Research White Paper

Outcome Measure Harmonization and Data Infrastructure for Patient-Centered Outcomes Research in Depression: Report on Registry Configuration

Research White Paper

Outcome Measure Harmonization and Data Infrastructure for Patient-Centered Outcomes Research in Depression: Report on Registry Configuration

Prepared for:

Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
5600 Fishers Lane
Rockville, MD 20857
www.ahrq.gov

Contract No. 75Q80119C00005

Prepared by:

OM1, Inc.,[1] with subcontractors:
American Board of Family Medicine[2]
American Psychiatric Association[3]

Investigators:
Michelle B. Leavy, M.P.H.[1]
Danielle Cooke, B.S.[1]
Sarah Hajjar, M.P.A., P.M.P.[2]
Eric Bikelman, P.M.P.[2]
Bailey Egan, B.S.[2]
Diana Clarke, Ph.D.[3]
Debbie Gibson, M.S.c.[3]
Barbara Casanova, B.S.[3]
Richard Gliklich, M.D.[1]

AHRQ Publication No. 21-EHC003
November 2020

This report is based on research conducted by OM1 with subcontractors the American Board of Family Medicine and American Psychiatric Association under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. 75Q80119C00005). This work is supported by the Office of the Secretary Patient-Centered Outcomes Research Trust Fund under Interagency Agreement 18-596R-18. The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

None of the investigators have any affiliations or financial involvement that conflicts with the material presented in this report.

The information in this report is intended to help healthcare decision makers—patients and clinicians, health system leaders, and policymakers, among others—make well-informed decisions and thereby improve the quality of healthcare services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances presented by individual patients.

AHRQ appreciates appropriate acknowledgment and citation of its work. Suggested language for acknowledgment: This work was based on a Research White Paper, Outcome Measure Harmonization and Data Infrastructure for Patient-Centered Outcomes Research in Depression, by OM1, Inc., with subcontractors the American Board of Family Medicine and American Psychiatric Association through the Agency for Healthcare Research and Quality (AHRQ).

Suggested citation: Leavy MB, Cooke D, Hajjar S, Bickelman E, Egan B, Clarke D, Gibson D, Casanova B, Gliklich R. Outcome Measure Harmonization and Data Infrastructure for Patient-Centered Outcomes Research in Depression. Research White Paper. (Prepared by OM1, Inc., with subcontractors American Board of Family Medicine and American Psychiatric Association under Contract No. 75Q80119C00005.) AHRQ Publication No. 21-EHC003. Rockville, MD: Agency for Healthcare Research and Quality; November 2020. Posted final reports are located on the Effective Health Care Program search page.
DOI: https://doi.org/10.23970/AHRQEPCREGISTRYOUTCOME.

Acknowledgments

The authors gratefully acknowledge the following individuals for their contributions to this project: Ashish Deshpande, Philip Wickline, and Ariella Yosafat from OM1, Inc.; Lars Peterson, Robert Phillips, and Jill Shuemaker from the American Board of Family Medicine; Rachel Brodie from the Pacific Business Group on Health; Carolyn Turvey from the University of Iowa; and Elise Berliner and Mary Nix from AHRQ.

Structured Abstract

Background: Major depressive disorder is a common mental disorder. Many pressing questions regarding depression treatment and outcomes exist, and new, efficient research approaches are necessary to address them. The primary objective of this project is to demonstrate the feasibility and value of capturing the harmonized depression outcome measures in the clinical workflow and submitting these data to different registries. Secondary objectives include demonstrating the feasibility of using these data for patient-centered outcomes research and developing a toolkit to support registries interested in sharing data with external researchers.

Methods: The harmonized outcome measures for depression were developed through a multi-stakeholder, consensus-based process supported by AHRQ. For this implementation effort, the PRIME Registry, sponsored by the American Board of Family Medicine, and PsychPRO, sponsored by the American Psychiatric Association, each recruited 10 pilot sites from existing registry sites, added the harmonized measures to the registry platform, and submitted the project for institutional review board review

Results: The process of preparing each registry to calculate the harmonized measures produced three major findings. First, some clarifications were necessary to make the harmonized definitions operational. Second, some data necessary for the measures are not routinely captured in structured form (e.g., PHQ-9 item 9, adverse events, suicide ideation and behavior, and mortality data). Finally, capture of the PHQ-9 requires operational and technical modifications. The next phase of this project will focus collection of the baseline and follow-up PHQ-9s, as well as other supporting clinical documentation. In parallel to the data collection process, the project team will examine the feasibility of using natural language processing to extract information on PHQ-9 scores, adverse events, and suicidal behaviors from unstructured data.

Conclusion: This pilot project represents the first practical implementation of the harmonized outcome measures for depression. Initial results indicate that it is feasible to calculate the measures within the two patient registries, although some challenges were encountered related to the harmonized definition specifications, the availability of the necessary data, and the clinical workflow for collecting the PHQ-9. The ongoing data collection period, combined with an evaluation of the utility of natural language processing for these measures, will produce more information about the practical challenges, value, and burden of using the harmonized measures in the primary care and mental health setting. These findings will be useful to inform future implementations of the harmonized depression outcome measures.

Contents

Introduction

Major depressive disorder is a common mental disorder. Many pressing questions regarding depression treatment and outcomes exist, and new, efficient research approaches are necessary to address them. Connecting data across registries and other data collection efforts would yield a robust national data infrastructure to help address these questions, but a lack of harmonization in the outcome measures currently collected in research programs and clinical practice hinders the ability to connect these data sources. The Agency for Healthcare Research and Quality recently funded a stakeholder-driven effort to use the Outcome Measures Framework (OMF) to develop a minimum set of harmonized outcome measures for use in depression registries and clinical practice. The minimum measure set was finalized in 2018.

The primary objective of this project is to demonstrate the feasibility and value of capturing the harmonized depression outcome measures in the clinical workflow and submitting these data to different registries. Secondary objectives include demonstrating the feasibility of using these data for patient-centered outcomes research and developing a toolkit to support registries interested in sharing data with external researchers.

OM1, in collaboration with the American Board of Family Medicine's PRIME Registry and the American Psychiatric Association's Psychiatric Patient Registry Online (PsychPRO), seeks to demonstrate that data can be captured within the clinician workflow and directly from patients, exchanged seamlessly with multiple registries, and provided back to clinicians in a usable format to inform decision making. The project team will accomplish this by calculating six harmonized measures developed for depression (response, remission, recurrence, adverse events, suicide ideation and behavior, and mortality) within the registries, developing Fast Healthcare Interoperability Resources (FHIR) resources to extract and send the relevant data from electronic health records (EHRs) and any ancillary patient-reported outcome (PRO) systems to the registries, and building and implementing a Substitutable Medical Applications, Reusable Technologies (SMART) on FHIR app to provide information back to the clinician.

The purpose of this document is to describe the approach and actions taken to enable calculation of six harmonized outcome measures within the PRIME Registry and PsychPRO. Key findings from this practical implementation of the measures are presented here, along with a summary of the next steps.

