but only future studies, using longitudinal designs, can address the primary limitation of this study. For example, this analysis included operative and nonoperative patients and a longitudinal study would be necessary to determine whether these subgroups are useful in determining who truly benefits from surgery. A secondary limitation of this study is the convenience sampling which included only those that completed the PROMIS measures in the analysis. This means we could not compare demographic and clinical features of those that did not complete PROMIS measures to those that were included in this analysis. Therefore, the generalizability of these subgroups may be limited if there were drastic demographic and clinical differences for those that did not complete the PROMIS measures. Another notable limitation is that we do not have relevant clinical information to describe these subgroups further, including region-specific questionnaires, medical diagnosis, and comorbidity data. This limitation prevents us from being able to characterize the subgroups beyond the PROMIS measures, sociodemographic features, and clinical speciality where care was being sought. Therefore we do not know if these subgroups are relevant for homogenous patient populations that might be recruited into a clinical trial (e.g. adults undergoing elective hip arthroplasty). Finally, another limitation to consider when interpreting these findings is that these subgroups were derived from two cohorts recruited from the same health system. Therefore, these subgroups may not be generalizable to other settings. A high priority for future research would be to investigate these subgroups in other clinical settings to determine if they can be reproduced in other settings.

These findings support the construct validity of four PROMIS-based patient subgroups for patients seeking orthopaedic care in an academic tertiary care setting. An advantage of these subgroups is that they are derived from PROMIS measures, which are psychometrically sound and can be implemented widely across orthopaedic patient populations. The four derived subgroups were similar in composition across two cohorts and robust to different PROMIS instruments (short forms and CAT) and domain numbers (8 and 4). These findings indicate a new way to integrate patient-reported outcomes into decision-making for individuals seeking orthopaedic care. The use of PROMIS measures allows for groupings to be formed based on shared patient-reported characteristics across mental and physical health domains supplementing the way patients are traditionally grouped in orthopaedics (e.g. medical diagnosis, imaging findings, and procedure type). These subgroups provide information in an important area that is often overlooked in routine medical encounters – objective data representing how an individual's overall quality of life is impacted by the orthopaedic conditions for which they are seeking care.

These findings also inform the number of PROMIS domains needed to determine subgroups in orthopaedic practice. In our first cohort eight PROMIS domains generated four subgroups, and the misclassification analysis indicated that the depression and anxiety domains were most important for determining subgroup membership. In the second cohort four PROMIS domains also generated four subgroups, and the misclassification analysis indicated that depression was the most important domain (anxiety was not administered in the second cohort). As expected the misclassification rates for the eight PROMIS domains was much lower than those for the four PROMIS domains. In fact, misclassification rates give a strong indication that four PROMIS domains are the minimum number needed to generate these four subgroups. The removal of any one of the four PROMIS domains results in almost a 25% misclassification rate for subgroup determination – a rate that is likely too high to be useful clinically. Interestingly, the rank order of domain importance was the same in both cohorts for subgroup misclassification; depression, physical function, pain interference, and sleep (in decreasing order of importance for subgroup determination). Collectively, these findings suggest that four PROMIS domains can reliably generate the same four subgroups as from eight PROMIS domains; and with the same weighting of domain importance for determining accuracy of subgroup membership. These are encouraging findings because they indicate that these subgroups are relatively stable even when addressing administrative burden by reducing the number of PROMIS domains. However, as the number of PROMIS domains is reduced it is important to include a minimum number of measures that best differentiate these subgroups. For example, in a recent systematic review it was most common for orthopaedic literature to report PROMIS measures for physical function and pain interference domains.[18] Our current findings suggest that the two most commonly implemented PROMIS measures in orthopaedics would not be sufficient to distinguish between the *Impaired Function* and *Impaired Function, Distressed* subgroup. Since this is a differentiation of high clinical importance our findings suggest that physical function and pain interference must be augmented with additional PROMIS measures for depression and sleep quality in order to support better clinical decision-making. It is speculative, but there is potential to improve clinical decision-making by using these subgroups to collectively characterize function, pain, and psychosocial distress to improve patient care plans. For example, patients in the *Mild Impairment* and *Impaired Function* subgroups may benefit from the way care is structured in existing care pathways but those in the *Impaired Function, Distress* subgroup may need care that is structured differently (e.g. inclusion of psychological or behavioural services). There are also research applications for these subgroups, where they could be used in cohort studies or clinical trials to further characterize patients beyond the medical diagnosis, anatomical region, and/or procedure type.

