The Office of the National Coordinator for
Health Information Technology

Pilot of a Data Quality Framework to Support Patient Matching

August 2019

TABLE OF CONTENTS

Executive Summary

High-quality patient demographic data are fundamental to accurate patient identification and matching. Similarly, accurate patient identification and matching is pivotal to interoperability, patient safety, and research, such as patient-centered outcomes research (PCOR). With the adoption of rapid developments in health information technology (IT) and advancements in electronic health record (EHR) systems, the scale of growth of the data captured, stored, and exchanged continues to increase. Because of the constant growth in patient populations and their diversity, along with increasing complexity of health care networks, health care staff must have the capabilities and best practices to capture high-quality data on the front line.

The Department of Health and Human Services (HHS) Office of the Assistant Secretary for Planning and Evaluation (ASPE) engaged the Office of the National Coordinator for Health Information Technology (ONC) to study and advance patient matching, aggregation, and linking (PMAL, or the PMAL Project) through the PCOR Trust Fund. ASPE oversees federal health programs funded through the PCOR Trust Fund to build data capacity for research. PCOR studies are designed to produce new scientific evidence that informs and supports the health care decisions of patients, families, and their health care providers. PCOR studies examine the effectiveness of prevention and treatment options while taking into consideration the preferences, values, and questions that are important to patients when they make health care choices.

The PMAL Project took a multi-faceted approach to addressing obstacles to linking patient data across research, claims, and clinical data sets. The project was intended to study and advance methods of accurately linking and aggregating patient records from various sources to increase the volume and quality of data available to support research. Accurate patient matching and data aggregation offers additional benefits, such as increased patient safety and improved health and well-being.

Under the auspices of this project, ONC tested the application of a Patient Demographic Data Quality (PDDQ) Framework,[1] a health care-specific demographic data management model that is based on the Capability Maturity Model Integration Institute (the CMMI Institute)Data Maturity Model℠.[2] The PDDQ Framework provides guidance for the standardization of policies, processes, procedures, and practices both locally at the clinic level and across an entire health care system. The framework includes standards and processes to help health care organizations improve the quality of patient demographic data by preventing the creation of duplicate patient records and helping them better manage existing duplicate records. The PDDQ Framework can be used in full or as an abbreviated version specific to ambulatory care settings. ONC used the Ambulatory Guide for this project.

From December 2016 through June 2017, ONC collaborated with OCHIN (a health IT support and service provider for safety-net clinics), and the Kaiser Permanente Center for Health Research (CHR), to test the

---

[1] Office of the National Coordinator for Health Information Technology, Patient Demographic Data Quality Framework. Retrieved from www.healthit.gov/playbook/pddq-framework/.

[2] Capability Maturity Model Institute, Data Management Maturity (DMM). Retrieved from https://cmmiinstitute.com/data-management-maturity.

PDDQ Framework's Ambulatory Guide[3] at participating OCHIN-affiliated community health centers. The pilot study results suggest that using the Ambulatory Guide has the potential to achieve improvements in data quality with a modest investment. The implementation considerations in this paper are intended to assist organizations in implementing the PDDQ Framework (or the Ambulatory Guide) and to encourage their contributions to developing additional resources to improve the quality of demographic data and patient matching. Further study is warranted to establish a correlation between the application of the PDDQ Framework and positive impact on metrics for tracking duplicates.

---

[3] Patient Demographic Data Quality Framework Ambulatory Guide. Retrieved from https://www.healthit.gov/playbook/ambulatory-guide/.

Introduction

This project sought to test a framework to improve the capture and management of demographic information and improve patient matching.

The rapid evolution of health information technology (IT) has increased the diversity and number of electronic health data sources. The proliferation of data sources and exchanges has exponentially increased the need to seamlessly exchange data between systems to better assure patient identification and care coordination. Increased data exchange raises the number of times that separate systems must mutually establish the identity of a patient through matching. Accurate patient matching ensures that patient records are not mistakenly matched and combined.[4] When records are correctly matched and then aggregated, a single picture of the patient emerges, which accordingly supports safe and efficient care as well as research.

High-quality data are fundamental to interoperability and safe and efficient care.

Accurate patient matching is vital for the delivery of appropriate care because patients may receive care from various unaffiliated organizations. Knowing what treatment was provided elsewhere can be critical to proper treatment in the moment.

Patient matching and data aggregation are also necessary to support health care research. Patient-centered outcomes research (PCOR) is designed to produce scientific evidence that informs and supports the health care decisions of patients, families, and their health care providers. The Patient-Centered Outcomes Research Trust Fund, overseen by the Office of the Assistant Secretary for Planning and Evaluation (ASPE), supports these projects within the Department of Health and Human Services (HHS). PCOR studies focus on evaluating the effectiveness of prevention and treatment options in consideration of the preferences, values, and questions patients face when making health care choices.[5] Through accurate patient matching, researchers can find all records relevant to an individual and aggregate that information into a single picture to study. Improved patient matching facilitates tracking patient outcomes over time and across locations, such as tracking patient improvements or the results of using a specific pharmaceutical over the course of years.[6]

---

[4] US Government Accountability Office, Health Information Technology: Approaches and Challenges to Electronically Matching Patients' Records Across Providers, GAO-19-197 (January 2019), available at https://www.gao.gov/assets/700/696426.pdf.

[5] See, e.g., Office of the Assistant Secretary for Planning and Evaluation, Patient-Centered Outcomes Research Trust Fund (home page). Retrieved from Patient-Centered Outcomes, https://aspe.hhs.gov/patient-centered-outcomes-research-trust-fund. Last accessed August 21, 2019.

[6] Pew Charitable Trusts, Report: Enhanced Patient Matching is Critical to Achieving Full Promise of Digital Health Records (October 2, 2018), available at https://www.pewtrusts.org/en/research-and-analysis/reports/2018/10/02/enhancedpatient-matching-critical-to-achieving-full-promise-of-digital-health-records.

Poor capture and low quality of electronic demographic data can impede successful patient matching. Variations at the organizational level that affect successful patient matching include differences in care settings and organizational structures, data entry and exchange practices, and inconsistent priorities across health care environments.

ONC is committed to creating interoperable health IT infrastructure for patient-centered outcomes research that improves individual, community, and population health.

Technical impediments to accurate patient matching include lack of standardization in data elements and limitations in IT resources.[7] Moreover, demographic data can often be incomplete and inaccurate, which leads to duplication of patient records. Lack of uniformity in the entry, format, storage, and update of patient demographic data challenges the accurate analysis and reporting on patient health across systems or the seamless linkage of data from different systems. Variation in electronic health record (EHR) product implementations may further introduce data heterogeneity within and across institutions. For example, different locations may use different formats to capture hyphenated names.[8] Establishing standardized demographic data elements, collection methods, and transparent patient matching tools is necessary for successful patient matching. Such standardization of data and processes must adequately address variations in the environment, technical capacity, patient population, and workflow of any given setting.

The Office of the National Coordinator for Health Information Technology (ONC) supports the creation of longitudinal health records and the sharing of patient data across organizations for the purposes of improving health care delivery and generating evidence through research. ONC undertook the Patient Matching, Aggregation, and Linking (PMAL) project, which employed a multi-faceted approach to addressing obstacles to linking patient data across research, claims, and clinical data sets.

This activity under the PMAL project tested one application of a Patient Demographic Data Quality (PDDQ) Framework.[9] The PDDQ Framework is a health care-specific, demographic data management model that ONC developed in partnership with the Capability Maturity Model Integration Institute (the CMMI® Institute). It is based on the CMMI Institute's Data Maturity Model℠.[10] The PDDQ Framework provides guidance for standardization of policies, procedures, and practices both locally at the clinic level and across an entire health care system to support improvements to the quality of patient demographic data. The framework's guidance focuses on preventing creation of duplicate patient records and how to better manage existing duplicate records. Users can apply the PDDQ Framework in full or use an abbreviated

---

[7] The Sequoia Project. (2018). A Framework for Cross-Organizational Patient Identity Management. Retrieved from A Framework for Cross-Organizational Patient Identity Management, available at: https://sequoiaproject.org/resources/patient-matching/2015; Agency for Healthcare Research and Quality. (2018). Improving Data Collection across the Health Care System. Retrieved from http://www.ahrq.gov/research/findings/final-reports/iomracereport/reldata5.html.

[8] US Government Accountability Office, Health Information Technology: Approaches and Challenges to Electronically Matching Patients' Records Across Providers, GAO-19-197 (January 2019), available at https://www.gao.gov/assets/700/696426.pdf .

[9] Office of the National Coordinator for Health Information Technology, Patient Demographic Data Quality Framework. Retrieved from www.healthit.gov/playbook/pddq-framework/.

[10] Capability Maturity Model Institute Data Management Maturity (DMM). Retrieved from https://cmmiinstitute.com/data-management-maturity.

version specific to ambulatory settings (the Ambulatory Guide). For this pilot test, ONC used the Ambulatory Guide.

From December 2016 through June 2017, ONC collaborated with OCHIN[11] (a health IT support and service provider for safety-net clinics) and the Kaiser Permanente Center for Health Research (CHR),[12] to pilot test an abbreviated version of the PDDQ Framework (the Ambulatory Guide) at participating OCHIN-affiliated community health centers. The goal of the pilot testing was to determine whether the PDDQ Framework could improve the collection of demographic data and either reduce or successfully manage duplicate patient records at small community health centers. This report provides a summary of the pilot's design, findings, limitations, and implementation considerations to help institutions interested in testing or implementing either the full PDDQ Framework or an abbreviated version such as the Ambulatory Guide.

Aims and Objectives

The overall goal of the PDDQ Framework pilot project was to pilot test the PDDQ Framework's ability to improve patient matching and demographic data quality. The PDDQ Framework was designed to provide a way to effectively create and implement demographic data management processes and procedures so health care organizations might better evaluate and improve their patient demographic data quality.

The PDDQ Framework pilot executed the following study objectives**:**

(1) Identification of best practices for the collection and management of patient demographic data and patient matching methodologies through a literature review
(2) Baseline assessment of pilot sites' current policies and practices for collecting and managing patient demographic data
(3) Implementation of the PDDQ Framework's Ambulatory Guide at the pilot sites and applying the PPDQ Framework scoring methodology to assess each site's PDDQ Ambulatory Guide score pre- and post-pilot to measure change
(4) Evaluation of the PDDQ Framework Ambulatory Guide's performance as a data management method for patient demographic data

Literature Review

PURPOSE AND METHODOLOGY

The CHR team conducted a literature review to build on the literature review contained in the 2014 ONC Patient Identification and Matching Final Report.[13] The CHR Team used 36 search terms related to patient

---

[11] OCHIN is a 501(c)(3) not-for-profit community-based health IT collaborative based in Portland, Oregon. OCHIN's mission is to pioneer the use of health IT in caring for the medically underserved. OCHIN serves community health centers (CHC), including Federally Qualified Health Centers (FQHC), rural and school-based health centers, safety-net providers, and public health and correction facilities across the nation. Retrieved from https://ochin.org/.