Harmonized Outcome Measures for Depression

The harmonized outcome measures for depression were developed through a multistakeholder, consensus-based process in the prior project.[1] This pilot project is the first practical implementation of the measures. The six measures selected for this project are presented in Table 1.

Table 1. Harmonized outcome measures selected for pilot project

OMF Category	Outcome Measure	Definition
Survival	Death from suicide	Patient with a diagnosis of major depression or dysthymia who died from suicide, reported in 12-month intervals. *This should be captured where feasible; however, it should be noted that this information may not be recorded accurately or available to all providers.*
Clinical Response	Improvement in Depressive Symptoms—Response	Patient age 18 or older with a diagnosis of major depression or dysthymia and an initial Patient Health Questionnaire (PHQ-9)* score > 9 who demonstrates a response to treatment defined as a PHQ-9 score that is reduced by 50 percent or greater from the initial PHQ-9 score. *The PHQ-9, or another brief, publicly available, validated patient-reported instrument with empirically derived cutpoints equivalent to the PHQ-9 cutpoints for remission and response and for which an evidence-based crosswalk to the PHQ-9 exists, should be used to measure clinical response. Other measures may be used in addition for research or other purposes. Timeframe for measurement: • 6 months (+/- 60 days) • 12 months (+/- 60 days) *In some implementations, it would be beneficial to capture earlier responses and remissions and to obtain higher degrees of followup. Additional measurements outside of the windows listed above are recommended as supplemental measures.*
Clinical Response	Improvement in Depressive Symptoms—Remission	Patient age 18 or older with a diagnosis of major depression or dysthymia and an initial PHQ-9* score > 9 who demonstrates remission defined as a PHQ-9 score less than 5. *The PHQ-9, or another brief, publicly available, validated patient-reported instrument with empirically derived cutpoints equivalent to the PHQ-9 cutpoints for remission and response and for which an evidence-based crosswalk to the PHQ-9 exists, should be used to measure clinical response. Other measures may be used in addition for research or other purposes. Timeframe for measurement: • 6 months (+/- 60 days) • 12 months (+/- 60 days) *In some implementations, it would be beneficial to capture earlier responses and remissions and to obtain higher degrees of followup. Additional measurements outside of the windows listed above are recommended as supplemental measures.*

OMF Category	Outcome Measure	Definition
Clinical Response	Worsening in Depressive Symptoms—Recurrence	Patient age 18 or older with a diagnosis of major depression or dysthymia and an initial PHQ-9* > 9 who demonstrates remission (defined as a PHQ-9 score < 5) of at least 2 months' duration and subsequently experiences a recurrence of a depressive episode, defined as a 50 percent increase in PHQ-9 score or defined as a PHQ-9 score > 9 OR hospitalization for depression or suicidality.** *The PHQ-9, or another brief, publicly available, validated patient-reported instrument with empirically derived cutpoints equivalent to the PHQ-9 cutpoints for remission and response and for which an evidence-based crosswalk to the PHQ-9 exists, should be used to measure clinical response. Other measures may be used in addition for research or other purposes. **This definition was proposed by the workgroup. Data accruing from ongoing registries are needed to assess the feasibility of using this definition to capture recurrence. Timeframe for measurement: • 6 months (+/- 60 days) • 12 months (+/- 60 days) *In some implementations, it would be beneficial to capture earlier responses and remissions and to obtain higher degrees of followup. Additional measurements outside of the windows listed above are recommended as supplemental measures.*
Events of Interest	Suicide Ideation & Behavior	Selection of "several days," "more than half the days," or "nearly every day" option on PHQ-9 item 9 ("Thoughts that you would be better off dead or of hurting yourself in some way"). *Supplemental assessments of suicide ideation and behavior should be completed for patients who screen positive for suicide ideation on the PHQ-9 or when a clinician has concerns about suicidality. Supplemental assessments should be completed using an appropriate, brief, validated instrument, such as the Concise Health Risk Tracking scale. Includes nonfatal suicide attempts/suicide attempt behaviors, planning/preparatory acts, and active suicidal ideation.* Reported in 12-month intervals (in conjunction with the PHQ-9 suicide item).
Events of Interest	Adverse Events	Depression treatment-related adverse events. Use of a brief, publicly available, validated measurement tool to capture adverse events is recommended. Reported in 12-month intervals.

This pilot project will test the feasibility of calculating these measures using the standardized definitions developed under the prior project.[2] For each measure, the standardized definition defines the initial population for measurement (e.g., all depression patients), the outcome focused population (patients who experienced the outcome of interest), and the data criteria and value sets. The purpose of the standardized definitions is to enable the measures to be extracted consistently from EHRs and other data sources. An example of a standardized definition is presented in Appendix A.

Measure Calculation in the PRIME Registry

The PRIME Registry, sponsored by the American Board of Family Medicine (ABFM), was established to help provide family physicians and primary care clinicians a faster, easier way to evaluate practice performance, with built-in tools for population health, risk stratification, empanelment and more, all designed to improve primary care practice and patient outcomes and reduce the burden of reporting for the Centers for Medicare & Medicaid Services payment programs. The PRIME Registry has over 2,500 active clinicians participating from 47 States and data on 42 million patients.

The PRIME Registry captures individual-level clinical data that is generated and documented during the course of patient treatment and care. Data are extracted electronically from EHRs and online portals. Data include: patient demographic data; diagnosis(es) and interventions (e.g., medications); encounter data; patient-reported outcomes; and limited provider details. Data are used primarily to support a practice's quality improvement activities and quality reporting to the Center for Medicare and Medicaid Services. Only de-identified data can be secondarily used for research and this project. As such, patient informed consent and institutional review board (IRB) approval are not required. However, participating sites are not precluded from seeking IRB approval.

To participate in this project, the PRIME Registry recruited 10 pilot sites from existing registry sites, added the measures to the registry platform, and submitted the project for IRB review. These steps are described further below.

Pilot Site Recruitment

Recruitment of PRIME Registry sites for this project occurred in two phases. First, ABFM identified sites that were already using the Patient Health Questionnaire-9 (PHQ-9) through their EHR systems and contacted them by email. Some sites responded to this initial outreach. To expand the pool of participating sites, ABFM extended the invitation to all sites participating in the PRIME Registry and obtained commitments to participate from 10 registry sites (note, due to the COVID-19 public health emergency, some sites were unable to participate in the pilot, and two new sites were added). Table 2 provides descriptive information about the PRIME Registry pilot sites.