Statistical differences in sociodemographic factors were detected across the subgroups, many likely due to the large sample size of this cohort. The *Impaired Function, Not Distressed* subgroup was clearly the oldest of the four clusters, however the age of all derived subgroups was within a five-year range. Therefore, this difference in age is not likely to be clinically relevant. The *Impaired Function, Distress* subgroup had the highest percentage of females (71.8%) which may be clinically relevant because the *Normal Function* subgroup had a lower percentage of females (53.8%). Finally, there were also subgroup differences in race, ethnicity, marital status, rural/urban living, and primary insurance provider. However, these findings were difficult to interpret because there were no consistent patterns for subgroups and many of the variable categories had small cell sizes (e.g. certain race categories). Collectively the reported sociodemographic differences suggest there are no obvious confounders from these factors and their clinical relevance remains to be determined in future studies.

Initial proof of concept for clinical application of these four subgroups was demonstrated by classifying patients across eight different orthopaedic clinical specialties. Subgroup frequency across service lines with higher volume suggested patterns that converged with patients seeking care from the included clinical specialties (Table IV). For example, the *Normal Function* subgroup had highest frequency in the Hand clinical speciality, which would be an expected finding given that PROMIS

measures are designed to assess overall health quality of life and not the specific nuances of hand function. Similarly, the *Mild Impairments* group had highest frequency in the Sports Medicine speciality, which reflects the overall higher health quality of life that would be associated with these patient populations. Finally, the higher frequencies of the *Impaired Function, Distressed* subgroup in the Spine and Joint Reconstruction specialities converges with patient populations that are seeking speciality care that may involve surgical options for treatment due to notable decrease in overall health quality of life. These data were reported for descriptive purposes only, but they do provide additional preliminary support for the clinical application of these subgroups.

In summary, this study suggests clinically relevant orthopaedic patient subgroups can be created from PROMIS measures. The four subgroups empirically derived were 1) *Normal Function*; 2) *Mild Impairment*; 3) *Impaired Function, Not Distressed*; and 4) *Impaired Function, Distressed*. These subgroups resulted in different clinical profiles for standardized measures of physical function, pain interference, and psychosocial distress. These initial findings are encouraging, but longitudinal research is necessary to determine whether these subgroups have prognostic validity for patients seeking orthopaedic care.

Take home message

- Patient-Reported Outcomes Measurement Information System (PROMIS) measures were used to identify patient subgroups with similar characteristics across several patient-reported outcome domains (e.g. Physical Function, Pain Interference, Depression, and Sleep Quality).
- Similar patient subgroups were identified whether four or eight PROMIS domains were used in the statistical analysis.
- Future research is necessary to determine if the patient subgroups reported in this paper a) can also be identified in other care settings or b) are accurate in predicting patient outcomes.

