

Building the Data Capacity for Patient-Centered Outcomes Research: The 2018 Annual Report

Office of Health Policy
Assistant Secretary for Planning and Evaluation
U.S. Department of Health and Human Services



Suggested citation: Dullabh P, Dhopeswarkar R, Heaney-Huls K, Hovey, L, Rajendran N, Moriarty E, Steiner C. Building the Data Capacity for Patient-Centered Outcomes Research: The 2018 Annual Report. Prepared under Contract No. HHSP233201600020I. The task order number for the current Cost Plus Fixed Fee umbrella contract is: HHSP23337001T between the Department of Health and Human Services' Office of the Assistant Secretary for Planning and Evaluation Office of Health Policy and NORC at the University of Chicago.

Table of Contents

Executive Summary	1
I. Introduction	2
II. Background and Context	3
Active OS-PCORTF Funded Projects	3
Methods	7
III. 2018 Major Accomplishments	7
Enhancing Patient Engagement through the Emerging Use of APIs	7
Data Governance	8
Enhancing Distributed Research Networks	9
Data Standardization	10
Linking Data	11
IV. Agency for Healthcare Research and Quality (AHRQ)	13
Harmonization of Clinical Data Element Definitions for Outcome Measures in Registries	13
Project Purpose and Goals:	14
Contributions to the PCOR Data Infrastructure Functionalities:	14
Accomplishments and Deliverables:	15
Coordination with Other Federal Agencies:	16
Capstone for Outcomes Measures Harmonization Project	16
Project Purpose and Goals:	17
Contributions to the PCOR Data Infrastructure Functionalities:	17
Accomplishments and Deliverables:	17
Coordination with other Federal Agencies:	17
V. Centers for Disease Control and Prevention (CDC)	18
Childhood Obesity Data Initiative (CODI): Integrated Data for Patient-Centered Outcomes Research Project	18
Project Purpose and Goals:	19
Contributions to PCOR Data Infrastructure Functionalities:	19
Accomplishments and Deliverables:	20
Enhancing Identification of Opioid-Involved Health Outcomes Using Linked Hospital Care and Mortality Data	20
Project Purpose and Goals:	21
Contributions to PCOR Data Infrastructure Functionalities:	21
Accomplishments and Deliverables:	21
Coordination with other Federal Agencies:	22

Improving the Mortality Data Infrastructure for Patient-Centered Outcomes Research	22
Project Purpose and Goals:.....	22
Contributions to PCOR Data Infrastructure Functionalities:.....	23
Accomplishments and Deliverables:	23
Strengthening the Data Infrastructure for Outcomes Research on Mortality Associated with Opioid Poisonings	24
Project Purpose and Goals.....	25
Contributions to PCOR Data Infrastructure Functionalities:.....	25
Accomplishments and Deliverables:	26
VI. Centers for Medicare and Medicaid Services (CMS)	27
Improving Beneficiary Access to Their Health Information through an Enhanced Blue Button Service	27
Project Purpose and Goals:.....	27
Contributions to PCOR Data Infrastructure Functionalities.....	28
Accomplishments and Deliverables	28
VII. Food and Drug Administration (FDA)	29
Collection of Patient-Provided Information through a Mobile Device Application for Use in Comparative Effectiveness and Drug Safety Research	29
Project Purpose and Goals:.....	30
Contributions to PCOR Data Infrastructure Functionalities:.....	30
Accomplishments and Deliverables:	31
Cross-Network Directory Service	32
Project Purpose and Goals:.....	32
Contributions to the PCOR Data Infrastructure Functionalities:.....	33
Accomplishments and Deliverables:	33
Source Data Capture from Electronic Health Records: Using Standardized Clinical Research Data	34
Project Purpose and Goals:.....	35
Contributions to PCOR Data Infrastructure Functionalities:.....	35
Accomplishments and Deliverables:	36
Coordination with Other Federal Agencies:.....	36
Standardization and Querying of Data Quality Metrics and Characteristics for Electronic Health Data	37
Project Purpose and Goals:.....	37
Contributions to PCOR Data Infrastructure Functionalities.....	37
Accomplishments and Deliverables:	37
Coordination with other Federal Agencies:	38

Utilizing Data from Various Data Partners in a Distributed Manner	38
Project Purpose and Goals:.....	39
Contributions to PCOR Data Infrastructure Functionalities:.....	39
Accomplishments and Deliverables:	40
VIII. National Institutes of Health (NIH).....	41
Creation of LOINC Equivalence Classes	41
Project Purpose and Goals:.....	42
Contributions to PCOR Data Infrastructure Functionalities:.....	42
Accomplishments and Deliverables:	42
Emergency Medicine Opioid Data Infrastructure: Key Venue to Address Opioid Morbidity and Mortality	44
Project Purpose and Goals:.....	44
Contributions to PCOR Data Infrastructure Functionalities:.....	45
Accomplishments and Deliverables:	45
Use of the ADAPTABLE Trial to Strengthen Methods to Collect and Integrate Patient-Reported Information with Other Data Sets and Assess Its Validity	46
Project Purpose and Goals:.....	46
Contributions to the PCOR Data Infrastructure Functionalities:.....	46
Accomplishments and Deliverables:	47
Coordination with Other Federal Agencies:.....	48
IX. Office of the National Coordinator for Health Information Technology (ONC)	49
Conceptualizing a Data Infrastructure for the Capture and Use of Patient-Generated Health Data.....	49
Project Purpose and Goals:.....	50
Contributions to PCOR Data Infrastructure Functionalities:.....	50
Accomplishments and Deliverables:	50
Security and Privacy Standards for Patient Matching, Linking, and Aggregation.....	51
Project Purpose and Goals:.....	52
Contributions to PCOR Data Infrastructure Functionalities:.....	53
Accomplishments and Deliverables:	53
X. Cross-Agency Funded Projects	55
Advancing the Collection and Use of Patient-Reported Outcomes through Health Information Technology	55
Project Purpose and Goals:.....	56
Contributions to PCOR Data Infrastructure Functionalities:.....	56
Accomplishments and Deliverables:	56
Coordination with Other Federal Agencies:.....	57

Assessing and Predicting Medical Needs in a Disaster	57
Project Purpose and Goals:.....	58
Contributions to the PCOR Data Infrastructure Functionalities:.....	58
AHRQ Accomplishments and Deliverables:	59
ASPR Accomplishments and Deliverables:.....	59
Developing a Strategically Coordinated Registry Network to Support Research on Women’s Health Technologies	59
Project Purpose and Goals:.....	60
Contributions to PCOR Data Infrastructure Functionalities:.....	61
Accomplishments and Deliverables:	61
Coordination with Other Federal Agencies:.....	63
Development of a Natural Language Processing Web Service for Public Health Use	63
Project Purpose and Goals:.....	64
Contributions to PCOR Data Infrastructure Functionalities:.....	64
Accomplishments and Deliverables:	65
Coordination with Other Federal Agencies:.....	67
Enhancing Data Resources for Researching Patterns of Mortality in Patient Centered Outcomes Research: Projects 1 - 4	68
Project Purpose and Goals:.....	69
Contributions to PCOR Data Infrastructure Functionalities:.....	70
Accomplishments and Deliverables:	70
Harmonization of Various Common Data Models and Open Standards for Evidence Generation	73
Project Purpose and Goals:.....	74
Contributions to the PCOR Data Infrastructure Functionalities:.....	75
Accomplishments and Deliverables:	75
Cross-Agency Collaboration:.....	77
Privacy and Security Blueprint, Legal Analysis and Ethics Framework for Data Use, and Use of Technology for Privacy	77
Project Purpose and Goals:.....	78
Contributions to PCOR Data Infrastructure Functionalities:.....	79
CDC Accomplishments and Deliverables:.....	79
ONC Accomplishments and Deliverables:	80
Coordination with Other Federal Agencies:.....	81
Technologies for Donating Medicare Beneficiary Claims Data to Research Studies	82
Project Purpose and Goals:.....	82
Contributions to PCOR Data Infrastructure Functionalities:.....	82
Accomplishments and Deliverables:	83
Disseminated Products:.....	83

XI. Conclusion	84
Appendix A. OS-PCORTF Project Portfolio	85
Appendix B. Abbreviations	88
Appendix C. Glossary.....	91
References.....	94

List of Exhibits

- Exhibit 1.** OS-PCORTF 25 Active Projects..... 4
- Exhibit 2.** Concluding Projects' Contribution to the Core Functionalities of the Strategic Framework.. 13
- Exhibit 3.** AHRQ Active Projects..... 13
- Exhibit 4.** CDC Active Projects 18
- Exhibit 5.** CMS Active Projects 27
- Exhibit 6.** FDA Active Projects 29
- Exhibit 7.** NIH Active Projects 41
- Exhibit 8.** ONC Active Projects 49
- Exhibit 9.** Cross-Agency Funded Active Projects 55

List of Tables

- Table A1.** Active OS-PCORTF Projects..... 85
- Table A2.** Completed OS-PCORTF Projects 87

Executive Summary

This 2018 Annual Report takes inventory of the Office of the Secretary Patient-Centered Outcomes Research Trust Fund (OS-PCORTF) portfolio projects active in 2018, in terms of goals, accomplishments, and contributions to building data capacity for patient-centered outcomes research (PCOR). The projects that make up the OS-PCORTF portfolio span a diverse set of topics that include data standardization at the point of care; improving the utility of disease registries and public health surveillance systems; empowering patients by improving access to their own health data and the ability to direct it towards research; and leveraging national standards to develop new applications to link data systems; as well as addressing urgent health and population concerns, such as disaster preparedness and the opioid crisis.

This annual report encompasses 25 projects funded and/or active in federal fiscal year (FY) 2018. These projects were funded to advance one or more of the five core research data infrastructure functionalities necessary for conducting PCOR:

- Collection of participant-provided information;
- Standardized collection of standardized clinical data;
- Linking of clinical and other data for research;
- Use of clinical data for research; and
- Use of enhanced publicly-funded data systems for research.

To illustrate the specific contributions of the OS-PCORTF projects to data infrastructure, the report includes an “at a glance” synthesis of the portfolio’s major accomplishments for 2018. This discussion highlights the ways in which 11 completed projects advance and enhance PCOR data infrastructure, offering high value solutions to researchers and clinicians. This involves innovations in:

- Use of application programming interfaces (APIs) to improve capture and use of health data;
- Enhancements to data governance structures to better support PCOR;
- Solutions to the well-known challenges associated with reliable patient matching;
- Facilitators for research being conducted across distributed networks;
- Data standardization to facilitate research and clinical care; and
- Improved mechanisms to produce reliable linkages between records.

In the final section of this report, project descriptions are provided for every project in the OS-PCORTF portfolio that were active in FY18. Each project description provides an overview of the project goals and objectives, summarizes the core research functionalities the project addresses, highlights major project accomplishments, and describes dissemination activities intended to raise awareness of the project’s work to the larger PCOR community. The project work performed in the past year contributes in a substantial way to building data capacity for PCOR.

I. Introduction

Harnessing the power of data for research holds tremendous potential to improve health care value, quality, and outcomes in the United States. Over the past decade, multiple laws have been enacted by Congress to build data capacity and expand research on the outcomes and effectiveness of interventions used in health care.

These laws recognized the need for better scientific evidence to inform the real world decisions of individual patients, providers, and policy makers about the benefits and risks of health care interventions. Furthermore, Congress recognized that the growing volume of existing health data could be better utilized for research that discovers new treatments and assesses the long-term health outcomes of therapies.

