

ORIGINAL RESEARCH

Incorporating PROMIS Symptom Measures into Primary Care Practice—a Randomized Clinical Trial

Kurt Kroenke, MD^{1,2,3}, Tasneem L. Talib, PhD², Timothy E. Stump, MS⁴, Jacob Kean, PhD⁵, David A. Haggstrom, MD^{1,2,3}, Paige DeChant, BA², Kittie R. Lake, BS², Madison Stout, BS², and Patrick O. Monahan, PhD⁴

¹Division of General Internal Medicine and Geriatrics, Indiana University School of Medicine, Indianapolis, IN, USA; ²Regenstrief Institute, Indianapolis, IN, USA; ³Veterans Affairs Health Services Research and Development Center for Health Information and Communication, Indianapolis, IN, USA; ⁴Department of Biostatistics, Fairbanks School of Public Health, Indianapolis, IN, USA; ⁵University of Utah School of Medicine, Salt Lake City, UT, USA.

BACKGROUND: Symptoms account for more than 400 million clinic visits annually in the USA. The SPADE symptoms (sleep, pain, anxiety, depression, and low energy/fatigue) are particularly prevalent and undertreated.

OBJECTIVE: To assess the effectiveness of providing PROMIS (Patient-Reported Outcome Measure Information System) symptom scores to clinicians on symptom outcomes.

DESIGN: Randomized clinical trial conducted from March 2015 through May 2016 in general internal medicine and family practice clinics in an academic healthcare system.

PARTICIPANTS: Primary care patients who screened positive for at least one SPADE symptom.

INTERVENTIONS: After completing the PROMIS symptom measures electronically immediately prior to their visit, the 300 study participants were randomized to a feedback group in which their clinician received a visual display of symptom scores or a control group in which scores were not provided to clinicians.

MAIN MEASURES: The primary outcome was the 3-month change in composite SPADE score. Secondary outcomes were individual symptom scores, symptom documentation in the clinic note, symptom-specific clinician actions, and patient satisfaction.

KEY RESULTS: Most patients (84%) had multiple clinically significant (T-score ≥ 55) SPADE symptoms. Both groups demonstrated moderate symptom improvement with a non-significant trend favoring the feedback compared to control group (between-group difference in composite T-score improvement, 1.1; $P=0.17$). Symptoms present at baseline resolved at 3-month follow-up only one third of the time, and patients frequently still desired treatment. Except for pain, clinically significant symptoms were documented less than half the time. Neither symptom documentation, symptom-specific clinician

actions, nor patient satisfaction differed between treatment arms. Predictors of greater symptom improvement included female sex, black race, fewer medical conditions, and receiving care in a family medicine clinic.

CONCLUSIONS: Simple feedback of symptom scores to primary care clinicians in the absence of additional systems support or incentives is not superior to usual care in improving symptom outcomes.

TRIAL REGISTRATION: clinicaltrials.gov identifier: NCT02383862.

KEY WORDS: patient-reported outcomes; symptoms; primary care; feedback; clinical trial.

J Gen Intern Med 33(8):1245–52
DOI: 10.1007/s11606-018-4391-0
© Society of General Internal Medicine 2018

Symptoms account for over half of all outpatient visits¹ and are associated with substantial impairments in health-related quality of life, work-related disability, and increased healthcare costs.^{1–3} Further, symptoms that are unexplained, multiple, or persistent lead to mutual patient and clinician dissatisfaction.^{4, 5} Nonetheless, symptoms have been underemphasized in research and clinician training, thereby leading to suboptimal recognition and management in patient care.⁶

The SPADE pentad (sleep problems, pain, anxiety, depression, and low energy/fatigue) is especially important for several reasons. First, the five SPADE symptoms are the most prevalent, chronic, disabling, and undertreated symptoms in both the general population^{7, 8} and clinical practice.^{2, 8–10} Second, they cause additive impairment and adversely affect treatment response of one another.^{2, 11–13} Third, the SPADE symptoms are ubiquitous across most medical and mental disorders. Fourth, these symptoms commonly cluster^{3, 14–23} so that clinically unbundling the SPADE cluster is both difficult and perhaps counterproductive.

Interest is building in incorporating patient-reported outcome measures (PROs) into clinical practice^{12, 24–26} based

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11606-018-4391-0>) contains supplementary material, which is available to authorized users.

Received October 5, 2017
Revised January 2, 2018
Accepted February 14, 2018
Published online April 5, 2018

upon the untested assumption that providing this information to clinicians and patients will change outcomes. Moreover, a number of PRO initiatives have occurred in specialty clinics which focus on a narrower range of diseases and outcomes. In contrast, the primary care clinician is responsible for managing all or most of a patient's acute and chronic conditions, and therefore is particularly challenged²⁷ in deciding how many and which PROs to administer. The PROMIS (patient-reported outcome measurement information system) measures are an extensively tested set of public domain PROs, and SPADE symptoms constitute 5 of the 7 domains assessed by the PROMIS 29-, 43-, and 57-item profiles (www.healthmeasures.net). The objective of this randomized clinical trial was to assess the effectiveness of providing PROMIS symptom scores to primary care clinicians on patient outcomes.

METHODS

Study Design and Participants

In this prospective, two-arm randomized clinical trial, patients were recruited from March 2015 through April 2016 from urban academic primary care clinics in which both faculty and residents provide care. Upon checking in for their clinic visit, patients were asked to complete a five-item symptom screener adapted from the MD Anderson Symptom Inventory²⁸ rating the severity of SPADE symptoms on a 0 to 10 scale. Patients were eligible if they were ≥ 18 years old and English-speaking, received care from a participating primary care clinician, and reported a severity score ≥ 4 for at least one SPADE symptom. The study was approved by Indiana University's institutional review board.