Table 2. PRIME Registry pilot sites

#	State	Practice Type	Number of Patients Served	Average Patient Age	Urban/Rural	EHR System
1	CO	Family Medicine	2,606	40.1	77.9 / 22.1	Amazing Charts
2	CO	Family Medicine	4,780	43.6	75.7 / 24.3	Aprima
3	CO	Family Medicine	496	57.1	82.0 / 18.0	Meditab
4	CO	Pediatrics	3,353	7.2	68.7 / 31.3	eMDs
5	VA	Family Medicine	6,394	47.9	83.7 / 16.3	eMDs
6	OK	Family Medicine	5,171	49.9	93.6 / 6.4	eMDs
7	GA	Family Medicine	3,221	56.1	5.4 / 94.6	eMDs

#	State	Practice Type	Number of Patients Served	Average Patient Age	Urban/Rural	EHR System
8	VA	Family Medicine	1,535	54.1	65.4 / 34.6	eMDs
9	OH	Internal Medicine	1,950	56.5	76.6 / 23.4	Amazing Charts
10	RI	Family Medicine	739	44.2	69.0 / 31.0	Amazing Charts
11	MO	Family Medicine	1,764	35	52.4 / 47.6	eClinical Works
12	WI	Family Medicine	985	47	46.5 / 53.5	Aprima

Adding Measures to the Registry Platform

The project team began the process of adding the six harmonized measures to the registry platform by comparing the harmonized measure definitions with the registry data dictionary. Several questions were identified related to the measure definitions, availability of data, and workflow. Because these questions are relevant to both registries, they are discussed further in the Key Findings section below.

Once the questions around the measure data were resolved, ABFM developed specifications for how and where the measures would be displayed within the registry platform, taking into account both the provider platform and the patient-facing PRO portal. ABFM worked with the registry technical vendor to extract the necessary data, configure the PRO tool to capture the PHQ-9 at the appropriate intervals, and calculate and display the measures in the appropriate locations. The registry technical vendor made the changes in the development environment and then moved the changes to a quality assurance environment, where ABFM registry staff completed user acceptance testing. Identified issues were reported back to the registry technical vendor, and corrections were made through an iterative, sprint process until all issues were resolved. The changes were moved to the production environment on February 27, 2020.

Appendix B provides screenshots of the measures as displayed with in the PRIME Registry platform.

IRB Review

ABFM provided the study protocol and IRB documentation to its local IRB for review on January 24, 2020. The project was determined to not be human subjects research for the following reasons: there are no human subjects involved, collection of data is for clinical practice and quality improvement purposes only, and analyses use de-identified data only. The determination was obtained on February 8, 2020.

Measure Calculation in PsychPRO

PsychPRO, a national mental health registry, was established to help psychiatrists and mental health professionals validate quality patient care through measurement and analysis to discover opportunities for improvement, avoid payment penalties and instead achieve bonuses for meeting quality reporting requirements, deploy cutting-edge technology to minimize the burden of data collection and reporting, and achieve optimal patient outcomes using tools to measure, chart, and benchmark clinical care. The PsychPRO Registry has over 600 active clinicians participating from 46 States and data on over 180,000 patients.

PsychPRO captures individual level clinical data that is generated and documented during the course of patient treatment and care. Data are electronically extracted directly from EHRs and from online portals. Data fields and elements vary with respect to standardization and include structured and unstructured data. Data include: patient demographic data; diagnosis(es) and intervention(s) (e.g., medications, therapy); encounter data; patient-reported outcomes; and limited provider details.

Data in PsychPRO are collected during routine assessment and clinical care of patients and used primarily to support a practice's quality improvement activities and quality reporting to the Center for Medicare and Medicaid Services. Only de-identified data can be secondarily used for research and this project. As such, patient informed consent and IRB approval are not required. However, participating sites are not precluded from seeking IRB approval.

To participate in this project, PsychPRO recruited 10 pilot sites from existing registry sites, added the measures to the registry platform, and submitted the project for IRB review. These steps are described further below.

Pilot Site Recruitment

The American Psychiatric Association (APA) focused on recruiting sites that are existing participants in the registry, are currently submitting data, and are using the PHQ-9 as part of their usual patient care. These inclusion criteria were important to have sites that were ready to begin data collection in March 2020. Because the registry includes a range of practice types and settings, this process was expected to return some variation in sites targeted for recruitment. APA identified eligible sites, invited them using an email invitation, and obtained commitments from 10 sites to participate. Table 3 provides descriptive information about the PsychPRO pilot sites.

Table 3. PsychPRO pilot sites

#	State	Practice Type	Number of Patients Served	EHR System
1	TX	State-designated Local Mental Health and Intellectual and Developmental Disability (IDD) Authority	17,800	Cerner
2	FL	Integrated Behavioral Healthcare center	2,770	InSync
3	FL	Adult and Geriatric Psychiatry	2,540*	Valant
4	CA	Freestanding outpatient psychiatric facility	400	Compulink
5	NY	Telemental Health Services	900	Valant
6	OR	Psychiatry Private Practice	70	Valant
7	OR	General Psychiatry Practice	100	Valant

#	State	Practice Type	Number of Patients Served	EHR System
8	TX	Mental Health Provider Network	790	Cerner; Now NetSmart
9	MD	General Psychiatry Practice	210	Valant
10	PA	Collaborative behavioral healthcare services integrated with Primary Care settings	1,620	Valant

*Patient count from 2018

Adding Measures to the Registry Platform

As with the PRIME Registry, APA began the process of adding the measures to the registry platform by comparing the harmonized measure definitions to the registry data dictionary. Questions identified during this process are discussed further in the Key Findings section below.

APA then provided specifications to the registry technical vendor and worked with the vendor to extract the necessary data and calculate and display the measures in the appropriate locations. Because APA and ABFM work with the same registry technical vendor, the process is similar to the steps described above.

IRB Review

APA provided the study protocol and IRB documentation to its local IRB for review on February 12, 2020. The project was determined to not be human subjects research for the following reasons: there are no human subjects involved, collection of data is for clinical practice and quality improvement purposes only, and analyses use de-identified data only. The determination was received on February 20, 2020, and a final letter of approval was obtained on March 6, 2020.

Key Findings

The primary objective of this project is to understand whether it is feasible to calculate these measures using the data that are routinely captured in different care settings (family medicine, primary care, mental health, etc.). The participating registries represent different care settings, as shown in Tables 2 and 3 above. The process of preparing each registry to calculate the harmonized measures produced three major findings:

1. Some clarifications were necessary to make the harmonized definitions operational
2. Some data necessary for the measures are not routinely captured in structured form
3. Capture of the PHQ-9 requires operational and technical modifications

These findings are discussed further below.

Harmonized Outcome Measure Definition Clarifications

Review of the standardized definitions by each registry resulted in some questions. First, the measure definitions did not specify the historical period (or look-back period) for capturing the major depression condition or dysthymia condition. A diagnosis of major depression or dysthymia is necessary to be included in the eligible patient population, but it was unclear how far back to look in a patient's record for that diagnosis. After review of the workgroup activities and the related quality measures, the project team clarified that the look-back period should be 12 months.

The adverse events measure also required a modification to capture adverse events related to procedures. The measure previously only captured adverse events related to medication treatment, but, as new procedure-based treatments become more common (e.g., vagus nerve stimulation therapy), it is important to capture adverse events related to these as well. Additional procedure codes were added to the Adverse Event value sets.

