VALUE IN HEALTH ▮ (2017) ▮▮▮–▮▮▮

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Scale Linking to Enable Patient-Reported Outcome Performance Measures (PRO-PMs) Assessed with Different Patient-Reported Outcome Measures

Irene L. Katzan, MD, MS[1,*], *Youran Fan, PhD*[1], *Sandra D. Griffith, PhD*[1], *Paul K. Crane, MD, MPH*[2], *Nicolas R. Thompson, MS*[1], *David Cella, PhD*[3]

[1]*Neurological Institute, Cleveland Clinic, Cleveland, OH, USA;* [2]*Division of General Internal Medicine, University of Washington, Seattle, WA, USA;* [3]*Department of Medical Social Sciences, Northwestern University, Chicago, IL, USA*

ABSTRACT

Background: Patient-reported outcome performance measures (PRO-PMs) incorporate outcomes from the patient's perspective into performance measures and may have great potential to impact health care. The various patient-reported outcome measures (PROMs) used to assess the same outcome challenge widespread use of PRO-PMs. A potential solution is to statistically link PROMs to provide equivalent PRO-PM conclusions to be drawn regardless of which PROM was used. **Objectives:** To determine the level of agreement in the performance of two depression-related PRO-PMs assessed using the nine-item Patient Health Questionnaire (PHQ-9) depression scale and the eight-item Patient-Reported Outcomes Measurement Information System (PROMIS) Depression short form and the PHQ-9 cocalibrated on the PROMIS metric. **Methods:** We conducted a retrospective cohort study of patients who visited one of eight ambulatory neurological and psychiatric clinics at the Cleveland Clinic between January 23 and June 15, 2012, and who completed both the PHQ-9 and PROMIS Depression scales at the same visit. The level of agreement was measured between PRO-PM performance assessed with standard scoring of the PHQ-9, the PROMIS cocalibrated scoring of the PHQ-9, and the PROMIS score for two depression-related PRO-PMs. **Results:** Of the 5736 enrolled patients, 701 had PROMs from two or more visits. Differences in performance of the depression remission PRO-PM ranged from 0.4% to 2.1%, and differences in the progress toward remission PRO-PM ranged from 0.9% to 5.1%, depending on which depression score was used. **Conclusions:** There was a high level of agreement in the PRO-PM for depression when incorporating different PROMs. These findings support the ability to use linkage of scale scores to assess performance of PRO-PMs with different PROMs.

Keywords: depression, patient-centered outcomes research, performance measurement, quality improvement.

Introduction

Performance measurement has traditionally focused on processes (rather than on outcomes) of care and the patient's satisfaction with that care. There is increasing recognition that incorporating outcomes from the patient's perspective into performance measurement has great potential to impact health care and patients' health [1]. Patient-reported outcome performance measures (PRO-PMs) are a relatively new type of performance measurement that aggregate patient-reported outcomes—as opposed to data regarding care processes—to measure performance of health care [2]. PRO-PMs incorporate patient-reported outcome measures (PROMs) into the standard health care quality performance measure format that includes a numerator and a denominator. The PRO-PM numerator typically includes the number of patients meeting a prespecified threshold of improvement in the PROM score. The denominator includes the number of patients eligible for the measure. The National Quality Forum (NQF) has endorsed several PRO-PMs related to depression management including the following: Depression Remission at 6 Months (NQF #0711) and at 12 Months (NQF #0710) and Progress toward Depression Remission at 6 Months (NQF #1884) and at 12 Months (NQF #1885) [3]. These PRO-PMs, which use the Patient Health Questionnaire 9 (PHQ-9) depression screen to identify the presence and monitor change in depressive symptoms, are used in national performance initiatives [4,5]. The PHQ-9 is a frequently used nine-item depression screen with standard total scores ranging from 0 to 27, with 0 indicating no depressive symptoms and 10 suggesting moderate depressive symptoms [6].

The various PROMs implemented in clinical practice to assess the same health outcome create an issue for the widespread use of PRO-PMs to measure performance. For example, some of the PROMs available to measure depression include Center for Epidemiologic Studies Depression Scale [7], PHQ-9, and Patient-Reported Outcomes Measurement Information System (PROMIS) Depression [8]. Health systems may be reluctant to change the

* *Address correspondence to:* Irene L. Katzan, Neurological Institute, Cleveland Clinic, 9500 Euclid Avenue, S80, Cleveland, OH 44195. E-mail: Katzani@ccf.org.

1098-3015$36.00 – see front matter

http://dx.doi.org/10.1016/j.jval.2017.05.012

PROMs they collect because of the resources that would be required to do so and difficulties in using new PROMs with their legacy data. The ability to use different PROMs to assess PRO-PM performance would have significant implications in the ability to use patient-centered outcomes in performance measurement. One solution is to establish a linkage between scores of different PROMs to a common standardized metric, which provides equivalent scores for different scales that measure the same health outcome. Methods have been developed to link different PROMs to the PROMIS metric [9], including measures of depression [10], and methods for scoring PHQ-9 item responses on the PROMIS metric have been published [10,11].

The PROMIS Depression eight-item short form measure version b is one of the tools available in PROMIS [8]. Scale scores are calibrated to the US general population and are standardized on the T scale in which the mean is 50 and the SD is 10 [12]. Higher scores indicate worse depressive symptoms. Choi et al. [10] have linked PHQ-9 scores to depression scores on the PROMIS metric using data from the general adult US population. Cocalibration of the PROMIS Depression and the PHQ-9 has also been performed by Gibbons et al. [11], using parameters derived from a clinical sample rather than from the general population.