[12] Based in the Northwest and Hawaii regions of Kaiser Permanente, the Center for Health Research pursues a vigorous agenda of public health research within large, diverse populations inside and outside Kaiser Permanente. Retrieved from https://research.kpchr.org/#6072017.

[13] Morris, G. et al., Patient Identification and Matching Final Report (February 7, 2014), available at https://www.healthit.gov/sites/default/files/patient_identification_matching_final_report.pdf.

matching and identification and focused on both industry artifacts and scholarly journals. Appendix A – Literature Review Methodology presents the search terms used in the literature review.

FINDINGS

Three main themes emerged from the literature review: (1) a need to improve data quality and integrity; (2) several considerations for implementing registration process standards within health care systems; and (3) suggestions to leverage data standards across health care systems. Additional themes addressed the potential for a more accurate patient record to develop trust between patients and clinicians and across data partners, and concerns about safety issues and costs associated with inaccurate or incomplete patient records.

The following patient matching concepts and issues were generally encountered in both industry and academic settings in the course of the literature review:

- **Safety –** Patient identification is a fundamental safety consideration in the health care industry. One study indicated that of the 195,000 deaths occurring each year due to medical errors, 10 out of every 17 result from patient identity errors.[14]
- **Volume of Duplication Rates –** Some industry stakeholders have estimated the rates of duplicate record in EHRs as between 8 to 20 percent of total records.[15] Because there is no standardized or industry-accepted method to measure or report duplicate record rates, the validity or comparative value of these estimates is not certain. McCoy, et al. noted that these findings suggest that existing approaches for managing patient identity do not adequately address safe creation of a single, longitudinal patient record.[16]
- **Cost Savings –** Cost savings in health care are possible if patients do not undergo duplicate tests or when resources are not expended to resolve duplicate or overlay records. One study suggested that resolving each duplicate record costs health care organizations about $96 per duplicate.[17]
- **Data Quality and Integrity –** Entering and updating patient information as correctly, completely, and consistently as possible is essential to data quality and critical to any successful patient matching and interoperability strategy. Strategies for managing patient demographic data that emphasize prevention, detection, removal, and error mitigation have succeeded in this area.[18] These strategies include ensuring consistent data formats to facilitate data exchange; addressing and correcting data discrepancies; accurately capturing racial, ethnic, and gender data; and proactively engaging patients to ensure that their data remains accurate over time.

---

[14] Smart Card Alliance. (2014). Effective Healthcare Identity Management: A Necessary First Step for Improving U.S. Healthcare Information Systems. Retrieved from https://www.securetechalliance.org/resources/pdf/Healthcare_Identity_Brief.pdf.

[15] Bipartisan Policy Center. (2012). Challenges and Strategies for Accurately Matching Patients to Their Health Data. Retrieved from Challenges and Strategies for Accurately Matching Patients to Their Health Data: https://www.bipartisanpolicy.org/library/challenges-and-strategies-accurately-matching-patients-their-health-data/.

[16] McCoy AB, W. A. (2013). Matching identifiers in electronic health records: Implications for duplicate records and patient safety. British Medical Journal of Quality & Safety, 22(3):219-224. doi: 10.113.

[17] Just BH, M. D. (2016). Why patient matching is a challenge: Research on master patient index (MPI) data discrepancies in key identifying fields. Perspectives in Health Information Management / AHIMA, American Health Information Manag.

[18] McCoy AB, W. A. (2013). Matching identifiers in electronic health records: Implications for duplicate records and patient safety. British Medical Journal of Quality & Safety, 22(3):219-224. doi: 10.113.

- **Workflow –** Developing and implementing consistent approaches to data capture and storage helps to improve data integrity. The registration process may occur at several locations throughout a facility, which creates the potential for different data capture processes[19] and highlights the importance of using a consistent approach.
- **Trust Among Patients, Clinicians, and Health Care Organizations –** Improving data quality and reducing errors improves confidence among patients, clinicians, and other stakeholders that the data received is accurate, reliable, and actionable. Following a defined set of principles has been shown to improve confidence in the integrity of the data and in the decisions made based on that data.[20]

Pilot Testing Approach

ONC collaborated with CHR and OCHIN to pilot test a data management approach based on the PDDQ Framework. CHR provided project management support while OCHIN staff led communication and coordination with pilot sites. In its capacity as the technical advisor, ONC provided subject matter expertise and guidance.

ONC, CHR, and OCHIN assessed the participating clinics at each pilot site throughout the pilot's period of performance. Appendix B – IT Data Report to Support Measurement presents a list of the data elements collected and reported in these assessments.

In addition to regular assessments, this pilot project trained the pilot site staff assigned to gather and/or manage patient demographic data and monitor site performance. Before an initial assessment of each participating site, CHR worked with the ONC and OCHIN project leads to develop a guiding document that focused on specific clinic practices during the patient registration process. The OCHIN staff conducted a qualitative survey of each site's practices, followed by a baseline assessment of patient registration and data collection practices at participating clinics. The OCHIN team monitored changes to duplicate record creation and existing duplicate rates. Data Quality Teams[21] were established at each pilot site to ensure that the pilot was implemented locally, involved the right stakeholders, and collected the appropriate data. The Data Quality Teams for each pilot site included IT staff, clinic managers, and billing staff identified by community health center management. Data Quality Teams were tasked with advocating for and implementing data management improvement measures within their respective clinics. They were provided with training materials and resources to support their pilot test of the PDDQ Framework's Ambulatory Guide.

PILOT DETAILS

Pilot sites received the PDDQ Framework's Ambulatory Guide, which narrows the overall PDDQ Framework guidance specifically to the ambulatory setting. The application of the framework was

---

[19] Lichtner V, W. S. (2008). The challenging nature of patient identifiers: An ethnographic study of patient identification at a London walk-in centre. Health Informatics Journal, 14(2):141-150. doi: 10.1177/1081180X08089321.

[20] Morris, G. et al., Patient Identification and Matching Final Report (February 7, 2014), available at https://www.healthit.gov/sites/default/files/patient_identification_matching_final_report.pdf

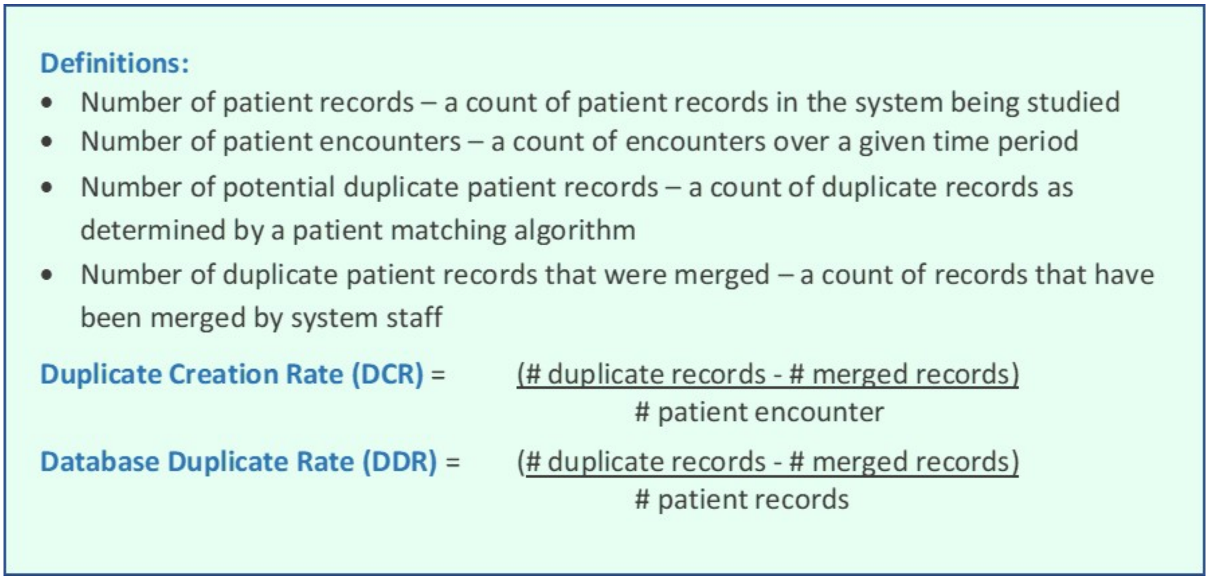
[21] The Data Quality Teams established at each pilot site included IT staff, clinic managers, and billing staff identified by community health center management. The teams were tasked with advocating for and implementing data management improvement measures within their respective centers.

intended to integrate with the current workflows at each pilot site and take advantage of existing processes and procedures.

The PDDQ Ambulatory Guide is organized into five categories: Data Governance, Data Quality, Data Operations, Platform and Standards, and Supporting Processes. These five categories contain 17 process areas and 22 evaluation questions. The Ambulatory Guide includes key items from the PDDQ Framework and a core set of questions aimed at small to medium ambulatory practices. The guide's scoring system allows practices to measure their baseline and improvements over time. This pilot testing approach calculated baseline and follow-up Ambulatory Guide scores. For entities interested in measuring demographic data quality at sites with broader clinical practices, additional sections of the PDDQ Framework beyond the Ambulatory Guide should be considered.

In addition to the PDDQ Ambulatory Guide score, two rates were calculated for each site (Duplication Creation Rate [DCR] and Database Duplication Rate [DDR]) to capture the presence of duplicate records in existing data and the rate at which duplicate records are created. The DCR is a ratio of duplicate records created to patient encounters. This rate measures the process of creating new records through the patient registration process, where staff rely on demographic data to identify individuals. The DDR is a measure of duplicate records that exist in the site's health IT systems under study. Figure 1 presents the formulas for calculating these rates.

Figure 1: Duplicate Creation Rate and Database Duplicate Rate Calculations



Definitions:

- Number of patient records – a count of patient records in the system being studied
- Number of patient encounters – a count of encounters over a given time period
- Number of potential duplicate patient records – a count of duplicate records as determined by a patient matching algorithm
- Number of duplicate patient records that were merged – a count of records that have been merged by system staff

Duplicate Creation Rate (DCR) = $\dfrac{\text{(\# duplicate records - \# merged records)}}{\text{\# patient encounter}}$

Database Duplicate Rate (DDR) = $\dfrac{\text{(\# duplicate records - \# merged records)}}{\text{\# patient records}}$

It should be noted that DCR and DDR both rely on patient matching algorithms to determine the number of duplicate records in a given system. Changing the matching algorithm or adjusting matching parameters would likely impact these rates.