Lastly, the death from suicide measure was updated to include additional International Classification of Diseases (ICD)-10 suicide condition codes.

The project team updated the library of standardized measures and shared these changes with the development team building the SMART on FHIR app.

Availability of Measure Data Elements

After clarifying the measure definitions, the registries compared the measure data elements to the data routinely captured by participating sites and available for extraction into the registry. Because both registries are designed primarily to support quality improvement activities (rather than research) and rely on routinely collected data, many sites participate without seeking IRB approval, and patients typically do not provide informed consent. Any change in the registry data collection that resulted in the need for IRB approval at the site level and possibly informed consent would introduce substantial burden and reduce the sustainability of the registries. Thus, the registries indicated that it was critical not to request any data that are not routinely captured as part of providing care for patients with depression. Potential challenges related to the availability of PHQ-9, adverse event, suicide, and mortality data are described below.

PHQ-9 Scores

Four of the six harmonized measures—remission, response, recurrence, and suicide ideation—rely on the PHQ-9. Consistent use of a validated instrument, such as the PHQ-9, is critical for providing measurement-based care for depression and is part of routine practice in many care settings. The sites recruited by PsychPRO are already using the PHQ-9 and will continue to do so for this project. Some PRIME Registry sites are already using the PHQ-9, while others will adopt it as part of this project and as part of a broader effort to provide high-quality care to patients with depression. Because the PHQ-9 is already widely used and accepted, adoption of the PHQ-9 was not identified as a barrier to calculation of the measures in this pilot project. However, broader adoption of the measures may result in challenges from practices that use other validated instruments, such as the Geriatric Depression Scale or Hamilton Depression Rating Scale (HAM-D), rather than the PHQ-9. While the measure definitions allow for use of other instruments provided a crosswalk is available, very few crosswalks are currently available. Further work is needed in this area.

While the pilot sites are committed to using the PHQ-9, extraction of the PHQ-9 score data still presents some challenges. Sites participating in the pilot administer the PHQ-9 in several ways; some sites use the registry patient portal, some capture the PHQ-9 through their EHR, and some capture the PHQ-9 on paper and scan a copy of the instrument into their EHR. The registries have ready access to the PHQ-9s completed through the patient portals. For other PHQ-9s, the data must be extracted from the EHR to support calculation of the harmonized measures. In PsychPRO, registry sites use a custom field within the EHR to document the PHQ-9 score, and the registry is able to extract the summary score for measurement purposes. In the PRIME Registry, extraction of PHQ-9 data is a new requirement for this pilot, and site training and technical configurations are necessary to extract these data. Specifically, sites must set up an appropriately named custom field in the EHR and enter the data in the custom field so it can be identified as PHQ-9 scores and extracted.

Extraction of the PHQ-9 data to support the suicide ideation measure is more challenging. Item 9 on the PHQ-9 captures information about suicide ideation. However, the pilot sites currently enter only the summary PHQ-9 score into the EHR, as opposed to individual item scores. As with the summary score, capturing the item 9 data will require sites to set up an appropriately named custom field in the EHR and modify their workflow to enter the item 9 data in the field.

While it is feasible technically to extract the necessary data, it may be challenging to implement the necessary workflow changes to document the PHQ-9 summary score and item 9 data within the EHR so it can be extracted for measurement purposes. For the pilot project, this will be addressed through training and ongoing communication with the sites throughout the data collection period. However, this issue may be more challenging as practices adopt the harmonized measures outside of the framework and support of the pilot study.

Adverse Events Data

The intent of the adverse events measure is to capture all adverse events related to depression treatment. While some adverse events may be documented in the patient's medical record, it is possible that patients experience other side effects that they either do not discuss with their provider or which are not noted because they are not significant enough to result in treatment

changes. The harmonization workgroup recommended "use of a brief, publicly available, validated measurement tool to capture adverse events," as a way to supplement the data found in the medical record. Specifically, the group suggested the Frequency, Intensity, and Burden Side Effects Rating Scale (FIBSER).[3] The FIBSER is a short, three-item patient-reported questionnaire that documents the frequency, intensity, and burden of side effects.

Through discussions with the registries and pilot sites, the project team learned that no sites currently use the FIBSER to capture adverse events. While there is evidence of use of the FIBSER in research settings, there is little to no evidence in the peer-reviewed literature of use of the FIBSER in routine clinical care. Because the FIBSER is not routinely used and would be added for the purposes of this study only, both registries expressed concerns that use of the FIBSER would require IRB approval of this study as a research study and possibly patient informed consent. This would introduce substantial burden for participating sites. Because of this concern, the adverse events measure will be calculated using data captured in the EHR only for this pilot study. Further work is needed to explore the utility of the FIBSER in routine clinical care.

Suicide Ideation and Behavior

The suicide ideation and behavior measure requires data on nonfatal suicide attempts/suicide attempt behaviors, planning/preparatory acts, and active suicidal ideation. The registries noted concerns about the availability of these data within the participating sites' EHRs. In some cases, these data may be documented in the EHR in unstructured form, making them challenging to extract for measurement purposes. A 2015 review of EHR data from primary care practices found that only 3 percent of patients with documentation of suicide ideation in unstructured clinical notes had a corresponding ICD-9 code, and only 5 percent of patients who indicated suicide ideation on the PHQ-9 (item 9) had a corresponding ICD-9 code. For suicide attempt, 19 percent of patients with a suicide attempt documented in the notes had a corresponding ICD-9 code.[4] While these findings may not be broadly generalizable to the sites participating in this pilot (particularly the PsychPRO sites), they do support the concerns expressed by the registries about the lack of structured documentation for suicide ideation and behavior. The registries also noted concerns about the possibility that these data will be missing entirely from the EHR. For example, patients may present to the emergency room rather than the primary care practice in these cases, leading to a gap in the patient's medical record.

While the suicide ideation and behavior measure is important to capture, it is equally important to ensure that all necessary data are captured and suicidal behaviors are not underreported. The harmonization workgroup emphasized that suicidal behaviors should be measured when systematic ascertainment from all possible sites of care is possible. Due to the practical challenges of systematic ascertainment, the pilot project will examine suicide ideation as measured using item 9 of the PHQ-9. The project team will examine the available data on suicidal behaviors in structured form and compare these data to data on suicidal behaviors extracted from notes using natural language processing (NLP). This information will be useful to inform future implementations of the measures.

Mortality Data

As with suicide ideation and behavior, the registries noted concerns about the availability of the necessary data to calculate the death from suicide measure. Death may not be recorded in the EHR, and, even when the date of death is recorded, information on the cause of death may not be readily available. In addition, in cases of suicide, a different cause of death may be listed because of the perceived stigma of suicide. While death from suicide is important to measure, it is equally important to ensure that all necessary data are captured and deaths are not underreported. Because of the practical challenges of systematic ascertainment, the pilot project will focus on examining differences in the information recorded in structured versus unstructured data using NLP. This examination may provide useful information to guide future implementations of the measures.