Randomization

After providing informed consent and completing the PROMIS measures on a touch-screen tablet, participants were allocated to the feedback or control group in randomly alternating computer-generated blocks of 2 and 4. Randomization occurred at the level of the patient in order to control for clinician factors likely to influence symptom evaluation and management.

Interventions

For patients randomized to the feedback arm, their clinician was provided, just before the encounter, a printed bar graph of PROMIS symptom scores (Fig. 1). The PROMIS numeric scores for all five SPADE symptoms were specified on the graph, and elevated scores (T-scores ≥ 55) were further highlighted by including threshold lines and making symptom bars that crossed the threshold line red.²⁹ Patients randomized to the control group completed the same study measures as the feedback group, but scores were not provided to their clinician.

Outcomes

The PROMIS³⁰ profile-29 includes 4-item scales for 7 domains; 5 of these domains were used for this study—sleep, pain, anxiety, depression, and fatigue—as they reflected the SPADE symptoms. Each PROMIS scale provides a raw score, ranging from 4 to 20. Raw scores can be converted to T-scores using the PROMIS conversion tables. A T-score of 50 on each PROMIS symptom scale represents the general population norm (i.e., mean), and each 10-point deviation represents one standard deviation (SD) from the population norm. A cut point of ≥ 55 was used to represent a clinically elevated symptom score as this is 0.5 SD worse than the population mean, which is traditionally considered a moderate effect size.³¹

The enrollment clinic visit note from the electronic health records was reviewed to assess clinical documentation of SPADE symptoms and SPADE-specific diagnostic and treatment actions. Coding criteria were adapted from previous chart review studies of symptoms,^{10, 32} and study team members were trained in use of the coding criteria. Every clinic note was independently coded by two study team members who were blinded to study group. Coding disagreements were arbitrated by a study investigator (KK).

Three months after the enrollment visit, participants completed a follow-up survey, selecting either a mailed or web-based version. Non-respondents were contacted up to five times to complete the survey by telephone. In addition to completing the PROMIS symptom scales, participants were asked to recall whether they had discussed any of the SPADE symptoms with their clinician during the enrollment visit (as well as reasons for not discussing) and whether they had received treatment for any of the symptoms. They were also asked if they currently desired treatment or a change in treatment for any of the SPADE symptoms. Satisfaction with the care of their symptoms was rated from 1 (excellent) to 5 (poor).³³

Statistical analysis

The trial was powered to detect a small to moderate effect size of 0.35 (T-score of 3.5 points on individual PROMIS scales and approximately 2.8 points on composite score). This required 131 patients per study group at an $\alpha = 0.05$ and $\beta = 0.20$ (power of 80%) or allowing for 10% attrition by 3 months, 146 per study group.

The primary hypothesis was that change in the composite PROMIS T-score from baseline to 3 months would be greater in the feedback group than in the control group. Multiple imputation was used to impute PROMIS scores for participants not completing the 3-month assessment. Secondly, complete cases and within-group changes were analyzed, as well as changes in the five individual symptom scores. All analyses were intent-to-treat (as randomized).

All-subsets multivariate regression analysis was used to explore whether certain patient factors (age, sex, race,

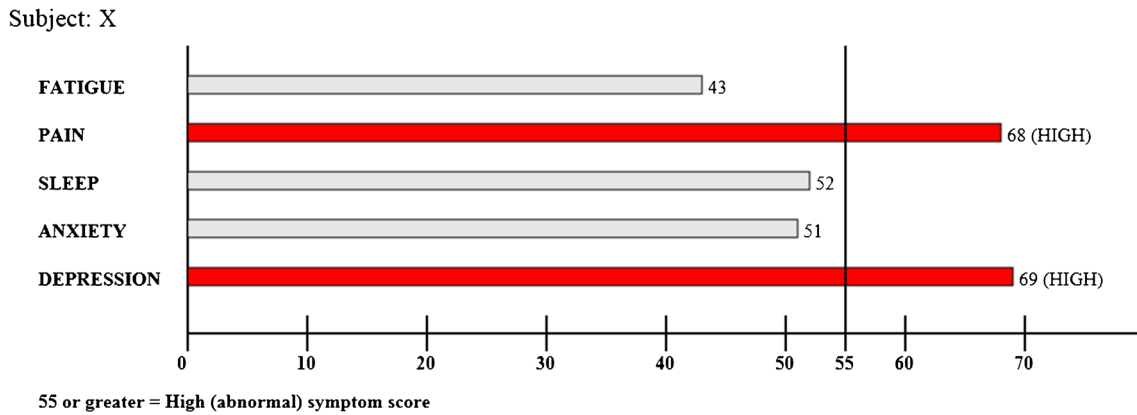


Fig. 1 Visual display of PROMIS symptom scores provided to physicians in feedback group.

education, number of comorbid medical conditions, and primary care discipline [internal medicine vs. family medicine] predicted symptom improvement, adjusting for study arm and baseline symptom severity.

RESULTS

Study Participants

Of 419 patients screened in the clinic, 374 (89%) screened positive for at least 1 of the 5 SPADE symptoms (Fig. 2). Symptom screening scores did not significantly differ between the 30 eligible patients who declined, 44 who were interested but unable to complete enrollment, and 300 who enrolled in the trial ($n = 300$). A total of 75 primary care clinicians (22 staff physicians, 2 nurse practitioners; 51 residents) had patients enrolled in the study, and of these, 61 received feedback on at least 1 patient.