More than 150 exabytes of data are generated by the US health care system and the volume is increasing exponentially because of breakthroughs in digital health like wearable devices, telehealth, genomics, and personalized medicine. Furthermore new technologies like artificial intelligence, machine learning, and blockchain are being applied to large data sets for researchers to find new ways to cure or manage costly health conditions such as cancer, heart disease, and Alzheimer's. With US health care spending reaching \$3.5 trillion in 2017, there is a growing recognition that big data and new technologies will transform the diagnosis, treatment, and prevention of disease in the 21 century health care system.

Since 2010, the Office of Health Policy (HP) of the Assistant Secretary for Planning and Evaluation (ASPE) has funded and supported a portfolio of approximately 40 projects to build data infrastructure on topics such as opioids, value-based care, mortality data, real world evidence, emergency preparedness/response, and interoperability of electronic health records. ASPE works with HHS agency leaders to oversee the development and approval of projects to address Department priorities that build data capacity for patient-centered outcomes research. Twelve HHS agencies and offices currently participate in the ASPE managed data infrastructure program under the auspices of the Office of the Secretary Patient-Centered Outcomes Research Trust Fund (OS-PCORTF) portfolio. The program facilitates intradepartmental collaborations that address HHS data and analysis priorities.

The general goal of the OS-PCORTF portfolio is to expand data capacity and infrastructure for patient-centered outcomes research (PCOR). By building data capacity for research, rigorous studies can be conducted that use existing data to understand the safety and effectiveness of interventions used in health care across different patient populations.

This annual report provides a synopsis of the 25 OS-PCORTF projects active in calendar year 2018. These 25 projects, many of which are multi-agency project collaborations, represent 40 individual project awards. The report details project goals, objectives, major accomplishments, illustrates how the projects contribute to the five core PCOR data research functionalities, and highlights cross-agency collaborations across HHS more broadly. The report also synthesizes the work of 11 projects that concluded activities in 2018. For these 11 projects, the report describes how the project products contribute to PCOR data research components and functionalities, and how the projects' resulting services, standards, and policies developed can be applied to PCOR more generally.

The products being developed across the OS-PCORTF portfolio encompass standards, tools, services, and resources to support PCOR. These products are in various stages of maturity—ranging from those in the planning stages to fully developed products available for implementation in real-world research settings. As such, the annual report is a key resource for monitoring ongoing progress and identifying future PCOR data infrastructure needs.

II. Background and Context

HHS agencies routinely collect, link, and analyze data that can be used to generate new scientific evidence that expands knowledge about the outcomes and effectiveness of health care treatments and interventions. The OS-PCORTF was created to help build and enable national data capacity and infrastructure and leverage existing clinical data and federal data for the conduct of PCOR.¹ These goals and data infrastructure investments support the mission, statutory authorities, and annual priorities of each HHS agency and the Department as a whole.

The Office of the Secretary of HHS delegated authority to ASPE to coordinate and fund OS-PCORTF projects to build data capacity for PCOR. Specifically, the Secretary is charged to:

... provide for the coordination of relevant Federal health programs to build data capacity for comparative clinical effectiveness research, including the development and use of clinical registries and health outcomes research networks, in order to develop and maintain a comprehensive, interoperable data network to collect, link, and analyze data on outcomes and effectiveness from multiple sources including electronic health records.²

In keeping with this charge, since 2011, ASPE has funded approximately 48 HHS agency projects that are designed to build data capacity for research on patient outcomes. A common cross-agency interest in a robust data infrastructure for patient-centered research brings together the expertise of HHS agency leaders, informaticians, technologists, and researchers to solve data governance, policy, and technical challenges and achieve the goals described for PCOR. The topics of these projects simultaneously respond to the Secretary's priorities—the opioid crisis, value-based care, health insurance reform, and drug pricing³—as well as major federal initiatives, and individual HHS agency priorities.

Active OS-PCORTF Funded Projects

The goal of this report is to describe each of the 25 OS-PCORTF projects active in the past calendar year (2018). These 25 projects represent 40 total project awards since FY13. The projects are organized by the federal agency responsible for administering the project. Eight of the 25 projects are collaborations between multiple HHS agencies. These jointly funded collaborations are summarized in the penultimate section of this report, "Cross-Agency Projects." Also included in this report is a section on major accomplishments, which presents a synthesis of the contributions of the 11 projects concluding in 2018.

This structure highlights the breadth of the federal partners involved and the applicability of building PCOR data capacity to each agency's core mission, as well as to the shared goals of improving research, quality, and patient outcomes through PCOR and the learning health system. Exhibit 1 provides a brief summary of the 25 active projects described in this Annual Report, organized alphabetically by agency and concluding with cross-agency projects.

Exhibit 1. OS-PCORTF 25 Active Projects

Active Project	Project Description
Agency for Healthcare Research and Quality (AHRQ)	
Capstone for Outcomes Measures Harmonization Project	<ul style="list-style-type: none"> ■ Improve collection and use of outcomes measures by linking clinical data to two different registries and pilot testing the bi-directional exchange of data between the registries and clinical sites.
Harmonization of Clinical Data Element Definitions for Outcome Measures in Registries	<ul style="list-style-type: none"> ■ Convene clinical topic-specific working groups to discuss current outcome measures and how their data definitions can be harmonized to promote the use of common definitions across systems. The resulting definitions are to be made publicly available to PCOR researchers and analysts.
Centers for Disease Control and Prevention (CDC)	
Childhood Obesity Data Initiative: Integrated Data for Patient-Centered Outcomes Research Project (CODI)	<ul style="list-style-type: none"> ■ Pilot and test enhanced linkage and de-duplication tools and services to link clinical data, weight management program intervention data, and community-level census information to expand the availability of data for researchers studying obesity.
Enhancing Identification of Opioid-Involved Health Outcomes Using Linked Hospital Care and Mortality	<ul style="list-style-type: none"> ■ Link three federal data sets—the National Health Care Survey, the National Death Index, and the National Vital Statistics System’s restricted-use mortality files on drug overdose death—to enhance the identification of opioid-related mortality.
Improving the Mortality Data Infrastructure for Patient-Centered Outcomes Research	<ul style="list-style-type: none"> ■ Improve the mortality infrastructure by supporting more timely delivery of state death records to the National Death Index database, and by linking National Death Index database records with nationally collected hospital datasets to obtain a more complete picture of patient care and outcomes.
Strengthening the Data Infrastructure for Outcomes Research on Mortality Associated with Opioid Poisonings	<ul style="list-style-type: none"> ■ Enhance the national vital statistics system through improved data collection and electronic data capture of opioid-related death data.
Centers for Medicare and Medicaid Services (CMS)	
Improving Beneficiary Access to their Health Information through an Enhanced Blue Button Service (Blue Button 2.0)	<ul style="list-style-type: none"> ■ Building upon a previous OS-PCORTF award, Blue Button 2.0 creates an interface that enables CMS beneficiaries to connect their MyMedicare.gov data to applications and services they trust, including research platforms to which the beneficiary may be interested in providing data.
Food and Drug Administration (FDA)	
Collection of Patient-Provided Information through a Mobile Device Application for Use in Comparative Effectiveness and Drug Safety Research (MyStudies App)	<ul style="list-style-type: none"> ■ Create the infrastructure for collecting data from patients through a mobile device application and linking to electronic health data from a single data partner that participates in the FDA’s Sentinel distributed network. The pilot study captures data from pregnant women who volunteer to participate, and the code and technical documentation for the mobile app, web configuration portal, and secure cloud-based storage environment will be publicly posted.

Active Project	Project Description
Cross-Network Directory Service (CNDS)	<ul style="list-style-type: none"> ■ Create an interoperable service that allows data partners to participate in multiple data research networks, query across the networks, and share analytic capabilities and knowledge across networks. The project will be piloted across two existing networks: FDA’s Sentinel and PCORnet.
Source Data Capture from Electronic Health Records: Using Standardized Clinical Research Data (OneSource)	<ul style="list-style-type: none"> ■ Create a tool to seamlessly integrate EHRs and electronic data capture systems through a single point using the Retrieve Form for Data Capture standard.
Standardization and Querying of Data Quality Metrics and Characteristics for Electronic Health Data	<ul style="list-style-type: none"> ■ Create and implement a metadata standard data capture and querying system for data quality and characteristics, data source and institutional characteristics, and “fitness for use.”
Utilizing Data from Various Data Partners in a Distributed Manner	<ul style="list-style-type: none"> ■ Develop and test the capability to conduct timely and secure distributed regression analysis in distributed data networks. Additionally, explore the feasibility of creating virtual linkages to utilize data from multiple data sources and data for one specific patient with information at different institutions.
National Institutes of Health (NIH)	
Creation of LOINC Equivalence Classes	<ul style="list-style-type: none"> ■ Create a flexible, extensible, and computable mechanism for rolling data into clinically relevant equivalence groups that enable more efficient processing of data from diverse health information technology systems. The primary focus of this work will be on laboratory tests.
Emergency Medicine Opioid Data Infrastructure: Key Venue to Address Opioid Morbidity and Mortality	<ul style="list-style-type: none"> ■ Identify, develop and test clinical data elements relevant to opioid use disorder in the emergency department setting. The goal is to facilitate opioid-related research through improved data standards, interoperability, and linkages between EHRs, research networks, and registries.
Use of the ADAPTABLE Trial to Strengthen Methods to Collect and Integrate Patient-Reported Information with Other Data Sets and Assess Its Validity	<ul style="list-style-type: none"> ■ Generate tools and data standards that could be broadly deployed to support PCOR by leveraging the infrastructure of the ADAPTABLE trial (Aspirin Dosing: A Patient-Centric Trial Assessing Benefits and Long Term Effectiveness). This trial is the first major randomized comparative effectiveness trial to be conducted by PCORnet.
Office of the National Coordinator for Health Information Technology (ONC)	
Conceptualization of a Data Infrastructure for the Capture and Use of Patient-Generated Health Data	<ul style="list-style-type: none"> ■ Develop a policy framework for the use of patient-generated data in research and care delivery that addresses data collection tools, data donation policies, regulatory gaps, combining data with medical record data, and interoperability of data across health information systems and devices.
Security and Privacy Standards for Patient Matching, Linking, and Aggregation (PMAL)	<ul style="list-style-type: none"> ■ Identify the best patient attributes to successfully link patients’ data across research, clinical, and claims data sets and establish a set of security profiles to support the standardization and sharing of information for PCOR.