Implementation of the PHQ-9

As noted above, four of the six measures depend on the PHQ-9, and PHQ-9 scores must be documented such that they can be extracted into the registry for measurement purposes. An additional challenge is the implementation of a workflow that allows for consistent capture of the PHQ-9 without overburdening clinicians, practice staff, or patients. While some sites participating in the pilot project already use the PHQ-9, some PRIME sites are adopting it as part of this project. The PRIME Registry developed a workflow in which eligible patients receive a link to complete the PHQ-9 approximately 6 to 8 weeks after a visit. This followup PHQ-9, which is sent directly to the patient, will enable clinicians to monitor a patient's depression symptoms (and possibly response to treatment) without requiring another office visit. However, implementation of this feature required careful planning so that clinicians would be alerted promptly if a patient reported severe symptoms or suicide ideation. The Registry instituted a process in which patients who report suicide ideation receive an immediate prompt to contact the National Suicide Hotline or their local emergency department for help; clinicians also receive a special message notifying them about the patient's responses.

These four measures also rely on patients' willingness to complete the PHQ-9. Most PsychPRO pilot sites understand their patients' willingness to complete a PHQ-9 and preferences, such as completing the survey at the time of the office visit or from home (outside of the office visit). However, the willingness of patients to complete the PHQ-9 through the PRIME Registry patient portal outside of an office visit has not yet been assessed. This project will provide information about patient response rates that may be useful to guide future implementations of the harmonized measures.

Next Steps

The next phase of this project will focus collection of the baseline and followup PHQ-9s, as well as other supporting clinical documentation. The project team will monitor data collection and site participation during this period to identify any workflow or technical issues and to understand completion rates for the PHQ-9 outside of an office visit. A series of listening sessions (1-hour webinars with a short presentation followed by open discussion) will be offered for sites on a regular basis throughout the data collection period to provide updates, reiterate key training messages, and discuss any issues. Registries will also provide information on the value and burden of the measures at the conclusion of the one-year data collection period.

In parallel to the data collection process, the project team will examine the feasibility of using NLP to extract information on PHQ-9 scores, adverse events, suicide, and suicidal behaviors from unstructured data. This information may be useful to guide future implementations of these measures.

Conclusions

The process of preparing the PRIME Registry and PsychPRO to calculate the harmonized outcome measures was completed in 10 months. One delay related to the MIPS submission period was encountered, but this delay was anticipated and will not affect the overall data collection timeframe.

The process yielded three major findings related to the harmonized definition specifications, the availability of the necessary data, and the workflow challenges related to implementation of the PHQ-9. The data collection period, combined with an evaluation of the utility of NLP for these measures, will produce more information about the practical challenges, value, and burden of using the harmonized measures in the primary care and mental health setting. These findings will be useful to inform future implementations of the harmonized depression outcome measures.

References

1. Gliklich RE, Leavy MB, Cosgrove L, et al. Harmonized Outcome Measures for Use in Depression Patient Registries and Clinical Practice. Ann Intern Med. 2020;172(12):803-809.
2. Gliklich RE, Leavy MB, Li F. AHRQ Methods for Effective Health Care. In: Standardized Library of Depression Outcome Measures. Rockville (MD): Agency for Healthcare Research and Quality (US); 2020.
3. Wisniewski SR, Rush AJ, Balasubramani GK, Trivedi MH, Nierenberg AA, Investigators S. Self-rated global measure of the frequency, intensity, and burden of side effects. J Psychiatr Pract. 2006;12(2):71-79.
4. Anderson HD, Pace WD, Brandt E, et al. Monitoring Suicidal Patients in Primary Care Using Electronic Health Records. The Journal of the American Board of Family Medicine. 2015;28:65-71.

Appendix A. Standardized Measure Definition

Outcome Measure	Definition	Population Criteria - Data Objects and Expression Logic			Data Criteria / Value Sets
		Initial Population (or Modified for Individual Outcome)	Outcome Focused Population	Exclusion	
Improvement in Depressive Symptoms – Response	Patient age 18 or older with a diagnosis of major depression or dysthymia and an initial PHQ-9* score > 9 who demonstrates a response to treatment defined as a PHQ-9 score that is reduced by 50% or greater from the initial PHQ-9 score. *The PHQ-9 or another brief, publicly available, validated patient-reported instrument with empirically derived cutpoints equivalent to the PHQ-9 cutpoints for remission and response and for which an evidence-based crosswalk to the PHQ-9 exists should be used to measure clinical response. Other measures may be used in addition for research or other purposes. Timeframe for measurement: • 6 months (+/- 60 days) • 12 months (+/- 60 days) *In some implementations, it would beneficial to capture earlier responses and remissions and to obtain higher degrees of follow-up. Additional measurements outside of the windows listed above are recommended as supplemental measures.*	o Condition ● code: in "Major Depression Condition" or "Dysthymia Condition" value set ● onset[x] ● clinicalStatus: any (active/inactive/remission/resolved, etc)	o AND: Observation ● Status: final ● Code: in "PHQ-9 Total Score Observable" ● effectiveDateTime ● valueQuantity o AND: ● PHQ-9 observation demonstrates appropriate change in score		Remission_Response_VS!A1