Although the concept is appealing, the ability to use different scales to measure the performance of a PRO-PM through scale cocalibration has not yet been demonstrated in practice. The objectives of this study were 1) to determine the level of agreement in depression-related PRO-PMs, assessed using the PHQ-9 standard scoring compared with the same PRO-PMs assessed using the PROMIS Depression short form as linked to the PHQ-9, and the PHQ-9 scored on the PROMIS metric via cocalibration and 2) to evaluate the characteristics that are associated with disagreement in the diagnosis of depression assessed with the PHQ-9 and the PROMIS Depression.

Methods

A retrospective cohort study of patients seen at the Cleveland Clinic from January 23 to June 15, 2012, in one of eight ambulatory condition-based neurological clinics (psychiatry, psychology, headache, pain, general neurology, sleep, epilepsy, and rehabilitation medicine) was conducted. To be included, patients had to have completed both a PHQ-9 scale and a PROMIS Depression short form during the same visit.

As part of routine care, patients seen in the Neurological Institute complete the PHQ-9 immediately before ambulatory visits unless it had been previously completed within the previous 28 days. Patients also complete condition-specific PROMs. The PROMIS Depression eight-item short form was also completed by patients in these eight clinics over the time period used in this analysis. Patient-reported data for this study were collected through the Cleveland Clinic Knowledge Program, an electronic platform for systematic collection of patient-reported information for patients seen in ambulatory clinics [13]. The patient-reported data are integrated within the electronic health record (Epic Corporation, Verona, WI) and are immediately available to the provider at the time of the clinical encounter. Best practice alerts are displayed in the electronic health record for standard total PHQ-9 scores at 15 or more.

Patient-Reported Outcome Performance Measures

The two PRO-PMs assessed in this study were adapted from depression-related PRO-PMs endorsed by the NQF: Depression Remission at 6 Months (#0711) and at 12 Months (#0710) and Depression Response—Progress toward Remission at 6 Months (#1884) and at 12 Months (#1885) [3]. The two adapted PRO-PMs in this study did not have a time component included in the four NQF-endorsed PRO-PMs. Presence of depression, the denominator of the adapted PRO-PM depression measures, was defined as the percentage of patients with a depression PROM score at the time of initial assessment exceeding the defined threshold. Threshold used for the PHQ-9 was explicitly stated in the NQF measures (PHQ-9 score >9). The equivalent PROMIS Depression, PHQ-9_{PROMIS} and PHQ-$9_{PROMIS(neuro)}$, was a T score of 59.9 or more [14]. Unlike the NQF-endorsed measures, a clinical diagnosis of depression was not a criterion for indicating the presence of depression.

1. *Depression remission* was defined as the percentage of patients with an initial PHQ-9 score of more than 9 or other PROM score of 59.9 or more who have a follow-up PHQ-9 score of less than 5 at the last visit during the study period. Corresponding scores for remission on the PROMIS metric, also obtained using crosswalk tables, were less than 52.5.
2. *Progress toward depression remission* was defined as the percentage of patients with an initial PHQ-9 score of more than 9 who had a follow-up PHQ-9 score during the study period that was reduced by 50% or more. The equivalent score on the PROMIS metric indicating 50% reduction was obtained by identifying the PHQ-9 score equivalent to the baseline PROMIS-based *T* scores, calculating a 50% reduction in this PHQ-9 score, and then converting it back to the corresponding *T* score on the PROMIS metric.

For all patients, data from the first visit, if collected within the allowed time window, were used to identify patients with an operational diagnosis of depression. Analysis of the PRO-PMs, which addressed improvement over time, included only those participants who had completed both the PHQ-9 and the PROMIS Depression 8b on at least two separate visits. If data on more than two encounters were available, the first and last scores were used.

Statistical Methods

Scale linking was performed using parameters provided by Choi et al. [10], on the basis of a fixed-parameter calibration method with response data from the general US population. A second set of links was derived from the clinical data collected in this study, as described by Gibbons et al. [11]. The results based on the Choi et al. [9] parameters are referred to as PHQ-9_{PROMIS}; those derived from the present sample are referred to as PHQ-$9_{PROMIS(neuro)}$. We included this second calibration method in our study to assess whether the derivation of linking parameters using data from the study cohort (PHQ-$9_{PROMIS(neuro)}$) as opposed to linking parameters based on the general population (PHQ-9_{PROMIS}) would provide a more accurate approximation of the PHQ-9 scores on the PROMIS metric.

SAS version 9.4 (SAS Institute Inc., Cary, NC) was used for all analyses. Descriptive statistics were calculated for patients in the study cohort and for patients with two or more visits. Spearman correlation coefficients quantified the association between the PHQ-9 scores and the PROMIS Depression 8b, PHQ-9_{PROMIS}, and PHQ-$9_{PROMIS(neuro)}$ scores. Finally, variation in the PROMIS Depression 8b and PHQ-9_{PROMIS} scores for each PHQ-9 score was plotted.

Agreement in the proportion of patients with depression and in PRO-PM performance between those assessed with the PHQ-9 and those assessed with other depression PROMs, at both the group and individual levels. Standard PRO-PM performance is assessed at the group level. Analyses were repeated at the individual level to evaluate agreement in categorization of each patient as achieving depression remission or progress toward remission. At the group level, differences were calculated