

Floor Effect of PROMIS Depression CAT Associated With Hasty Completion in Orthopaedic Surgery Patients

Jason M. Guattery MS, Agnes Z. Dardas BA, Michael Kelly MD, Aaron Chamberlain MD, MSc, Christopher McAndrew MD, MSc, Ryan P. Calfee MD, MSc

Received: 15 January 2017 / revised: 19 April 2017 / Accepted: 17 November 2017 / Published online: 7 February 2018
Copyright © 2018 by the Association of Bone and Joint Surgeons

Abstract

Background The Patient Reported Outcomes Measurement Information System (PROMIS) was developed to provide valid, reliable, and standardized measures to gather patient-reported outcomes for many health domains, including depression, independent of patient condition. Most studies confirming the performance of these measures were conducted with a consented, volunteer study population for testing. Using a study population that has undergone the process of informed consent may be differentiated from the validation group because they are educated specifically as to the purpose of the questions and they will not have answers recorded in their permanent health record.

Questions/purposes (1) When given as part of routine practice to an orthopaedic population, do PROMIS Physical Function and Depression item banks produce score distributions different than those produced by the populations used to calibrate and validate the item banks? (2) Does the presence of a nonnormal distribution in the PROMIS Depression scores in a clinical population reflect

a deliberately hasty answering of questions by patients? (3) Are patients who are reporting minimal depressive symptoms by scoring the minimum score on the PROMIS Depression Computer Adaptive Testing (CAT) distinct from other patients according to demographic data or their scores on other PROMIS assessments?

Methods Univariate descriptive statistics and graphic histograms were used to describe the frequency distribution of scores for the Physical Function and Depression item banks for all orthopaedic patients 18 years or older who had an outpatient visit between June 2015 and December 2016. The study population was then broken into two groups based on whether they indicated a lack of depressive symptoms and scored the minimum score (34.2) on the Depression CAT assessment (Floor Group) or not (Standard Group). The distribution of Physical Function CAT scores was compared between the two groups. Finally, a time-per-question value was calculated for both the Physical Function and Depression CATs and was

Research reported in time for one of the authors (AZD) when contributing to this publication was supported by Washington University Institute of Clinical and Translational Sciences grant UL1TR000448, subaward TL1TR000449, from the National Center for Advancing Translational Sciences of the National Institutes of Health (NIH).

All ICMJE Conflict of Interest Forms for authors and *Clinical Orthopaedics and Related Research*® editors and board members are on file with the publication and can be viewed on request.

This content is solely the responsibility of the authors and does not necessarily represent the official view of the NIH. The funding did not play a direct role in this investigation.

Clinical Orthopaedics and Related Research® neither advocates nor endorses the use of any treatment, drug, or device. Readers are encouraged to always seek additional information, including FDA-approval status, of any drug or device prior to clinical use.

Each author certifies that his or her institution approved or waived approval for the human protocol for this investigation and that all investigations were conducted in conformity with ethical principles of research.

Jason M. Guattery MS, Agnes Z. Dardas BA, Michael Kelly MD, Aaron Chamberlain MD, MSc, Christopher McAndrew MD, MSc, Ryan P. Calfee MD, MSc, Department of Orthopedic Surgery, Washington University in St Louis School of Medicine, St Louis, MO, USA

R. P. Calfee (✉), Washington University in St Louis School of Medicine, Department of Orthopedic Surgery, Campus Box 8233, 660 S Euclid, St Louis, MO 63110, USA, email: calfeer@wudosis.wustl.edu

compared between assessments within each group as well as between the two groups. Bivariate statistics compared the demographic data between the two groups.

Results Physical Function CAT scores in musculoskeletal patients were normally distributed like the distribution calibration population; however, the score distribution of the Depression CAT in musculoskeletal patients was nonnormal with a spike in the floor score. After excluding the floor spike, the distribution of the Depression CAT scores was not different from the population control group. Patients who scored the floor score on the Depression CAT took slightly less time per question for Physical Function CAT when compared with other musculoskeletal patients (floor patients: 11 ± 9 seconds; normally distributed patients: 12 ± 10 seconds; mean difference: 1 second [0.8–1.1]; $p < 0.001$ but not clinically relevant). They spent a substantially shorter amount of time per question on the Depression CAT (Floor Group: 4 ± 3 seconds; Standard Group: 7 ± 7 seconds; mean difference: 3 [2.9–3.2]; $p < 0.001$). Patients who scored the minimum score on the PROMIS Depression CAT were younger than other patients (Floor Group: 50 ± 18 SD; Standard Group: 55 ± 16 SD; mean difference: 4.5 [4.2–4.7]; $p < 0.001$) with a larger percentage of men (Floor Group: 48.8%; Standard Group 40.0%; odds ratio 0.6 [0.6–0.7]; $p < 0.001$) and minor differences in racial breakdown (Floor Group: white 85.2%, black 11.9%, other 0.03%; Standard Group: white 83.9%, black 13.7%, other 0.02%).