References

1. **Nelson EC**, **Eftimovska E**, **Lind C**, **Hager A**, **Wasson JH**, **Lindblad S**. Patient reported outcome measures in practice. *BMJ*. 2015;350:g7818.
2. **Hung M**, **Stuart AR**, **Higgins TF**, **Saltzman CL**, **Kubiak EN**. Computerized adaptive testing using the PROMIS physical function item bank reduces test burden with less ceiling effects compared with the short musculoskeletal function assessment in orthopaedic trauma patients. *J Orthop Trauma*. 2014;28(8):439–443.
3. **Hill JC**, **Dunn KM**, **Lewis M**, **et al**. A primary care back pain screening tool: identifying patient subgroups for initial treatment. *Arthritis Rheum*. 2008;59(5):632–641.
4. **Beneciuk JM**, **Bishop MD**, **Fritz JM**, **et al**. The start back screening tool and individual psychological measures: evaluation of prognostic capabilities for low back pain clinical outcomes in outpatient physical therapy settings. *Phys Ther*. 2013;93(3):321–333.
5. **Rundell SD**, **Pennings JS**, **Nian H**, **et al**. Adding 3-month patient data improves prognostic models of 12-month disability, pain, and satisfaction after specific lumbar spine surgical procedures: development and validation of a prediction model. *Spine J*. 2020;20(4):600–613.
6. **Bruce B**, **Fries J**, **Lingala B**, **Hussain YN**, **Krishnan E**. Development and assessment of floor and ceiling items for the PROMIS physical function item bank. *Arthritis Res Ther*. 2013;15(5):R144.
7. **Amtmann D**, **Cook KF**, **Jensen MP**, **et al**. Development of a PROMIS item bank to measure pain interference. *Pain*. 2010;150(1):173–182.
8. **Haefeli M**, **Elfering A**. Pain assessment. *Eur Spine J*. 2006;15(Suppl 1):S17–S24.
9. **Pilkonis PA**, **Choi SW**, **Reise SP**, **et al**. Item banks for measuring emotional distress from the patient-reported outcomes measurement information system (PROMIS®): depression, anxiety, and anger. *Assessment*. 2011;18(3):263–283.
10. **Yu L**, **Buysse DJ**, **Germain A**, **et al**. Development of short forms from the PROMIS™ sleep disturbance and sleep-related impairment item banks. *Behav Sleep Med*. 2011;10(1):6–24.
11. **Lai J-S**, **Cella D**, **Choi S**, **et al**. How item banks and their application can influence measurement practice in rehabilitation medicine: a PROMIS fatigue item bank example. *Arch Phys Med Rehabil*. 2011;92(10 Suppl):S20–S27.
12. **Pilkonis PA**, **Yu L**, **Dodds NE**, **Johnston KL**, **Maihoefer CC**, **Lawrence SM**. Validation of the depression item bank from the patient-reported outcomes measurement information system (PROMIS) in a three-month observational study. *J Psychiatr Res*. 2014;56(56):112–119.
13. **Hahn EA**, **DeWalt DA**, **Bode RK**, **et al**. New English and Spanish social health measures will facilitate evaluating health determinants. *Health Psychol*. 2014;33(5):490–499.
14. **HealthMeasures**. PROMIS. https://www.healthmeasures.net/score-and-interpret/interpret-scores/promis (date last accessed May 2021).
15. **Kortlever JTP**, **Leyton-Mange A**, **Keulen MHF**, **et al**. Promis physical function correlates with KOOS, JR in patients with knee pain. *J Knee Surg*. 2020;33(9):903–911.
16. **Moore MLG**, **Kortlever JTP**, **Keulen MHF**, **Brigati DP**, **Bozic KJ**, **Koenig KM**. Promis PF correlates with HOOS, jr in patients with hip pain. *J Orthop*. 2020;21:58–61.
17. **Brodke DJ**, **Saltzman CL**, **Brodke DS**. Promis for orthopaedic outcomes measurement. *J Am Acad Orthop Surg*. 2016;24(11):744–749.
18. **Horn ME**, **Reinke EK**, **Couce LJ**, **Reeve BB**, **Ledbetter L**, **George SZ**. Reporting and utilization of patient-reported outcomes measurement information System® (PROMIS) measures in orthopedic research and practice: a systematic review. *J Orthop Surg Res*. 2020;15(1):553.
19. **Jildeh TR**, **Lizzio VA**, **Meta F**, **Fidai MS**, **Kaat AJ**, **Makhni EC**. The correlation between PROMIS pain interference and vas pain in ambulatory orthopedic patients. *Orthopedics*. 2018;41(6):e813–e819.
20. **Cella D**, **Riley W**, **Stone A**, **et al**. The patient-reported outcomes measurement information system (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. *J Clin Epidemiol*. 2010;63(11):1179–1194.
21. **Rothrock NE**, **Hays RD**, **Spritzer K**, **Yount SE**, **Riley W**, **Cella D**. 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*. 2010;63(11) :1195–1204
22. **HealthMeasures**. PROMIS Score Cut Points. https://www.healthmeasures.net/score-and-interpret/interpret-scores/promis/promis-score-cut-points (date last accessed May 2021).
23. **Biber J**, **Ose D**, **Reese J**, **et al**. Patient reported outcomes - experiences with implementation in a University Health Care setting. *J Patient Rep Outcomes*. 2017;2:34.
24. **Voloshin I**. Editorial commentary: patient-reported outcomes measurement information system (PROMIS) is important part of future clinical research. *Arthroscopy*. 2020;36(5):1429–1430.