The feedback and control groups were similar at baseline (Table 1). Average age of the sample was 49.4 years with 72%

women and a similar proportion of white (45.0%) and African-American (49.3%) patients. The mean composite PROMIS T-score was 58.3. Participants typically had multiple SPADE symptoms; the proportion with 0, 1, 2, 3, 4, and 5 clinically significant symptoms (T-score ≥ 55) was 5, 11, 13, 18, 21, and 31%, respectively.

Symptom Outcomes

Follow-up data was collected from 256 (85.3%) of the study participants. Compared to participants with follow-up data, the 44 participants without follow-up data were younger (41.6 vs. 50.7 years, $P < 0.001$) but were otherwise similar with regard to recruitment site, sex, race, education and baseline PROMIS composite T-score.

As shown in Table 2, participants demonstrated significant small to moderate within-group T-score improvements for each of the individual symptoms as well as the composite T-score, with effect sizes in imputed analyses ranging from 0.17 to 0.52. Although feedback participants reported slightly greater within-group improvement than

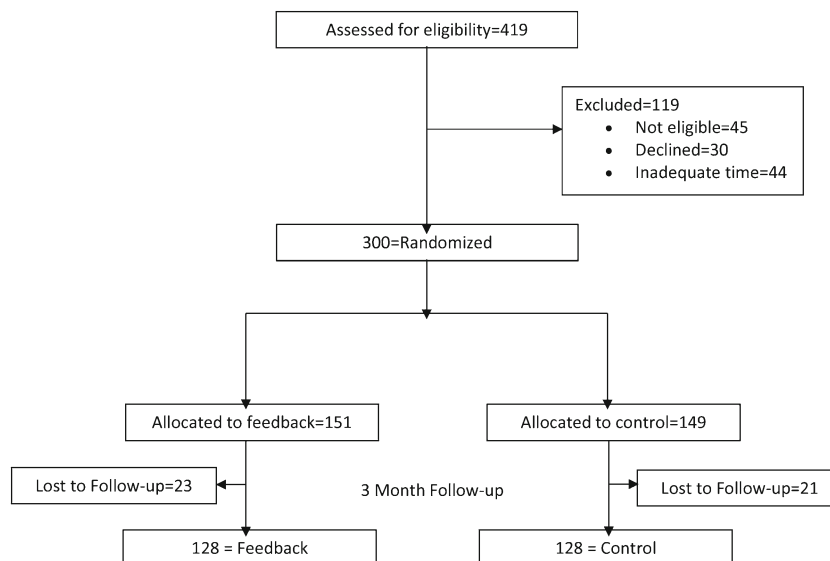


Fig. 2 CONSORT diagram showing participant flow.

Table 1 Baseline Characteristics of Patients Enrolled in SPADE Trial

| Characteristic | Total (n = 300) | | Study arm | | P value ^a | |
|-------------------------------------------|-----------------|--------|--------------------|-------------------|----------------------|--------|
| | | | Feedback (n = 151) | Control (n = 149) | | |
| Site, n (%) | | | | | .24 | |
| Internal medicine | 169 | (56.3) | 80 | (53.0) | 89 | (59.7) |
| Family medicine | 131 | (43.7) | 71 | (47.0) | 60 | (40.3) |
| Age, mean (SD) | 49.4 | (14.4) | 50.5 | (14.1) | 48.2 | (14.7) |
| Women, n (%) | 215 | (71.7) | 111 | (73.5) | 104 | (69.8) |
| Race, n (%) | | | | | | .50 |
| White | 135 | (45.0) | 65 | (43.1) | 70 | (47.0) |
| Black | 148 | (49.3) | 79 | (52.3) | 69 | (46.3) |
| Other | 17 | (5.7) | 7 | (4.6) | 10 | (6.7) |
| Education, n (%) ^b | | | | | | .62 |
| High school or less | 136 | (53.3) | 65 | (51.2) | 71 | (55.5) |
| Some college or trade school | 85 | (33.3) | 46 | (36.2) | 39 | (30.5) |
| College graduate | 34 | (13.3) | 16 | (12.6) | 18 | (14.0) |
| Comorbid diseases, mean (SD) ^b | 2.1 | (1.6) | 2.1 | (1.5) | 2.2 | (1.7) |
| PROMIS T-scores, mean (SD) | | | | | | .61 |
| Pain | 61.5 | (9.4) | 61.5 | (9.3) | 61.4 | (9.5) |
| Anxiety | 59.1 | (9.4) | 59.0 | (10.1) | 59.2 | (8.7) |
| Sleep | 58.2 | (9.0) | 58.3 | (9.4) | 58.1 | (8.7) |
| Fatigue | 57.0 | (10.0) | 56.8 | (10.1) | 57.2 | (9.9) |
| Depression | 55.9 | (9.8) | 55.8 | (10.4) | 56.0 | (9.1) |
| Composite | 58.3 | (7.0) | 58.3 | (7.6) | 58.4 | (6.4) |
| PROMIS T-score ≥ 55, n (%) | | | | | | .52 |
| Pain | 235 | (78.3) | 116 | (76.8) | 119 | (79.9) |
| Anxiety | 217 | (72.3) | 107 | (70.9) | 110 | (73.8) |
| Fatigue | 187 | (62.3) | 94 | (62.3) | 93 | (62.4) |
| Sleep | 182 | (60.7) | 91 | (60.3) | 91 | (61.1) |
| Depression | 178 | (59.3) | 86 | (57.0) | 92 | (61.7) |
| No. symptoms ≥ 55, n (%) | | | | | | .19 |
| 0 | 16 | (5.3) | 11 | (7.3) | 5 | (3.4) |
| 1 | 33 | (11.0) | 17 | (11.3) | 16 | (10.7) |
| 2 | 39 | (13.0) | 22 | (14.6) | 17 | (11.4) |
| 3 | 54 | (18.0) | 23 | (15.2) | 31 | (20.8) |
| 4 | 64 | (21.3) | 26 | (17.2) | 38 | (25.5) |
| 5 | 94 | (31.3) | 52 | (34.4) | 42 | (28.2) |

^aChi-square test used for categorical variables; *t* test used for continuous variables

^bTotal N for education was 255 (feedback = 127 and control = 128); total N for medical comorbidity was 255 (feedback = 128; control = 127)

the control group (3.48 vs. 2.38 decrease in PROMIS composite T-score), the between-group difference of 1.1 (effect size = 0.16) was not significant ($P = 0.17$). Likewise, between-group differences were not significant for any of the five individual symptom T-scores. Results of complete case analyses were similar.