Conclusions In an orthopaedic surgery population that is given PROMIS CAT as part of routine practice, the Physical Function item bank had a normal performance, but there is a group of patients who hastily complete Depression questions producing a strong floor effect and calling into question the validity of those floor scores that indicate minimal depression.

Level of Evidence Level II, diagnostic study.

Introduction

The Patient Reported Outcomes Measurement Information System (PROMIS) domain assessments were created to develop efficient [13, 19, 21, 34, 37], reliable [19–21, 29], responsive, and sensitive item banks [13, 14, 20, 29, 32, 37] that can be used to measure domains such as physical function across diagnoses [8] with less patient burden [19, 34, 37]. The PROMIS measures produce a t-score based on a distribution with a mean score of 50 that represents the mean for the US population based on Census data from 2000 [7]. Higher scores mean more of the domain measured and a difference of 10 points from the mean score equates to 1 SD higher or lower than that mean. In addition to the standardized scoring, PROMIS is distinguished from legacy patient-reported measures by the ability to utilize

Computer Adaptive Testing (CAT) technology with the PROMIS domain item banks, which is a large factor in the benefits PROMIS delivers while also enabling immediate scoring for consideration while delivering care [4], which can help providers focus on pressing patient symptoms [38] and have a positive effect on patient-provider dialogue [10, 18].

The PROMIS Depression CAT was developed to provide a metric to measure emotional distress that is applicable to a population with a wide range of health conditions [30, 33]. In addition to the numerous general benefits of utilizing CAT, research has already defined the relationship of the Depression CAT scores with preexisting mental health measures [3], has provided a bridge between these measures [11], and can be accessed at websites such as PROSetta Stone[®] (<http://www.prosettastone.org/Pages/default.aspx>). Mental health, including depression, and its impact on patient-reported outcomes [24, 25, 36], perceived disability [26], prognosis [22], and satisfaction with care [2, 5], has been of increasing interest in orthopaedic surgery. This research is part of a growing discussion regarding the measurement of depressive symptoms in nonpsychiatric care [1, 15, 28, 35]. The clinically relevant impact and high prevalence of psychologic issues among patients seeking care for other reasons have influenced new requirements and recommendations emerging in health care regarding screening patients for depression [1, 28, 35]. In addition to the US Preventative Services Task Force recommendations, the Centers for Medicare & Medicaid Services, through its electronic health record incentive programs, has listed depression screening as one of their Adult Recommended Core measures [9]. This has directly translated into requirements for orthopaedic programs to include emotional distress reporting as a component of bundled payment documentation for large joint arthroplasty.

In June 2015 the Department of Orthopedic Surgery at Washington University in St Louis began department-wide administration of the PROMIS CAT for three different domains: Physical Function, Pain Interference, and Depression. The use of the PROMIS instruments was adopted as part of routine practice for all outpatient visits for patients older than 18 years of age. Although capture and completion rates have been consistently between 98% and 100%, registration personnel have reported that some patients have questioned the relevance of, and exhibited a general reluctance to complete, the Depression CAT. Because the normative population that was used in the development of these measures was a research population that is comparatively highly educated through the informed consent process, how they complete the questions may be different than a population given these measures as part of routine care, which may affect the overall performance of the measure.

Therefore, we asked: (1) When given as part of routine practice to an orthopaedic population, do PROMIS Physical Function and Depression item banks produce score distributions different than those produced by the populations used to calibrate and validate the item banks? (2) Does the presence of a nonnormal distribution in the PROMIS Depression scores in a clinical population reflect a deliberately hasty answering of questions by patients? (3) Are patients who are reporting minimal depressive symptoms by scoring the minimum score on the PROMIS Depression CAT distinct from other patients according to demographic data or their scores on other PROMIS assessments?

Patients and Methods

This is a retrospective analysis of PROMIS Physical Function and Depression CAT scores completed as part of routine practice for all outpatient orthopaedic surgery visits. Score data came from patient visits of all types at outpatient orthopaedic clinics of a single academic tertiary center between the dates of June 22, 2015, and August 1, 2016.

Data were gathered from patient visits of all types where the patient was 18 years or older and had scores for both the PROMIS Physical Function CAT as well as the PROMIS Depression CAT. Visits with completion times for either the Physical Function or the Depression CAT longer than 600 seconds were excluded because these outliers reflected survey completion interruption by other steps of the visit such as obtaining radiographs or failure to complete the survey before interaction with the treating surgeon.

A total of 77,211 patient visits were initially selected for analysis. A total of 637 visits were excluded as a result of completion times longer than 600 seconds ($n = 604$) or because of a technical error in recording the time ($n = 33$). The final study population included 76,574 visits (Table 1). The mean age of the study population

was 54 years (range, 18-99 years; SD 17 years; Table 1). The study population comprised 44,022 visits with women (57%) and 32,552 visits with men (43%). A total of 64,486 visits were with white patients (84%) and 10,142 (13%) with black patients. The remaining 2.54% of patient visits were split among all other racial groups including patients who described themselves as Asian (1.04%), other (0.87%), unknown/decline (0.31%), Native American (0.14%), multiracial (0.14%), Hispanic (0.03%), and Pacific Islander (0.01%).