Active Project	Project Description
Cross-Agency Funded Projects	
Advancing the Collection and Use of Patient-Reported Outcomes through Health Information Technology	<ul style="list-style-type: none"> ■ Develop technical tools for collecting and integrating patient-reported outcomes assessments into EHRs or other health information technology products. ■ AHRQ, ONC
Assessing and Predicting Medical Needs in a Disaster	<ul style="list-style-type: none"> ■ Create the infrastructure to collect and query information to assess and predict medical utilization and need to aide in disaster response and recovery operations. ■ AHRQ and the Office of the Assistant Secretary for Preparedness and Response (ASPR)
Developing a Strategically Coordinated Registry Network to Support Research on Women's Health Technologies (WHT-CRN)	<ul style="list-style-type: none"> ■ Create a coordinated registry network for women's health technologies that will collect patient reported outcomes and employ structured data capture from EHRs for data collection and exchange. ■ FDA, NIH National Library of Medicine (NLM), ONC
Development of a Natural Language Processing Web Service for Public Health Use	<ul style="list-style-type: none"> ■ Develop a natural language processing service that will be accessible and publicly available to researchers on the Public Health Community Platform to address public health priority areas like improving population health outcomes and health equity. ■ CDC, FDA
Enhancing Data Resources for Researching Patterns of Mortality in Patient Centered Outcomes Research: Projects 1 - 4	<ul style="list-style-type: none"> ■ Successfully link National Death Index data on fact, cause, and manner of death to several federal population-based health data platforms in order to demonstrate the feasibility, better enable PCOR on patterns and correlates of mortality, and to facilitate federal collaboration on strengthening the infrastructure and methods for linking health care data to mortality data. ■ CDC, CMS, FDA
Harmonization of Various Common Data Models and Open Standards for Evidence Generation	<ul style="list-style-type: none"> ■ Build data infrastructure for conducting PCOR using data from routine clinical settings, including insurance billing claims, EHRs, and patient registries. This project intends to harmonize several existing common data models, potentially including PCORnet and other networks. ■ FDA, NIH National Cancer Institute (NCI), NIH National Center for Advancing Translational Sciences (NCATS), NLM, ONC
Privacy and Security Blueprint, Legal Analysis and Ethics Framework for Data Use, and Use of Technology for Privacy	<ul style="list-style-type: none"> ■ Develop a privacy and security data infrastructure blueprint, legal analysis, and ethical framework to address legal and privacy and security related policy issues that affect the use of data for various types of PCOR. ■ CDC, ONC
Technologies for Donating Medicare Beneficiary Claims Data to Research Studies	<ul style="list-style-type: none"> ■ Leverage the Sync for Science and Blue Button application programming interface programs to enable Medicare beneficiaries to donate their medical claims data for scientific research studies. ■ CMS, NIH

Methods

The project profiles that follow were constructed based on a review of five main sources of documentation reflecting project activities in the past calendar year. These sources included: 1) project statements of work describing the projects as originally conceived; 2) quarterly progress reports submitted by the awardees to ASPE; 3) meeting summaries of the quarterly project status calls between ASPE project officers and awardees; 4) final project reports, when available; and 5) information publicly posted to HHS agency websites. When possible, project deliverables submitted in tandem with the quarterly progress reports were also reviewed, allowing us to summarize completed deliverables and progress towards milestone deliverables. Finally, for projects that concluded in calendar year 2018, available final reports were reviewed and their conclusions incorporated into the summaries and synthesis of major accomplishments that follow.

III. 2018 Major Accomplishments

Across the portfolio, the OS-PCORTF projects have made meaningful contributions to the data infrastructure available to support research. ASPE is charged with coordinating the OS-PCORTF including providing guidance “to develop and maintain a comprehensive, interoperable data network to collect, link, and analyze data on outcomes and effectiveness from multiple sources, including electronic health records.”⁴

To illustrate the impact and scope of the OS-PCORTF contributions to data infrastructure, the report highlights key areas in which the 11 projects that concluded activities this year offer solutions to the needs of researchers and clinicians. These areas involve innovations in:

- the use of application programming interfaces (APIs) to improve capture and use of health data
- enhancements to data governance structures to better support PCOR
- solutions to the well-known challenges associated with linking data from different sources
- facilitators for research being conducted across distributed networks
- data standardization to facilitate research and clinical care
- improved mechanisms to produce reliable linkages between records

The featured OS-PCORTF projects within these areas represent work completed that is now available for broad use by the PCOR and research communities. Project Final Reports are posted to ASPE’s website in the “PCORTF Reports” tab: <https://aspe.hhs.gov/patient-centered-outcomes-research-trust-fund-reports>.

Enhancing Patient Engagement through the Emerging Use of APIs

Two projects that concluded activities in FY18 make new services available to patients in the form of enhanced control over their data via APIs, the FDA project, *Collection of Patient-Provided Information through a Mobile Device Application for Use in Comparative Effectiveness and Drug Safety Research* (MyStudies App) and the CMS project, *Improving Beneficiary Access to Their Health Information through an Enhanced Blue Button Service* (Blue Button 2.0). Patient engagement can take the form of both capture and reporting of Participant Provided Information (PPI), use of health-supportive apps, and the

donation of data to research programs. API-enabled apps also improve the services available to patients who wish to engage with their health data.

The FDA project created a mobile device application (app), the FDA MyStudies App, and patient data storage environment in which APIs: 1) facilitate secure and transparent recording of patient-reported outcomes (PRO) data, and 2) enable data linkage to existing electronic health data in distributed or centralized studies, trials, or registries. The app was piloted with an integrated health care delivery organization that participates in both Sentinel and PCORnet, two existing national distributed research networks that support comparative effectiveness and drug safety studies. The FDA's mobile app fills an important gap in clinical data by enhancing the collection of PPI from a complex and understudied population (pregnant women) on an important health issue (medication exposure). The FDA mobile app can be customized for other study populations beyond pregnancy to serve a variety of health care conditions and outcomes. Moreover, the app can be used in other distributed research networks beyond PCORnet and Sentinel. The app code and supporting documentation are publicly available at: <https://www.fda.gov/Drugs/ScienceResearch/ucm624785.htm>.

The original CMS Blue Button project made data available to patients for view and download; the next iteration, "Blue Button 2.0," enhanced these services, allowing patients to access their claims data through APIs based on Fast Healthcare Interoperability Resource (FHIR®) standards, open standards designed to promote standardized and more efficient data access and transfer. Blue Button 2.0 also allows patients to transfer standardized data to a range of third party health care platforms or apps of their choosing. It can be connected to a variety of different apps; for example, allowing patients to share health information with apps offering diabetes management tools or other disease-specific support.

Summary. Both the FDA MyStudies App and CMS Blue Button projects demonstrate how data collected through APIs can be transmitted to third parties for research—through mechanisms like direct data sharing with providers and research institutions, enrolling in research studies of interest, and/or data donation to registries or research networks. When patient data are standardized (e.g., using FHIR®), their clinical and research applications can be further enhanced by linking them to other data sources, as demonstrated by these products.

- FDA Collection of Patient-Provided Information through a Mobile Device Application for Use in Comparative Effectiveness and Drug Safety Research
 - ASPE-posted Final Report: <https://aspe.hhs.gov/system/files/pdf/259016/FDAMyStudiesAppAPatientCenteredOutcomes.pdf>
- CMS Improving Beneficiary Access to Their Health Information through an Enhanced Blue Button Service
 - Blue Button Website: <https://bluebutton.cms.gov/>

Data Governance

Patient-level data must be made available to researchers in a way that ensures the protection of patient privacy, while providing sufficient granularity to allow meaningful research questions to be assessed. This requires suitable policies and governance that are well known, understood, and implemented. Two portfolio projects enumerate and clarify common areas of confusion over the laws, policies, and technical

infrastructure needed to support the use of patient data for PCOR, and they have created practical solutions to support and demonstrate patient-generated health data (PGHD) collection and use.

The *ONC/CDC Privacy and Security Blueprint, Legal Analysis and Ethics Framework for Data Use and Use of Technology for Privacy (Privacy and Security Blueprints)* was developed from an in-depth analysis of the current privacy and security data infrastructure, and the legal, ethical, and policy issues that apply to PCOR researchers. The blueprints are designed to offer practical advice to researchers on how to safely use and protect health data, including guidance on the relevant federal and state laws, issues such as consent procedures for special populations and proper data de-identification practices. Such guidance better enables the use of clinical data for research and enhances the utility of publicly funded data systems for research.

The *ONC Conceptualizing a Data Infrastructure for the Capture and Use of Patient-Generated Health Data project* was a multi-pronged effort to combine PGHD data with medical record data across multiple health information systems and devices. This began with an assessment of the necessary data collection tools, technical barriers, data donation policies, and regulatory gaps. It then demonstrated successful capture and use of PGHD in two pilot sites. One site focused on app-driven capture of PGHD and its integration into physician workflow to achieve better care coordination and population management for diabetes patients. The other site tested a technical platform capturing PGHD to support care for orthopedic surgery, behavioral health, bariatric surgery, and stroke.

Summary. These projects fill gaps in the policies, governance, and technical infrastructure needed to facilitate data flow for PCOR, and address confusion in the industry that can create its own barriers to research. The Privacy and Security (ONC), Legal and Ethical (ONC, CDC), and PGHD (ONC) frameworks can be applied across the research landscape, and the PGHD project further demonstrates the necessary framework and technology for the successful capture, use, and sharing of PGHD in clinical care delivery and research models, improving care for diverse medical conditions and populations.

- ONC Privacy and Security Blueprint, Legal Analysis and Ethics Framework for Data Use, and Use of Technology for Privacy
 - ONC Final Legal Architecture: <https://www.healthit.gov/sites/default/files/page/2018-02/Legal-and-Ethical-Architecture-for-PCOR-Data-Final.pdf>
- CDC Privacy and Security Blueprint, Legal Analysis and Ethics Framework for Data Use, and Use of Technology for Privacy
 - CDC Final Blueprint: https://aspe.hhs.gov/system/files/pdf/259016/PCOR_Legal_508_2.pdf
- ONC Conceptualizing a Data Infrastructure for the Capture and Use of Patient-Generated Health Data Final White Paper
 - https://www.healthit.gov/sites/default/files/onc_pghd_final_white_paper.pdf

Enhancing Distributed Research Networks

Information on patient health care services is captured across various data sources, therefore tools are needed to link data across health care databases. Linkages provide more robust cross-sectional or longitudinal patient profiles to enhance secondary uses of electronic health care information for research

purposes, and to improve access to diverse clinical information. Distributed research networks support data querying and other functions across institutions without requiring any transfer of protected data. There are various technical challenges to supporting research across distributed networks, which two portfolio projects attempted to address with different approaches for querying and running analyses (tested in both Sentinel and PCORnet).

The FDA's project, *Utilizing Data from Various Data Partners in a Distributed Manner*, developed enhanced analytic capabilities and fully automated distributed linear regression analysis of patient data across organizations. This functionality is enabled through an open source software application, PopMedNet™, which allows stakeholders within a distributed data network to access and analyze data within a single dataset, and it can be adapted for non-PopMedNet™ applications. This enhances the use of federal and non-federal data systems for research, allowing network members to easily access disparate data sources.

FDA's *Cross-Network Directory Service (CNDS)* attempts to make distributed queries easier by creating an open source interoperable service that allows data partners to participate in multiple data research networks, query across the networks, and share analytic capabilities and knowledge across networks. The project was piloted in FDA's Sentinel System and PCORnet. Both project teams developed adaptable solutions to a real-world research problem and have also made available the technical and user documentation to encourage its implementation by others.

Summary. Both projects address a persistent problem in clinical practice and research by easing the access and use of data from multiple data sources with unique populations, as well as data for one specific patient with information housed at different institutions. These solutions enhance the shareability and queryability of data, while allowing research networks to maintain control of patient-level data, providing a balance between analytic requirements, patient privacy and confidentiality, and proprietary considerations. This increases the availability of data to answer research questions and is applicable across research networks.

- FDA Utilizing Data from Various Data Partners in a Distributed Manner
 - Final Technical Report: <https://aspe.hhs.gov/system/files/pdf/259016/DistributedRegressionAnalysisFinalTechnicalReport.pdf>
- FDA Cross Network Directory Service
 - Final Report: https://aspe.hhs.gov/system/files/pdf/259016/CNDS_Final_Report.pdf

Data Standardization

Health data being collected as part of a clinical encounter have the potential to support hypothesis generation and large scale clinical research studies, as well as care quality improvements. However, differences in clinical definitions across institutions and the resulting variability can challenge the meaningful interpretation of study results and use of the results to improve patient outcomes. Two projects, from the NIH and AHRQ, respectively, address the need for clinically relevant standards with which to capture data and uniform application of these standards across institutions.

