

Clinical Research

Determining the Generalizability of the PROMIS Depression Domain's Floor Effect and Completion Time in Patients Undergoing Orthopaedic Surgery

David N. Bernstein MBA, MA, Joshua Atkinson BS, Kathleen Fear PhD, Judith F. Baumhauer MD, MPH, Addisu Mesfin MD, Paul T. Rubery MD, Warren C. Hammert MD

Received: 10 January 2019 / Accepted: 2 April 2019 / Published online: 23 July 2019
Copyright © 2019 by the Association of Bone and Joint Surgeons

Abstract

Background Depression is common and can decrease the likelihood that a patient undergoing an orthopaedic procedure will improve after that intervention. Research at a single institution using the Patient-reported Outcomes Measurement Information System (PROMIS) showed that the PROMIS Depression domain had a substantial floor effect (lowest depressive symptoms measurable) in patients with orthopaedic conditions, although the instrument is designed to follow a population-based normal distribution. Patients with the lowest scores (floor) completed the questionnaire more quickly than those who did not have the lowest scores, which suggests that patients may not want to report or discuss depressive symptoms with their orthopaedic surgeon.

Questions/purposes (1) Do the scores of a large population of patients with orthopaedic conditions completing the PROMIS Physical Function (PF), Pain Interference (PI), and Depression questionnaires follow the normal population distribution used to calibrate the PROMIS instrument? (2) Do the scores of a large population of patients with foot and ankle, spine, upper extremity, multiple orthopaedic subspecialty, or other orthopaedic conditions completing the PROMIS Depression questionnaire follow the normal population distribution used to calibrate the PROMIS instrument? (3) Do all orthopaedic patients with the lowest possible score (floor effect) on the PROMIS Depression questionnaire answer questions more quickly than those who do not?

DNB reports grants from Alpha Omega Alpha outside the submitted work; JFB reports other sources of support from PROMS Health Organization outside the submitted work; AM reports grants from AO Spine, grants from OMeGA, grants from CoreLink, and grants from the LES Society outside the submitted work; PTR reports grants from AO Spine outside the submitted work; WCH reports being an ASSH Hand-e Editor outside of the submitted work; JA and KF certify that they have no commercial associations (e.g. consultancies, stock ownership, equity interest, patent/licensing arrangements, etc.) that might pose a conflict of interest in connection with the submitted article. Each author certifies that his or her institution approved the human protocol for this investigation and that all investigations were conducted in conformity with ethical principles of research. The work was performed at the Department of Orthopaedics and Rehabilitation, University of Rochester Medical Center, Rochester, NY, USA.

D. N. Bernstein, J. Atkinson, University of Rochester School of Medicine and Dentistry, Rochester, NY, USA

K. Fear, University of Rochester Medical Center, Rochester, NY, USA

J. F. Baumhauer, MD, P.T. Rubery, A. Mesfin, W.C. Hammert, Department of Orthopaedics and Rehabilitation, University of Rochester Medical Center, Rochester, NY, USA

Warren C. Hammert MD (✉), Department of Orthopaedics and Rehabilitation, University of Rochester Medical Center, 601 Elmwood Ave, Box 665, Rochester, NY 14642, Email: Warren_Hammert@urmc.rochester.edu

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.

Methods Between February 12, 2015 and July 19, 2018, a total of 513,042 patients with orthopaedic conditions at a single urban, tertiary academic medical center were asked to complete the PROMIS PF, PI, and Depression domains. The PROMIS PF, PI, and Depression questionnaires were administered at 433,939 patient encounters (85%), and 325,251 of 433,939 (75%) completed all three questionnaires and were included in our final sample. The remaining encounters were excluded from all analyses because of incomplete PROMIS data. Using this large sample size, we determined the distributions of the PROMIS PF, PI, and Depression domains for all patients with orthopaedic conditions by visual inspection. We calculated the percentage of patients with the lowest possible score on the PROMIS Depression domain (“floor group”) for all patients with orthopaedic conditions, as well for patients presenting to orthopaedic subspecialists (foot and ankle, spine, upper extremity, multiple subspecialties, or other specialists). We compared patient and PROMIS characteristics between patients in the floor group and all others (“standard group”) using a bivariate analysis, including the total time it took to complete the questionnaires and average time it took to answer each question.

Results Of the three PROMIS domains analyzed, only the PROMIS PF demonstrated a normal distribution on visual inspection. PROMIS PI scores had a measurable ceiling effect (more pain symptoms) (15,520 of 325,251 patient encounters; 5%), while PROMIS Depression scores demonstrated a clear floor effect (65,226 of 325,251; 20%). When analyzed by subspecialty, there was a clear floor effect for the PROMIS Depression in the specialties of foot and ankle (6031 of 29,273 patient encounters; 21%), spine (7611 of 58,266; 13%), upper extremity (9851 of 42,864; 23%), multiple subspecialties (276 of 1662; 17%), and other (41,454 of 193,185; 22%). Generally, those in the floor group spent less time per question on the PROMIS Depression questionnaire than did those in the standard group (5.3 seconds \pm 2.6 seconds versus 8.1 seconds \pm 4.6 seconds; mean difference, 2.8 seconds; $p < 0.001$).