Multivariate analysis showed that independent predictors of improvement in the SPADE composite T-score at 3 months were female sex (1.7 points greater improvement in T-score, $P = .036$), black race (2.5 points greater improvement, $P < .001$), fewer than 2 comorbid medical diseases (2.5 points greater improvement, $P = .001$), and having a family medicine provider (1.9 points greater improvement, $P = .013$). Age and education were not significant predictors.

Symptoms were more likely to persist than resolve (online Appendix, eTable 1). Of the 256 patients with 3-month follow-up data who had a threshold-level symptom at baseline, persistence at 3 months was 78% (157/201) for pain, 76% (139/182) for anxiety, 70% (105/149) for depression, 65% (101/156) for fatigue, and 56% (86/154) for sleep problems; thus, less than one third (254/842) of symptoms resolved. Of patients without a given symptom at baseline, the 3-month incidence was 5% for pain, 7% for anxiety and sleep problems, and 9% for depression and fatigue.

Symptom Documentation and Symptom-Specific Clinician Actions

Baseline visit notes were available to review for 292 patients, of which 26 (9%) were new patient visits and 266 (91%) were patients previously seen by the primary care clinician. In the feedback group, PROMIS scores were directly mentioned in only 1 of 147 notes. Patients with threshold-level PROMIS T-scores (i.e., ≥ 55) were more likely to have SPADE symptoms documented in the medical record (Fig. 3). However, even threshold-level symptom documentation varied substantially by symptom type, ranging from 81% for pain to 16% for fatigue. Overall, threshold-level, non-pain SPADE symptoms were documented $< 50\%$ of the time. Documentation rates did not differ between feedback and control groups.

SPADE symptom-specific clinician actions are summarized in eTable 2 (online Appendix). Since patients often had multiple SPADE symptoms, the actions shown in the table are for any SPADE symptom. The most common clinician actions were medication for 65.7% of study participants, another type of treatment (e.g., education) for 35.3%, and specialty referrals for 28.1%. With the exception of one category (diagnostic tests other than laboratory tests or imaging), clinician actions did not differ between the feedback and control groups. Medication prescriptions and referrals (but not other clinician actions) increased with symptom burden.

Table 2 Symptom Outcomes at 3 Months in Imputed (n=300) and Complete (n=256) Cases

| T-score | Within-group changes | | | Between-group changes | | | | | |
|----------------|--------------------------------------------|---------|--------------------------|-------------------------------------------|---------|--------------------------|-------------------------------------------|---------|--------------------------|
| | Feedback group T-score change ^a | P value | Effect size ^b | Control group T-score change ^a | P value | Effect size ^b | Difference in T-score change ^c | P value | Effect size ^b |
| Composite | | | | | | | | | |
| Imputed | 3.48 | <.0001 | .46 | 2.38 | <.0001 | .37 | 1.10 | .165 | .16 |
| Complete cases | 3.65 | <.0001 | .48 | 2.39 | <.0001 | .37 | 1.25 | .103 | .18 |
| Sleep | | | | | | | | | |
| Imputed | 4.88 | <.0001 | .52 | 4.04 | <.0001 | .46 | 0.84 | .425 | .09 |
| Complete cases | 5.16 | <.0001 | .55 | 3.98 | <.0001 | .46 | 1.18 | .271 | .13 |
| Pain | | | | | | | | | |
| Imputed | 2.77 | <.0001 | .30 | 2.12 | .007 | .22 | 0.65 | .539 | .07 |
| Complete cases | 2.89 | <.0001 | .31 | 2.10 | .001 | .22 | 0.79 | .463 | .08 |
| Anxiety | | | | | | | | | |
| Imputed | 2.96 | .0002 | .29 | 2.13 | .002 | .24 | 0.83 | .471 | .09 |
| Complete cases | 3.03 | .0001 | .30 | 2.33 | .006 | .27 | 0.69 | .539 | .07 |
| Depression | | | | | | | | | |
| Imputed | 3.08 | <.0001 | .30 | 1.59 | .040 | .17 | 1.49 | .174 | .15 |
| Complete cases | 3.14 | .0001 | .30 | 1.89 | .013 | .21 | 1.25 | .252 | .13 |
| Fatigue | | | | | | | | | |
| Imputed | 3.68 | <.0001 | .36 | 2.01 | .043 | .20 | 1.67 | .222 | .17 |
| Complete cases | 4.02 | <.0001 | .40 | 1.77 | .091 | .18 | 2.25 | .095 | .22 |

^aT-score change = baseline - 3 months (thus, positive change score = improvement)

^bEffect size for within-group change = within-group change / SD of that group at baseline. Effect size for between-group change = difference in change scores / pooled SD of total sample at baseline

^cFeedback group T-score change - control group T-score change

Patient-Reported Discussion and Treatment of SPADE Symptoms

At 3-month follow up, patients reported whether they had discussed symptoms and received treatment at the baseline clinic visit (online Appendix, eTable 3). The level of clinician action (not discussed vs. discussed but not treated vs. treated) increased with symptom severity whether measured as the mean symptom T-score or as a threshold-level symptom (T-score ≥ 55). There were no differences, however, between feedback and control group patients. The proportion of threshold-level symptoms not discussed was lowest for pain (12%), intermediate for sleep and fatigue (22% each), and highest for depression (35%) and anxiety (36%). The level

of patient-reported clinician action was not associated with patient demographics, medical comorbidity, specialty (internal medicine vs. family medicine), or overall satisfaction with symptom care.