DCRs, DDRs, and PDDQ Ambulatory Guide scores were collected and calculated before, during, and after the pilot project.

SETTING

Pilot sites were recruited through OCHIN. OCHIN provides and supports centrally hosted EHR and practice management services to 100 community health centers that together manage 500 clinics nationwide. Community health centers are community-based, non-profit health organizations that provide health care to under-insured, underserved, and vulnerable patients. Community health centers usually operate several clinics at different locations. In addition to primary care clinics, some community health centers provide integrated access to pharmacy, dental, mental health, and substance use disorder services.

Clinics from three community health centers initially agreed to participate in the pilot and engaged in the pre-pilot site assessment. Pilot sites did not receive compensation for participation. In some cases, pilot sites invested modest resources to participate in the project. After the initial assessment, one community health center decided not to continue with the project because of competing priorities and resource limitations.

The first participating community health center operated three primary care clinics and two mental health clinics. From July 2016 to July 2017, this community health center created just under 6,000 unique patient records, completed more than 680 patient record merges, and had 72,500-plus patient encounters. Two of its three primary care clinics participated in the project and are referred to in this report as pilot site A.

The second participating community health center operated nine primary care clinics, one mobile clinic, and a scheduling call center. From July 2016 to July 2017, this community health center created more than 12,000 unique patient records, completed more than 5,800 patient record merges, and had 162,000-plus patient encounters. Two of its nine primary care clinics and the call center participated in the pilot and are referred to in this report as pilot site B.

INITIAL ASSESSMENT

A CHR team visited six clinic locations to assess patient registration workflows and the handling of patient demographic data, including monitoring for and correction of duplicate records. These visits included two sites each from pilot sites A and B as well as two from the community health center that declined to continue in the pilot. This CHR team also visited the scheduling call center from pilot site B. Each visit included an introduction and orientation of the clinic by the practice manager, observation of patient registration and scheduling staff at work, and interviews with the registration and scheduling staff. The CHR team used an ethnographic data collection technique called the Rapid Assessment Process[22] in the in-person interviews and direct observation of registration and scheduling staff.

QUALITATIVE QUESTIONNAIRE

To better understand existing clinic processes and develop qualitative comparisons of the processes before and after pilot testing, the CHR team developed a questionnaire for registration and scheduling staff at each clinic's baseline site visit. Most questions had multiple-choice or yes/no response options and allowed for additional open-ended comments. Questions covered staffing, training, and the frequency of patient demographic data updates; safeguards and processes in place to prevent duplicate record creation; and procedures for merging duplicates when found. The CHR team used a list of 43 demographic data elements to determine whether listed data elements were required, optional, or not

---

[22] Ash, J. S. (2008). A rapid assessment process for clinical informatics interventions. *Annual Symposium proceedings* (pp. 26-30). AMIA.

collected at the site. Two qualitative questions were also included: “What do you feel are the greatest challenges to an accurate patient record?” and “Do you have anything you would like to add or any ideas you want to share about changes to processes or tools that could improve the patient registration process?” At baseline, the CHR team administered the questionnaire in person to two to five staff members at each pilot site, including the scheduling call center. A follow-up phone questionnaire was administered at the end of the pilot testing period.

APPLICATION OF THE PDDQ FRAMEWORK

ONC, CHR, and OCHIN met with the pilot site Data Quality Teams in December 2016 to describe and discuss the PDDQ Framework and the use of the PDDQ Framework’s Ambulatory Guide in the pilot testing approach. Data Quality Team members were asked to describe any documentation, policies, procedures, and reports that could be leveraged to apply and test the framework. After discussion at the Data Quality Team meetings, each process area was scored in each category at each pilot site. Scoring was determined by the method contained in the PDDQ Framework (please refer to Appendix F – PDDQ Ambulatory Guide Scoring Sheet).

Data Quality Improvement Training

As part of the application of the PDDQ Framework’s Ambulatory Guide at the pilot sites, training was provided in understanding the framework as well as the DCR and DDR reports that would be generated and improving data quality. The Data Quality Improvement Training effort therefore consisted of training material development and assistance in developing data quality plans, initial training, baseline scoring, ongoing training, and follow-up scoring. During training, sites were provided with individualized patient duplicate and data completion rate reports. Pilot site staff received instructions on how to access and review duplicate patient record reports that were made available within the sites’ EHRs. This allowed each clinic to monitor its duplicate patient reports and resolve any patient records identified as true duplicates. To ensure that pilot sites could review progress during the pilot, OCHIN provided sites with monthly, individual data assessment reports that broke out the number of unique patient records, the number of patient encounters, the number of potential duplicate patient records (as determined by a patient matching algorithm), and the number of duplicate patient records that were merged.

The pilot sites also received training in methods for improving their patient demographic data collection and management capabilities. Accordingly, the Data Quality Teams received the following materials:

- **Business Glossary Template.** The pilot sites received a business glossary template as an example as they created their own glossaries. Business glossary templates are intended to align data and language usage across an enterprise to ensure consistency and common understanding. Ensuring common terms and usage of terms is important in a project to improve data matching across different clinics within one community health center. The Data Quality Teams were trained on how to research and complete the elements of the glossary. The sites’ business glossaries were reviewed regularly. Appendix C presents an excerpt from the business glossary template.
- **Training Inventory Template.** A sample training inventory template provided a single location to document the trainings available at each pilot site.
- **Data Quality Plan.** This plan assisted the development of pilot site-specific data quality plans. Appendix D shows a sample data quality plan.
- **Additional Resources.** The Data Quality Teams also received a “toolkit” (the Data Quality Team Toolkit) that contained additional resources to further advance data quality at sites. Although

these documents were provided during the pilot, they were not always implemented for this project because of time and resource constraints. Appendix E presents an excerpt from the Data Quality Team Toolkit.

Additional supplementary support was provided to Data Quality Team staff via monthly teleconferences for six months. Discussion included such topics as PDDQ Framework concepts, development of demographic data quality improvement processes and procedures, and registration training for deployment at the pilot sites.

Scoring

Scores were captured during the baseline assessment on a monthly basis during and at the conclusion of the pilot. The final PDDQ Framework Ambulatory Guide score for each pilot site was calculated in May 2017. In addition to the training and discussion of baseline and monthly scores, CHR and ONC provided to each pilot site the results of the final score, a discussion of changes since baseline, and suggestions on how to continue efforts to improve patient demographic data.

Findings

QUALITATIVE QUESTIONNAIRE

A total of 20 participants were interviewed at baseline among the three pilot sites. Ten participants from pilot sites A and B participated in the post-pilot interview. From those 10 participants, one person had been hired after the pre-pilot interview. The total number of responses varied by question. Most of the baseline questionnaires were conducted in person while questions from follow-up questionnaires were conducted by phone.

Questionnaire Responses

The following summaries describe the baseline questionnaire responses on Training, Patient Identification, and Demographic Updates. Responses to some questions are not reported because of low response rates.

Training

- In response to a question asking who provided the training and how the staff were trained in collecting patient data, all respondents highlighted the importance of shadowing other staff as part of their training. The responses about training method were varied. Some indicated they received one-on-one training or organization-specific training or materials.
- Regarding the frequency of the training received on patient registration tools, technology, and workflow, 45 percent of respondents reported receiving some training when they started their job. The frequency of training varied, with several respondents reporting ad hoc or irregular training and a few reporting they had never been trained.

Patient Identification

- Forty-two percent of respondents reported no healthcare provider at their clinic ever had difficulty identifying a patient’s health record because of a duplicate record.
- Sixty-five percent of respondents reported that their clinic scanned and saved photo ID cards in patient records.

- More than 90 percent of respondents reported that the following demographic data fields were required at their clinics: first and last names, gender assigned at birth, permanent home address, home/primary phone number, insurance plan name, Social Security Number (SSN) and Medicaid/Medicare insurance ID numbers.
- Over 50 percent of respondents reported that the following demographic data fields were also required: preferred language, race, ethnicity, mobile phone number, EHR ID, and gender identity.
- Fields reported as optional by at least half of respondents included prefixes (e.g., Mr., Ms., and others), suffixes (e.g., Jr., III, and others), hyphens in last names, nicknames/aliases, preferred names, full middle names and middle initials, previous names, confidential addresses, and e-mail addresses.
- Seventy-eight percent of respondents reported being involved in review of patient records. Seventy percent of the time, OCHIN Site Specialists (OCHIN technical support representatives) were most commonly identified as a resource for merging duplicates. Other resources identified included the clinic billing staff, EHR trainers, and IT staff.
- Only 15 percent of respondents knew approximately the number of duplicate records in each clinic's EHR.
- Only 27 percent of respondents reported that their clinics had policies and processes in place for merging duplicate records. Another 27 percent reported that their clinic did not have such measures in place, and 39 percent did not know. Seven percent did not respond or provided an alternative response.
- Twenty-five percent of respondents reported personal knowledge of accidental merging of the wrong patient records, resulting in an overlay.

Demographic Updates

- All respondents indicated that their clinics took walk-in patients and 31 percent indicated that differences in registration procedures for walk-in and scheduled patients.
- Staff ranked the most reliable sources of demographic information as follows (because more than one response was permitted, percentages do not sum to 100 percent): Patients (85 percent of respondents), ID cards (65 percent), insurance cards (60 percent) and existing information in the EHR (50 percent).
- The most frequently reported ways to register new patients or update demographic information on existing patients were the EHR (80 percent) and the electronic patient portal (65 percent).

Key Responses from Follow-up Questions

- Consistent with baseline qualitative questionnaire responses, 91 percent of respondents reported that demographics were updated at every unique patient visit.
- While only 27 percent reported at baseline that their clinic had policies and processes for merging duplicate patient records, at follow-up 100 percent replied affirmatively to this question.
- Somewhat contradictory to the first two follow-up responses, while 77 percent of interviewees reported at baseline that their clinics had "policies and processes in place for quality assurance of electronic medical records, including demographics," only 45 percent responded affirmatively to this question at follow-up.

Beyond the initial qualitative questionnaire, Data Quality Team meetings provided additional opportunities to assess and evaluate registration and demographic data collection process similarities and differences between the two pilot sites. The differences identified included methods for gathering patient information and updates (verbally or via patient written responses on a form), whether pictures of patients were kept on file, and whether patients were asked to self-identify race and ethnicity. The similarities identified consisted of a lack of standardized policies to update patient data, lack of training on managing patients reluctant to provide demographic data, verification of changes in patient name, and allowing clinical staff to make certain determinations related to demographics (sexual orientation and gender identity) in conjunction with the patient in lieu of front desk staff.