Conclusions The substantial floor effect of the PROMIS Depression in the setting of hasty completion raises the question of whether such results truly indicate that these patients have minimal depression or whether they simply do not report their true mental health status in an orthopaedic setting. When such patients are seen in clinic, surgeons may benefit from using the PROMIS Depression as an educational opportunity to explain the importance of mental health in orthopaedic care and ensure the questionnaire is completed accurately. This study builds on previous work by suggesting these findings apply to other general and subspecialty orthopaedic patient populations at academic medical centers. Future work may seek to determine the best way of ensuring that the PROMIS

Depression questionnaire is completed accurately in orthopaedic clinics.

Level of Evidence Level II, diagnostic study.

Introduction

Depression is predicted to outpace ischemic heart disease as the most common cause of disease worldwide by 2030 [28]. Currently, in the US, depression has a 12-month prevalence of 10.4% among adults [14]. Within orthopaedic surgery, the prevalence of depressive symptoms has been shown to be as high as 45% in an orthopaedic trauma population [9]. Depression and other psychological factors have been shown to be associated with poorer patient-reported outcome measures (PROMs) for satisfaction [1], physical function [18, 26], and perceived disability [22] in orthopaedic patients. In athletes, depression has likewise been associated with delays in return to play after ACL reconstruction surgery [3]. These findings demonstrate a need to determine how to measure depression, which exists on a continuum and impacts functional outcomes along a continuum as well, accurately in patients who undergo orthopaedic surgery. Only then can appropriate interventions be introduced to address mental health concerns in the setting of orthopaedics.

A recent study, which we seek to replicate with an independent sample, began to evaluate whether routinely collected PROMs accurately measure depression in patients who undergo orthopaedic surgery [13]. The authors used the Patient-reported Outcomes Measurement Information System (PROMIS), a validated, general PROM developed by the National Institutes of Health in 2004 that uses computerized adaptive testing based on item response theory [7]. Despite an underlying assumption that the PROMIS follows a population-derived, normal distribution, in that previous analysis of nearly 80,000 patients who visited an orthopaedic clinic, the authors found that a large group of patients reported the minimal number of depressive symptoms (that is, the floor effect). This differed from the distribution used to calibrate the PROMIS Depression domain [13] and suggests depressive symptoms may not have been accurately measured. The same patients also completed the PROMIS Depression domain in less time. Hasty completion, which was associated with a clear floor effect, may indicate that patients were generally reluctant to complete the questionnaire and/or discuss mental health with orthopaedic surgeons, which is worrisome because of the impact depression can have on functional outcomes, as well as PROMs. If a patient's depression status cannot be assessed accurately, orthopaedic surgeons may miss an opportunity to improve care and outcomes. Replication is important to good science because it helps to ensure the generalizability and

consistency of results [21]. Given the association between depression and health outcomes and the importance of PROMs as the US’s health care system transitions to a system that rewards value—defined as health outcome achieved per dollar spent [25]—we believe it is critical to independently validate Guattery et al.’s [13] findings to determine their generalizability. In addition to replicating portions of this previous study, we also sought to answer whether other PROMIS domains (Physical Function [PF] and Pain Interference [PI]) demonstrated similar floor effects and whether the PROMIS Depression floor effect was present in orthopaedic subspecialties.

Therefore, we asked: (1) Do the scores of a large population of patients with orthopaedic conditions completing the PROMIS PF, PI, and Depression questionnaires follow the normal population distribution used to calibrate the PROMIS instrument? (2) Do the scores of a large population of patients with foot and ankle, spine, upper extremity, multiple orthopaedic subspecialty, or other orthopaedic conditions completing the PROMIS Depression questionnaire follow the normal population distribution used to calibrate the PROMIS instrument? (3) Do all orthopaedic patients with the lowest possible score (floor effect) on the PROMIS Depression questionnaire answer questions more quickly than those who do not?

Patients and Methods

Patient Sample and Data Collection

The appropriate institutional review board approved this retrospective study of PROMs collected as part of routine

practice. Between February 12, 2015 and July 19, 2018, we asked all patients 18 years or older seeking orthopaedic care at a single urban, tertiary academic medical center to complete the PROMIS PF, PI, and Depression domains as part of routine clinical care [24]. PROMIS questionnaires were offered in either English or Spanish based on patient preference.

A total of 513,042 patient encounters were eligible for inclusion. All three PROMIS domains of interest (PF, PI, and Depression) were administered in 433,939 of the 513,042 patient encounters (85%). A total of 82,593 of 433,939 patients (19%) declined to answer or ignored the PROMIS questionnaires at the onset, and an additional 17,690 of the 433,939 patients (4%) stopped answering the PROMIS questionnaires before beginning the PROMIS Depression domain. A total of 333,656 of 433,939 (77%) PROMIS PF, PI, and Depression administrations ultimately reached the PROMIS Depression domain. However, for an additional 1857 of 433,939 patient encounters (0.4%), the PROMIS Depression domain was reached but the patient ultimately declined to fill it out or ignored or skipped it. A final 6548 of 433,939 patient encounters (2%) were removed because the PROMIS Depression domain was started but not completed. Our final sample for all analyses was 325,251 of 433,939 patient encounters (75%) (Fig. 1). This final sample was larger than the sample of an earlier single-institution study [14] by one-quarter of a million patient encounters. Using Apple iPads (Apple, Cupertino, CA, USA), patients completed PROMIS computerized adaptive tests, which have been shown to be superior to traditional pen and paper for collecting PROM data [30]. PROMIS computerized adaptive tests were completed in the following order each time: PF, PI, and