Reasons for not discussing the symptom were provided by 140 patients. The most common perceived reasons were more pressing medical issues to discuss (n = 68; 49%) or the patient did not need (n = 66; 47%) or want (n = 40; 29%) treatment, followed by the doctor not bringing the symptom up (n = 30; 21%), the patient (n = 22; 16%) or doctor (n = 13; 9%) not feeling comfortable talking about the symptom, or the doctor seeming too busy (n = 10; 7%).

Table 3 shows the proportion of patients who still desired treatment for symptoms at 3-month follow-up which ranged from 23% for depression to 40% for pain. Patients who still desired treatment had more severe symptoms at 3 months as measured by either the mean symptom T-score or a threshold-level (T-score ≥ 55) symptom, less improvement in their symptom from baseline to 3-month follow-up, lower satisfaction with their overall symptom care, and greater medical comorbidity (latter not shown in table). Desire for treatment did not differ between feedback and control groups, and also was not associated with patient demographics or primary care specialty.

Treatment Satisfaction

Overall satisfaction with symptom care was rated as excellent by 18% of participants, very good by 24%, good by 32%, fair by 19%, and poor by 8%. Satisfaction did not differ between study groups. However, participants who still desired treatment for their symptoms at 3 months were less likely to rate their satisfaction as excellent or very good (Table 3).

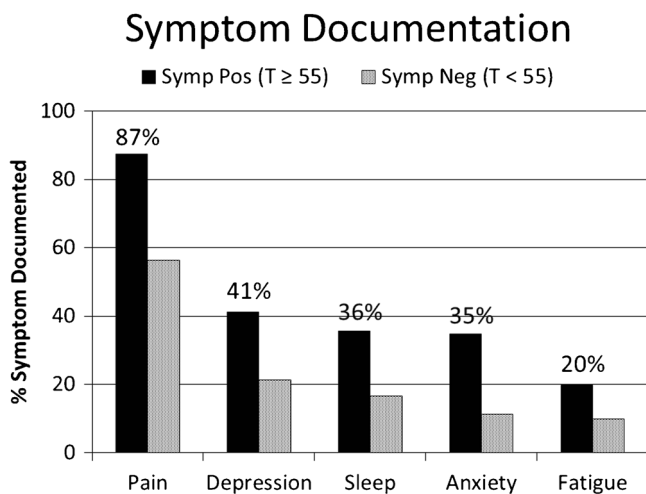


Fig. 3 Symptom documentation in clinic visit note.

Table 3 Patient-Reported Desire for SPADE Treatment at 3-Month Follow-up (n = 256)

| Symptom treatment desired at 3-month follow-up | Number (%) | 3-month symptom T-score (mean) | Symptom T-score change ^a (mean) | High patient satisfaction ^b (%) | Treatment desired by whether symptom is threshold level (T-Score \geq 55) ^c (%) | |
|------------------------------------------------|------------|--------------------------------|--------------------------------------------|--------------------------------------------|----------------------------------------------------------------------------------------------|-----------------|
| | | | | | Not threshold | Threshold level |
| Pain treatment | | | | | n = 87 | n = 168 |
| Not desired | 154 (60.4) | 55.2 | 4.06 | 52.0 | 85.1 | 47.6 |
| Desired | 101 (39.6) | 65.0 | 0.21 | 26.0 | 14.9 | 52.4 |
| P value | | <.0001 | .0004 | <.0001 | <.0001 | |
| Sleep treatment | | | | | n = 153 | n = 102 |
| Not desired | 176 (68.8) | 50.2 | 5.69 | 49.4 | 88.2 | 39.2 |
| Desired | 80 (31.2) | 60.8 | 2.27 | 25.0 | 11.8 | 60.8 |
| P value | | <.0001 | .003 | <.0001 | <.0001 | |
| Fatigue treatment | | | | | n = 132 | n = 123 |
| Not desired | 165 (64.5) | 50.3 | 4.64 | 48.5 | 82.6 | 44.7 |
| Desired | 91 (35.5) | 60.7 | -0.15 | 29.7 | 17.4 | 55.3 |
| P value | | <.0001 | .0006 | .0004 | <.0001 | |
| Anxiety treatment | | | | | n = 98 | n = 157 |
| Not desired | 191 (74.6) | 53.5 | 3.38 | 48.7 | 96.9 | 60.5 |
| Desired | 65 (25.4) | 64.7 | 0.79 | 21.5 | 3.1 | 39.5 |
| P value | | <.0001 | .045 | <.0001 | <.0001 | |
| Depression treatment | | | | | n = 127 | n = 127 |
| Not desired | 197 (77.0) | 50.5 | 3.15 | 47.7 | 93.7 | 59.8 |
| Desired | 59 (23.0) | 62.4 | 0.68 | 22.0 | 6.3 | 40.2 |
| P value | | <.0001 | .054 | <.0001 | <.0001 | |

^aT-score change = baseline - 3 months (thus, positive change score = improvement)

^bValues are % reporting excellent to very good for overall satisfaction with symptom care, but P values are for five-category satisfaction (excellent, very good, good, fair, poor)

^cPercent desiring treatment for a symptom at 3 months by whether patient did or did not have threshold level (T-score \geq 55) of symptom at 3 months. For example, treatment for pain was desired by 52.4% of the 168 patients with threshold-level pain at 3 months compared to 14.9% of the 87 patients without threshold-level pain

DISCUSSION

Our trial has several important implications for the real-world implementation of symptom measures in clinical practice. First, simple feedback of PROMIS symptom scores to primary care clinicians was inadequate to significantly enhance symptom improvement at 3-month follow-up. A minimal clinically important change in PROMIS T-scores is generally in the 2 to 4 point range³⁴⁻³⁶ which corresponds to the within-group changes in both study arms, but not the between-group difference in our trial. Second, SPADE symptoms other than pain were infrequently documented in the clinician's note. Third, a substantial proportion of patients reported persistent symptoms at follow-up for which they desired treatment.