Appendix B. Measure Display Within the PRIME Registry

Figure B-1. Measures as displayed at the top level of the registry dashboard

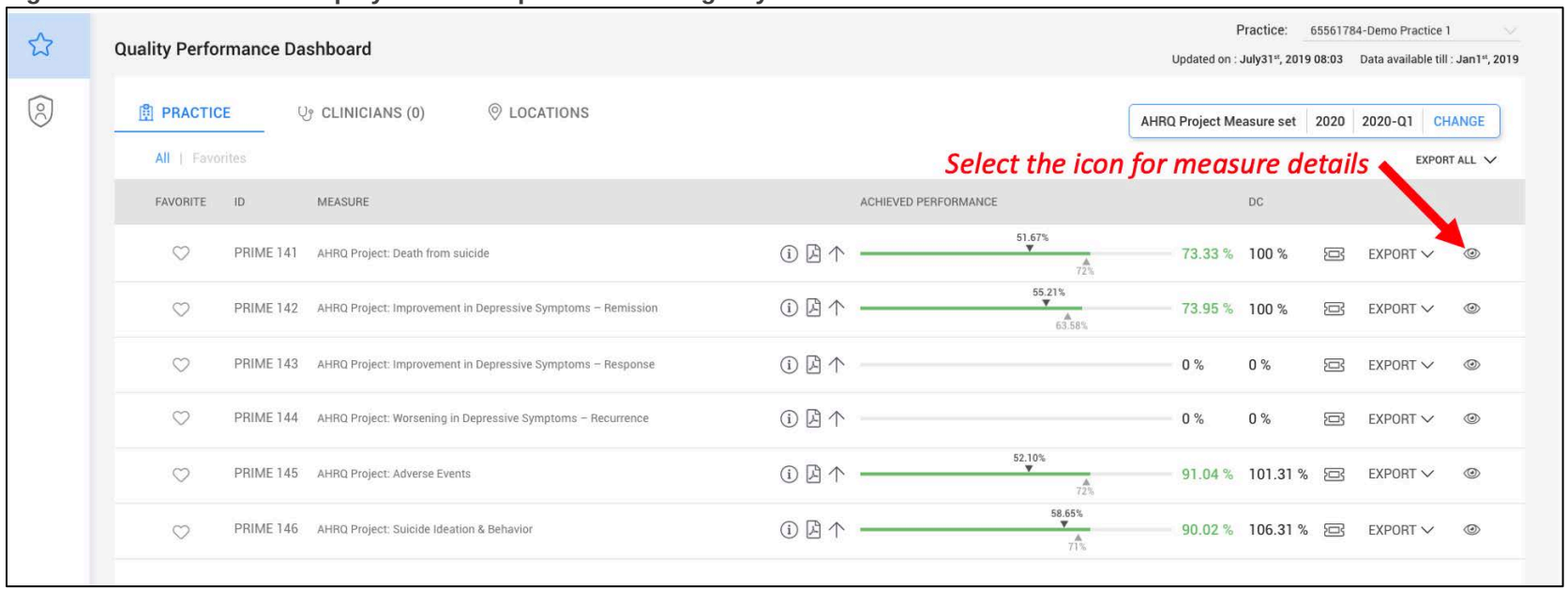

Figure B-2. Measure detail view in the registry dashboard

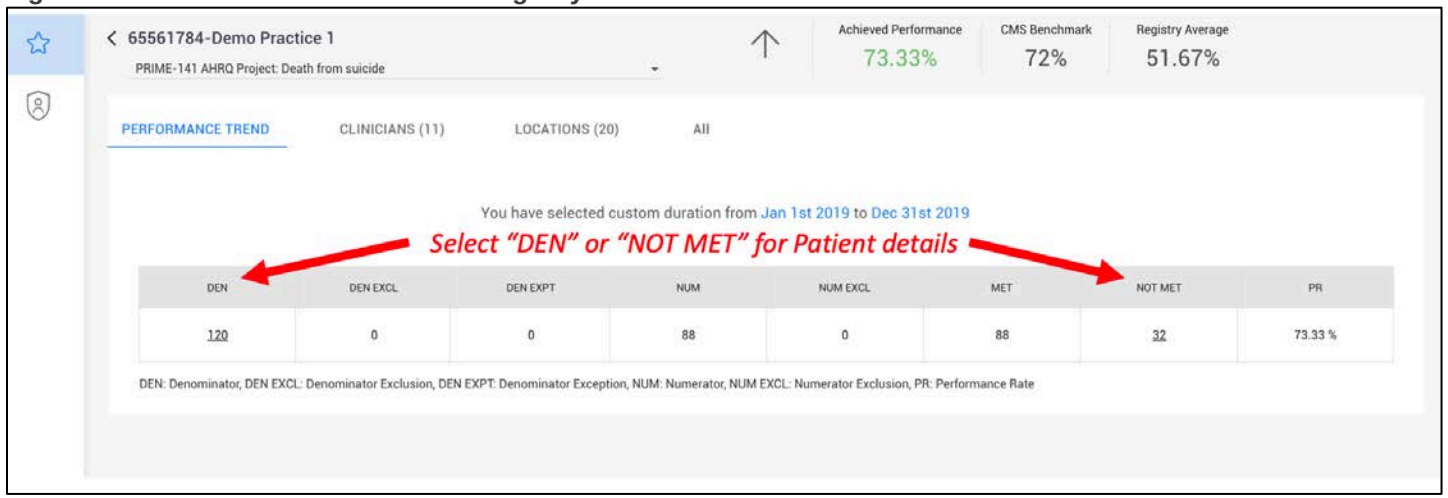