NIH's project, *Creation of LOINC Equivalence Classes*, endeavors to simplify the use of LOINC standards through "equivalencing" or rolling-up a hierarchy. This gathers similar LOINC codes together in clinically

relevant groups to enable more efficient processing and aggregation of data from diverse health information technology (health IT) systems. This process involved the development of an enhanced software tool to help create priority subsets of codes that were officially incorporated into the LOINC library and publicly disseminated. Creation of useable and widely disseminated standards for data collection helps improve the quality of data collected and thereby enhances the quality of the research that can be conducted using such data.

AHRQ's *Harmonization of Clinical Data Element Definitions for Outcome Measures in Registries* is an effort to create and pilot a conceptual model for developing standard outcome measures. The project developed a consensus set of clinical data element definitions for outcome measures for each of five clinical topic areas: atrial fibrillation, asthma, depression, lung cancer and lumbar spondylolisthesis. These standards are to be used in registries and as a pilot demonstration of best practices for governance of data element definition libraries, and best practices for harmonization of outcome measures between different registries, EHRs, and reporting requirements.

Summary. Both projects contribute to the availability of standards for clinical data collection and research, as well as standardizing the process for developing outcome measures with which to gauge clinical improvement. This has direct application to the need for tools and processes to standardize data across multiple clinical conditions and settings, and improves the data and reporting for the five specified clinical conditions with opportunities for continued expansion into other diseases.

- NIH Creation of LOINC Equivalence Classes
 - The LOINC release artifact was published along with LOINC version 2.64 and is available at: <https://loinc.org/groups/>
 - The latest LOINC Groups content available via an API based on the FHIR® standard specification is available at: <https://loinc.org/fhir/>
- AHRQ Harmonization of Clinical Data Element Definitions for Outcome Measures in Registries
 - A draft of the final report: <https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/data-elements-draft-report.pdf>

Linking Data

Three portfolio projects addressed the need for robust data and linkages to support research, through different approaches. One developed services to better link and aggregate data from disparate sources, while two others enhanced high value data sources (e.g., the National Death Index [NDI]) in anticipation of such linkages. One of the projects also addressed the absence of widely adopted standards that consistently address patient identity, authorization, and consent, which are necessary to capture and use electronic health data for PCOR.

ONC's project, *Security and Privacy Standards for Patient Matching, Linking, and Aggregation (PMAL)*, improves the tools available to match patient data across diverse records and aggregate records into a single data set, and created a privacy and security "layer" that ensures that the data are being matched properly; this means: 1) with the appropriate consents; 2) by an authorized user; and 3) with authentication. The project team pilot tested the standards and services with clinical data research networks and integrated additional data from the National Plan and Provider Enumeration System (NPPES) to improve patient matching across sources.

The CDC's effort *Improving the Mortality Data Infrastructure for Patient-Centered Outcomes* addressed three gaps in mortality infrastructure by successfully: 1) increasing the electronic reporting of mortality data from states to the CDC; 2) piloting national standards for data exchange between EHRs and state vital records systems, and state vital records systems and the NDI; and 3) linking the NDI with mortality datasets from the National Hospital Care Survey (NHCS) of in-patient and emergency departments. Given that death is a significant outcome measure when assessing care quality, these activities improved the quality, timeliness, and completeness of data in the NDI that can be used for research.

The CDC project, *Enhancing Data Resources for Researching Patterns of Mortality in PCOR* expanded the availability of data on cause and manner of death that is available in the NDI (i.e., adding Medicare data from 2008-2016 and Medicaid data from 2007-2015). This involved the development of standard, repeatable, and efficient technical solutions for linking the NDI's death and cause of death data to large, commercially and publicly insured populations. The project also demonstrated the feasibility of using these tools for linkages and subsequent research, completing a research use case that assessed the association between mortality and certain medications.

Summary. The projects demonstrate the importance and the complexity of solutions to support data linkages in research. PMAL provides a set of standards to ensure high quality linking—addressing key issues in data linkages like consent and data provenance, which have broad applicability in supporting further linkages of clinical data with mortality data. The projects also enhanced existing data sources to ensure that high quality data is available for linkage and subsequent research. Mortality is an important outcome for any public health crisis, including the opioid crisis, and therefore having access to more accurate, timely, and complete data via NDI is a boon for researchers, providers, and patients alike.

- ONC Security and Privacy Standards for Patient Matching, Linking, and Aggregation
 - Tool Specifications Document—Design & Component Development: <https://github.com/mitreid-connect/OpenID-Connect-Java-Spring-Server/>
 - API profile specifications: <https://openid.bitbucket.io/HEART/>
 - The Health Data Provenance Challenge: <https://www.cccinnovationcenter.com/challenges/provenance-challenge/>
 - The Move Health Data Forward Challenge: <https://www.healthit.gov/buzz-blog/interoperability/announcing-winners-move-health-data-challenge/>
- CDC Improving the Mortality Data Infrastructure for Patient Centered Outcomes
 - Final Report: https://www.cdc.gov/nchs/data/nvss/evital/PCOR_200-20012-50804_July_2017.pdf
- CDC Enhancing Data Resources for Researching Patterns of Mortality in PCOR
 - Methodological Report: <https://www.cdc.gov/nchs/data-linkage/nhcs-ndi.htm>

These 11 projects contribute substantially to the core functionalities put forth in the Strategic Framework, most contributing to more than one functionality. Exhibit 2 displays the project contributions in terms of these five core functionalities.

Exhibit 2. Concluding Projects' Contribution to the Core Functionalities of the Strategic Framework

Core Functionalities	Projects Contributing Tools and Services
Collection of Participant-Provided Information	3
Standardized Collection of Standardized Clinical Data	5
Linking of Clinical and Other Data for Research	6
Use of Clinical Data for Research	4
Use of Enhanced Publicly-Funded Data Systems for Research	4

In the sections that follow, the reports presents the 25 active portfolio projects, describing their objectives, progress, and contributions to PCOR data infrastructure and the above functionalities in detail.

IV. Agency for Healthcare Research and Quality (AHRQ)

AHRQ is administering four active OS-PCORTF-funded projects (Exhibit 3).

Exhibit 3. AHRQ Active Projects

AHRQ-Funded Projects
Harmonization of Clinical Data Element Definitions for Outcome Measures in Registries
Capstone for Outcomes Measures Harmonization Project
Advancing the Collection and Use of Patient-Reported Outcomes through Health Information Technology*
Assessing and Predicting Medical Needs in a Disaster*

* Denotes a cross-agency funded project that involves more than one federal agency; these projects are described in the "Cross-Agency Funded Projects" section.

Harmonization of Clinical Data Element Definitions for Outcome Measures in Registries

EHRs and registries contain a variety of rich clinical information including demographics, diagnoses, medications, allergies, and laboratory values. These data have the potential to support hypothesis generation and large-scale clinical research studies. To this end, vocabulary and standardization of data, and methods to electronically exchange information between EHRs and registries are a priority for federal stakeholders, developers, and researchers alike. Less focus has been placed on differences in clinical definitions across institutions and on efforts to harmonize these definitions; however, it is essential to address the variability in clinical definitions in order to meaningfully interpret

AHRQ has developed standardized libraries of outcome measures in five condition areas: atrial fibrillation, asthma, depression, lung cancer and lumbar spondylolisthesis. These condition areas represent diverse populations and care settings, different treatment modalities, and different levels of harmonization.

results of studies and use the results to improve patient outcomes. One example is variability within quality measures and public reporting requirements related to care processes, use of medications, mortality, and readmission rates after myocardial infarction. While existing measures provide standard definitions for each process and outcome being measured, they do not address variability in the clinical definition of myocardial infarction used by individual health care institutions. For example, one study investigating the impact of differing clinical definitions on study findings showed that in one registry, the rate of reported myocardial infarction was 7.2 percent when using the biomarker CKMB as the criteria, compared to 24.3 percent when using the biomarker troponin as the criteria, representing a three-fold difference in ascertainment.⁵

This project supported the creation of harmonized outcome measures and the AHRQ Outcome Measures Framework—a conceptual model for developing standard outcome measures. Together, these activities aimed to harmonize clinical data element definitions thereby supporting the comparison of data across registries housed in the AHRQ Registry of Patient Registries (RoPR), and create a model for future outcome measure harmonization efforts between registries, EHRs, and reporting requirements.⁶

Project Purpose and Goals:

The purpose of this project was to convene a series of clinical topic-specific working groups to discuss the various definitions for outcome measures currently in use and how these definitions can be harmonized into a common set to be used across data collection and reporting systems. The working groups solicited input from a broad stakeholder community, including registry holders, informaticists, and clinical subject matter experts; policymakers developing quality measures and other types of mandatory reporting measures; as well as clinicians, health systems, industry representatives, federal health and human service agencies, and patients.

The project objectives were to:

- Develop a consensus set of clinical data element (CDE) definitions that can be consistently used to represent specific outcome measures for each of five clinical topic areas: atrial fibrillation, lumbar spondylolisthesis (lower back pain), lung cancer, asthma, and depression.
- Develop best practices for governance of data element definition libraries and harmonization between registries. The definitions and best practices have been made publicly available for use by registry owners and PCOR researchers.

Contributions to the PCOR Data Infrastructure Functionalities:

- *Linking of Clinical and Other Data for Research.* There is significant variation in both the types and definitions of outcome measures used in patient registries, even within the same clinical area. These heterogeneous data reduce the utility of registries to collect, link, and aggregate data. By using the AHRQ libraries, data captured in one registry will have the same clinical meaning as another so it becomes easier to aggregate data collected for varying primary purposes. This project also addressed issues of data quality by reducing data heterogeneity to increase the availability of consistent quality data for PCOR.
- *Standardized Collection of Standardized Clinical Data.* This project was also aimed at standardizing clinical data collection. By adopting the core common outcome measures/data element sets, the data available for research will be standardized and comparable across different registries and health care organizations that collect the information.

Accomplishments and Deliverables:

The five clinical topic focus areas this project addresses were selected because they represent different clinical conditions and a wide variety of patient populations. In addition to the active clinical working groups advising the project, a stakeholder group of payers, patient representatives, and health system leaders were assembled to discuss challenges and provide feedback on the harmonization effort.

The draft libraries of clinical definitions for the topics of interest—atrial fibrillation, depression, asthma, lung cancer and lumbar spondylolisthesis—are complete and posted for peer review. These libraries contain not only the clinical definitions of the outcome measures, but also the value sets, which are critical to implementation and interoperability. Value sets use standard clinical coding terminologies (i.e., RxNorm, ICD-10, LOINC, and SNOMED) to define the clinical concepts (in this context, the outcome measure itself) and to help users implement the codes in EHRs. The team is also working on submission to peer reviewed journals and presentations at scientific meetings in the clinical topics.

AHRQ is working with relevant professional societies (e.g., Heart Rhythm Society, the American Medical Association) to publish and seek endorsement for the AHRQ definitions in these organizations' registries. Some professional societies have expressed interest in seeing the clinical data elements and value sets submitted to the NIH's National Library of Medicine (NLM) Value Set Authority Center, a repository of publicly created value sets. AHRQ is also pursuing inclusion in the NIH's. The NIH CDE Repository is a platform that contains data element definitions to use in research and other purposes. The NIH CDE Repository definitions include "machine-readable" definitions (i.e., those that can be interpreted by an EHR). Adapting the AHRQ clinical data elements in machine-readable format will facilitate standard data collection at the point of care.

The Spine Society has invited AHRQ to speak at their annual conference to disseminate their results among its members. In addition, the American College of Cardiology is exploring a project similar to AHRQ's OS-PCORTF Capstone for Outcomes Measures Harmonization Project, to improve the collection of registry-ready data at the point of care through EHRs.