All patients completed both Physical Function (Version 1.2) and Depression (Version 1.0) PROMIS CATs. Patients are uniformly given the assessments at the time of registration. Registration staff provide a general statement explaining the assessment while the assessment loads on a tablet computer (iPad mini; Apple, Cupertino, CA, USA) before patients independently complete the PROMIS surveys. Assessment scores are immediately routed into the patient's electronic medical record in an outcomes table.

The variables gathered for analysis in this study were the PROMIS CAT scores for the Physical Function and Depression item banks. Administrative technical reports provided the number of questions answered by the patient during each assessment as well as the time it took the patient to complete the assessment. All of the variables were created and gathered electronically. Although the scores and the administrative data were gathered from separate health IT systems, both the scores and the administrative data are linked together through a unique visit identifier that is used by both systems ensuring the data are properly aligned.

Statistical Analysis

Univariate descriptive statistics and graphic histograms were created to describe the frequency distribution for the entire population of study visits ($n = 76,574$) for both the Physical Function and Depression CATs.

Table 1. Demographics of standard and floor groups

Characteristic	Population	Standard group (N = 58,631)	Floor group (N = 17,943)	Mean difference or odds ratio	p value
Age (years; mean \pm SD)	54 \pm 17	55 \pm 16	50 \pm 18	5 (4-5)	< 0.001
Gender				0.6 (0.6-0.7)	< 0.001
Men	32,552 (43%)	23,362 (40%)	9190 (49%)		
Women	44,022 (58%)	35,269 (60%)	8753 (51%)		
Race					
White	60,486 (84%)	45,195 (84%)	15,291 (85%)		< 0.001
Black	10,142 (13%)	8005 (14%)	2137 (12%)		
Other	1946 (3%)	1431 (2%)	515 (3%)		

After recognition of a nonnormal distribution of Depression CAT scores, the study population was split into two separate groups based on Depression CAT scores. The Standard Group includes all visits in which the patient scored above the minimum on the Depression CAT (34.2) suggesting that the patients in this group may have been experiencing some elements of depression in the previous 7 days. The Floor Group includes the remaining visits in which the patients scored the minimum score (34.2) on the Depression CAT indicating that the patients in this group did not experience any elements of depression in the previous 7 days. Subsequent analysis was conducted to explore differences between those groups. Both mean values and the distribution of the PROMIS Physical Function CAT scores were compared between the Standard and Floor Groups. Using published minimally clinically important differences on the PROMIS Physical Function as a proxy for a relevant between-group difference, we presumed a 5-point difference in mean score would be a meaningful difference [39]. The mean time per question (TPQ) spent answering the Physical Function and Depression questions was analyzed using t-tests within the Standard and Floor Groups. Comparisons of the mean TPQ spent on Physical Function and Depression between the Standard and Floor Groups were done by t-tests as well. We presumed that a 30% reduction in time spent per question would represent a clinically meaningful change in the TPQ between groups. Demographic data of age, gender, and race were compared with bivariate statistics.

Results

When given as a part of routine practice to an orthopaedic population, the scores of the Physical Function CAT were normally distributed with a mean score of 39.9 (SD 9.6), 1 SD worse than the normal population (Fig. 1). There was a nonnormal distribution of the study population's scores for the Depression CAT with a spike at 34.2 (floor score) followed by a normal distribution with a mean of 47.8 (SD 10.5) (Fig. 2). After removing the Floor Group scores that comprise the floor spike, the remaining Depression CAT score distribution of the Standard Group is normal with a mean score of 52.0 (SD 8.3) (Fig. 3).

The presence of a nonnormal distribution in the PROMIS Depression scores is reflective of a deliberately hasty answering of questions by patients in the Floor Group (Table 2). The Floor Group spent a slightly lower mean TPQ for Physical Function CAT when compared with other musculoskeletal patients from the Standard Group (Floor Group: 11 ± 9 seconds; Standard Group 12 ± 10 seconds; mean difference: 1 [0.8-1.1]; $p < 0.001$ but not clinically important). However, the Floor Group spent much less time on the Depression CAT than the Standard Group (Floor Group: 4 ± 3 seconds; Standard Group: 7 ± 7 seconds; mean difference: 3 [2.9-3.2]; $p < 0.001$).

Patients who scored the minimum on PROMIS Depression were younger than other patients (Floor Group: 50 ± 18 ; Standard Group: 55 ± 16 ; mean difference: 5 [4-5]; $p < 0.001$) and included a larger percentage of men