BASELINE SITE ASSESSMENT

The Site Assessment consisted of understanding each site's registration process and identifying potential factors affecting the creation of duplicate records.

Registration Process

Both pilot sites used the same EHR system, but followed clinic-specific policies, practices, and processes to identify patient records and to confirm and update patient demographics. Both pilot sites used the following required fields to identify patients in the system: first name, last name, and date of birth (DOB). Registration staff then relied on address, phone number, and emergency contact information to confirm a patient's identity if two or more patients shared the same required elements.

Some pilot site clinics asked patients to recite their current demographic information aloud, while other clinics read what was displayed on the screen and asked patients to confirm the accuracy of the information. Only one clinic asked if there were any changes to a patient's address or phone number(s) since the patient's last visit. There were different policies and procedures at each clinic for handling the registration process when the EHR system was not available. When the system was unavailable at one clinic, patient rosters were printed out daily and copies were made for registration staff to hand write information for entry into the EHR later. One clinic indicated that there was a procedure logged in a binder for dealing with a system shutdown, but registration staff were not generally aware of it. Another clinic noted that, in the event of a power outage, they would close the clinic because they had no procedures in place for that situation.

Different types of staff managed the adjudication and merging of duplicate patient records at each clinic; however, the procedure for adjudication and merging were not well documented at any pilot site. Typically, registration and billing staff would identify and flag duplicate records to be investigated and possibly merged by OCHIN Site Specialists or IT staff. Site specialists from OCHIN typically served as clinical application specialists, informaticists, and coordinators.

Interviews with clinic staff revealed different perspectives on the problem of patient demographics accuracy and duplicate record creation. At one clinic, registration staff reported that they did not perceive that duplicate record creation was problematic and felt that creating a new patient identity introduced less risk than selecting an incorrect patient from existing records. For example, registration staff reported that there were several patients at their clinic with the same first and last name. In such cases, staff would often opt to err on the side of duplication and create a new record rather than trust other additional information, such as DOB or SSN, to confirm a match.

Potential Factors Affecting the Creation of Duplicate Records

Findings from the baseline site assessments enabled ONC, CHR, and OCHIN to generate a series of hypotheses on issues that could contribute to the creation of duplicate records at the sites. Some issues were related to processes and tools at the clinic level while others related to the differences in the populations served by the clinics and each site's ability to adapt its technology to meet the needs of its patient population. Given the low observed DCR at sites, discussed in the Duplicate Reports section, it is difficult to assert what impact these issues ultimately have on creation of duplicate records. The CHR team identified the following specific issues influencing the collection and creation of demographic data present at the sites:

- **Compliance with Standardized Procedures and Processes.** One potential issue was whether the site staff used standardized processes and procedures to identify patients and confirm patient demographic data. Some registration staff followed standardized processes in how they asked patients for name, DOB, phone number, and address. Other registration staff collected the information in ways that diverged from the prescribed process.
- **Staff Longevity.** Staff longevity affected how pilot sites implemented and maintained a standardized approach to collecting and updating demographic information. For example, at one clinic, registration staff that had long tenure were observed to disregard new procedures. At another clinic, registration staff of long standing tended to know patients on sight and did not feel the need to follow the prescribed procedures. While this may not cause new records to be created, it may prevent updating demographic information which leads to future record-matching issues. High registration and scheduling staff turnover also may result in a higher number of untrained personnel. For example, one clinic employed only temporary staff (with a contract tenure of only six months each) at its front desk.
- **Time Limitations.** Implementation of standardized processes may suffer when there are issues with time limitations (perceived or actual) at clinics. Front office and other patient registration environments are often hectic, time constrained, and demanding,[23] leading to deviations from processes as registration staff attempt to complete required tasks as quickly and efficiently as possible.
- **Lack of Complete, Standardized Training.** Clinic registration and scheduling staff generally reported very little to no formal training for collecting, entering, and updating patient demographic data, although some web-based training opportunities were available to clinic staff. Most registration staff reported never being formally trained and simply shadowed other registration staff to learn on the job. Clinic staff reported receiving additional training only when there were updates to the EHR. The EHR developer led these subsequent trainings, which only addressed updates to the technology data collection process.
- **Lack of Awareness of Advanced Software Options and Lack of Trust in Software.** Registration and scheduling staff admitted to very little knowledge about expanded options for finding a patient record such as the 'Sounds Like' feature. Use of this feature may allow discovery of a patient record and potentially reduce duplicate creation. Some of the staff aware of this type of option did not trust it to work well or found the features unhelpful.

Potential Patient Factors

Several clinic staff reported that a portion of their patient population often does not have or cannot communicate all required demographic data. This may increase the likelihood of creating duplicate patient records. Some of the reasons for the lack of accurate data included:

---

[23] Rivera-Rodriguez, A. J.-T. (2010). Interruptions and distractions in healthcare: review and reappraisal. Quality & Safety in Health Care, 304-312.

- **Unknown or imprecise DOB.** Patients do not always know their DOB or know only their approximate year of birth. Because the system requires entering a valid date, some registration staff enter 1/1/ and year of birth, resulting in an overrepresentation of non-unique DOBs.
- **Inconsistency of last names.** Last names may be listed differently in different places, patients may provide different names, or names may have been misspelled or changed. For example, a patient with two last names may have the name recorded both with and without a hyphen, the names may be listed in different order, or only one of two last names listed may have been captured in the EHR.
- **Lack of SSN.** Many patients do not have or do not wish to provide their SSN. A common practice in this case is to repeat one number, such as 999-99-999.
- **Illiteracy / Language Barrier.** Illiterate or non-English-speaking patients sometimes do not know or may have difficulty communicating their information. Also, registration and scheduling staff may encounter difficulty in communicating to patients the purpose and importance of regularly confirming and updating their demographic data.

DUPLICATE REPORTS

For the pilot, monthly tracking of duplicate creation rates ran from January through June 2017 and tracking of database duplicate rates ran from February to June 2017.

Figure 2 shows the monthly DCRs captured for pilot site A (PSA) and pilot site B (PSB). Overall, DCR decreased from 0.26 percent to 0.24 percent (a relative decrease of 7.7 percent) at pilot site A and a decrease from 0.16 percent to 0.11 percent (a relative decrease of 31.3 percent) at pilot site B. In practical terms, a clinic seeing 4000 patients per month would create about 11 duplicates per month with a DCR of 0.26 percent and about 10 per month with a DCR of 0.24 percent. These DCR reports contain data on all duplicates created across a community health center. This means that a reduction in DCR may be conservative because the rate includes data from sites that did not participate in the pilot.

Figure 2: Monthly Duplicate Creation Rate (DCR) by Pilot Site

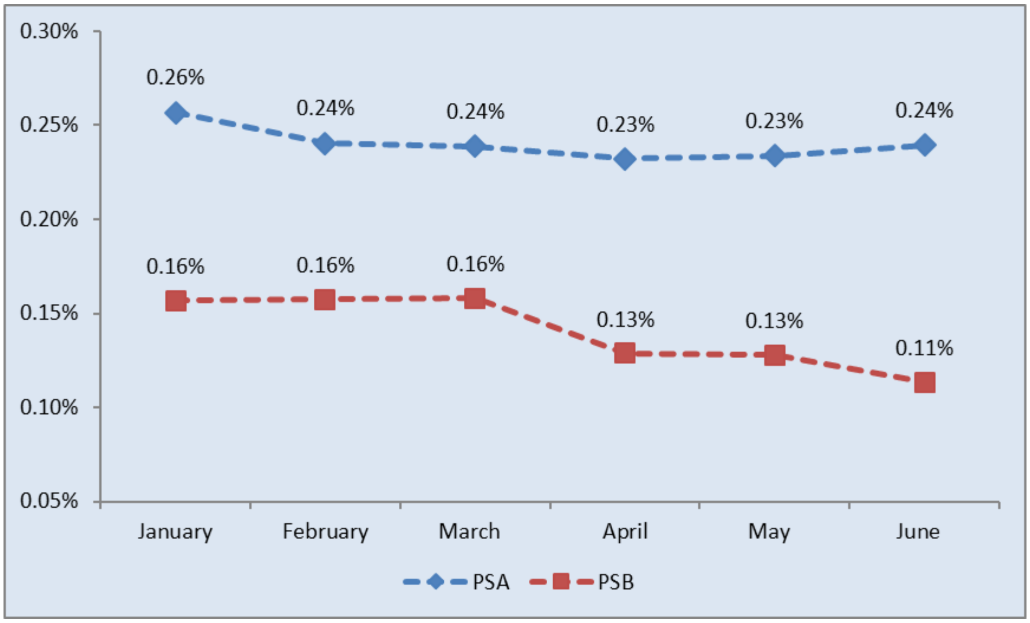
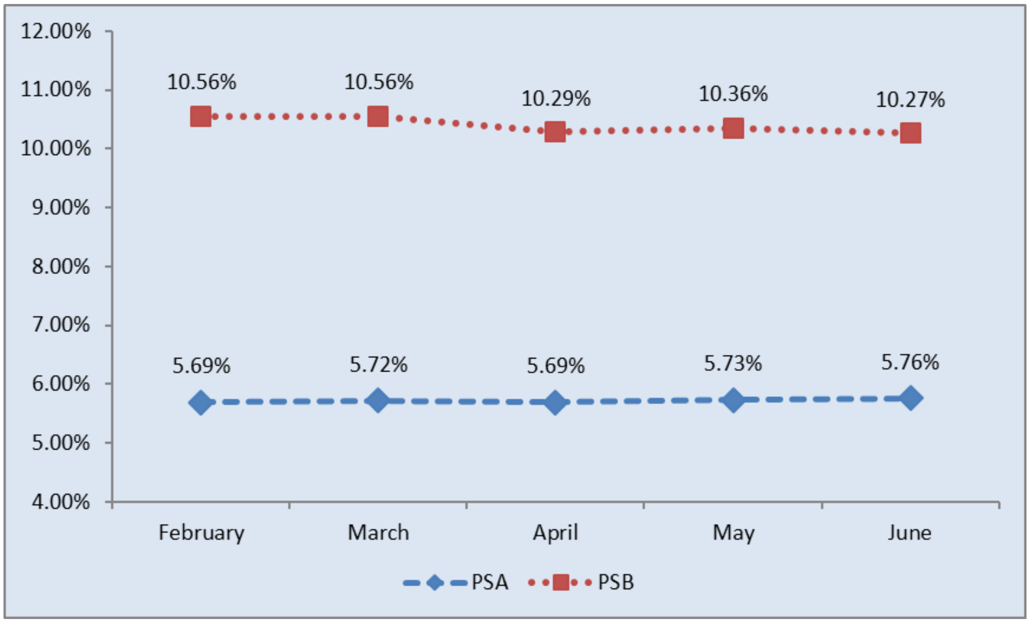


Figure 3 shows the monthly DDR for pilot site A and pilot site B. The DDR showed a slight decrease from 10.56 percent to 10.27 percent at pilot site B and a slight increase from 5.69 percent to 5.76 percent at pilot site A. The DDR downward trend at pilot site B suggests there may be a decrease in DDR rates if efforts initiated during the pilot study were to continue and expand.