Author information:

- S. Z. George, PhD, FAPTA, Department of Orthopaedic Surgery and Duke Clinical Research Institute, Duke University, Durham, North Carolina, USA.
- X. Yan, PhD, Associate Professor
- S. Luo, PhD, Post-Doctoral Research Fellow Biostatistics & Bioinformatics and Duke Clinical Research Institute, Duke University, Durham, North Carolina, USA.
- S. A. Olson, MD, Professor, Department of Orthopaedic Surgery, Division of Trauma, Duke University, Durham, North California, USA.
- E. K. Reinke, PhD, Research Program Leader, Department of Orthopaedic Surgery, Division of Sports Medicine, Duke University, Durham, North California, USA.
- M. P. Bolognesi, MD, Professor, Department of Orthopaedic Surgery, Division of Adult Reconstruction, Duke University, Durham, North California, USA.
- M. E. Horn, DPT, MPH, PhD, Assistant Professor, Department of Orthopaedic Surgery, Division of Physical Therapy, Duke University, Durham, North California, USA.

Author contributions:

- S. Z. George: Provided final approval of the version to be published, Conceptualized the work, Interpreted the data, Drafted the article.
- X. Yan: Provided final approval of the version to be published, Analyzed and interpreted the data, Drafted the article.
- S. Luo: Provided final approval of the version to be published, Analyzed and interpreted the data, Critically revised the article.
- S. A. Olson: Provided final approval of the version to be published, Critically revised the article.
- E. K. Reinke: Provided final approval of the version to be published, Critically revised the article.

- M. P. Bolognesi: Provided final approval of the version to be published, Critically revised the article.
- M. E. Horn: Provided final approval of the version to be published, Conceptualized the work, Collected, analyzed, and interpreted the data, Drafted the article.

Funding statement:
- The author or one or more of the authors have received or will receive benefits for personal or professional use from a commercial party related directly or indirectly to the subject of this article.

ICMJE COI statement:
- Each author certifies that he or she has no commercial associations (eg, consultancies, stock ownership, equity interest, patent/licensing arrangements, etc) that might pose a conflict of interest in connection with the submitted article. M. E. Horn reports institutional grants from NIH and DoD, unrelated to this study. M. P. Bolognesi reports royalties, consultancy fees, and stock from TJO, royalties, speaker fees, and research support from Zimmer, grants and royalties from Biomet, research support grants from Depuy and Exactech, and is a director of EOA and AAHKS, all unrelated to this study. S. Z. George reports grants from NIH, royalties from Rehab Essentials, and travel and honorarium payments for their service on Med Risk's Scientific Advisory Board, all outside the submitted work.

Acknowledgements:
- The Department of Orthopaedic Surgery Total Joint Arthroplastly Learning Health Unit provided support for the statistical analysis. SZG received support from NIH/NIAMS AR0055899 while working on this manuscript. We would like to thank the Duke Department of Orthopaedic Surgery Comprehensive Outcomes in Orthopaedics and Rehabilitation Data System (COORDS) for supporting the infrastructure for data collection. We would also like to thank Jennifer Gagnon for her contribution to data acquisition. Finally, we would like to thank Richard C. Mather and Jonathan O'Donnell for their assistance in determining the PROMIS measures to be used for routine data collection.

Ethical review statement:
- This study was determined to be exempt by the Duke University Institutional Review Board.