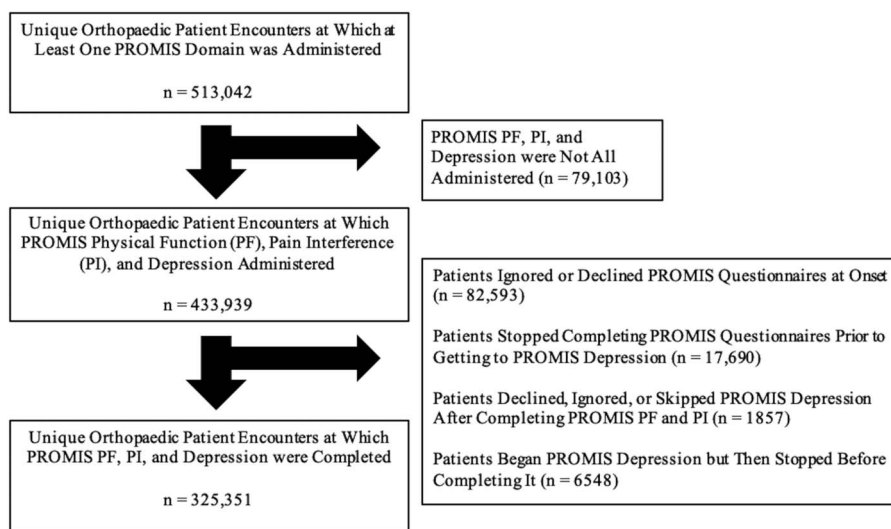


Fig. 1 This STROBE diagram illustrates the pathway from the total possible patient sample to the final number used for all analyses.

Depression. Patients were able to complete the PROMIS domains in the waiting room while waiting to be called to the examination room. At our institution, we have found that routinely collecting PROMIS questionnaires allows us to have a high proportion of patients complete the surveys as part of routine clinical care [24]. The time to complete each PROMIS questionnaire is gathered automatically via our self-designed institutional software that integrates PROMIS data directly into each patient's electronic health record [4]. This approach to gathering PROMIS data was reviewed and optimized in orthopaedic surgery before being offered to all clinical ambulatory services at our institution [4]. PROMIS questionnaires have been shown to be efficient [10, 12, 16, 17, 20, 27, 29] and reliable [16, 17], and are correlated strongly with historically used PROMs, such as the DASH [2, 5, 6, 8, 20, 22, 23, 29]. For the PROMIS PF domain, higher scores indicate better physical function, while lower scores suggest the opposite. For the PROMIS PI and Depression scales, higher scores indicate decreased activity because of pain and poorer mental health, respectively.

For our analyses, we included only patients who completed all the requested PROMIS questionnaires (PF, PI, and Depression). We excluded patients who filled out every questionnaire but skipped questions for one or more domains.

Only a small number of patients completed the Spanish version of the PROMIS questionnaire (724 of 325,251 patient encounters; 0.22%) (Table 1). PROMIS Depression version 1.0, PI version 1.1, and PF version 1.2 or 2.0 were used (Table 1). Our patients had a mean age of 53 years \pm 17 years and most were women (57%) (Table 2). Patients who were white constituted the largest portion of our sample (87%). The mean overall PROMIS PF, PI, and Depression scores were 41 \pm 9, 60 \pm 8, and 49 \pm 10, respectively (Table 2). The patient characteristics of those who did not complete the PROMIS Depression domain were not different from those who did (Table 3).

Primary and Secondary Study Endpoints

Our primary study endpoint was to determine the distribution of PROMIS PF, PI, and Depression scores in an orthopaedic population to determine if these instruments follow a normal distribution, testing this by visualizing Q-Q plots (Fig. 2A-C). Given the large sample size in our study, determining normality by visualization is an appropriate approach [19].

Our secondary study endpoints were to determine if the floor effect of the PROMIS Depression remained when patient data were analyzed by orthopaedic subspecialty and

Table 1. The PROMIS domains and versions used

Domain	Instrument	Assessments	Total number of questions answered	Mean \pm SD number of questions answered per assessment	Mean \pm SD active duration per assessment (seconds)	Mean \pm SD active duration per answered question (seconds)
Depression	PROMIS banco version 1.0-depresión	724	3753	5.2 \pm 2.5	64 \pm 32	13 \pm 5.8
	PROMIS bank version 1.0-depression	324,527	2,137,667	6.6 \pm 3.5	46 \pm 29	7.5 \pm 4.4
	Total	325,251	2,141,420	6.6 \pm 3.5	46 \pm 29	7.6 \pm 4.4
Pain Interference	PROMIS banco version 1.1-efectos del dolor	724	3200	4.4 \pm 1.6	65 \pm 28	15 \pm 5.5
	PROMIS bank version 1.1-pain interference	324,527	1,464,711	4.5 \pm 1.8	44 \pm 23	10 \pm 4.6
	Total	325,251	1,467,911	4.5 \pm 1.8	44 \pm 23	10 \pm 4.6
Physical Function	PROMIS banco version 1.2-capacidad de funcionamiento físico	724	3092	4.3 \pm 1.2	71 \pm 29	17 \pm 5.2
	PROMIS bank version 1.2-physical function	221,912	963,425	4.3 \pm 1.2	53 \pm 23	12 \pm 4.9
	PROMIS bank version 2.0-physical function	102,615	437,173	4.3 \pm 1.0	52 \pm 23	12 \pm 4.8
	Total	325,251	1,403,690	4.3 \pm 1.2	53 \pm 23	12 \pm 4.9