Figure 3: Monthly Database Duplicate Rate (DDR) by Pilot Site



PDDQ Framework's Ambulatory Guide Scores

Using the PDDQ Framework's Ambulatory Guide Scoring Sheet provided in Appendix F, baseline scores were calculated in December 2016 and follow-up scores were calculated May 2017. The baseline PDDQ Framework Ambulatory Guide scores for pilot sites A and B were 9 out of a possible 22 and 7 out of a possible 22, respectively. Differences in the initial scoring between the two sites were mainly due to more developed data quality practices and procedures at pilot site A compared to pilot site B.

Both sites experienced increases in scores from baseline to follow-up. Between the baseline scoring and the follow-up scoring, pilot site A's score rose from 9 out of 22 to 22 out of 22. The most significant improvements from baseline to follow-up were attributed to Data Governance and Data Operations efforts. Pilot site A has been working to improve procedures and documentation through community health center-organized efforts in addition to the Data Quality Team activities. These efforts included the creation of a specific process for producing and reviewing data quality reports and changes to workflow based on information learned from the reports. In the area of Data Operations, pilot site A received full credit for all categories and process areas in this section during the follow-up scoring and improved its baseline score from 1.5 to 4. This increase was directly due to the development of the business glossary, which provides a mechanism to align and synchronize data elements across an enterprise.

Pilot site B was also able to improve its score by 3.5 points, from 7 out of 22 to 10.5 out of 22. Improvements were attributed to the creation of a business glossary and data requirements, then linking the business glossary and data requirements together as well as to other metadata.

Three additional observations about the PPDQ Ambulatory Guide scoring merit mention:

- Both pilot sites were able to complete a draft of the Training Inventory; however, neither was able to create a final version during the project period.
- Data Quality Teams did not deploy the full set of training resources provided to them, such as front desk staff training.
- Neither participating pilot site had a clinic staff member whose sole responsibility was to review the duplicate records report, conduct the necessary research and analysis, and merge true duplicate patient records. Both sites had an ad hoc process in which duplicate records were addressed when discovered by front desk or billing staff.

Discussion

IMPACT OF THE PDDQ FRAMEWORK APPLICATION

This project tested the ability of a version of the PDDQ Framework to improve the quality of demographic data in health IT systems. Over the course of the pilot test, the PDDQ Ambulatory Guide improved scores at both sites. Based on follow-up questionnaires, the increase in Ambulatory Guide scores corresponds to greater staff awareness about demographic data management procedures.

This project used the DCR and DDR metrics to evaluate the performance of the PDDQ Framework. Initial indications from DCR are positive. Both pilot sites showed a decrease in DCR throughout the pilot. This decrease should be interpreted with caution. Given the low rate at which duplicates are currently created—approximately 0.0024 and 0.0011 per encounter—changes in the rate reflect very small differences. Longer studies will be required to confirm lasting effects. Finally, it should be recognized that the changes to DCR are likely conservative. DCR was reported across the community health center in general, not just the pilot site.

The DDR metric provided mixed results. Pilot site A demonstrated a decrease in DDR while its counterpart showed an increase. In both cases, the changes in DDR were small and do not support or refute the ability of the PDDQ Framework to improve demographic data quality.

LIMITATIONS

The following limitations may influence the interpretation of the project's findings and may inform the design and implementation of future projects:

- **Constraints on available staff and resources to conduct the pilot.** The ability to devote staff and resources to the PDDQ Framework pilot project presented a key limitation. A community health center is a demanding environment because it must provide care for a complex patient population on relatively limited resources. In this project, one site dropped out after the baseline assessment citing insufficient resources. Had more community health centers participated in the pilot project, there would have been more variability in the data collected and greater diversity of input and feedback on the PDDQ Framework. The project's Data Quality Teams might have developed more resources to handle the input, and in turn, may have laid the foundation for wider generalizability of the findings.

Resource constraints also limited the composition of the Data Quality Teams to middle management staff. No front desk staff or clinic leadership participated. High staff turnover caused differences in interviewers' ability to compare between baseline and post-application, with fewer registration and scheduling staff interviewed post pilot. Because of such resource constraints, the Ambulatory Guide was only partially implemented.

For these reasons, the pilot's findings should be interpreted carefully. The results do suggest, however, the importance of ascertaining the experience and knowledge of all staff who are involved with demographic data collection and management.

- **Training.** Front desk training materials were shared with the Data Quality Teams but were not delivered to the front desk staff. Also, the Training Inventory was not completed due to a lack of time and resources at the clinics.
- **Data specificity.** The data reports developed for this pilot included data from all clinics at each community health center, whether or not a center's clinic participated in the pilot. This lack of data specificity may skew the results of any analysis.

In general, this pilot test provides some insight into how the PDDQ Framework's Ambulatory Guide made improvements in the quality of demographic data. Given the lack of clinic resources and inability to refine data reports, further study of the PDDQ Framework is warranted, with implementation in its entirety and also at a larger subset of clinics that have or are provided the necessary resources to implement its full processes or conduct robust pilot testing of the PDDQ Ambulatory Guide.

Summary and Implementation Considerations

CONSIDERATIONS FOR PDDQ FRAMEWORK IMPLEMENTATION

The intended modular implementation of the PDDQ Framework helps organizations build their capability to improve demographic data quality by establishing best practices according to their respective priorities, resources, and schedules. Based on findings from the pilot, health care organizations considering implementation of the PDDQ Framework would benefit from the following four PPDQ Framework considerations:

PPDQ Framework Implementation Consideration 1

Review Reporting and Tracking Processes for Demographic Data and Duplicate Creation

The PDDQ Framework can guide in the identification and prioritization of data quality processes for development or improvement. For example, an organization is advised to review its reporting and tracking processes for creating demographic data and avoiding duplicate patient records. Routine data quality reports should be readily available and provide a clear understanding of the quality (e.g., completeness, validity, and distinctness) of demographic data elements that are key to accurate patient matching. These data quality reports can help prioritize data elements for targeting quality improvement efforts. Data elements that have high accuracy when populated but are infrequently recorded will be of limited value in decreasing the frequency of duplicate records. Similarly, well-populated data elements (i.e., complete), but do not contain patient-specific information (e.g., SSNs recorded as 999-99-999 or sex recorded as "unknown") are of limited value.

A DCR report may be helpful in monitoring the frequency of duplicate records created and tracking the impact of data quality improvement efforts on these rates. Ongoing review of the DCR will demonstrate natural variability over time and whether quality improvement efforts are having an impact on the rate. If not already in place, the DCR report may require a short-term dedication of programming resources to compile monthly counts of patient encounters, potential duplicate patient records created (identified by one or more record-matching algorithms), and duplicate record merges.

PPDQ Framework Implementation Consideration 2

Identity Responsible Staff and Available Resources for Demographic Data Management

An organization should identify the staff responsible to carry out the demographic data management effort and the resources available for this effort. Organizational leadership and the staff should review current staffing roles, training policies and processes, and available and accessible training resources. Future applications of the PDDQ Framework should also allow more time for implementation than what was available in this pilot and should consider the necessary resources to support organizations and staff in transitioning and expanding their data quality roles.

PPDQ Framework Implementation Consideration 3

Provide Additional (or Revised) Structured Fields and Customized Formats for Accurate Patient Matching

Findings from discussions with the Data Quality Teams suggest providing additional (or revised) structured fields and formats for data elements especially relevant for accurate patient matching. For example, it may be advisable to allow for additional fields to record variations in naming conventions among a diverse patient population. Such EHR customizations were not part of this pilot.

PPDQ Framework Implementation Consideration 4

Ability to Monitor and Evaluate Outcomes Over Time

The data quality infrastructure developed through this process should allow for monthly monitoring, tracking, and evaluation of data quality and duplicate record creation to assess outcomes over time. It will be important to document temporal changes occurring internally or externally to the health system (e.g., to the EHR, workflow, staffing structure, or in key patient characteristics), to account for any impact they may have on the interpretation of reports.

CONSIDERATIONS FOR ONGOING DEVELOPMENT OF THE PDDQ FRAMEWORK

Future development of the PDDQ Framework would benefit from increased support described in the following three Ongoing Development Considerations.

Ongoing Development Consideration 1

Establish a National Priority in Demographic Data Quality Improvement

Encouraging continued discussion about the difficulty of data quality improvement can increase the visibility of the patient matching problem at a national level. Raising awareness regarding the critical role that high-quality data demographic data plays in patient matching could clarify the need for sufficient resources to carry out quality improvement work as part of safety or research initiatives.

Ongoing Development Consideration 2

Seek Additional Insights into the Validity of the PPDQ Framework in Various Settings

Widespread adoption of the PDDQ Framework could illuminate the validity of its application in different settings. Obtaining additional insights from these implementations as well as stakeholder experiences and recommendations may inform refinements of the framework and associated resources.

Ongoing Development Consideration 3

Continue Development of Health Care Industry-Specific Resources for Demographic Data Quality Improvement

The PDDQ Framework will benefit from continuing development of health care industry-specific resources. These resources should recognize the specific challenges and burdens associated with demographic data quality improvement and should address these challenges at the patient interface and multiple levels within an organization, including the leadership, management, clinical, front desk, scheduling, and IT staff. These PDDQ Framework materials should facilitate coordination and increased visibility of priorities and quality improvement activities between these various stakeholders. In addition, the development of tailored guidance could assist sites in identifying processes that maximize the impact of quality improvement efforts in their local settings. For example, sites relying on call centers for scheduling patients may want to focus initial efforts on call center staff, while other sites may find it more optimal to focus on front desk staff. In the end, both must work in tandem to create, update, and maintain accurate and up-to-date patient demographic information.

CONCLUSION

High-quality patient demographic data is fundamental to accurate patient identification and matching, which in turn is pivotal to interoperability, patient safety, and research, such as PCOR. The continuing challenges of accurately identifying patients and matching patient records make it essential to stay abreast of recent findings, observations, and best practices that can inform strategies and effective solutions. With the adoption of rapid developments in health IT and advancements in EHRs, the scale of growth of the data captured, stored, and exchanged increases. The constant growth in patient populations and their diversity, along with the increasing complexity of health care networks, means that staff must have the necessary training, infrastructure, and ability to apply best practices to capture high-quality data on the front line.