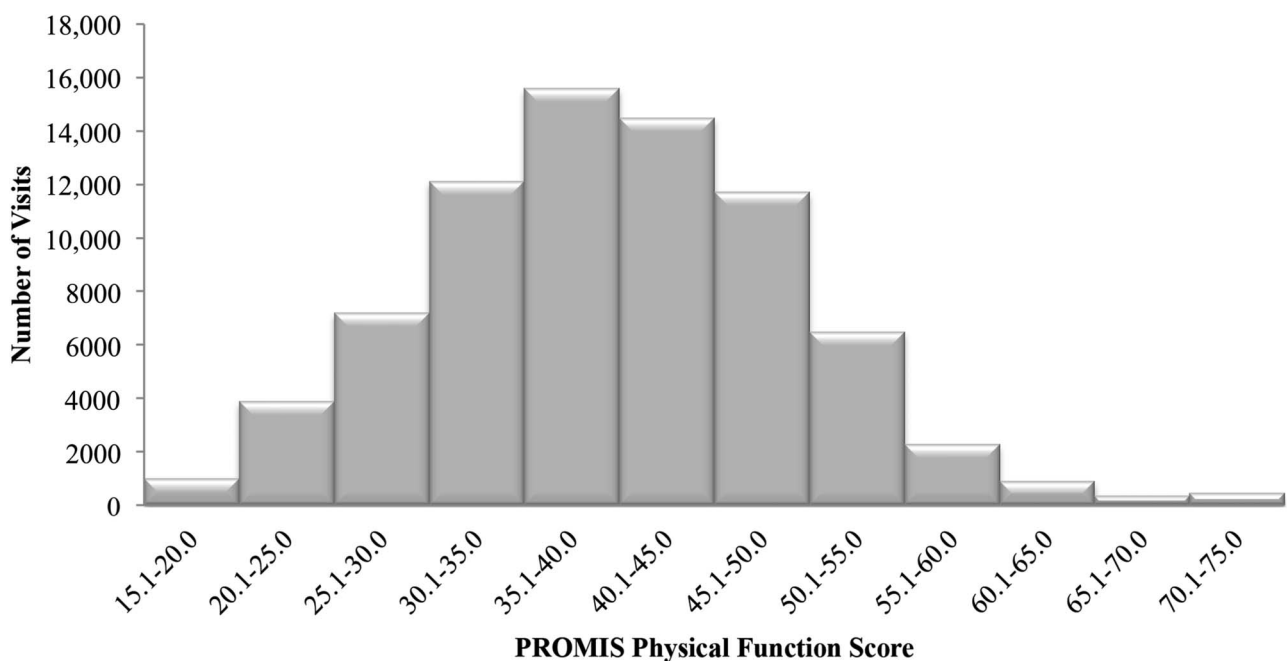


Fig. 1 The graph shows a frequency distribution of scores for the PROMIS Physical Function CAT in the study population.

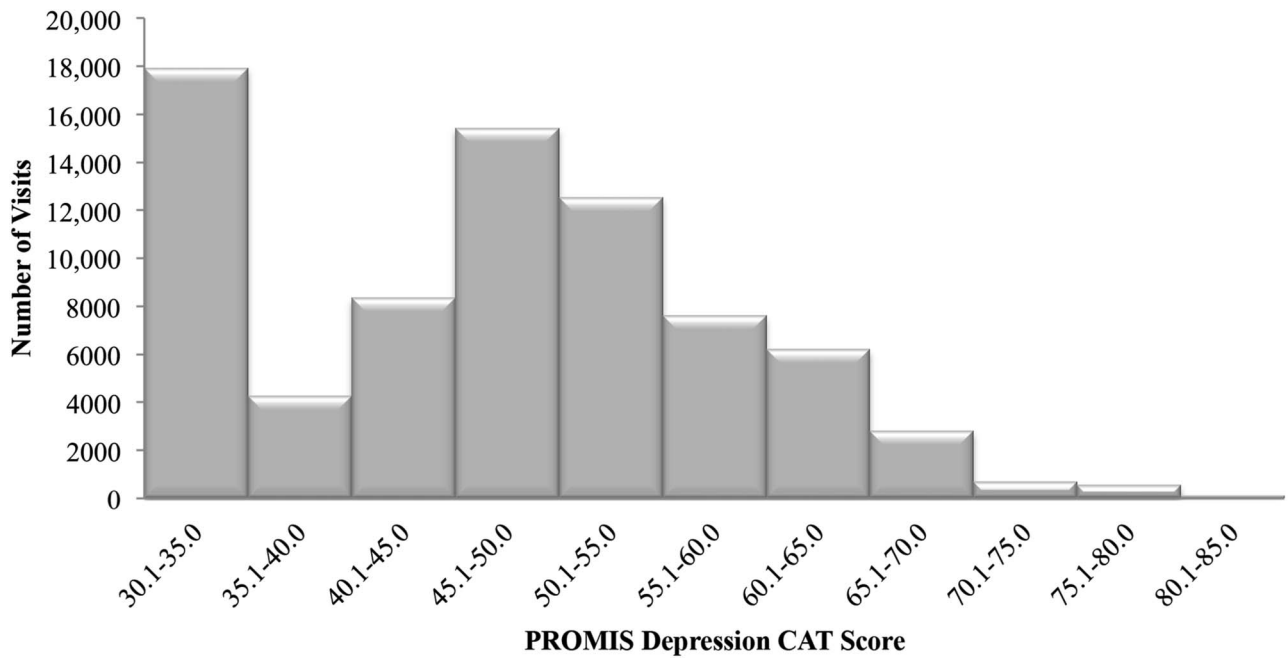


Fig. 2 The graph shows a frequency distribution of scores for the PROMIS Depression CAT in the study population.