Table 2. Patient and PROMIS domain characteristics

Characteristic	Number of patients
Total	325,251
Patient	
Age (years)	53 ± 17
Gender	
Men	139,519 (43%)
Women	185,731 (57%)
Unknown	1 (0%)
Race*	
White	281,315 (87%)
Black	31,613 (9.7%)
Other	10,899 (3.4%)
Unknown race	1424 (0.4%)
Evaluation	
Total active duration (seconds)	142 ± 58
Total questions answered	15 ± 4.7
Physical Function domain	
Score	41 ± 9
Total duration (seconds)	53 ± 23
Questions answered	4.3 ± 1.2
Time per question (seconds)	12 ± 4.9
Pain Interference domain	
Score	60 ± 8
Total duration (seconds)	44 ± 23
Questions answered	4.5 ± 1.8
Time per question (seconds)	10 ± 4.6
Depression domain	
Score	49 ± 10
Total duration (seconds)	46 ± 29
Questions answered	6.6 ± 3.5
Time per question (seconds)	7.6 ± 4.4

Data are presented as the mean ± SD.

*Determined from patient intake forms and located in patient records

to determine if patients with orthopaedic conditions who have the lowest possible PROMIS Depression score (floor effect) answered questions more quickly than those who did not have the lowest scores. We tested the secondary endpoint by determining the number of patients seeking foot and ankle, spine, upper extremity, multiple orthopaedic subspecialty, or other orthopaedic subspecialty care who had the lowest possible PROMIS Depression score (floor effect). We also tested the secondary endpoint by dividing the total time to complete the PROMIS Depression questionnaire in both the floor and standard groups by the number of questions answered; next, we compared the average time per question between the two groups.

Statistical Analysis

We calculated descriptive statistics for the complete patient population, including age, gender, race (as reported by patients in our electronic records), and characteristics of the completed PROMIS questionnaires. The PROMIS questionnaire characteristics for each domain included the mean score, total duration to complete the questionnaire (seconds), questions answered, and time per question.

We then divided the patients into two subgroups: one group included all patients for whom the lowest possible PROMIS Depression score was recorded, and the other group included all other patients. The group of patients with the lowest possible PROMIS Depression score was denoted as the “floor group,” while all other patients who did not have the lowest possible PROMIS Depression score were classified into the “standard group.” PROMIS domains, including the PROMIS Depression, were designed and validated following a population-based normal distribution with a mean t-score of 50 and SD of 10 [15]. This supports the idea that depression exists on a continuum. Similar to a study by Guattery et al. [13], we included patients with PROMIS Depression scores of or below 34.2 in the floor group. For categorical variables, we used chi-squared tests. For continuous variables, we used t-tests.

Patients in the floor group completed the PROMIS Depression questionnaire by answering “never” to 12 questions. We compared patients in the floor group with those in the standard group, as well as with a subgroup of patients in the standard group who did not have the lowest scores for the PROMIS Depression domain but nonetheless completed 12 questions to receive t-scores.

Results

When analyzing all patients with orthopaedic conditions, we found that the PROMIS PF scores demonstrated a normal distribution without a floor or ceiling effect. In contrast, PROMIS PI scores demonstrated a ceiling effect (15,520 of 325,251 patient encounters; 5%) and the PROMIS Depression demonstrated a floor effect (65,226 of 325,251 patient encounters; 20%).

When analyzing large populations of patients with foot and ankle, spine, upper extremity, multiple orthopaedic subspecialty, or other orthopaedic conditions completing the PROMIS Depression questionnaire, we found that there was a floor effect for the PROMIS Depression in all groups (foot and ankle: 6031 of 29,272 patient encounters [21%]; spine: 7611 of 58,265 [13%]; upper extremity: 9851 of 42,863 [23%]; multiple orthopaedic subspecialties: 276 of 1662 [17%]; other: 41,454 of 193,185 [22%]) (Table 4).

Table 3. Characteristics of patients who did and did not complete the PROMIS Depression domain

Characteristic	Completed	Declined the entire evaluation	Quit before reaching PROMIS Depression	Declined to fill out PROMIS Depression	Started but did not complete the PROMIS Depression
Total	325,251	82,593	17,690	1857	6535
Age (years; mean ± SD)	54 ± 17	60 ± 16	61 ± 17	59 ± 17	60 ± 17
Gender					
Men	139,520 (43%)	34,745 (42%)	7490 (42%)	792 (43%)	3043 (47%)
Women	185,731 (57%)	47,848 (58%)	10,200 (58%)	1065 (57%)	3492 (53%)
Unknown	1 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Race*					
White	281,316 (87%)	68,854 (83%)	14,480 (82%)	1627 (88%)	5676 (87%)
Black	31,613 (9.7%)	8432 (10%)	2313 (13%)	166 (8.9%)	591 (9.0%)
Other	10,899 (3.4%)	4696 (5.7%)	804 (4.5%)	57 (3.1%)	228 (3.5%)
Unknown	1424 (0.4%)	611 (0.7%)	93 (0.5%)	7 (0.4%)	40 (0.6%)

*Determined from patient intake forms and located in patient records

Generally, those in the floor group spent less time answering each question of the PROMIS Depression questionnaire than did those who were in the standard group (5.3 seconds ± 2.6 seconds versus 8.1 seconds ± 4.6 seconds, mean difference: 2.8 seconds [95% CI, 2.8-2.8 seconds]; $p < 0.001$) (Table 3). However, patients in the floor group answered more questions than those in the standard group did (12 questions versus 5.2 questions; $p < 0.001$) and took a longer time to complete the PROMIS Depression questionnaire (64 seconds ± 31 seconds versus 41 seconds ± 27 seconds, mean difference: 22 seconds [95% CI, 22-23 seconds]; $p < 0.001$) (Table 5).