Disseminated Products:

- A draft of the final report is posted for peer review: <https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/data-elements-draft-report.pdf>
- The project team is continuing to submit abstracts to meetings of methodologists and researchers (e.g., DIA, ISPOR, AcademyHealth) and the clinical community in each of the five clinical areas.
- The project team received final Office of Management and Budget approval for the next version of the RoPR, the database of registry-specific information that promotes collaboration, redundancy, and transparency. They will soon upgrade RoPR, which will allow any registry owner to upload their definitions of outcome measures, and will also importantly allow people to compare and retrieve the harmonized definitions. The registry is available here: <https://patientregistry.ahrq.gov/>
- The project team is discussing with the American Medical Association [Integrated Health Model Initiative](#) about collaborations, and presented in a joint webinar for the OS-PCORTF team.
- AHRQ gave a talk to the Medicaid Medical Directors to raise awareness about the asthma library.
- AHRQ publicly posts each clinical definition in a library on their website for peer review. These libraries are available here: <https://effectivehealthcare.ahrq.gov/topics/registry-of-patient-registries/outcome-measures-framework>

- AHRQ presented a panel titled "Harmonizing Outcome Measures to Increase the Utility of Patient Registries: A Case Study in Atrial Fibrillation" at the American Medical Informatics Association (AMIA) Joint Summits on Translational Science in 2017.

Coordination with Other Federal Agencies:

AHRQ continues to work with the NIH/NLM to add the AHRQ-developed clinical data elements, when complete, to the NIH's CDE Repository.

Period of Performance:

4/29/16 – 4/28/18

Federal Point of Contact:

Elise Berliner

Capstone for Outcomes Measures Harmonization Project

The AHRQ RoPR contains almost 4,000 registries for research, quality improvement, public reporting, and post market surveillance as of April 2018.⁷ These registries have considerable potential to be used for PCOR. However, because these data do not share standardized definitions of outcomes measures, analysis of data derived across registries is challenging. AHRQ has an existing OS-PCORTF funded project to harmonize measures in five clinical areas titled "*Harmonization of Clinical Data Element Definitions for Outcome Measures in Registries*," in which stakeholders identified three major barriers to the implementation of measures: 1) burden on clinical sites to collect data; 2) disruption to clinical care and challenges in extracting data from the clinical records; and 3) challenges with working with EHRs. While stakeholders recognize the importance of harmonized outcomes measures, these barriers contribute to reluctance to implement.

The main goal of this project is to address these barriers, using the depression topic as an example. Researchers will collect complex outcome measure information on depression from clinical sites and will transfer those data to existing patient registries. This project will utilize many PRO and structured data collection tools developed by the OS-PCORTF funded project "*Advancing the Collection and Use of Patient-Reported Outcomes through Health Information Technology*." These tools include a mobile and web-based platform with reminders for patients to fill out forms at specific time intervals and automatic generation of outcome measures that are presented back to physicians through integration with EHRs. While these tools can address the barriers, they have not yet been broadly implemented.

Researchers will implement outcome measures for depression into a variety of settings by linking clinical data to two different registries and testing the exchange of data back from the registries to participating clinical sites. The project will collect outcome measures through three different methods: 1) by extracting data already available in the EHR; 2) from those calculated from items in the nine item Patient Health Questionnaire (PHQ-9); and 3) using new data collection from structured data capture or through natural language processing of clinical notes,

This project will work with clinical sites of an integrated health system and two distinct registries (a primary care focused registry and a specialty care registry) to address and overcome challenges while developing registry enhancements. Knowledge gained from this project will address barriers with accessible tools so that other registries can enable collection and use of data that are cost effective for sites and will reasonably fit into their clinical workflow.

Project Purpose and Goals:

This project expands data capacity for PCOR by: 1) collecting electronic health data from multiple clinical sites on patient outcomes associated with treatment for depression; 2) transferring the collected data to electronic patient registries for research; and 3) developing infrastructure and tools to allow replication. The overall goal of the project is to examine whether this approach to data collection enhances the use of registries of research on patient outcomes by completion of the following outcomes:

- Tools for clinicians and patients to facilitate integration of the harmonized depression outcome measures into EHRs and registries so that these data will be available for clinical research, PCOR, quality improvement and implementation research.
- Proof of concept for a standards-based approach for collecting and reporting patient outcomes information to clinicians within their workflow and simultaneously transmitting the data to registries to make it available for research.
- Tools, such as instructions and pieces of code, to make it easier for researchers and registry developers to integrate registries with clinical systems.

Contributions to the PCOR Data Infrastructure Functionalities:

- *Linking of Clinical and Other data for Research.* AHRQ's work will support the connection between EHRs and data registries. This will improve patient outcomes research by broadening the scope of core outcome measures by linking EHR data to two existing registries
- *Use of Enhanced Publicly-Funded Data Systems for Research.* This project will link EHRs to registries. The linkage will allow researchers to share knowledge across networks.
- *Use of Clinical Data for Research.* The ability to link EHR and registry data provides broader patient profiles and supports analyses across populations that may not have been otherwise possible.

Accomplishments and Deliverables:

Since the project began, AHRQ has finalized and posted the Request for Proposal for the project. Upon contract award, planning and recruitment of clinical sites will begin.

Coordination with other Federal Agencies:

AHRQ will work with NLM to develop standardized implementation models in year 1 of the contract. A standardized implementation model is a reproducible model that improves the likelihood of clinical uptake through standardized steps. These models are externally validated to enable sharing and implement across environments and healthcare databases.

Period of Performance:

6/1/18 – 5/1/21

Federal Point of Contact:

Kristen Huntley

V. Centers for Disease Control and Prevention (CDC)

CDC is administering seven active OS-PCORTF-funded projects (Exhibit 4).

Exhibit 4. CDC Active Projects

CDC-Funded Projects
Childhood Obesity Data Initiative: Integrated Data for Patient-Centered Outcomes Research Project
Enhancing Identification of Opioid-Involved Health Outcomes Using Linked Hospital Care and Mortality Data
Improving the Mortality Data Infrastructure for Patient-Centered Outcomes Research
Strengthening the Data Infrastructure for Outcomes Research on Mortality Associated with Opioid Poisonings
Development of a Natural Language Processing Web Service for Public Health Use*
Enhancing Data Resources for Researching Patterns of Mortality in PCOR*
Privacy and Security Blueprint, Legal Analysis and Ethics Framework for Data Use, and Use of Technology for Privacy*

* Denotes a cross-agency funded project that involves more than one federal agency; these projects are described in the “Cross-Agency Funded Projects” section.

Childhood Obesity Data Initiative (CODI): Integrated Data for Patient-Centered Outcomes Research Project

Childhood obesity affects over 20 percent of children age 2-19 years old.⁸ In 2017, the U.S. Preventative Services Task Force (USPSTF) issued a recommendation to screen children for obesity and refer children with obesity to intensive weight management programs (WMPs) to improve physical and mental health.⁹ The USPSTF identified gaps in childhood obesity research related to the ability to personalize interventions for children and families according to their health, sociodemographic, and social factors.

Data to study factors that influence the health outcomes and effectiveness of interventions used to prevent or treat childhood obesity are often unavailable for research. Thus, a priority of those working in PCOR is the ability to link EHR data (e.g., weight or laboratory assessments) with weight management intervention data (e.g., type and duration of services provided) and community-level census information to support comparative effectiveness research (CER). An expanded health information dataset and system to link EHR and community-level data would allow researchers to answer critical questions such as how to personalize interventions for diverse children and families.¹⁰

Currently, distributed PCOR networks routinely gather data collected in health care settings, and structure these data in a common way (i.e., with a Common Data Model [CDM]). These networks, though capable of combining patient-level health intervention and community-level data, lack coding for children’s data, so these types of linkages have been limited. By building linkages and more advanced tools, CODI will help researchers fill the evidence gaps identified by the USPSTF in 2017. These data will help researchers answer questions such as whether all children are being screened appropriately for obesity, whether disparities exist by demographic groups, and will assist in the identification of WMPs that are most effective.

The project will pilot enhanced tools and services (e.g., patient record linkage and deduplication services) in the Colorado Health Observation Regional Data Service (CHORDS), a PCORnet Clinical Data Research Network (CDRN). CHORDS has initiatives to link patient health records to other data sources. To date, coding improvements and implementation of linkage services in large networks has been limited due to low resources for this purpose. The CODI project is designed to build data capacity of related research by linking data from health records such as weight height and blood pressure with WMP interventions and communities. This will expand the availability of data for childhood obesity PCOR that help health professionals develop and tailor interventions that are specific to the needs of children and.

This project will take place in three phases: 1) align efforts and resources; 2) expand electronic health IT infrastructure to capture, standardize, and query linked data; and 3) identify, develop, document and support a local implementation of CODI in Denver, Colorado. These tools will allow local researchers to combine patient-level EHR, WMP intervention, and community (census) data from multiple Denver health care and public health institutions that have existing collaborations for childhood obesity PCOR.

Project Purpose and Goals:

The purpose of this project is to link pediatric clinical EHR data, WMP intervention data, and community-level census information to expand the availability of data for researchers.

The objectives of the CODI project are:

- Establish an end-user collaborative of 15 subject matter experts, childhood obesity researchers, and network representatives to capture the data needs of end-users and technical requirements needed to facilitate the exchange and linkage of the data identified.
- Expand and standardize patient-level EHR and WMP intervention data, as well as community-level census data by expanding to capture CODI clinical and intervention data elements in the PCORnet CDM.
- Identify and describe the current business and technical processes and tasks for capturing the required childhood obesity data within health care and WMP intervention workflows, and then design the future ideal processes for the CODI project.
- Expand linkage and de-duplication tools for integrating childhood obesity data. The CODI record linkage and deduplication data tools, services, and implementation guide will be publicly available on CDC's cloud-based Surveillance Data Platform.
- Pilot the enhanced CODI IT services and implementation guide to test the ability to link and query data in CHORDS to produce datasets for PCOR researchers to analyze.

Contributions to PCOR Data Infrastructure Functionalities:

- *Use of Clinical Data for Current Research:* This project will use a range of data sources to capture location and demographic specific data on interventions.
- *Standardized Collection of Standardized Clinical and Claims Data:* This project will expand and standardize information that is available in a distributed research network to study the effectiveness of interventions used for the prevention and treatment of childhood obesity.
- *Linking of Clinical and Other Data for Research:* This project will connect EHR data on height, weight, and blood pressure with program interventions and community-level data to create a broader picture of how to best target interventions.

Accomplishments and Deliverables:

The project began in earnest in September 2018, with contract execution for the development of the end-user collaborative, identification of the business and technical process for capturing childhood obesity data, enhancing the linkage and de-duplication tool, and hosting the CODI resources on the Surveillance Data Platform, along with a cooperative agreement award to manage, govern, develop, and implement CODI.

Disseminated Products:

- Project team members presented twice at the Public Health Informatics Conference, August, 2018, Atlanta, Georgia:
 - Linking Clinical & Community Systems: Child Obesity Data Initiative
 - Linking Clinical and Community Systems: Integrating EHR and Community-health Data to Address Childhood obesity and Food Insecurity

Period of Performance:

5/30/18- 9/30/21

Federal Point of Contact:

Aly Goodman

Enhancing Identification of Opioid-Involved Health Outcomes Using Linked Hospital Care and Mortality Data

National-level statistics on opioid-related hospitalizations are limited and often incomplete. From 2005 to 2014, the rate of opioid emergency department (ED) visits increased by an estimated 99.4 percent.¹¹ It was also reported that opioid overdose deaths in the ED increased 27.7 percent from 2015 to 2016.¹² These estimates do not identify specific drugs causing the rise in ED visits and overdoses. EHR data contain much more detail such as clinical notes and laboratory results, which allow a wider perspective on the hospitalization. These data will allow health policy makers, clinicians, and researchers to develop effective prevention strategies and improve patient care.