(Floor Group: 9190 [49%]; Standard Group: 23,362 [40%]; odds ratio: 0.6 [0.6-0.7]). There were only minor differences in the racial composition of the two groups (Floor Group: 15,291 [85%] white, 2137 [12%] black, and 515 [3%]; Standard Group: 49,195 [84%] white, 8005 [14%] black, and 1431 (2%) other; $p < 0.001$) (Table 1).

Discussion

The development of the PROMIS tools provides clinicians a standardized metric to measure patient-reported outcomes independently of disease. In addition to a standardized metric, PROMIS item banks and CAT allow for greater precision in measurement while lowering patient

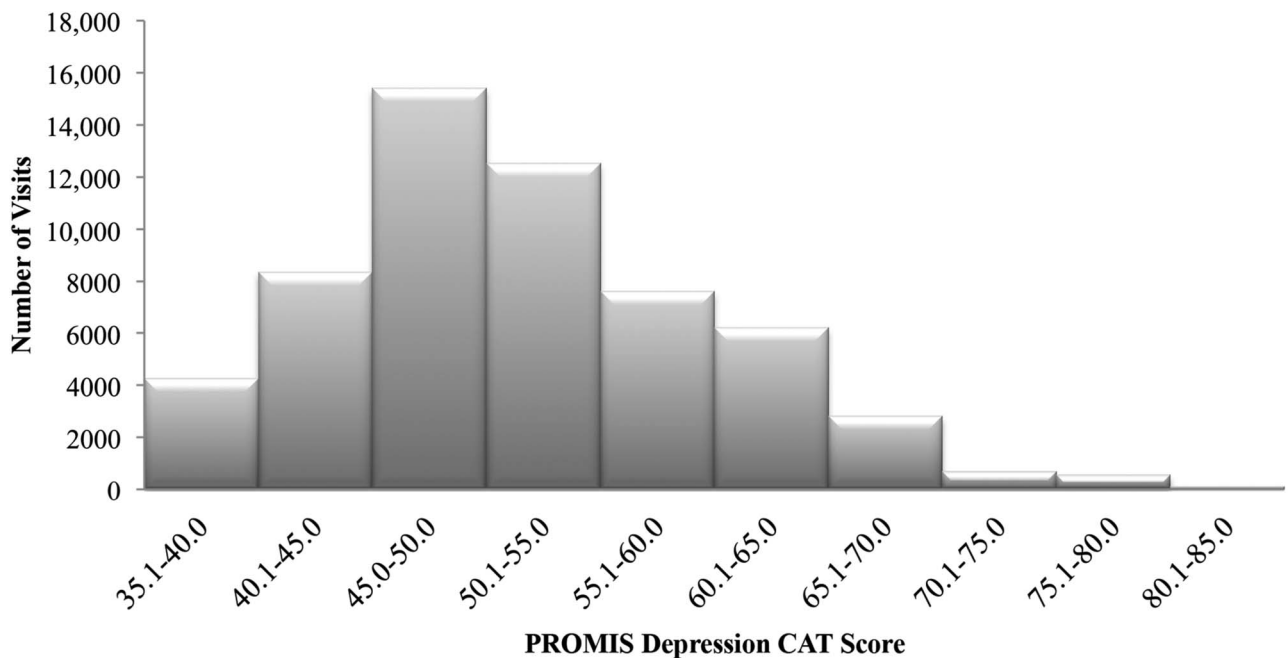


Fig. 3 The graph shows a frequency distribution of scores for the PROMIS Depression CAT in Group 1.

Table 2. TPQ between the standard and floor groups

Assessment	Standard group (N = 58,631)	Floor group (N = 17,943)	Mean difference (95% CI)	p value
Physical function (seconds)	12 ± 10	11 ± 9	1 (0.8-1.1)	< 0.001
Depression (seconds)	7 ± 7	4 ± 3	3 (2.9-3.2)	< 0.001

Values are mean ± SD.

TPQ = time per question; CI = confidence interval.

burden. Although the instruments have demonstrated validity and reliability in previous studies, these studies have involved controlled, voluntary, and potentially more motivated populations that may result in different performance than when the assessments are given to the general clinical population as part of routine practice. Examining the performance of the Physical Function and Depression CATs when given to the general orthopaedic population, we found that the Physical Function CAT performs as expected with a normal distribution; however, the Depression CAT exhibited a nonnormal score distribution that is likely a result of patients completing the assessment with a deliberate hastiness that is shown by the much faster mean TPQ Floor Group patients spent on the Depression CAT than they spent on the Physical Function CAT. This means TPQ for the Depression CAT is so low that we speculate that patients are speeding through the assessment that may be affecting the accuracy of the scores.