Figure B-3. Patient detail view accessible through the registry dashboard

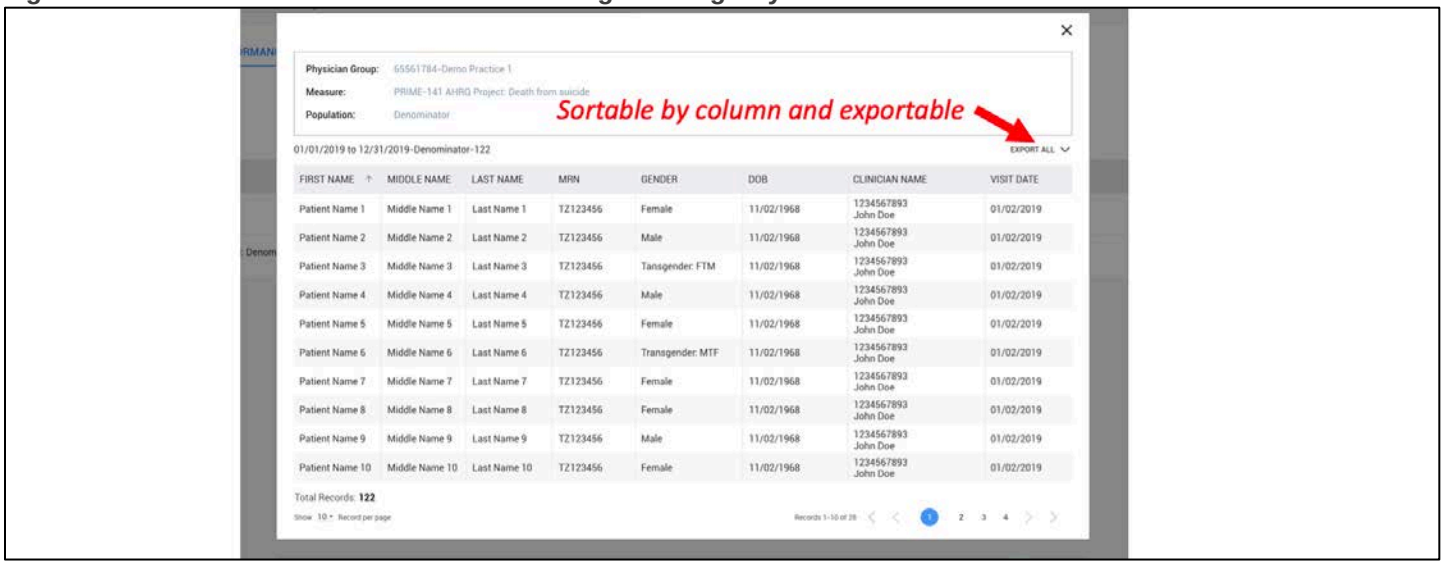

Figure B-4. Patient survey in the patient-facing PRO portal

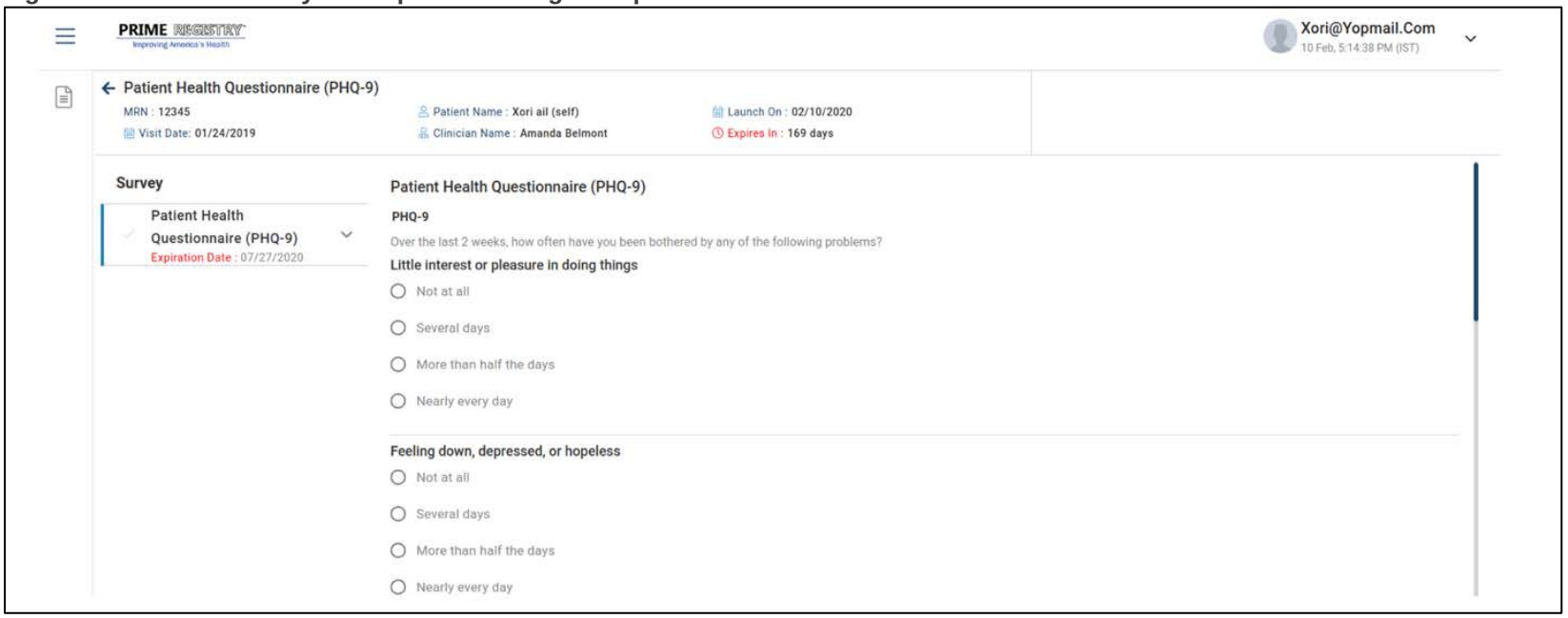

Figure B-5. Practice administrator dashboard view

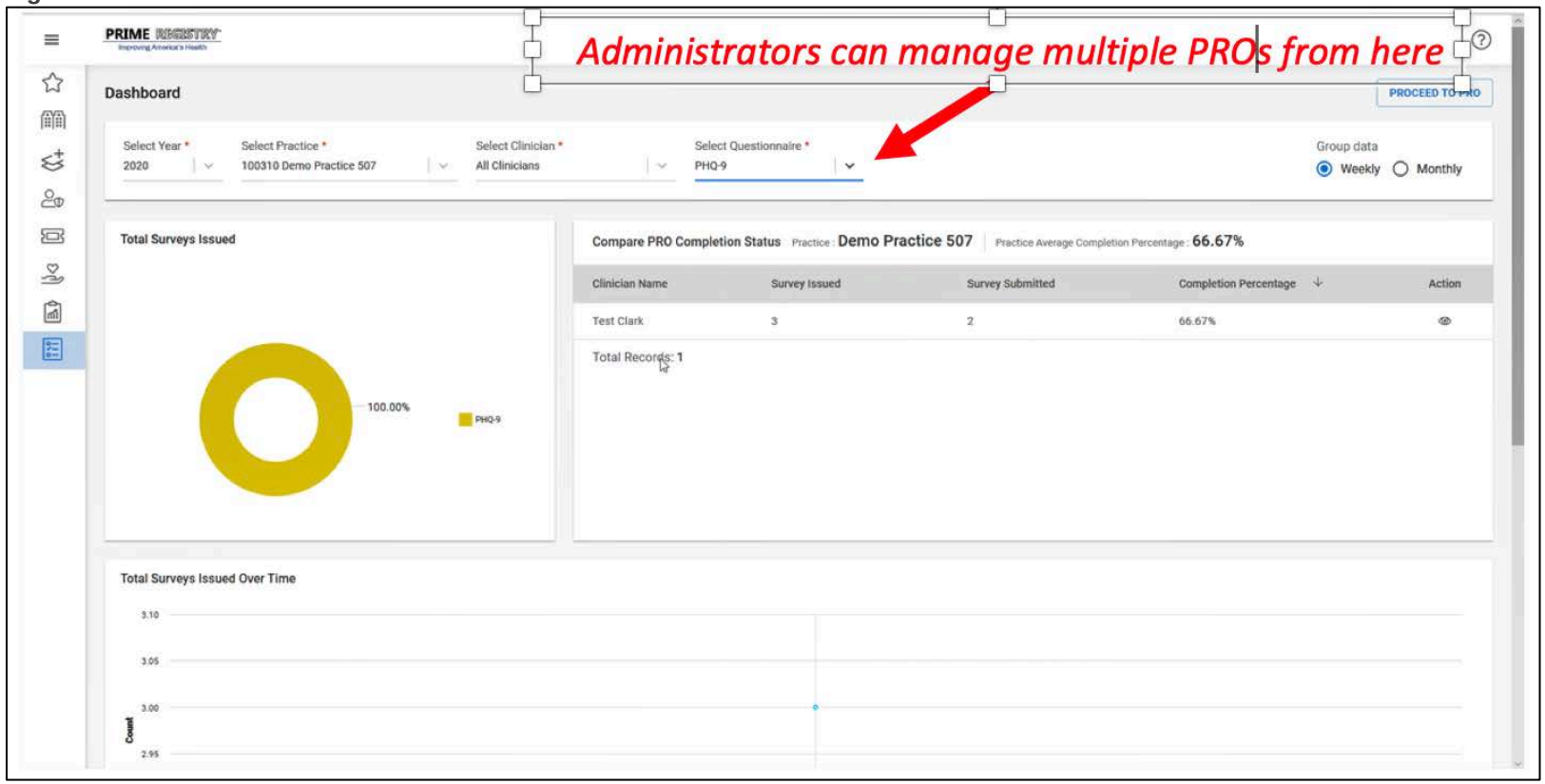