The National Center for Health Statistics (NCHS) houses three data sources that, when combined, will offer broad, national-level data on hospital care and death related to opioid-involved drug overdose. The first source, the National Hospital Care Survey (NHCS), collects inpatient, ED, and outpatient claims and EHR data from over 500 participating hospitals. The NDI, the second source, includes all deaths occurring within the U.S. along with cause of death. The third source, The National Vital Statistics System restricted-use mortality files on drug overdose death (NVSS-M-DO) includes information on specific drugs involved in overdose death. Each of these three sources has limitations such as identification of specific opioids and inclusion of deaths occurring outside of a hospital setting. Prior projects in the OS-PCORTF portfolio including “*Enhancing Data Resources for Researching Patterns of Mortality in Patient Centered Outcomes in Research: Projects 1-4*”, and “*Improving the Mortality Data Infrastructure for Patient-Centered Outcomes Research*” project have addressed improved specificity within each source. Now, the combination of these three data sources will allow researchers to follow patients with an opioid event from presentation at a hospital to death (if applicable), and retroactively analyze previous encounters for more information. The project will produce several data files that will be available to researchers through the NCHS Research Data Center (RDC) network.

Project Purpose and Goals:

This project expands data capacity for PCOR on opioid use by: 1) creating a new research data file with specific opioid names in ED visits, hospitalizations, and deaths; and 2) developing data collection and reporting tools to support research on hospital encounters involving opioids. The overall goal of the project is to improve surveillance and expand researchers' access to data on hospital care patterns and risk factors associated with opioid overdose deaths.

The project objectives are to:

- Create a merged data set for research by linking data between hospital care and mortality by merging the NHCS, NDI, and NVSS-M-DO. Analysis of this data set will result in a report exploring characteristics of individuals who have opioid-related events, patterns of hospital use in months prior to death, and comparison of patients and services.
- Improve researchers' ability to identify opioid-specific hospital encounters and deaths by improving existing techniques and developing new methods in vocabulary and procedure coding.
- Enrich the opioid-specific hospital care data available in the linked NHCS, NDI, and NVSS-M-DO dataset with enhanced hospital and death certificate opioid identification.
- Invest in the infrastructure to improve the collection and reporting of hospital data, and disseminate the methodologies, analyses and knowledge obtained to promote the use of the enhanced data set for PCOR. Researchers will build a secure web-based portal for participating NHCS hospitals to compare submitted data to aggregate data across all hospitals. This portal will allow visualization and analysis of opioid-involved hospital encounters and, by the second year, will implement a notification system for communicating EHR standards for data collection back to hospitals.

Subtasks to enhance the ability to identify opioids in the hospital setting include the creation of an updated 2016 Drugs Mentioned with Involvement (DMI) vocabulary and the application of code-based algorithms to identify opioids in 2016 NHCS encounter diagnosis and procedure codes.

Contributions to PCOR Data Infrastructure Functionalities:

- *Standardized Collection of Standardized Clinical and Claims Data:* CDC aims to link and aggregate data across sources and networks.
- *Improved Use of Clinical Data for Current Research:* This improved source will collect more clinical data, including data relevant to assessing safety, efficacy, and adherence. This project will capture information from a variety of care settings and pathways, more accurately capturing the range of variables that may influence health outcomes.

Accomplishments and Deliverables:

Since work on the project began in April of 2018, the project team has made notable progress on several tasks in support of deliverable completion.

- The 2014 NHCS and NDI linked data was merged with the 2014 NVSS-M-DO file. Analysts are reviewing the merge and creating documentation for the RDC.

- Analysis began on the merged NHCS/NDI/NVSS-M-DO file, this includes developing research questions to understand the numbers of opioids found in hospitals and death certificates.
- The first Technical Expert Panel (TEP) was held on September 10, 2018.

Disseminated Products:

- This project was presented at the 2018 Academy Health Annual Research Meeting on June 25, 2018. During the presentation, "Using Linked Data and Natural Language Processing to Support Patient-Centered Outcomes Research on Opioids," the speaker solicited input on ongoing PCOR projects that apply natural language processing to linked EHRs and death certificates, and identify specific opioid agents leading to ED use, hospitalization, and death.¹³

Coordination with other Federal Agencies:

NCHS will collaborate with Substance Abuse and Mental Health Services Administration (SAMHSA), FDA, National Institute on Drug Abuse (NIDA), and ASPE in the development of identification algorithms and dissemination.

Period of Performance:

4/15/18 - 4/14/20

Federal Point of Contact:

Carol DeFrances

Improving the Mortality Data Infrastructure for Patient-Centered Outcomes Research

The timely reporting of death information is critical for public health surveillance and PCOR. Composed of all U.S. mortality events since 1979, the NDI database allows researchers to match death entries in the NDI with their research populations (e.g., those participating in longitudinal clinical and epidemiologic studies) to determine both death "status" and cause of death. A significant challenge with the NDI is the lag between the occurrence of an individual's death and the availability of their record for matching purposes. This has limited the NDI's utility for timely patient follow-up and survival outcomes determination.

Project Purpose and Goals:

The CDC's NCHS is working to enhance the mortality data infrastructure on multiple fronts, including improving the timeliness of state and local death records (which provide cause of death) being incorporated into the NDI database, and obtaining a more complete picture of patient care by linking NDI records with nationally collected hospital datasets.¹⁴

In this project, the CDC engaged in three inter-related tasks to enhance the national mortality system:

- *Strengthening existing state mortality data collection infrastructure.* Currently, 48 jurisdictions have Electronic Death Registration Systems (EDRS); however, not all are statewide and they do not cover all death events in a given region. This project envisioned a comprehensive EDRS network that enables electronic data transfer of mortality data to NCHS. The intended outcomes were for states to report 80 percent of deaths electronically within 10 days of event to NCHS to enable more timely availability of clinical data for research; and to transmit deaths related to state-specified causes of deaths to the State epidemiologists within 1 day of the registration of the death certificate or receipt of the specified cause.

- *Conducting inter-system exchanges between EHRs and an EDRS in 1 state using draft national Health Level Seven International (HL7) standards.* Integrating EHR systems with vital records systems provides another opportunity to improve quality and timeliness of mortality data collection and distribution through standardization and reduction of duplicative data entry. The CDC used the HL7 standards to support the bi-directional exchange of mortality data from Epic EHRs to California's EDRS, and the EDRS to NCHS. The exchange of data in this format demonstrated the feasibility of implementation, and if implemented more broadly will support improvement in the quality and timeliness of mortality data, clinical care assessments, and PCOR.
- *Piloting the linkage of the NCHS inpatient and emergency department (ED) data with the NDI.* The pilot assessed the feasibility, validity, and reliability of linking NCHS data on inpatient and post-discharge mortality to the NDI with the goal of enhancing the data available for clinical research. This linkage provides additional information on patient survival after hospitalization and emergency care to better allow researchers to identify patterns of care, and individual and provider characteristics that may be associated with in-hospital and post-discharge mortality.

Contributions to PCOR Data Infrastructure Functionalities:

- *Standardized Collection of Standardized Clinical Data.* The project modified state systems to enable the use of HL7 messages and piloting standards to govern the transmission of death-related information from EDRS to the NDI. The result was better quality data from state-level systems, better electronic reporting, and therefore more complete mortality data for researchers to easily utilize standardized components. This included publishing the HL7 Version 2.6 Implementation Guide: Vital Records Death Reporting, Release 1 - US Realm to enhance standardization.
- *Linking of Clinical and Other Data for Research.* CDC's work also supports the enhancement of strategic publicly-funded data systems by linking NHCS data (inpatient and ED) with NDI data to facilitate analysis of trends in hospital mortality. In addition, the improved linkage between state-level EDRS to the NDI will improve mortality reporting and the data's completeness data for research purposes.

Accomplishments and Deliverables:

The CDC NCHS project successfully addressed three objectives:

Objective 1: Strengthen mortality data infrastructures of the states and the NCHS/Division of Vital Statistics' (DVS) for more timely delivery of state records to the NDI database. Nineteen states received awards to improve their transmission of mortality data, overcoming challenges primarily related to: stakeholder engagement, the need to enhance technology, staff training, and making upgrades to internal systems and processes. Despite challenges facing local and state vital records office, there is evidence of substantial improvement and a foundation for continued improvement: almost every jurisdiction's timeliness improved; and 10/19 awarded states exceed the national average of 52 percent; reporting 73 percent of their mortality records within 10 days and the total average of all 19 awarded states was 55 percent; again exceeding the national reporting average.

Objective 2: Conduct inter-system exchanges between EHRs and an EDRS in 1 state (California) using draft national HL7 standards. This resulted from a partnership between NCHS, DVS, the California Department of Public Health and its vendor for mortality reporting (University of California Davis

Health System), and the EHR vendor Epic, who collaborated on the necessary interface, testing, and piloting.

Objective 3: Link the NHCS inpatient and ED data with the NDI data to measure within and post-hospital mortality. Successful linkage included securing essential support from NHCS sampled hospitals to participate in the effort; collecting and linking the NHCS and NDI data, including the storage and analysis of data in compliance with policies that regulate the governance of data; and conducting a matching process via a rigorous scientific methodology that ensured the fidelity of process. This demonstrated both the feasibility and mechanism to advance the linking of mortality data to further support opportunities for research to be conducted, comparing patterns of care in multiple healthcare settings.

The project findings demonstrate successful technical and non-technical mechanisms to improve the timeliness and quality of state EDRS at the state and local level, given the importance of mortality as an outcome to clinical research and PCOR.

Disseminated Products:

- Established a Medical Examiner and Coroner Directory to support local engagement with these stakeholders
- HL7 Version 2.6 Implementation Guide: Vital Records Death Reporting, Release 1 - US is available here: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=209
- Published the “Electronic Death Reporting System Online Reference Manual” as a resource guide for jurisdictions to support development and sustainability of EDRS. This manual includes an analysis and reporting on timeliness success factors, including the identification of critical factors that lead to success. The online manual is available here: <https://www.cdc.gov/nchs/data/dvs/edrs-online-reference-manual.pdf>

Period of Performance:

4/3/15 - 7/1/18

Federal Point of Contact:

H. Mac McCraw

Strengthening the Data Infrastructure for Outcomes Research on Mortality Associated with Opioid Poisonings

Cause of death information from death certificate data is often used by researchers and those in public health for programmatic, policy, and PCOR research needs. During this project, the Vital Statistics Rapid Release (VSRR) and Medical Mortality Data System (MMDS), both subsets of the National Vital Statistics System (NVSS), will be redesigned and enhanced to improve the quality of death information data and release. The redesign of the MMDS to code and process a larger percentage of death certificate records including deaths involving opioids will improve timeliness and accuracy of data thus improving the quality of data available to researchers. The enhancement of the VSSR will capture a broader array of geographic and demographic data in records.

The MMDS codes all death records in the United States. Developed in the 1980s, it uses algorithms to assign underlying cause of death, multiple cause of death, and other fields using data inputs from the medical and demographic portions of the death certificate. Seventy-nine percent of all records are coded electronically, but only 33 percent of records with a drug overdose death are coded electronically. Additionally, almost 15 percent of death certificates do not specify the drugs involved in the death. Previous research has shown that the quality of data on death certificates improved when the physician

completes the certificate using their EHR.¹⁵ Medical examiners and coroners typically certify death in the case of drug overdose, so integrating their case management systems used by coroners with state EDRS is expected to improve the quality of data for the deaths they certify.