The initial analysis of these data is limited by the fact that the population examined only includes visits in which both the Physical Function and Depression CATs were completed. However, because the completion rate for the department was 97%, the number of excluded visits was small. The order in which the PROMIS CATs are delivered represents a potential confounding variable. The Depression CAT was given at the end of the assessment battery for the majority of the analysis timeframe and patient burden could contribute to some percentage of the results. However, this limitation may be unavoidable: the order of the assessments was determined in response to negative feedback given regarding the Depression CAT being at the beginning of the assessment battery. As a preemptive measure, to improve patient compliance, the Depression CAT was moved to the end of the assessment battery. However, we feel that responder burden is unlikely to be a substantial confounder because the mean total time spent to complete these assessments was only 2 minutes 55 seconds and required completion of fewer questions than traditionally administered to our patients when using static written patient-reported outcome surveys. Our study was also limited in that we only delivered the PROMIS CAT assessments. Although the measurement of the meant TPQ provides us with circumstantial evidence that the patients are completing the Depression CAT in a much more hasty

way than the Physical Function CAT, this limitation prevents us from concretely knowing whether the hastiness is a result of any deliberate subversion of the assessment. It is also possible that our findings could change had patients been provided an explanation regarding the importance or purpose of the depression questions, but we did not provide this because our research question focused on the delivery of this survey as routine care. Finally, we do not know if the same survey-taking behavior would manifest with use of an alternative depression screening tool. We did not deliver a second depression survey because linkage tables exist to crosswalk scores between PROMIS and the often referenced Patient Health Questionnaire-9 so there was no anticipated clinical benefit from screening for depressive symptoms with redundant surveys.

To our knowledge, there has not been any large-scale examination of the score distribution of the Depression CAT scores when given as part of routine practice and our analysis indicates a striking irregularity in the performance of the PROMIS Depression CAT when given to a population of outpatients seeking orthopaedic care. Rather than normally distributed scores, our patient score distribution has a secondary spike in the floor score. The anomaly in the distribution of CAT scores is limited to the Depression CAT and, within the study population, the performance of the Physical Function CAT appears normal, providing a normal distribution of scores that matches what is expected from the development of the instruments.

Patients contributing to the floor effect completed the questions in a hasty fashion by rapidly selecting the most symptom-minimizing answer. All of the visits in the floor score group required 12 questions to be answered with the most symptom-minimizing answer to get the floor score for the Depression CAT. Additionally, this answer option is also at the top of the list of possible response options shown to the patients for all of the Depression CAT questions. This layout makes rapid answering possible and we speculate this is how the Floor Group completed their assessments because they had such a lower mean TPQ for the Depression CAT than the Standard Group. We hypothesize, based on anecdotal evidence from patient responses during the implementation process, that potential reasons for this behavior would include concern

over the stigma of having depression recorded in their medical record, a possible perceived lack of relevance to the orthopaedic condition for which they are seeking care, or not wanting to have to discuss their mental health with their orthopaedic surgeon. When a patient comes in for a visit and scores 34.2 on the Depression CAT, the provider should not immediately presume that the patient has no depressive symptoms and should remain vigilant for clinical clues of depression that the patient may be reluctant to discuss.

Other than a greater proportion of men, our data suggest that patients contributing to this floor effect of depression do not seem to be substantially different than other patients with the exception of their answers on the Depression CAT and spending substantially less time per question (and with less variance in TPQ) on it. Demographic data show the age between of the Standard Group is 4 years older than the Floor Group; however, it is unlikely to play a large role in the TPQ spent because this difference does not separate the mean ages of these groups into distinct demographic categories that mirror any age cutoffs used to define clinically distinct groups in other orthopaedic or depression research. Adult orthopaedic research largely focuses on the differences between geriatric (65 years or older) and nongeriatric patients (< 65 years old) [16, 17, 23]. Additionally, research on Major Depressive Episodes (MDEs) and demographic factors has found some mixed evidence for differences in the incidence of major depressive episodes, but this link between age and incidence of MDEs varies based on other demographic factors such as the country measuring the MDEs [6]. One notable demographic difference between the two groups was composition of the groups broken down by gender. The Floor Group consisted of more men. The higher prevalence of men in the Floor Group supports other research indicating that men have a greater degree of reluctance to discuss or seek consultation for depressive symptoms than women [12, 27, 31].

Initial data compiled from nearly 80,000 patient encounters over the first 13 months of PROMIS delivery at one tertiary center indicate that approximately 20% of the orthopaedic population may not accurately complete the Depression CAT. Because the Depression CAT has been found to be valid and reliable, it is unlikely this is a reflection on the questions within the item bank of the CAT. It is more likely this substantial floor effect is the result of other patient-related reasons behind a reluctance to report depressive symptoms or to carefully consider questions regarding depression. In future research, it will be necessary to perform further investigation into those patients contributing to the floor effect to determine which patients' scores are unreliable and how to manage the interpretation of those PROMIS Depression scores in both clinical practice and research.