The pilot study results described in this report suggest the potential for improvements in data quality with a modest investment. These implementation considerations are intended to guide organizations in the implementation of the PDDQ Framework, whether the full version or an abbreviated version such as the Ambulatory Guide, and to encourage their contributions to developing the set of guidance resources for improvement of the quality of demographic data and patient matching.

Appendix A – Literature Review Methodology

Search terms used for the literature review are included in the following list. We found that after executing these search terms many of the same articles were presented by the search engine, Google.

- Data algorithms
- Data attributes – patient data attributes and matching errors
- Data entry errors
- Data Parsing, Matching and De-duplication
- Deterministic matching
- Patient matching errors
- Patient matching metrics
- Probabilistic matching
- Duplicate patient record retention
- True negative
- True positive
- Identification authentication and patient matching
- Patient matching criteria
- Duplicate patient detection
- (Patient Matching) AND (Overlays OR False Positives)
- (Patient Matching) AND (False Negative OR False Negative Errors)
- (Patient Matching OR Patient Identification) AND (Probabilistic Algorithm OR Probabilistic Matching)
- (Patient Matching) AND (Matching Rate) AND (Data Attributes OR Data Standardization)
- (Patient) AND (Enrollment Process) AND (Data Governance)
- Data maturity model
- Data maturity scale
- Data quality
- Duplicate detection
- Electronic health records
- Front office staff
- Health information exchange
- Health information technology
- Interoperability
- Master patient index
- OCHIN
- Patient identification
- Patient identity
- Patient matching
- Patient registration systems
- Patient safety
- Privacy and security

The literary search employed two approaches. The first approach was conducting a search of “industry” artifacts such as white papers, reports, journal articles, conference proceedings, survey results, and marketing materials to analyze industry initiatives that are addressing patient matching since the publication of the ONC Patient Identification and Matching Final Report. The second approach included published materials from “academic research papers and journal articles” of research conducted to analyze and test a variety of patient matching processes and algorithms as well as other related issues.

Appendix B – IT Data Report to Support Measurement

Figure 4: Example IT Data Report

Site					
Date of Go-Live					
Transition from paper or migration from another system?					
No.	**Data Element or Metric**	**Baseline**	**Date** ______	**Date** ______	**Date** ______
I	Number of unique patient records (including possible unidentified or unreconciled duplicates)				
II	Number of (possible) duplicate patient records (not pairs) as identified by EHR algorithm				
III	Number of patient records merged in EHR				
IV	Total duplicates with merges performed				
V	Number of patient encounters in database				
VI	Duplicate Creation Rate (DCR) Database Duplicates with merges (DD) x 100 (IV) Number of Patient Encounters (V)				
VII	Number of months since go live (as of 07/05/2016)				
VIII	Average number of encounters per month (V/month)				
IX	Average number of merges performed per month (III/month)				
X	Average number of duplicates created per month (including merges performed)				
XI	Duplicate Creation Rate (DCR) **not including merges performed ** Database Duplicates (DD) x 100 (II) Number of Patient Encounters (VIII)				
XII	Average number of duplicates created per month ** not including merges performed **				
XIII	Database Duplicate Rate (DDR) DDR = Database Duplicates (DD) x 100 (IV) Database Size (DS) (I)				

Appendix C – Sample Business Glossary

Table 1: Excerpt of Sample Business Glossary

Data Element	Definition	Notes	Data Format	Activity	Required / Optional
PATIENT NAME	All names bestowed to patient when they are born, including all first given names, middle names (where applicable), and surnames or married names (where applicable).	When creating a patient in your EHR, please enter all last names (comma) all first names (space) all middle names (where applicable) (space) suffix (where applicable). In your EHR, anything that is entered after the comma is considered a first or middle name.	LAST, FIRST MIDDLE_SUFFIX	Registration	Required
		When creating or updating a patient in your EHR, please enter the patient's full middle name (if they have one), not just their middle initial.			
		Please do not enter hyphens or apostrophes in a patient's name, unless these symbols are reflected on their insurance card.			
		If a patient’s name is spelled differently than what is listed on their insurance card, add the correct spelling in the alias field and ask the patient to contact their insurance company to correct the spelling on their card and update their record once their card accurately reflects the spelling of their name.			
		When searching for a patient by their last name, search by all possible last names individually.			

Data Element	Definition	Notes	Data Format	Activity	Required / Optional
		Use the "3,3" rule when searching for a patient using the 3 letters of the last name (comma), and then the first 3 letters of the first name (zzz,zzz).			
		If a patient's name contains hyphens or apostrophes, please search for the patient's name with and without hyphens and apostrophes if you are unable to find the patient.			
		Ask the patient for alternate first and last names and search for those. Ask: "Have you been known by any other last names or first names?"			
		If you do not find a patient match, use the Sounds-like feature to search for possible matches.	LAST, FIRST MIDDLE_SUFFIX	Registration	Required
		Patient name is an important data element for patent matching and the prevention of duplicate records - especially when names change, are different than those on a patient's insurance card or ID, when a patient uses different names in different contexts, etc.			
SUFFIX	A part of a patient's name that follows the last name that provides additional information about that patient.	Suffixes typically indicate professional or academic status (e.g., MD or PhD) or may indicate birth order or a generational title (Senior, Junior, II, III, etc.).	LAST, FIRST MIDDLE_SUFFIX	Registration	Optional

Data Element	Definition	Notes	Data Format	Activity	Required / Optional
		Please do not separate the suffix with a comma when creating or updating a patient's name. Instead, please separate a suffix by a space.			
		Suffix is an important data element for patent matching and the prevention of duplicate records - especially with regard to generational titles.			
ALIAS	Any other first, middle, or last name that a patient may be known as now or has been known as in the past.	Examples include maiden name, married name, name after divorce, some cultures have a different baptismal name and a given name, some immigrants change their name or have their name changed upon entering the country, etc.	LAST, FIRST MIDDLE_SUFFIX	Registration	Optional
		If a patient notifies you that they have changed their name for whatever reason, please enter the name that they were previously known as in the alias field.			
	Any other first, middle, or last name that a patient may be known as now or has been known as in the past.	Alias is an important data element for patent matching and the prevention of duplicate records - especially when names change, are different than those on a patient's insurance card or ID, and when a patient uses different names in different contexts.	LAST, FIRST MIDDLE_SUFFIX	Registration	Optional
PREFERRED NAME	The name that a patient prefers to be called; typically, a first name.	Preferred names are often referred to as nicknames, e.g., Brad instead of Bradley, Ben instead of Benjamin.		Registration	Optional

Data Element	Definition	Notes	Data Format	Activity	Required / Optional
		Some preferred names can be aliases but not all aliases are preferred names.			
		Please ensure that a patient's preferred name is entered in this field and not in the "patient name" field.			
CONFIDENTIAL NAME	The name a patient may be known as in cases where the patient requests that their identity and associated health and financial records remain confidential.	Confidential names may be used for patients with sensitive diagnoses, domestic violence, or age-sensitivity (e.g., adolescents), to name a few examples.		Registration	Optional
DATE OF BIRTH	The known and confirmed, or estimated date of birth of a patient.	Your EHR automatically calculates the age of the patient.	MM/DD/YYYY	Registration	Required
		Only front office or call center staff should enter date of birth. Please ensure that administrative privileges to update date of birth are limited to these staff.			
	The known and confirmed, or estimated date of birth of a patient.	• *Date of birth is an important data element for patent matching and the prevention of duplicate records.*			
		If patient does not know their date of birth: Clarify that the date refers to the date of their birth, not the date of their baptism, or date of immigration, or any other date that is culturally relevant			

Data Element	Definition	Notes	Data Format	Activity	Required / Optional
		Reference insurance or other identification and copy that exact date as represented			
		Ask patient to estimate their year of birth or estimate their age and then calculate their year of birth (from January 1 of this year), then enter 01/01/YEAR.			
PATIENT NAME	All names bestowed to patient when they are born, including all first given names, middle names (where applicable), and surnames or married names (where applicable).	When creating a patient in your EHR, please enter all last names (comma) all first names (space) all middle names (where applicable) (space) suffix (where applicable). In your EHR, anything that is entered after the comma is considered a first or middle name.	LAST, FIRST MIDDLE_SUFFIX	Registration	Required
		When creating or updating a patient in your EHR, please enter the patient's full middle name (if they have one), not just their middle initial.			
		Please do not enter hyphens or apostrophes in a patient's name, unless these symbols are reflected on their insurance card.			
		If a patient’s name is spelled differently than what is listed on their insurance card, add the correct spelling in the alias field and ask the patient to contact their insurance company to correct the spelling on their card and update their record once their card accurately reflects the spelling of their name.			

Data Element	Definition	Notes	Data Format	Activity	Required / Optional
		When searching for a patient by their last name, search by all possible last names individually.			
		Use the "3,3" rule when searching for a patient using the 3 letters of the last name (comma), and then the first 3 letters of the first name (zzz,zzz).			
		If a patient's name contains hyphens or apostrophes, please search for the patient's name with and without hyphens and apostrophes if you are unable to find the patient.			
		Ask the patient for alternate first and last names and search for those. Ask: "Have you been known by any other last names or first names?"			
		If you do not find a patient match, use the Sounds-like feature to search for possible matches.	LAST, FIRST MIDDLE_SUFFIX	Registration	Required
		Patient name is an important data element for patent matching and the prevention of duplicate records - especially when names change, are different than those on a patient's insurance card or ID, when a patient uses different names in different contexts, etc.			
SUFFIX	A part of a patient's name that follows the last name that provides additional information about that patient.	Suffixes typically indicate professional or academic status (e.g., MD or PhD) or may indicate birth order or a generational title (Senior, Junior, II, III, etc.).	LAST, FIRST MIDDLE_SUFFIX	Registration	Optional

Data Element	Definition	Notes	Data Format	Activity	Required / Optional
		Please do not separate the suffix with a comma when creating or updating a patient's name. Instead, please separate a suffix by a space.			
		Suffix is an important data element for patent matching and the prevention of duplicate records - especially with regard to generational titles.			
ALIAS	Any other first, middle, or last name that a patient may be known as now or has been known as in the past.	Examples include maiden name, married name, name after divorce, some cultures have a different baptismal name and a given name, some immigrants change their name or have their name changed upon entering the country, etc.	LAST, FIRST MIDDLE_SUFFIX	Registration	Optional
		If a patient notifies you that they have changed their name for whatever reason, please enter the name that they were previously known as in the alias field.			
	Any other first, middle, or last name that a patient may be known as now or has been known as in the past.	Alias is an important data element for patent matching and the prevention of duplicate records - especially when names change, are different than those on a patient's insurance card or ID, and when a patient uses different names in different contexts.	LAST, FIRST MIDDLE_SUFFIX	Registration	Optional
PREFERRED NAME	The name that a patient prefers to be called; typically, a first name.	Preferred names are often referred to as nicknames, e.g., Brad instead of Bradley, Ben instead of Benjamin.		Registration	Optional

Data Element	Definition	Notes	Data Format	Activity	Required / Optional
		Some preferred names can be aliases but not all aliases are preferred names.			
		Please ensure that a patient's preferred name is entered in this field and not in the "patient name" field.			
CONFIDENTIAL NAME	The name a patient may be known as in cases where the patient requests that their identity and associated health and financial records remain confidential.	Confidential names may be used for patients with sensitive diagnoses, domestic violence, or age-sensitivity (e.g., adolescents), to name a few examples.		Registration	Optional
DATE OF BIRTH	The known and confirmed, or estimated date of birth of a patient	Your EHR automatically calculates the age of the patient.	MM/DD/YYYY	Registration	Required
		Only front office or call center staff should enter date of birth. Please ensure that administrative privileges to update date of birth are limited to these staff.			
DATE OF BIRTH (cont.)	The known and confirmed, or estimated date of birth of a patient.	• *Date of birth is an important data element for patent matching and the prevention of duplicate records.* *If patient does not know their date of birth:* Clarify that the date refers to the date of their birth, not the date of their baptism, or date of immigration, or any other date that is culturally relevant			

Data Element	Definition	Notes	Data Format	Activity	Required / Optional
		Reference insurance or other identification and copy that exact date as represented			
		Ask patient to estimate their year of birth or estimate their age and then calculate their year of birth (from January 1 of this year), then enter 01/01/YEAR.			