The literal text portion of the death certificate contains information about drugs that caused or contributed to death. Software developed by the NCHS and the Food and Drug Administration (FDA) will be used as an enhanced prototype to strengthen the MMDS coding process through collaboration with the NLM. The enhancements will identify drug information found in the literal text field, assign the literal text data to drug vocabularies and classifications used in the research community, and include the coded supplemental information in NVSS' restricted-use multiple cause of death mortality files (NVSS-M) and the NDI.

The VSRR Program currently supports drug overdose death surveillance through quarterly estimates of national death rates for leading causes of death and monthly counts of drug overdose deaths by state for a limited set of drugs identifiable from ICD-10 codes. The data system used to produce VSRR reports will be enhanced to include geographic, demographic, and drug details in the death, information not currently captured by provisional monthly and quarterly releases, as well as to automate the production of standard and ad hoc VSRR reports.

Project Purpose and Goals

The overall goal of this project is to strengthen mortality data infrastructure for outcomes research on deaths associated with opioid poisoning. This project will also lay the foundation for research on other causes of death in the future.

The project objectives are to:

- Redesign the MMDS to electronically code and incorporate specific drug information captured in the literal text fields of death certificate records.
- Incorporate supplemental drug information from the literal text fields of death certificate records, especially information related to deaths involving opioids, as new variables in the NDI and the NVSS-M.
- Annually produce the NDI and the NVSS-M data files containing the supplemental information for deaths involving drugs such as opioids for use by approved researchers.
- Improve the specificity of drug information on death certificates supplied by states by developing and pilot testing a FHIR[®] API for the exchange of information from between medical examiner and coroner case management systems and state EDRS.
- Improve the depth and timeliness of national reporting on drug deaths involving opioids by re-architecting the data system to produce and release more in-depth information about drug overdose data (e.g., specific drugs, demographic information) on a monthly basis for public health surveillance and research.
- Establish an advisory committee of the NCHS Board of Scientific Advisors to align changes in the mortality data system with end-users' (i.e., researchers') needs.

Contributions to PCOR Data Infrastructure Functionalities:

- *Use of Enhanced Publicly-Funded Data Systems for Research:* This project will focus on improving the quality, availability, and timeliness of mortality data for use in research on drug-involved mortality, including that of opioid deaths.

Accomplishments and Deliverables:

In the first three months of this project, the team made a great deal of progress in setting up the systems, contracts, and processes needed to achieve the project goals.

- Various machine learning models for the coding of the causes of deaths have been tested. A hybrid model using rule-based coding and machine-learning coding has shown to be the most effective. In pilot testing, the hybrid model was able to identify and accurately code 2,000 of the most frequently coded ICD-10 codes for underlying cause of death with 88.7 percent accuracy.
- Contracts were awarded to six vital registration jurisdictions (CA, FL, GA, MI, NH, NY) to participate with the Implementer's Group, a set of states specifically focused on establishing data interoperability between medical examiner/coroner offices, the vital registration offices and at least one state public health surveillance system.
- A contract was awarded to support a statistician and a database programmer to work on the effort to restructure the provisional data system for the VSRR program.

Disseminated Products:

- Presented "Future NCHS-Linked Data: Using Linked Data and Natural Language Processing to Support Patient-Centered Outcomes Research on Opioids" at the Academy Health 2018 Annual Research Meeting and at the NCHS Board of Scientific Counselors meeting to ensure alignment and utility of mortality infrastructure improvements with end-users' needs.
- DVS obtained support from communications specialists within CDC to disseminate information about our efforts to modernize death reporting. Website stories written during the current quarter include:
 - "Drugs, Death, and Data" available here: <https://www.cdc.gov/surveillance/blogs-stories/drugs-death-data.html>; and
 - "Project - Improving the Quality and Timeliness of Data on Drug Overdose Deaths" available here: <https://www.cdc.gov/surveillance/projects/improving-data-on-drug-overdose-deaths.html>.
- NCHS released the monthly provisional drug overdose death counts on July 11, 2018, August 15, 2018, and September 12, 2018. The provisional data presented include: (a) the reported and predicted provisional counts of deaths due to drug overdose occurring nationally and in each jurisdiction; (b) a U.S. map of the percentage changes in provisional drug overdose deaths for the current 12 month-ending period compared with the 12-month period ending in the same month of the previous year, by jurisdiction; and (c) the reported and predicted provisional counts of drug overdose deaths involving specific drugs or drug classes occurring nationally and in selected jurisdictions.

Period of Performance:
5/1/18- 5/14/20

Federal Point of Contact:
Kate Brett

VI. Centers for Medicare and Medicaid Services (CMS)

CMS is administering three active OS-PCORTF-funded projects (Exhibit 5).

Exhibit 5. CMS Active Projects

CMS-Funded Projects
Improving Beneficiary Access to Their Health Information through an Enhanced Blue Button Service
Technologies for Donating Medicare Beneficiary Claims Data to Research Studies*
Enhancing Data Resources for Researching Patterns of Mortality in Patient Centered Outcomes Research*

* Denotes a cross-agency funded project that involves more than one federal agency; these projects are described in the “Cross-Agency Funded Projects” section.

Improving Beneficiary Access to Their Health Information through an Enhanced Blue Button Service

The original CMS Blue Button service established in 2010 to allow CMS beneficiaries to download their data via MyMedicare.gov. However, the download functionality supported in the Blue Button beneficiary portal presented patient data in formats (e.g., plain text and PDF) that made it difficult for patients to use and share their data with applications, researchers, and other service providers they trust.

CMS undertook a conceptual redesign of Blue Button to be a “data as a service” platform, allowing data access on demand. To accomplish this redesign, CMS utilized the HL7 FHIR® standard.

Project Purpose and Goals:

The enhanced Blue Button service, Blue Button 2.0, is an application programming interface (API) that enables CMS beneficiaries to connect their Medicare Parts A, B and D claims data to applications and services they trust, including research platforms. CMS Blue Button API uses HL7 FHIR®, a standard for exchanging healthcare information electronically. FHIR® is designed to simplify implementation without sacrificing information integrity. CMS utilizes FHIR® for Blue Button 2.0 in order to ensure that data are in a structured format and easily able to integrate with other applications and services. Research study investigators can direct interested beneficiaries to a web page or app that guides the beneficiary through the process of granting permission to investigators to access their Medicare Parts A, B, and D data. This dramatically simplifies acquisition of beneficiary claims information to support participation in clinical research studies. The Blue Button API allows researchers to selectively pull individual beneficiaries’ data for specific research needs. For example, queries could be constructed to access all Part D medication claims or all claims occurring since the last data request.

“As of September 2018, 1,200 developers and 785 unique organizations joined the Developer Preview program for CMS Blue Button 2.0.”

The project objectives were to:

- Develop a Blue Button API 2.0 and publish the code as open source software
- Launch full production of the Blue Button 2.0 API
- Promote the availability of Blue Button 2.0 API to external sources

Contributions to PCOR Data Infrastructure Functionalities

- *Use of Enhanced Publicly-Funded Data Systems for Research.* This project focused on facilitating the access, use, and retrieval of CMS data by beneficiaries to support their participation in research.

Accomplishments and Deliverables

The Blue Button API, Blue Button 2.0, officially launched in March 2018 and organizations are now able to request access to the production API, which will allow Medicare beneficiaries to connect their claims data to the health applications they choose to trust. Blue Button 2.0 contains four years of Medicare Part A, B, and D data for 53 million Medicare beneficiaries. The data contains a variety of information about a beneficiary's health including type of Medicare coverage, drug prescriptions and primary care treatment and cost.

The Blue Button project is comprised of two platforms: a front-end platform to manage consumer and developer access using OAuth 2.0 protocols for authorization and a back-end platform providing a standard FHIR® database to provide claims information in the FHIR® Explanation of Benefit Resource format.

The project initiated the developer application vetting process and on-boarded the first cohort of Developer Applications into the production environment. The CMS Blue Button API developer preview environment with synthetic data is available at <https://bluebutton.cms.gov>. A prototype with synthetic data has been available in Amazon Web Service (AWS) since April 2016. In December 2017 this data set was replaced with a newer synthetic data set that was certified as de-identified data and represented a pool of 30,000 synthetic beneficiaries with realistic claims information available via Patient, Coverage and Explanation of Benefit resources. As of September 2018, there are 1,200 developers and 785 organizations that have joined the Developer Preview program for CMS Blue Button 2.0.

In August 2018, CMS hosted the first-ever Blue Button® 2.0 Developer Conference in Washington, D.C. The conference brought together developers to learn, build software, and share insights on how to leverage Medicare claims data to improve health outcomes. The event included hands-on sessions for the technical community to create or build upon their Blue Button 2.0 application and lighting talks on topics like FHIR® and OAuth 2.0.

In June 2018, Blue Button 2.0 was honored as a winner at the Fourth Annual FedHealthIT Innovation Awards for Federal Programs. The Blue Button 2.0 API Team also received an Administrator's Achievement Award for excellence in creating the technology and infrastructure that enables CMS beneficiaries to be the true owners of their healthcare data.

Disseminated Products:

- The CMS Blue Button API developer environment is open to all interested parties and can be accessed via the CMS webpage at <https://bluebutton.cms.gov>.
- Open source codes published to the Blue Button repositories on GitHub: <https://github.com/CMSgov>. Note: codes are updated on a continuous basis.
- Project team members participated in a series of HL7 FHIR® Connect-a-thons and Health Datapalooza code-a-thons to advance the development work of the Blue Button API.

- The project team presented “The Power of Consumer-Directed Data” presentation at Health Datapalooza in Washington, D.C. (April 2017).

Period of Performance:
4/30/16 – 4/29/18

Federal Point of Contact:
Lori Pettebone-Koraganie

VII. Food and Drug Administration (FDA)

FDA is administering nine active OS-PCORTF-funded projects (Exhibit 6).

Exhibit 6. FDA Active Projects

FDA-Funded Projects
Collection of Patient-Provided Information through a Mobile Device Application for Use in Comparative Effectiveness and Drug Safety Research
Cross-Network Directory Service
Source Data Capture from Electronic Health Records: Using Standardized Clinical Research Data
Standardization and Querying of Data Quality Metrics and Characteristics for Electronic Health Data
Utilizing Data from Various Data Partners in a Distributed Manner
Developing a Strategically Coordinated Registry Network to Support Research on Women’s Health Technologies*
Development of a Natural Language Processing Web Service for Public Health Use*
Harmonization of Various Common Data Models and Open Standards for Evidence Generation*
Enhancing Data Resources for Researching Patterns of Mortality in Patient-Centered Outcomes Research*

* Denotes a cross-agency funded project that involves more than one federal agency; these projects are described in the “Cross-Agency Funded Projects” section.