References

1. Alattas SA, Smith T, Bhatti M, Wilson-Nunn D, Donell S. Greater pre-operative anxiety, pain and poorer function predict a worse outcome of a total knee arthroplasty. *Knee Surg Sports Traumatol Arthrosc.* 2016 Oct 12. [Epub ahead of print]
2. Ali A, Lindstrand A, Sundberg M, Flivik G. Preoperative anxiety and depression correlate with dissatisfaction after total knee arthroplasty: a prospective longitudinal cohort study of 186 patients, with 4-year follow-up. *J Arthroplasty.* 2017;32:767–770.
3. Amtmann D, Kim J, Chung H, Bamer AM, Askew RL, Wu S, Cook KF, Johnson KL. Comparing CESD-10, PHQ-9, and PROMIS depression instruments in individuals with multiple sclerosis. *Rehabil Psychol.* 2014;59:220–229.
4. Bjorner JB, Rose M, Gandek B, Stone AA, Junghaenel DU, Ware JE Jr. Method of administration of PROMIS scales did not significantly impact score level, reliability, or validity. *J Clin Epidemiol.* 2014;67:108–113.
5. Blackburn J, Qureshi A, Amirfeyz R, Bannister G. Does pre-operative anxiety and depression predict satisfaction after total knee replacement? *Knee.* 2012;19:522–524.
6. Bromet E, Andrade LH, Hwang I, Sampson NA, Alonso J, de Girolamo G, de Graaf R, Demyttenaere K, Hu C, Iwata N, Karam AN, Kaur J, Kostyuchenko S, Lepine JP, Levinson D, Matschinger H, Mora ME, Browne MO, Posada-Villa J, Viana MC, Williams DR, Kessler RC. Cross-national epidemiology of DSM-IV major depressive episode. *BMC Med.* 2011;9:90.
7. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, Amtmann D, Bode R, Buysse D, Choi S, Cook K, Devellis R, DeWalt D, Fries JF, Gershon R, Hahn EA, Lai JS, Pilkonis P, Revicki D, Rose M, Weinfurt K, Hays R; PROMIS Cooperative Group. 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;63:1179–1194.
8. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, Ader D, Fries JF, Bruce B, Rose M; PROMIS Cooperative Group. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. *Med Care.* 2007;45 (Suppl 1):S3–S11.
9. Centers for Medicare & Medicaid Services. 2014 Clinical Quality Measures (CQMs) Adult Recommended Core Measures. Available at: https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Downloads/2014_cqm_Adult_Recommend_CoreSetTable.pdf. Accessed October 18, 2016.
10. Chen J, Ou L, Hollis SJ. A systematic review of the impact of routine collection of patient reported outcome measures on patients, providers and health organisations in an oncologic setting. *BMC Health Serv Res.* 2013;13:211.
11. Choi SW, Schalet B, Cook KF, Cella D. Establishing a common metric for depressive symptoms: linking the BDI-II, CES-D, and PHQ-9 to PROMIS Depression. *Psychol Assess.* 2014;26:513–527.
12. Corney RH. Sex differences in general practice attendance and help seeking for minor illness. *J Psychosom Res.* 1990;34:525–534.
13. Doring AC, Nota SP, Hageman MG, Ring DC. Measurement of upper extremity disability using the Patient-Reported Outcomes Measurement Information System. *J Hand Surg Am.* 2014;39:1160–1165.
14. Fries J, Witter J, Rose M, Cella D, Khanna D, Morgan-DeWitt E. Item response theory, computerized adaptive testing, and