Discussion

There is an increasing interest in using PROMs in orthopaedic surgery to improve patient satisfaction, increase shared decision-making, and provide surgeons with deeper insight into a patient's health that comes directly from the patient. Compared with historically used instruments such as the DASH, PROMIS is a universal PROM that has been shown to perform better than other PROMs in terms of decreased ceiling and floor effects, and it takes less time to complete [6, 10, 11]. Many previous studies have focused on the physical function domains [10, 11] and have not specifically evaluated the potential limitations of the PROMIS Depression domain. One previous study evaluated the performance of the PROMIS Depression and found a large floor effect across all patients with orthopaedic conditions [13]. Good science requires repetition to ensure that the results of one study are consistent across other independent samples. Thus, using a larger sample of patients with orthopaedic conditions, we sought to replicate

this previous work by determining the distribution of PROMIS Depression scores, as well as PROMIS PF and PI scores. In addition, we built on a prior study [13] by evaluating the distribution of PROMIS Depression scores in patients with foot and ankle, spine, upper extremity, multiple orthopaedic subspecialty, or other orthopaedic conditions. Lastly, replicating a prior study [13], we evaluated whether patients in the floor group completed the PROMIS Depression questionnaire faster than those in the standard group did. We found that the PROMIS Depression domain had a floor effect among all patients with orthopaedic conditions, as well as among patient scores when subanalyzed by orthopaedic subspecialty. Among all patients, the PROMIS PF domain demonstrated a normal distribution, while the PROMIS PI had a ceiling effect. Patients in the floor group for the PROMIS Depression completed the questionnaire in a shorter time than those in the standard group did. Overall, we believe our findings suggest that patients with the lowest possible PROMIS Depression scores may not be willing to discuss mental health concerns with orthopaedic surgeons or do not see the value of such a discussion in an orthopaedic setting. Alternatively, patients may experience fatigue from answering questions and be more inclined to answer without truly thinking about the most accurate response. When surgeons witness such a result, they may wish to ask patients about their mental health to ensure accurate reporting, because depression can negatively impact patient PROMs for satisfaction [1], physical function [18, 26], and perceived disability [22].

Our findings must be interpreted in consideration of the limitations of our work. First, we only included patients who were offered and ultimately completed all three PROMIS domain questionnaires (PF, PI, and Depression);