Our findings that feedback alone was insufficient to improve symptom outcomes is consistent with multiple trials showing that the provision of additional information to primary care clinicians in a busy setting with many competing demands—without also providing additional time or resources—is relatively ineffective.³⁷ This phenomenon has been best demonstrated for depression,^{38, 39} and several studies have shown that simply providing pain or anxiety scores to clinicians does not change outcomes.⁴⁰⁻⁴² To our knowledge, the effect of feedback regarding fatigue or sleep problems has not been previously studied. Research suggesting feedback of symptom scores may be beneficial have largely demonstrated improved processes of care (e.g., documentation of symptoms, discussions with patients, treatment actions) rather than symptom outcomes and, where outcomes have improved, this has

occurred predominantly in specialty settings (e.g., cancer centers, palliative care) with additional clinical team members and extra patient contacts.^{37, 43-49} The movement to implement PROs into clinical practice and electronic health records^{24, 50, 51} may have limited impact unless simultaneous consideration is given to the systems support necessary to facilitate clinical actions, monitor outcomes, and adjust treatment.³⁹ However, the lack of systems support may not be the only explanation for our study findings. It is also possible that the type or number of symptoms chosen made clinical actions or symptom improvement more challenging, that the method of feedback used was suboptimal, or that PRO feedback was not particularly conducive (or necessary) to the primary care setting in which the intervention was implemented.

Most patients had more than one threshold-level SPADE symptom. The fact that multiple symptoms is the norm was also found in a trial involving 250 primary care patients with chronic pain in which the proportion with 0, 1, 2, 3, 4, and 5 SPADE symptoms was 10, 20, 16, 23, 12, and 20%, respectively.¹⁴ Admittedly, selection bias might play some role in that eligibility for our study required that patients screen positive for at least one symptom. Still, of the 419 patients screened for our trial, only 11% did not screen positive for at least 1 symptom, suggesting study participants were not a highly selected sample. Also, other studies have shown that patients reporting one symptom typically have other symptoms as well.⁶

Despite the prevalence of symptoms, documentation of threshold-level symptoms (i.e., T-score \geq 55) in the visit note was only 20-41% for the four non-pain SPADE symptoms,

suggesting substantial limitations in using EHR data from unstructured clinical notes for the secondary purposes of symptoms research or quality improvement. Under-documentation may be due to the time constraints and competing demands of primary care, as well as the lack of incentives for evaluating and managing symptoms. Also, patients frequently noted that symptoms were not discussed because there were more pressing issues or they did not want treatment. Finally, PROs may detect a higher frequency of symptoms (including less bothersome symptoms) than symptoms spontaneously reported by patients.¹

The decision about which symptoms warrant treatment must weigh symptom severity, availability of evidence-based therapies, patient and provider prioritization of symptoms, and treatment preferences. Optimal treatment for the SPADE symptoms, particularly when chronic, typically includes non-pharmacological therapies (e.g., cognitive-behavioral therapy, exercise, mindfulness-based treatments) rather than medications alone.⁶ However, several obstacles exist to broader implementation of these treatments, including an insufficient number of healthcare professionals trained in these non-pharmacological therapies, reimbursement barriers, and motivating patients to engage in these treatments. Moreover, even if such treatments had been provided, the 3-month follow-up assessment used in our trial may have been an inadequate period of time for patients to receive a sufficient intensity and duration of non-pharmacological therapy to experience optimal symptomatic improvement.

Symptoms present at a threshold-level at baseline persisted in half to three-quarters of patients at 3-month follow-up, and patients frequently still desired treatment. This suggests that symptom severity and persistence coupled with patient expectations^{5, 52, 53} might be one approach to balancing overtreatment vs. patient-centered treatment of common symptoms. Other factors influencing management might include whether the symptom is secondary to another medical condition or treatment, the presence of competing health concerns, the relative role of clinical judgment vs. PRO scores in determining clinician actions, and the option of watchful waiting to distinguish persistent from self-limited symptoms.⁴⁷ Shared decision-making between the clinician and patient is core to navigating these factors.⁵⁴

A study strength in terms of generalizability was the relatively balanced distribution of patients among the two principal disciplines providing primary care for adults: general internal medicine and family medicine. Second, the patient sample had a good distribution of age, race, and medical comorbidity. Third, the participation rate among eligible patients was reasonably high, minimizing refusal as a major source of selection bias.

Several study limitations should be noted. Three-month follow-up data could not be obtained for 14.6% of the study participants. However, multiple imputation using the full sample of 300 participants and analysis of the 256 complete cases produced similar results. Second, secondary outcomes assessed by patient report at 3 months or by chart review are susceptible to recall or rater bias, respectively. The latter,

however, was reduced by rater training, explicit coding criteria, independent review of all notes by two raters, and rater blinding to study group. Third, 61 clinicians received feedback on one or more of the 151 patients in the feedback arm, meaning that most physicians received feedback on only a few patients in the trial. Receiving symptom feedback on more patients over a longer period of time might lead to greater attention to SPADE symptoms. Fourth, the trial was conducted in academic clinics staffed by both faculty and residents who were providing care to an underserved population, and findings should be replicated more broadly.