Figure B-6. Practice administrator link to patient responses from dashboard

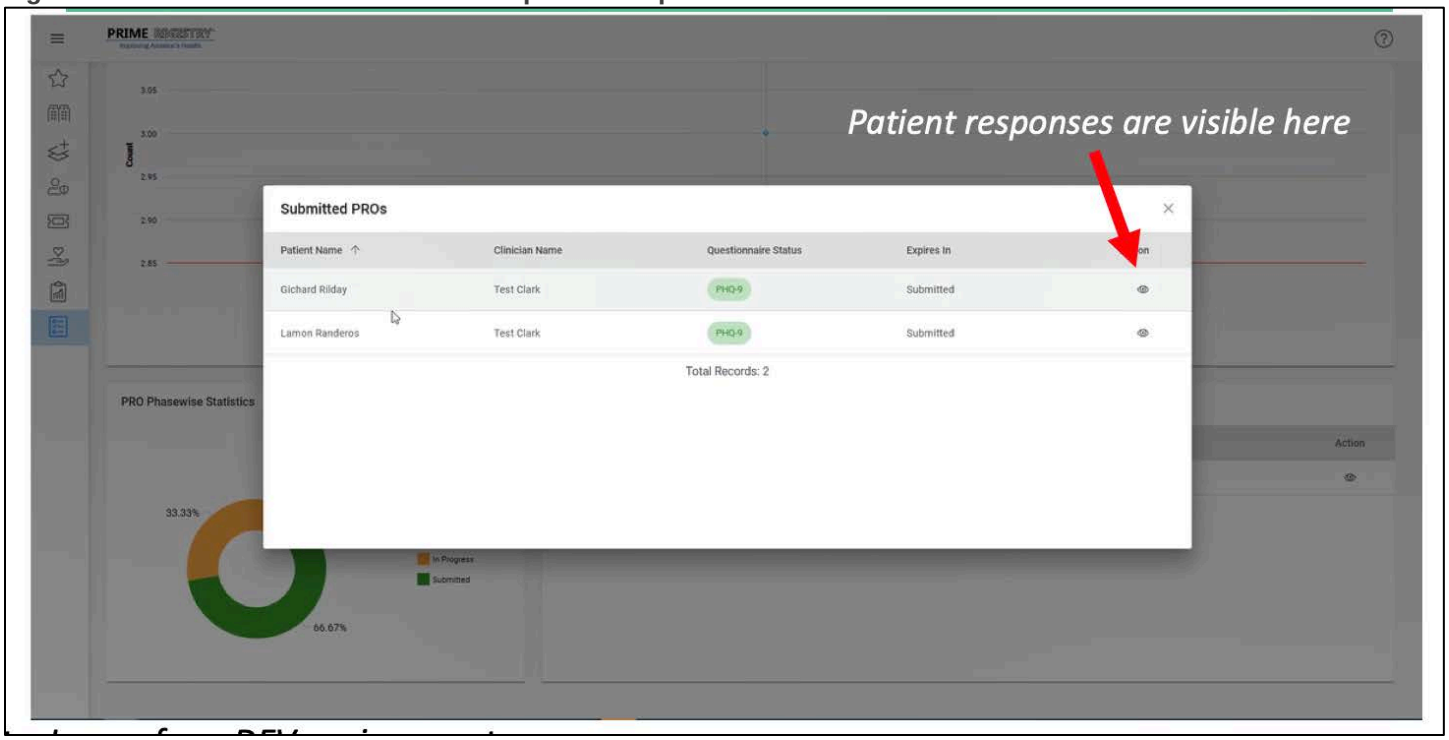

Figure B-7. Practice administrator view of survey responses for individual patient

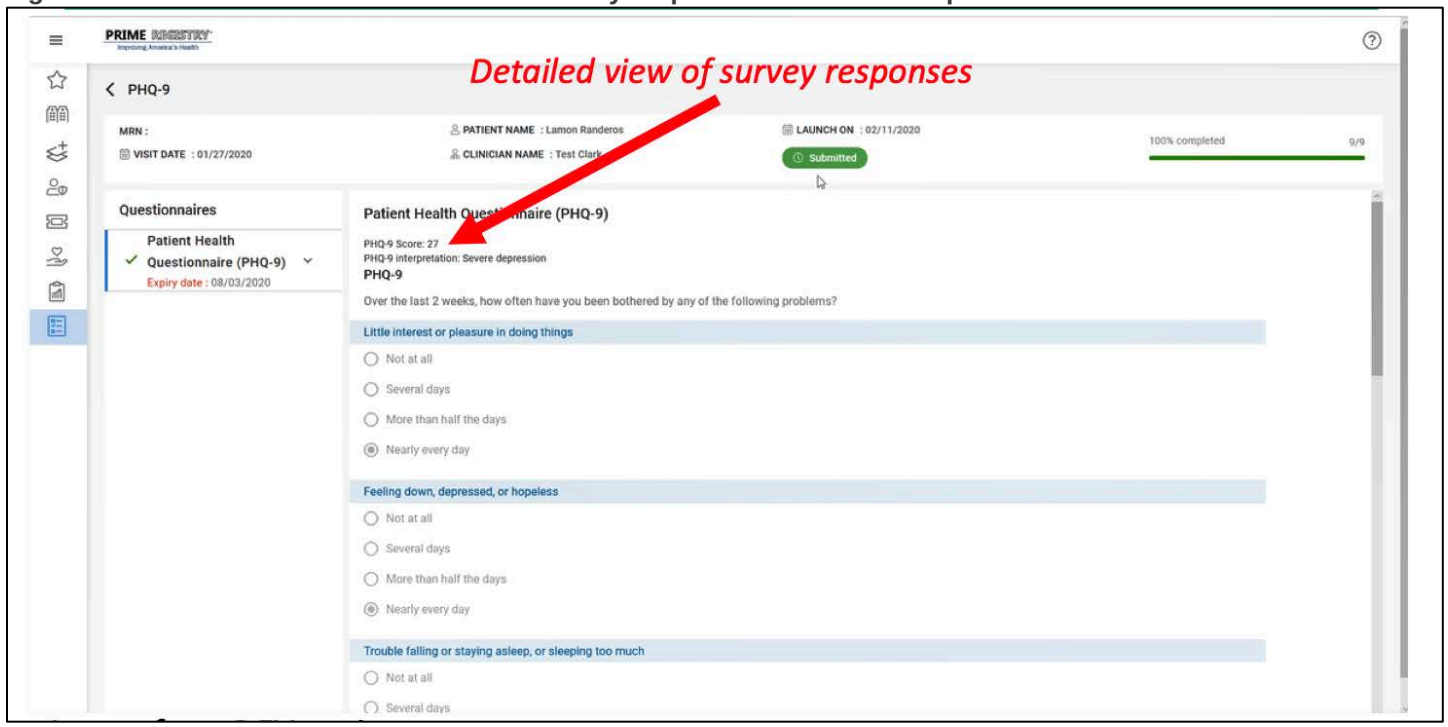

Figure B-8. Patient browser view for practice administrator