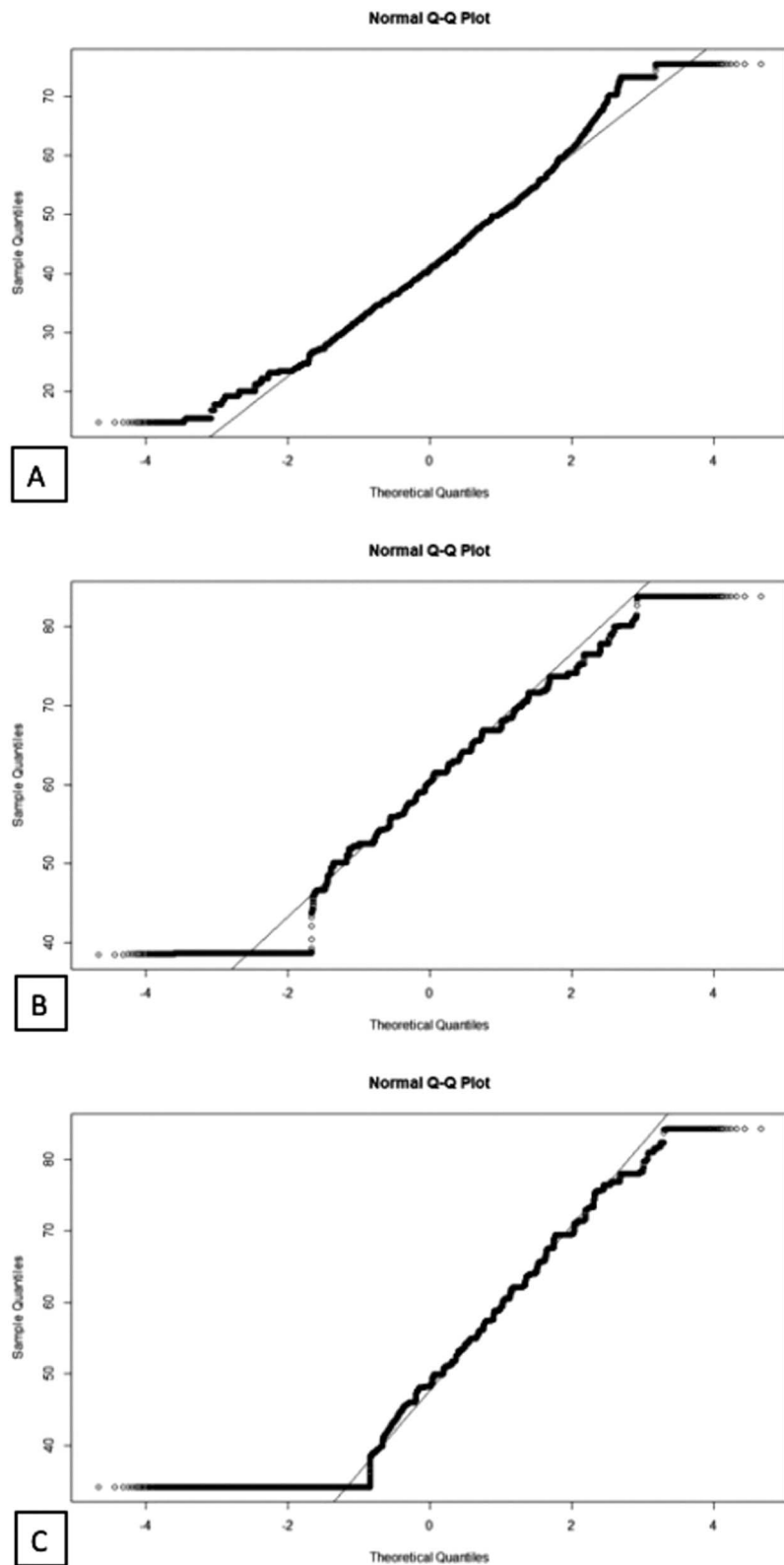


Fig. 2 This figure shows (A) a Q-Q plot for the PROMIS PF, which demonstrates a normal distribution; (B) a Q-Q plot for the PROMIS PI, which has a ceiling effect; and (C) a Q-Q plot for the PROMIS Depression, which has a floor effect.

Table 4. Floor effect of the PROMIS Depression by orthopaedic subspecialty

Subspecialty	Patients in floor group, n (%)
Foot and ankle (n = 29,272)	6031 (21)
Upper extremity (n = 42,863)	9851 (23)
Spine (n = 58,265)	7611 (13)
Multiple subspecialties (n = 1662)	276 (17)
Other (n = 193,185)	41,454 (22)

however, we included a large number of patients. We offered the three PROMIS questionnaires to 433,939 of the 513,042 patients (85%), 325,251 (75%) of whom

completed all of the domains. While our findings may have differed if the offer and completion rates had been 100%, we believe that this large number represents our patient population. Furthermore, even if all excluded patients completed the PROMIS Depression questionnaire and their results fell within the normal distribution, our finding of a floor effect would remain.

Second, at our institution, we ask patients to complete the PROMIS domains in the waiting room in the following order: PF, PI, and Depression. Thus, patients may have become fatigued while answering the PROMIS questionnaires or been concerned that other patients could see their answers. Although the total time to complete the Depression domain was greater for the floor group than for the standard group, patients in the floor group answered

Table 5. A comparison of patient and PROMIS domain characteristics by subgroup—standard group versus floor group

Characteristic	Overall population	Standard group	Floor group	Mean difference	p value
Total	325,251	260,021	65,226		
Patient demographics					
Age (years)	53 ± 17	54 ± 16	49 ± 18	4.7 (4.8-4.6)	< 0.001
Gender					
Men	139,519 (43%)	104,679 (40%)	34,840 (53%)		< 0.01
Women	185,731 (57%)	155,345 (60%)	30,386 (47%)		
Unknown	1 (0%)	1 (0%)	0 (0%)		
Race*					
White	281,315 (87%)	223,155 (86%)	58,160 (89%)		< 0.001
Black	31,613 (9.7%)	26,734 (10.3%)	4,879 (7.5%)		
Other	10,899 (3.4%)	8965 (3.4%)	1,934 (3%)		
Unknown	1424 (0.4%)	1171 (0.5%)	253 (0.4%)		
Evaluation					
Total active duration (seconds)	142 ± 58	137 ± 56	160 ± 63	23 (22.7-23.3)	< 0.001
Total questions answered	15 ± 4.7	14 ± 3.2	22 ± 3.9	8.3 (8.3-8.3)	< 0.001
Physical Function domain					
Score	41 ± 9.3	40 ± 8.8	46 ± 9.6	6.5 (6.4-6.6)	< 0.001
Total duration (seconds)	53 ± 23	53 ± 22	53 ± 25	0.2 (0.2-0.2)	0.96
Questions answered	4.3 ± 1.2	4.2 ± 0.9	4.7 ± 1.9	0.5 (0.5-0.5)	< 0.001
Time per question (seconds)	12 ± 4.9	13 ± 4.9	12 ± 4.6	1.1 (1.1-1.0)	< 0.001
Pain Interference domain					
Score	60 ± 8.4	61 ± 7.8	54 ± 8.4	6.9 (7.0-6.8)	< 0.001
Total duration (seconds)	44 ± 23	44 ± 22	44 ± 26	0.5 (0.2-0.7)	< 0.001
Questions answered	4.5 ± 1.8	4.3 ± 1.4	5.3 ± 2.9	1 (1-1.1)	< 0.001
Time per question (seconds)	10 ± 4.6	10 ± 4.7	8.8 ± 4.1	1.5 (1.5-1.5)	< 0.001
Depression					
Score	49 ± 10	52 ± 8.4	34.2 ± 0	18.2 (18.2-18.1)	< 0.001
Total duration (seconds)	46 ± 29	41 ± 27	64 ± 31	22.3 (22.1-22.6)	< 0.001
Questions answered	6.6 ± 3.5	5.2 ± 2.4	12 ± 0	6.8 (6.8-6.8)	< 0.001
Time per question (seconds)	7.6 ± 4.4	8.1 ± 4.6	5.3 ± 2.6	2.8 (2.8-2.8)	< 0.001

Data are presented as the mean ± SD.