Appendix D – Sample Data Quality Plans

Table 2: Data Quality Plan (Example 1): Prioritized list of areas that you want to improve

Area	Issue	Actions for Improvement
Patient data capture	Variations in spelling of patient names	• If staff is unsure about name spelling, ask patient to slowly spell their name aloud • Staff should confirm variations in name spellings (e.g., Gomes vs Gomez) with patient • If patient's name is spelled differently on insurance card, enter name exactly as it is on card, then enter correct name spelling under Alias field until corrected • Remind patient of their need to contact their insurance company to correct the spelling of their name on their card • Collect all variations of name spellings if vary between identification cards or other documents, with the spelling of their name on their insurance card to serve as the primary source • Educate patients on the importance of maintaining correct and up-to-date information

Table 3: Data Quality Plan (Example 2): Targeted Data Quality Plan

Objective	Indicator	Baseline	Target 1 (within 6 months)	Target 2 (within 1 year)	Measurement	Actions for Improvement
Capture correct patient address	Patient record duplicates and billing issues related to incorrect patient address	Correct patient address captured and maintained 65% of the time	Capturing correct patient address more than 80% of the time	Capturing correct patient address more than 95% of the time	Manual review of charts, random quality checks on patient demographics, report review.	• Staff should rely on insurance card and identification cards for correct spelling of address • At each and every visit, staff should confirm that they have the most up-to-date and correct address with patient • If patient's address is different on insurance card or other ID, at current and correct address is, then enter that address as current address • Ensure that patient is web-enabled for patient portal or other online access to EHR and that they have the ability to update their contact information online • Educate patients on the importance of maintaining correct and up-to-date information
Accuracy (e.g., correctly spelling data elements)						
Timeliness (e.g., how often and quickly is a data element updated when it changes, like an address change)						
Uniqueness (e.g., duplicate patient record rates)						

Appendix E – Excerpt from the Data Quality Team Toolkit

Table of Contents

The purpose of this toolkit is to provide examples of tools and documents that Data Quality Teams may adapt to their own environment.

1. **Building a Data Quality Team**
    a. Create a Data Quality Team (Data Quality Team)
        1. Example Guide to Identify Team Members and Set Roles/Responsibilities for the Data Quality Team
    b. Hold regular Data Quality Team meetings
        1. Example Monthly Data Quality Team Meeting – Sample Agenda
        ii. Review reports, compare month to month, track improvements over time
        iii. Increase awareness among all staff

2. **Supporting Processes**
    a. Front Desk/Call Center Training
    b. Measurement and Analysis

Example Guide to Identify Team Members and Set Roles/Responsibilities for Data Quality Team

The Data Quality Team is responsible for …
Managing quality assessments

- ✓ Reviewing duplicate patient reports from OCHIN
- ✓ Identifying workflow and technical problems that lead to duplicates and troubleshooting and addressing issues
- ✓ Sharing responsibility for data quality outcomes

Raising awareness across your practice about data quality

- ✓ Coordinating with call center, billing, IT and other staff on data quality efforts
- ✓ Contacting Data Quality Team Lead & Site Specialist when duplicate records are identified and when other issues arise

Facilitating implementation of best practices

- ✓ Training, guiding, and reminding all staff of best practices
- ✓ Distributing training and other knowledge resources

Position	Role	Data Quality Team Roles/Responsibilities	Data Quality Team Strengths	Names
Front Office Staff / Registrars	• Welcome patients as they contact the clinic personally or by telephone • Direct walk-in patients • Register all patients per registration protocols and collect all documentation • Reviews and verifies patient coverage of insurance • Updates patient demographic and insurance information when requested / needed	• Reviews and verifies patient contact information • Updates patient demographic and insurance information when requested/needed • Notifies Data Quality Team Lead/Site specialist of duplicate patient records	• Interacts with patients directly • Establishes and maintains trust and ongoing relationship with patients • Collects, updates, and maintains identification and contact information for patients	• __________ • __________ • __________
Appointment schedulers / Call Center staff	• Answer patient calls for appointments • Schedule new and change appointments • Register all patients per registration protocols and collect all documentation • Reviews and verifies patient coverage of insurance • Updates patient demographic and insurance information when requested/ needed • Call and remind patient of his/her appointment	• Reviews and verifies patient contact information • Updates patient demographic and insurance information when requested/needed • Notifies Data Quality Team Lead/Site specialist of duplicate patient records	• Interacts with patients over the phone • Collects, updates, and maintains identification and contact information for patients	• __________ • __________ • __________

Position	Role	Data Quality Team Roles/Responsibilities	Data Quality Team Strengths	Names
Billing staff	• Corrects, completes, and processes claims of all payer codes and reviews all postings before claim submission • Assures accuracy of billing for all services rendered in patients account • Updates patient demographic and insurance information when requested/needed	• Reviews and verifies patient contact information • Updates patient demographic and insurance information when requested/ needed • Notifies Data Quality Team Lead/Site specialist of duplicate patient records	• May interact with patients over the phone • Collects, updates, and maintains identification, contact, and insurance information for patients	• ___________ • ___________ • ___________

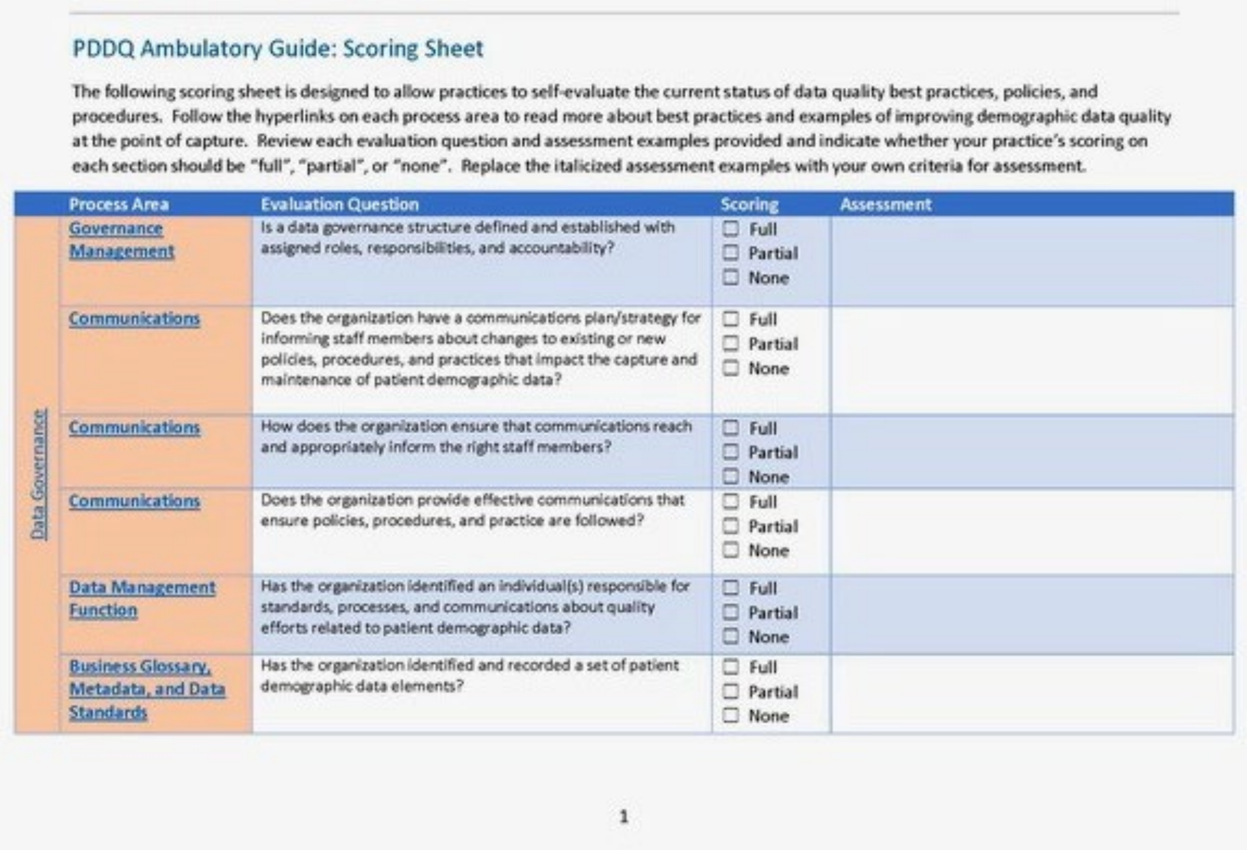
Example Monthly Data Quality Team Meeting Agenda
Date:

Time	Team Member	Item to Discuss	Action Items
12:00	Data Quality Team Lead and Site Specialist	Latest Data Quality Reports	Review latest duplicates report • Identify problem areas • Discuss workflow modifications • Troubleshoot technical issues if needed • Compare this month to last month duplicates/issues • Prioritize areas for improvement
12:30	Data Quality Team	Feedback on Front Desk Training	Short report outlining feedback to Data Quality Team Lead & Site Specialist • What has worked • What can be improved • Coordination between staff, clinics, and sites

Appendix F – PDDQ Ambulatory Guide Scoring Sheet

Figure 5a, Figure 5b, and Figure 5c present the PDDQ Ambulatory Guide Scoring Sheet.