- PROMIS: assessment of physical function. *J Rheumatol*. 2014; 41:153–158.
15. Golkari S, Teunis T, Ring D, Vranceanu AM. Changes in depression, health anxiety, and pain catastrophizing between enrollment and 1 month after a radius fracture. *Psychosomatics*. 2015; 56:652–657.
 16. Hinds RM, Capo JT, Kakar S, Roberson J, Gottschalk MB. Early complications following osteosynthesis of distal radius fractures: a comparison of geriatric and nongeriatric cohorts. *Geriatr Orthop Surg Rehabil*. 2017;8:30–33.
 17. Hinds RM, Garner MR, Tran WH, Lazaro LE, Dines JS, Lorch DG. Geriatric proximal humeral fracture patients show similar clinical outcomes to non-geriatric patients after osteosynthesis with endosteal fibular strut allograft augmentation. *J Shoulder Elbow Surg*. 2015;24:889–896.
 18. Ho B, Houck JR, Flemister AS, Ketz J, Oh I, DiGiovanni BF, Baumhauer JF. Preoperative PROMIS scores predict postoperative success in foot and ankle patients. *Foot Ankle Int*. 2016; 37:911–918.
 19. Hung M, Baumhauer JF, Brodsky JW, Cheng C, Ellis SJ, Franklin JD, Hon SD, Ishikawa SN, Latt LD, Phisitkul P, Saltzman CL, SooHoo NF, Hunt KJ; Orthopedic Foot & Ankle Outcomes Research (OFAR), Orthopedic Foot & Ankle Society (AOFAS). Psychometric comparison of the PROMIS Physical Function CAT with the FAAM and FFI for measuring patient-reported outcomes. *Foot Ankle Int*. 2014;35:592–599.
 20. Hung M, Franklin JD, Hon SD, Cheng C, Conrad J, Saltzman CL. Time for a paradigm shift with computerized adaptive testing of general physical function outcomes measurements. *Foot Ankle Int*. 2014;35:1–7.
 21. 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 orthopedic trauma patients. *J Orthop Trauma*. 2014;28:439–444.
 22. Li S, Qi M, Yuan W, Chen H. The impact of the depression and anxiety on prognosis of cervical total disc replacement. *Spine (Phila Pa 1976)*. 2015;40:E266–271.
 23. Little MT, Berkes MB, Lazaro LE, Sculco PK, Cymerman RM, Pardee N, Helfet DL, Lorch DG. Comparison of supination external rotation type IV ankle fractures in geriatric versus non-geriatric populations. *Foot Ankle Int*. 2013;34:512–517.
 24. London DA, Stepan JG, Boyer MI, Calfee RP. The impact of depression and pain catastrophization on initial presentation and treatment outcomes for atraumatic hand conditions. *J Bone Joint Surg Am*. 2014;96:806–814.
 25. Mayo BC, Massel DH, Bohl DD, Narain AS, Hijji FY, Long WW, Modi KD, Basques BA, Yacob A, Singh K. Preoperative mental health status may not be predictive of improvements in patient-reported outcomes following an anterior cervical discectomy and fusion. *J Neurosurg Spine*. 2017;26:177–182.
 26. Menendez ME, Bot AG, Hageman MG, Neuhaus V, Mudgal CS, Ring D. Computerized adaptive testing of psychological factors: relation to upper-extremity disability. *J Bone Joint Surg Am*. 2013;95:e149.
 27. Moller-Leimkuhler AM. Barriers to help-seeking by men: a review of sociocultural and clinical literature with particular reference to depression. *J Affect Disord*. 2002;71:1–9.
 28. Oflazoglu K, Mellema JJ, Menendez ME, Mudgal CS, Ring D, Chen NC. Prevalence of and factors associated with major depression in patients with upper extremity conditions. *J Hand Surg Am*. 2016;41:263–269.e1-7.
 29. Oude Voshaar MA, Ten Klooster PM, Glas CA, Vonkeman HE, Krishnan E, van de Laar MA. Relative performance of commonly used physical function questionnaires in rheumatoid arthritis and a patient-reported outcomes measurement information system computerized adaptive test. *Arthritis Rheumatol*. 2014;66: 2900–2908.
 30. Pilkonis P, Choi S, Reise SP, Stover AM, Riley W, Cella D. Item banks for measuring emotional distress from the Patient Reported Outcomes Measurement Information System (PROMIS): depression, anxiety, and anger. *Assessment*. 2011;18:263–283.
 31. Rickwood DJ, Braithwaite VA. Social-psychological factors affecting help-seeking for emotional problems. *Soc Sci Med*. 1994; 39:563–572.
 32. Rose M, Bjorner JB, Gandek B, Bruce B, Fries JF, Ware JE Jr. The PROMIS Physical Function item bank was calibrated to a standardized metric and shown to improve measurement efficiency. *J Clin Epidemiol*. 2014;67:516–526.
 33. Schalet BD, Pilkonis P, Yu L, Dodds N, Johnston KL, Yount S, Riley W, Cella D. Clinical validity of PROMIS Depression, Anxiety, and Anger across diverse clinical samples. *J Clin Epidemiol*. 2016;73:119–127.
 34. Senders A, Hanes D, Bourdette D, Whitham R, Shinto L. Reducing survey burden: feasibility and validity of PROMIS measures in multiple sclerosis. *Mult Scler*. 2014;20:1102–1111.
 35. Siu AL, US Preventive Services Task Force, Bibbins-Domingo K, Grossman DC, Baumann LC, Davidson KW, Ebell M, Garcia FA, Gillman M, Herzstein J, Kemper AR, Krist AH, Kurth AE, Owens DK, Phillips WR, Phipps MG, Pignone MP. Screening for depression in adults: US Preventive Services Task Force Recommendation Statement. *JAMA*. 2016;315:380–387.
 36. Skeppholm M, Fransson R, Hammar M, Olerud C. The association between preoperative mental distress and patient reported outcome measures in patients treated surgically for cervical radiculopathy. *Spine J*. 2016 Mar 22. [Epub ahead of print]
 37. Tyser AR, Beckmann J, Franklin JD, Cheng C, Hon SD, Wang A, Hung M. Evaluation of the PROMIS physical function computer adaptive test in the upper extremity. *J Hand Surg Am*. 2014; 39:2047–2051.e4.
 38. Wagner LI, Schink J, Bass M, Patel S, Diaz MV, Rothrock N, Pearman T, Gershon R, Penedo FJ, Rosen S, Cella D. Bringing PROMIS to practice: brief and precise symptom screening in ambulatory cancer care. *Cancer*. 2015;121:927–934.
 39. Yost KJ, Eton DT, Garcia SF, Cella D. Minimally important differences were estimated for six Patient-Reported Outcomes Measurement Information System-Cancer scales in advanced-stage cancer patients. *J Clin Epidemiol*. 2011;64:507–516.