*Determined from patient intake forms and located in patient records

more questions and took less time to answer each question. This may indicate that patients completed the PROMIS Depression domain inaccurately or hastily without much consideration. However, the PROMIS is completed more quickly than historically used instruments such as the DASH are, and because patients complete it on an iPad, we do not feel that concerns about fatigue or privacy likely caused our results.

Third, the floor effect of the PROMIS Depression may differ based on procedure and/or diagnosis. Although we did not conduct such an analysis in our study, we believe future research can seek to determine if this is the case. Nonetheless, we believe that our study offers valuable insight at the aggregate and subspecialty levels, the latter of which has not been reported to our knowledge.

Finally, we are unsure if the mode of questionnaire completion (iPad) or order in which the domains were completed affected the results. Although we were unable to evaluate the socioeconomic status of the patients who did not complete all three domains, we feel that such factors may have affected whether the questionnaires were completed using technology. However, because of the high completion rate and large sample, we feel this did not have a large impact on our overall findings.

Among all patients in our study, we found a floor effect for the PROMIS Depression and ceiling effect for the PROMIS PI; however, the PROMIS PF was normally distributed. This research builds on prior work by Guattery et al. [13] analyzing 76,574 patient visits to determine the generalizability of their findings. The current study used a larger independent sample to answer similar questions. Our finding of a floor effect for the PROMIS Depression in 20% of an orthopaedic patient population is consistent with the results of Guattery et al. [13], who found that 23% of their sample were in the floor group. The consistency of our PROMIS Depression findings with Guattery et al.'s [13] suggests that our results are generalizable to other large academic medical centers in the US. However, our study is different because we also reported that a proportion of patients had the highest possible score on the PROMIS PI domain. This suggests a ceiling effect, because higher PROMIS PI scores are associated with an increased effect of pain on function, and there is no larger effect that pain can have on function at this level. Future work can seek to understand if certain characteristics better define this cohort of patients; this knowledge may help surgeons understand who may or may not answer PROMIS PI appropriately.

To build on Guattery et al.'s work [13], we analyzed the distribution of the PROMIS Depression scores by orthopaedic subspecialty (foot and ankle, spine, upper extremity, multiple orthopaedic subspecialty, or other orthopaedic conditions) and found similar findings to that seen in the overall orthopaedic population. Although we found a floor

effect in patients with general spine conditions in this study, previous research analyzing 227 visits of patients with spine deformities found no such floor effect for the PROMIS Depression [6]. One possible reason for this finding may be that those with a spine deformity are more apt to discuss all elements of their health in an orthopaedic setting because of their more frequent engagement with the health care system compared with patients with general spine conditions. We are not aware of similar research in smaller patient populations in different orthopaedic subspecialties, apart from that study on patients with spinal deformities [6]. However, the difference found in a subset of patients with spine conditions suggests that differences may exist, and further research is warranted.

When we evaluated the mean time to complete each PROMIS Depression question, we found that patients hastily completed questions in a similar fashion to that shown in prior work [13]. In our opinion, this may mean one of two things: patients may not be willing to discuss their mental health with orthopaedic surgeons or patients do not see the need to discuss their mental health with their orthopaedic surgeon. Regardless of the underlying reason, it is crucial that surgeons explain to patients why it is important for the surgeon to have a sense of the patient's mental health because it can affect orthopaedic clinical outcomes such as function and pain. Further, it is crucial that surgeons provide an open and supportive environment for mental health discussions because there continues to be a stigma around mental illness. In general, we believe that demonstrating compassion and empathy in addition to providing patient education can help break down barriers to discussing mental health in the orthopaedic clinic and in order to collect accurate PROMIS Depression scores.

In a large orthopaedic patient sample, we demonstrated that the PROMIS PI and PROMIS Depression did not follow a normal distribution despite being designed to do so. Among all patients with orthopaedic conditions, the PROMIS PI had a ceiling effect, while the PROMIS Depression had a floor effect. However, the PROMIS PF followed a normal distribution. When analyzed by orthopaedic subspecialty, the PROMIS Depression continued to have a floor effect, and patients in the floor group completed the PROMIS Depression questionnaire faster than those in the standard group did. Given these findings, surgeons who use the PROMIS Depression to gauge a patient's mental health status may wish to take additional time to ask patients with the lowest possible score if they completed the questionnaire as accurately as possible. To avoid accusation, we believe it is crucial for surgeons to use this opportunity as a teaching moment and explain why it is important for the surgeon to know the patient's mental health status because it may influence functional and pain outcomes. Indeed, depression and its impact on physical limitations fall on a continuum and surgeons should

emphasize this fact during patient encounters. Future work can also help determine whether certain characteristics of patients with orthopaedic conditions tend to be associated with more severe depressive symptoms; this knowledge may also help remind surgeons to ask certain patients about depressive symptoms. While we speculate that patients may be reluctant to discuss depression when seeking orthopaedic care because of stigma or because they do not feel there is an important relationship between depression and physical health, future research is needed to determine if this is true.