Diagnostic testing and procedures are unnecessary for the majority of patients with SPADE symptoms; instead, the history and physical examination coupled with communication strategies are more effective for symptom evaluation and management.⁶ Realigning incentives to enable more patient-centered approaches has the potential of improving symptom outcomes at lower cost. Making information from PROs readily actionable through sufficient training, time, and resources may be critical to the effective use of PROs by practicing clinicians.⁵⁵ At the same time, determining which PROs are valued by clinicians and patients, the optimal frequency of assessment and provision of results, and in which setting PROs can improve symptom outcomes are all appropriate steps prior to widespread PRO implementation.

Corresponding Author: Kurt Kroenke, MD; Regenstrief Institute, Indianapolis, IN, USA (e-mail: kkroenke@regenstrief.org).

Authors' Contributions There are no contributors who do not meet the criteria for authorship.

Funding This work was supported by Patient-Centered Outcomes Research Institute (PCORI) Contract ME-1403-12043.

Compliance with Ethical Standards:

Conflicts of Interest: The authors declare that they do not have a conflict of interest.

Prior Presentations: Part of this work was presented at the Health Measures User Conference, September 27, 2017, in Chicago, Illinois.

REFERENCES

1. **Kroenke K.** Studying symptoms: sampling and measurement issues. *Ann Intern Med.* 2001;134(9 Pt 2):844-853.
2. **Kroenke K, Spitzer RL, Williams JBW, et al.** Physical symptoms in primary care: predictors of psychiatric disorders and functional impairment. *Arch Fam Med.* 1994;3:774-779.
3. **Kroenke K.** Patients presenting with somatic complaints: epidemiology, psychiatric comorbidity and management. *Int J Methods Psychiatr Res.* 2003;12(1):34-43.
4. **Hahn SR.** Physical symptoms and physician-experienced difficulty in the physician-patient relationship. *Ann Intern Med.* 2001;134:897-904.
5. **Jackson JL, Kroenke K.** The effect of unmet expectations among adults presenting with physical symptoms. *Ann Intern Med.* 2001;134:889-897.
6. **Kroenke K.** A practical and evidence-based approach to common symptoms: a narrative review. *Ann Intern Med.* 2014;161(8):579-586.
7. **Schappert SM.** National Ambulatory Medical Care Survey: 1991 summary. *Adv Data.* 1993(230):1-16.