Figure 5a. PDDQ Ambulatory Guide Scoring Sheet (1 of 3)



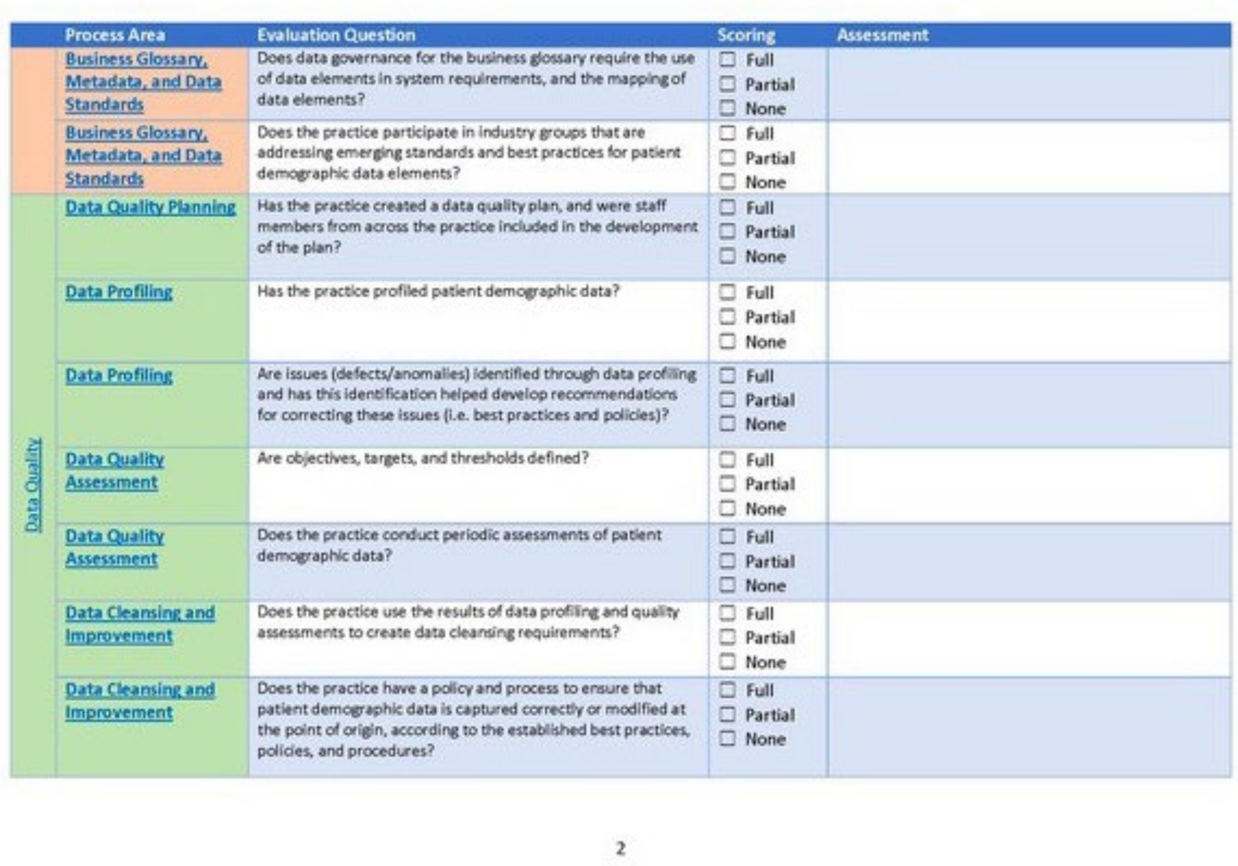
PDDQ Ambulatory Guide: Scoring Sheet

The following scoring sheet is designed to allow practices to self-evaluate the current status of data quality best practices, policies, and procedures. Follow the hyperlinks on each process area to read more about best practices and examples of improving demographic data quality at the point of capture. Review each evaluation question and assessment examples provided and indicate whether your practice's scoring on each section should be "full", "partial", or "none". Replace the italicized assessment examples with your own criteria for assessment.

	Process Area	Evaluation Question	Scoring	Assessment
Data Governance	Governance Management	Is a data governance structure defined and established with assigned roles, responsibilities, and accountability?	☐ Full ☐ Partial ☐ None	
	Communications	Does the organization have a communications plan/strategy for informing staff members about changes to existing or new policies, procedures, and practices that impact the capture and maintenance of patient demographic data?	☐ Full ☐ Partial ☐ None	
	Communications	How does the organization ensure that communications reach and appropriately inform the right staff members?	☐ Full ☐ Partial ☐ None	
	Communications	Does the organization provide effective communications that ensure policies, procedures, and practice are followed?	☐ Full ☐ Partial ☐ None	
	Data Management Function	Has the organization identified an individual(s) responsible for standards, processes, and communications about quality efforts related to patient demographic data?	☐ Full ☐ Partial ☐ None	
	Business Glossary, Metadata, and Data Standards	Has the organization identified and recorded a set of patient demographic data elements?	☐ Full ☐ Partial ☐ None	

1

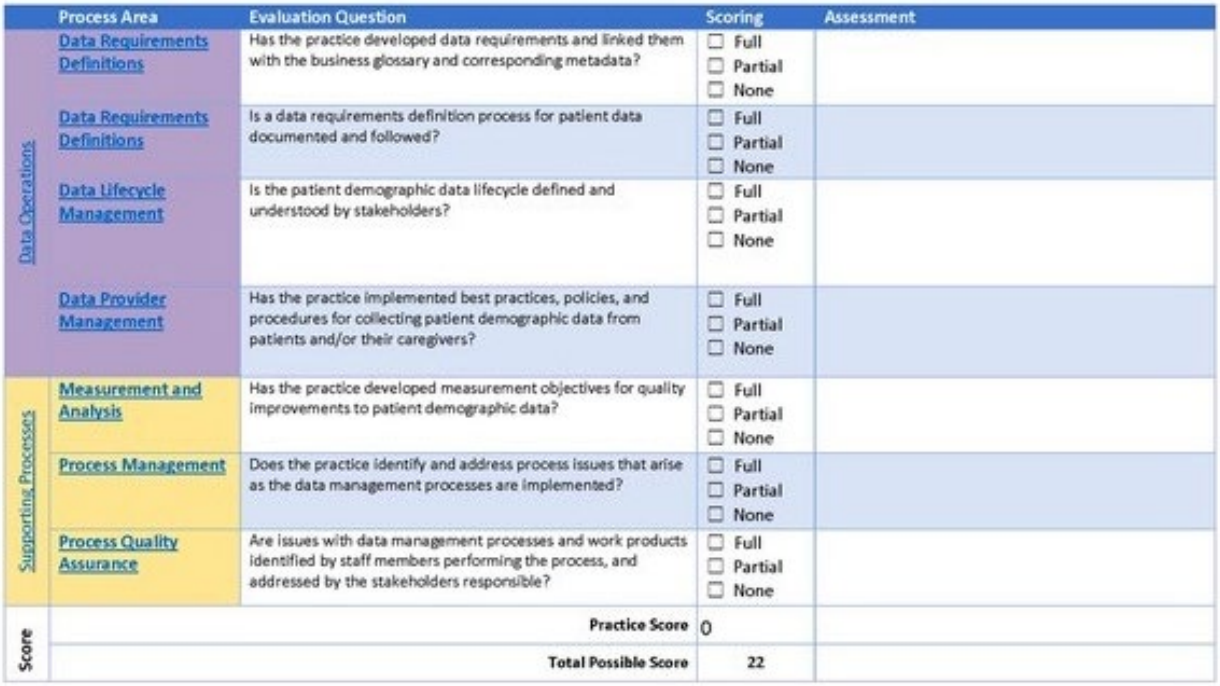
Figure 5b. PDDQ Ambulatory Guide Scoring Sheet (2 of 3)



	Process Area	Evaluation Question	Scoring	Assessment
	Business Glossary, Metadata, and Data Standards	Does data governance for the business glossary require the use of data elements in system requirements, and the mapping of data elements?	☐ Full ☐ Partial ☐ None	
	Business Glossary, Metadata, and Data Standards	Does the practice participate in industry groups that are addressing emerging standards and best practices for patient demographic data elements?	☐ Full ☐ Partial ☐ None	
Data Quality	Data Quality Planning	Has the practice created a data quality plan, and were staff members from across the practice included in the development of the plan?	☐ Full ☐ Partial ☐ None	
	Data Profiling	Has the practice profiled patient demographic data?	☐ Full ☐ Partial ☐ None	
	Data Profiling	Are issues (defects/anomalies) identified through data profiling and has this identification helped develop recommendations for correcting these issues (i.e. best practices and policies)?	☐ Full ☐ Partial ☐ None	
	Data Quality Assessment	Are objectives, targets, and thresholds defined?	☐ Full ☐ Partial ☐ None	
	Data Quality Assessment	Does the practice conduct periodic assessments of patient demographic data?	☐ Full ☐ Partial ☐ None	
	Data Cleansing and Improvement	Does the practice use the results of data profiling and quality assessments to create data cleansing requirements?	☐ Full ☐ Partial ☐ None	
	Data Cleansing and Improvement	Does the practice have a policy and process to ensure that patient demographic data is captured correctly or modified at the point of origin, according to the established best practices, policies, and procedures?	☐ Full ☐ Partial ☐ None	

2

Figure 5c. PDDQ Ambulatory Guide Scoring Sheet (3 of 3)



	Process Area	Evaluation Question	Scoring	Assessment
Data Operations	Data Requirements Definitions	Has the practice developed data requirements and linked them with the business glossary and corresponding metadata?	☐ Full ☐ Partial ☐ None	
	Data Requirements Definitions	Is a data requirements definition process for patient data documented and followed?	☐ Full ☐ Partial ☐ None	
	Data Lifecycle Management	Is the patient demographic data lifecycle defined and understood by stakeholders?	☐ Full ☐ Partial ☐ None	
	Data Provider Management	Has the practice implemented best practices, policies, and procedures for collecting patient demographic data from patients and/or their caregivers?	☐ Full ☐ Partial ☐ None	
Supporting Processes	Measurement and Analysis	Has the practice developed measurement objectives for quality improvements to patient demographic data?	☐ Full ☐ Partial ☐ None	
	Process Management	Does the practice identify and address process issues that arise as the data management processes are implemented?	☐ Full ☐ Partial ☐ None	
	Process Quality Assurance	Are issues with data management processes and work products identified by staff members performing the process, and addressed by the stakeholders responsible?	☐ Full ☐ Partial ☐ None	
Score		**Practice Score**	0	
		Total Possible Score	22	