Acknowledgment We thank the University of Rochester Health Laboratory and the faculty of the Department of Orthopaedics and Rehabilitation, University of Rochester.

References

- Adogwa O, Parker SL, Shau DN, Mendenhall SK, Bydon A, Cheng JS, Asher AL, McGirt MJ. Preoperative Zung depression scale predicts patient satisfaction independent of the extent of improvement after revision lumbar surgery. *Spine J*. 2013;13:501-506.
- 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.
- Ardem CL, Taylor NF, Feller JA, Whitehead TS, Webster KE. Psychological responses matter in returning to preinjury level of sport after anterior cruciate ligament reconstruction surgery. *Am J Sports Med*. 2013;41:1549-1558.
- Baumhauer JF, Bozic KJ. Value-based Healthcare: Patient-reported outcomes in clinical decision making. *Clin Orthop Relat Res*. 2016;474:1375-1378.
- Bernstein DN, Kelly M, Houck JR, Ketz JP, Flemister AS, DiGiovanni BF, Baumhauer JF, Oh I. PROMIS Pain Interference is superior vs numeric pain rating scale for pain assessment in foot and ankle patients. *Foot Ankle Int*. 2019;40:139-144.
- Bernstein DN, Papuga MO, Sanders JO, Rubery PT, Menga EN, Mesfin A. Evaluating the correlation and performance of PROMIS to SRS questionnaires in adult and pediatric spinal deformity patients. *Spine Deform*. 2019;7:118-124.
- Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, Ader D, Fries JF, Bruce B, Rose M. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. *Med Care*. 2007;45:S3-S11.
- 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.
- Crichlow RJ, Andres PL, Morrison SM, Haley SM, Vrahas MS. Depression in orthopaedic trauma patients. Prevalence and severity. *J Bone Joint Surg Am*. 2006;88:1927-1933.
- 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.
- Gausden EB, Levack A, Nwachukwu BU, Sin D, Wellman DS, Lorch DG. Computerized adaptive testing for patient reported outcomes in ankle fracture surgery. *Foot Ankle Int*. 2018;39:1192-1198.
- Gausden EB, Levack AE, Sin DN, Nwachukwu BU, Fabricant PD, Nellestein AM, Wellman DS, Lorch DG. Validating the Patient Reported Outcomes Measurement Information System (PROMIS) computerized adaptive tests for upper extremity fracture care. *J Shoulder Elbow Surg*. 2018;27:1191-1197.
- Guattery JM, Dardas AZ, Kelly M, Chamberlain A, McAndrew C, Calfee RP. Floor Effect of PROMIS Depression CAT associated with hasty completion in orthopaedic surgery patients. *Clin Orthop Relat Res*. 2018;476:696-703.
- Hasin DS, Sarvet AL, Meyers JL, Saha TD, Ruan WJ, Stohl M, Grant BF. Epidemiology of adult DSM-5 major depressive disorder and its specifiers in the United States. *JAMA Psychiatry*. 2018;75:336-346.
- HealthMeasures. PROMIS. 2018. Available at: <http://www.healthmeasures.net/score-and-interpret/interpret-scores/promis>. Accessed: December 26, 2018.
- 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. 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.
- 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;28:439-443.
- Jacobs CA, Burnham JM, Jochimsen KN, Molina Dt, Hamilton DA, Duncan ST. Preoperative symptoms in femoroacetabular impingement patients are more related to mental health scores than the severity of labral tear or magnitude of bony deformity. *J Arthroplasty*. 2017;32:3603-3606.
- Kim HY. Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restor Dent Endod*. 2013;38:52-54.
- Kortlever JT, Janssen SJ, van Berckel MM, Ring D, Vranceanu AM. What Is the most useful questionnaire for measurement of coping strategies in response to nociception? *Clin Orthop Relat Res*. 2015;473:3511-3518.
- Laine C, Goodman SN, Griswold ME, Sox HC. Reproducible research: moving toward research the public can really trust. *Ann Intern Med*. 2007;146:450-453.
- 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.
- Overbeek CL, Nota SP, Jayakumar P, Hageman MG, Ring D. The PROMIS physical function correlates with the QuickDASH in patients with upper extremity illness. *Clin Orthop Relat Res*. 2015;473:311-317.
- Papuga MO, Dasilva C, McIntyre A, Mitten D, Kates S, Baumhauer JF. Large-scale clinical implementation of PROMIS computer adaptive testing with direct incorporation into the electronic medical record. *Health Systems*. 2017;1-12.
- Porter ME. What Is value in health care? *N Engl J Med*. 2010;363:2477-2481.
- Potter MQ, Wylie JD, Sun GS, Beckmann JT, Aoki SK. Psychologic distress reduces preoperative self-assessment scores in femoroacetabular impingement patients. *Clin Orthop Relat Res*. 2014;472:1886-1892.

27. 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.
28. Tucci V, Moukaddam N. We are the hollow men: the worldwide epidemic of mental illness, psychiatric and behavioral emergencies, and its impact on patients and providers. *J Emerg Trauma Shock*. 2017;10:4-6.
29. 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.e2044.
30. Yaffe M, Goyal N, Kokmeyer D, Merrell GA. The use of an iPad to collect patient-reported functional outcome measures in hand surgery. *Hand (N Y)*. 2015;10:522-528.