8. **Kroenke K, Price RK.** Symptoms in the community: prevalence, classification, and psychiatric comorbidity. *Arch Intern Med.* 1993;153:2474-2480.
9. **Kroenke K, Arrington ME, Mangelsdorff AD.** The prevalence of symptoms in medical outpatients and the adequacy of therapy. *Arch Intern Med.* 1990;150:1685-1689.
10. **Khan AA, Khan A, Harezlak J, Tu W, Kroenke K.** Somatic symptoms in primary care: etiology and outcome. *Psychosomatics.* 2003;44(6):471-478.
11. **Bair MJ, Robinson RL, Katon W, Kroenke K.** Depression and pain comorbidity: a literature review. *Arch Intern Med.* 2003;163(20):2433-2445.
12. **Bair MJ, Poleshuck EL, Wu J, et al.** Anxiety but not social stressors predict 12-month depression and pain outcomes. *Clin J Pain.* 2013;29(2):95-101.
13. **Kroenke K, Wu J, Bair MJ, Krebs EE, Damush TM, Tu W.** Reciprocal relationship between pain and depression: a 12-month longitudinal analysis in primary care. *J Pain.* 2011;12:964-973.
14. **Davis LL, Kroenke K, Monahan P, Kean J, Stump TE.** The SPADE symptom cluster in primary care patients with chronic pain. *Clin J Pain.* 2016;32(5):388-393.
15. **Barsevick AM.** The concept of symptom cluster. *Semin Oncol Nurs.* 2007;23(2):89-98.
16. **Collen M.** The case for Pain Insomnia Depression Syndrome (PIDS): a symptom cluster in chronic nonmalignant pain. *J Pain Palliat Care Pharmacother.* 2008;22(3):221-225.
17. **Lee KS, Song EK, Lennie TA, et al.** Symptom clusters in men and women with heart failure and their impact on cardiac event-free survival. *J Cardiovasc Nurs.* 2010;25(4):263-272.
18. **Hunter Revell SM.** Symptom clusters in traumatic spinal cord injury: an exploratory literature review. *J Neurosci Nurs.* 2011;43(2):85-93.
19. **Donovan KA, Jacobsen PB.** Fatigue, depression, and insomnia: evidence for a symptom cluster in cancer. *Semin Oncol Nurs.* 2007;23(2):127-135.
20. **Fleishman SB.** Treatment of symptom clusters: pain, depression, and fatigue. *J Natl Cancer Inst Monogr.* 2004(32):119-123.
21. **Brown LF, Kroenke K.** Cancer-related fatigue and its association with depression and anxiety: a systematic literature review. *Psychosomatics.* 2009;50:440-447.
22. **Lowe B, Spitzer RL, Williams JB, Mussell M, Schellberg D, Kroenke K.** Depression, anxiety and somatization in primary care: syndrome overlap and functional impairment. *Gen Hosp Psychiatry.* 2008;30(3):191-199.
23. **Haftgoli N, Favrat B, Verdon F, et al.** Patients presenting with somatic complaints in general practice: depression, anxiety and somatoform disorders are frequent and associated with psychosocial stressors. *BMC Fam Practice.* 2010;11:67.
24. **Snyder CF, Aaronson NK, Chouchair AK, et al.** Implementing patient-reported outcomes assessment in clinical practice: a review of the options and considerations. *Qual Life Res.* 2012;21:1305-1314.
25. **Glasgow RE, Riley WT.** Pragmatic measures: what they are and why we need them. *Am J Prev Med.* 2013;45(2):237-243.
26. **Reeve BB, Wyrwich KW, Wu AW, et al.** ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res.* 2013;22:1889-1905.
27. **Kroenke K.** The many C's of primary care. *J Gen Intern Med.* 2004;19(6):708-709.
28. **Cleeland CS, Mendoza TR, Wang XS, et al.** Assessing symptom distress in cancer patients: the M.D. Anderson Symptom Inventory. *Cancer.* 2000;89(7):1634-1646.
29. **Snyder CF, Smith KC, Bantug ET, et al.** What do these scores mean? Presenting patient-reported outcomes data to patients and clinicians to improve interpretability. *Cancer.* 2017;123(10):1848-1859.
30. **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.
31. **Kazis LE, Anderson JJ, Meenan RF.** Effect sizes for interpreting changes in health status. *Med Care.* 1989;27:S178-S189.
32. **Kroenke K, Mangelsdorff AD.** Common symptoms in ambulatory care: incidence, evaluation, therapy, and outcome. *Am J Med.* 1989;86(3):262-266.
33. **Kroenke K, Evans E, Weitlauf S, et al.** Comprehensive vs. Assisted Management of Mood and Pain Symptoms (CAMMPS) trial: Study design and sample characteristics. *Contemp Clin Trials.* 2018;64:179-187.
34. **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(5):507-516.
35. **Deyo RA, Katrina R, Buckley DI, et al.** Performance of a Patient Reported Outcomes Measurement Information System (PROMIS) Short Form in Older Adults with Chronic Musculoskeletal Pain. *Pain Med.* 2016;17(2):314-324.
36. **Beaumont JL, Fries JF, Curtis JR, Cella D, Yun H.** Minimally important differences for Patient-Reported Outcomes Measurement Information System (PROMIS) fatigue and pain interference scores. *Value Health.* 2015;18(3):A165-A166.
37. **Boyce MB, Browne JP.** Does providing feedback on patient-reported outcomes to healthcare professionals result in better outcomes for patients? A systematic review. *Qual Life Res.* 2013;22(9):2265-2278.
38. **Gilbody S, Sheldon T, House A.** Screening and case-finding instruments for depression: a meta-analysis. *CMAJ.* 2008;178(8):997-1003.
39. **Kroenke K, Unutzer J.** Closing the False Divide: Sustainable Approaches to Integrating Mental Health Services into Primary Care. *J Gen Intern Med.* 2017;32(4):404-410.
40. **Mularski RA, White-Chu F, Overbay D, Miller L, Asch SM, Ganzini L.** Measuring pain as the 5th vital sign does not improve quality of pain management. *J Gen Intern Med.* 2006;21(6):607-612.
41. **Ahles TA, Wasson JH, Seville JL, et al.** A controlled trial of methods for managing pain in primary care patients with or without co-occurring psychosocial problems. *Ann Fam Med.* 2006;4(4):341-350.
42. **Mathias SD, Fifer SK, Mazonson PD, Lubeck DP, Buesching DP, Patrick DL.** Necessary but not sufficient: the effect of screening and feedback on outcomes of primary care patients with untreated anxiety. *J Gen Intern Med.* 1994;9(11):606-615.
43. **Valderas JM, Kotzeva A, Espallargues M, et al.** The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. *Qual Life Res.* 2008;17(2):179-193.
44. **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.
45. **Kotronoulas G, Kearney N, Maguire R, et al.** What is the value of the routine use of patient-reported outcome measures toward improvement of patient outcomes, processes of care, and health service outcomes in cancer care? A systematic review of controlled trials. *J Clin Oncol.* 2014;32(14):1480-1501.
46. **Basch E, Deal AM, Kris MG, et al.** Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. *J Clin Oncol.* 2016;34(6):557-565.
47. **Greenhalgh J.** The applications of PROs in clinical practice: what are they, do they work, and why? *Qual Life Res.* 2009;18(1):115-123.
48. **Basch E, Deal AM, Dueck AC, et al.** Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. *JAMA.* 2017;318(2):197-198.
49. **Kroenke K, Cheville AL.** Symptom improvement requires more than screening and feedback. *J Clin Oncol.* 2016;34(27):3351-3352.
50. **Glasgow RE, Kaplan RM, Ockene JK, Fisher EB, Emmons KM.** Patient-reported measures of psychosocial issues and health behavior should be added to electronic health records. *Health Affairs.* 2012;31:497-504.
51. **Nelson EC, Eftimovska E, Lind C, Hager A, Wasson JH, Lindblad S.** Patient reported outcome measures in practice. *BMJ.* 2015;350:g7818.
52. **Arroll B, Goodyear-Smith F, Kerse N, Fishman T, Gunn J.** Effect of the addition of a "help" question to two screening questions on specificity for diagnosis of depression in general practice: diagnostic validity study. *BMJ.* 2005;331(7521):884.
53. **Kroenke K, Krebs E, Wu J, et al.** Stepped Care to Optimize Pain Care Effectiveness (SCOPE) Trial: study design and sample characteristics. *Contemp Clin Trials.* 2013;34:270-281.
54. **Elwyn G, Frosch D, Thomson R, et al.** Shared decision making: a model for clinical practice. *J Gen Intern Med.* 2012;27(10):1361-1367.
55. **Kroenke K, Monahan PO, Kean J.** Pragmatic characteristics of patient-reported outcome measures are important for use in clinical practice. *J Clin Epidemiol.* 2015;68(9):1085-1092.