Collection of Patient-Provided Information through a Mobile Device Application for Use in Comparative Effectiveness and Drug Safety Research

A critical component of the FDA’s mission is to monitor the safety and effectiveness of the products it regulates. The FDA has developed a robust national electronic surveillance system, Sentinel, for monitoring FDA regulated medical products. Sentinel is a distributed research network, using existing electronic health care data from multiple sources to support FDA surveillance activities. In a distributed network, data are held locally at each data partner site and queried with analytics developed at a coordinating center. Currently, Sentinel is comprised largely of administrative and claims data from health insurance plans.¹⁶ PCORnet, another distributed network developed by the Patient-Centered Outcomes Research Institute (PCORI), contains electronic health record data from 13 Clinical Data Research Networks as well as administrative and claims data from two Health Plan Research Networks.¹⁷

With a heightened priority on patient-centered care, FDA saw an opportunity to develop and pilot a mobile application to capture the patient’s perspective and link it to established electronic health data available

through Sentinel, addressing a gap in the information currently collected for the purpose of medical product safety surveillance and CER. Patient reported information is especially valuable to PCOR as it generally provides information on the patients' experiences of symptoms and on what values are most important to them.¹⁸ In addition, the FDA is currently evaluating the role of Real World Evidence for regulatory purposes, and clinical trials often have endpoints such as functional status scales that depend on patient input. This data linkage can support multi-site trials, registries, and prospective observational studies. Additionally, the mobile application provides a mechanism for data collection that is significantly less time and resource intensive than traditional data collection with research cohorts.

Project Purpose and Goals:

This project created the infrastructure for collecting data from patients on medical product exposures, outcomes, risk factors and confounders through a mobile device application, allowing patient-generated data to be linked with a single data partner that participates in the Sentinel and PCORnet distributed networks. The project developed and piloted a mobile application to capture these types of data from pregnant women, who represent a traditionally understudied population and for whom significant gaps in research on drug safety exist. Examples of such information gaps include, but are not limited to, adherence to prescription medications or therapies, health outcomes that are not medically attended, and characteristics which are inconsistently recorded in electronic health data such as illicit drug use, tobacco use, vitamin and supplement use, race, socio-economic status, educational attainment, and over-the-counter medication use. As a result of the app, researchers were able to query both the new PGHD and the data routinely captured by the Sentinel data partner.

The project objectives were to:

- Leverage existing electronic health record algorithms to identify pregnant women in one collaborating health plan (“data partner”).
- Develop a generalizable HHS/FDA/Sentinel Operations Center mobile device application that can be transmitted to the pregnant women identified by the data partner.
- Develop Medical/Epidemiological data elements and interface to capture pregnancy exposures, outcomes, confounders, and risk factors, as well as a programming and hosting component.
- Link data provided by patients through the application with data from the health plan (data partner).

This effort served as a pilot for future applications. For example, the web-based interface has a customizable design; therefore, the mobile application can be re-configured to collect information from another patient population or a disease cohort so that other researchers or HHS agencies could answer other research questions of interest.

Contributions to PCOR Data Infrastructure Functionalities:

- *Collection of Participant-Provided Information.* This project supported the collection of PPI functionality by creating a secure mobile application to collect PPI from pregnant women and the app can be reconfigured for expansion to other patient populations.
- *Linking of Clinical and Other Data for Research.* This project supported this functionality demonstrating the integration of PPI information into the Sentinel research network simultaneously enabling integration into PCORnet.

Accomplishments and Deliverables:

In the section that follows, the report describes the FDA mobile application pilot and present opportunities to leverage these tools for PCOR now that the project has concluded.

The FDA project successfully pilot tested the mobile application, FDA My Studies App, to examine medication use and health outcomes of pregnant women with a data partner participating in both Sentinel and PCORnet. The FDA developed two apps using Apple's ResearchKit and Android's ResearchStack API frameworks. The app used a questionnaire developed by an expert workgroup convened by FDA to collect patient-provided information of interest to the study. The app enables patients to enroll and enter data on their iOS or Android mobile device, providing researchers with an invaluable resource to study desired populations' behaviors and outcomes.

Throughout the pilot, participants reported reasons for their various actions including discontinuing certain medications, extending over-the-counter medication use, and illicit drug use. The study was able to successfully demonstrate the technical, scientific, and governance procedures of the application and secure data storage solution in a distributed clinical research environment. With the goal of widely disseminating the results of this project, the developed software is publicly available on GitHub. As the MyStudies system is Federal Information Security Management Act (FISMA) compliant and capable of supporting a 21 CFR Part 11 compliant study, it is available for implementation by external research organizations and app developers. The mobile app and web configuration portal are equipped with an API layer so they are available to be used with different storage environments. Researchers also identified options for use in either an existing single database or an existing distributed database system.

In the context of a drug effectiveness and safety study, the app was able to quickly engage with the study population of women in their first trimester of pregnancy using data from the EHR, and capture sensitive information including continued alcohol, smoking, and illicit drug use during pregnancy. One of the cutting-edge aspects of this app is its ability to interact with the regulated and commonly distributed clinical research environment. This project was able to construct the first platform that enables research organizations to control almost all aspects of multiple smartphone-based studies at multiple clinical sites through a configurable interface. This includes the ability to collect and store data via a point and click web-based portal and single mobile app. The FDA sees value in continuing to develop the app, as it will likely enhance usability, allow for integration of other real world data sources, and increase the overall ease of including a patient's perspective in clinical research. In addition, as the source code is available in the public domain, external app developers will be able to contribute to added functionalities.

FDA is undertaking a follow-on demonstration project supporting two disease registries as well as an embedded clinical trial. In addition, a company that conducts real world trials and studies using disease-specific cohorts used public code from FDA to spin up the MyStudies app and Web Configuration Portal within days of public posting. It is currently exploring a potential real world use case to support value-based care using the mobile app to measure PROs related to drug products.

Disseminated Products:

- The final report is available here:
<https://aspe.hhs.gov/system/files/pdf/259016/FDAMyStudiesAppAPatientCenteredOutcomes.pdf>
- The source code and accompanying technical documents are publicly available on GitHub here:
<https://github.com/PopMedNet-Team/FDA-My-Studies-Mobile-Application-System>

- Mobile App and Web Configuration Portal Demonstration by Zac Wyner at the American Medical Informatics Association in San Francisco, California, on November 6, 2018
- Presentation by David Martin on Mobile Technologies – Opportunities and Regulatory Considerations, at the 34th International Conference on Pharmacoepidemiology and Therapeutic Risk Management in Prague, Czech Republic on August 24, 2018
- David Martin presented the FDA My Studies App within a larger presentation regarding real world data and evidence at the yearly European Commission/European Medicines Agency –US FDA Bilateral Meeting on June 18, 2018. The title of the presentation is “Real World Data and Evidence: “Deep Dive” updates regarding Drug Development and Post-Marketing Monitoring”
- David Martin and Zac Wyner presented the project to the Biologics and Biosimilars Collective Intelligence Consortium (BBCIC) on February 6th, 2018. BBCIC is considering the utility of the app for biologics and biosimilars studies
- Presented “Developing a Mobile App for Studies of Medication Safety” at the International Conference on Pharmacoepidemiology & Therapeutic Risk Management (August 2017). The presentation is available here: <https://www.sentinelinitiative.org/sites/default/files/Sentinel-ICPE-2017-Presentation-Mobile-App.pdf>
- Presented “Engaging Patients in Evidence Generation” at the National Academy of Medicine’s Clinical Effectiveness Research Innovation Collaborative (April 20, 2017). The presentation is available here: <https://nam.edu/wp-content/uploads/2017/04/Martin-NAS-042917.pdf>
- A poster was presented at the AMIA Joint Summits on Translational Science titled, “Collection of Patient Provided Information through a Mobile Device Application for Use in Medical Product Safety Surveillance” in March 2017.

Period of Performance:

7/15/18 – 6/30/18

Federal Point of Contact:

David Martin

Cross-Network Directory Service

The FDA was tasked with developing and implementing a Cross-Network Directory Service (CNDS) that addresses the stand-alone nature of existing distributed research networks and barriers to working across these networks. Distributed research networks facilitate large scale comparative and effectiveness studies by allowing researchers to send data queries to multiple organizations and networks, while those organizations being queried maintain possession and protection over the data they house. This project extended PopMedNet™ to support query creation and execution for cross-network queries. PopMedNet™ is a distributed research network platform for query composition and distribution. Numerous research networks already use this open source data model. CNDS enhances network scalability and allows each data partner to determine their own level of participation and governance rules.

Project Purpose and Goals:

This project created an open source interoperable service that allows data partners to participate in multiple data research networks, query across the networks, and share analytic capabilities and knowledge across networks. The project was piloted across two existing networks: FDA’s Sentinel System and PCORI’s national PCORnet.

The project objectives were to:

- Identify the key functionalities and an overarching technical design for a CNDS
- Develop and test a detailed design for the CNDS with at least two existing distributed research networks
- Conduct additional analyses of the robustness of the CNDS and produce user materials

This project produced the following deliverables: CNDS Software and Source Code; a report on the Failure Mode and Effects Analysis; technical documentation; and user documentation. Additionally, the information gained from the CNDS will help inform additional FDA projects, including the *Standardization and Querying of Data Quality Metrics and Characteristics for Electronic Health Data* project.

Contributions to the PCOR Data Infrastructure Functionalities:

- *Use of Enhanced Publicly-Funded Data Systems for Research.* This project built and implemented a CNDS that supports a scalable data network infrastructure to allow integration of HHS investments in health data networks. The resulting CNDS enables data partners to participate in multiple data research networks, seamless querying across networks, and sharing of analytic capabilities and knowledge across networks. Ultimately, the knowledge gleaned and tools developed from this project will inform other FDA initiatives and research.

Accomplishments and Deliverables:

All deliverables for this project have been successfully completed. Three reports were posted on the Sentinel website, Design and Technical Documentation, User Documentation, and the Final Report. Documentation of the design phase of this project process is reported in the User Documentation report. This report provides business and technical requirements of CNDS including a description of components including: 1) registration, which is the service that captures the information that informs the discovery process, 2) discovery, the functionality that enables users to find potential collaborators and data sources, 3) communication, the functionality that distributes queries across networks, and 4) governance, which supports the enforcement of the rules of use. The document highlights how users can create organization, investigator, and data source profiles, specifying visibility restrictions profile metadata. Users are then able to search within the metadata to facilitate potential collaboration with other users and organization.

The CNDS software application has been designed, built, released, and tested. Beta testing of the software application was performed by a Sentinel and a PCORnet data partner. In the first round of beta testing, the data partners successfully registered and entered their metadata. In the second and final round of testing, the data partners successfully completed a round trip through Discovery (completed a search of data sources) and Communication (were able to send and receive queries). Further, data partners were not able to discover data that the other partner did not actively make available. In early 2018, the project team prepared the technical and user documentation for the CNDS.

The “Cross Network Directory Service Project Design and Technical Documentation” (version 1.0) was posted on the Sentinel Initiative website on January 31, 2018. This report provides a detailed description of the architectural design and software requirements, highlighting the core concepts of system entities and metadata as well as the four user facing functions (registration, discovery, communication, and governance.).

A final technical report was published to the Sentinel website in September 2018. This report highlights similar concepts of the previous two reports, and provides an overview of the project's goals, objectives, and the accomplishments for each phase of the project. The final report highlights lessons learned, most of which are related to the metadata as well as areas of focus for future work. The project noticed that workgroup members had difficulty distinguishing data versus metadata. In future work, they hope to make differences more concrete. Additionally, the team believes they could benefit from the development of a taxonomy and robust metadata dictionary, which they will consider for future work.

Disseminated Products:

- "Cross Network Directory Service Final Report" was published on the Sentinel website on September 18, 2018. It can be found here: https://www.sentinelinitiative.org/sites/default/files/Methods/CNDS_Final_Report_v10.pdf and on ASPE's website here: https://aspe.hhs.gov/system/files/pdf/259016/CNDS_Final_Report.pdf
- The source code and documentation on how to implement the code in an existing instance of PopMedNet™ is publicly available on GitHub here: <https://github.com/PopMedNet-Team/cnds>
- Version 1.0 of the "Cross Network Directory Service Project Design and Technical Documentation" was published on the Sentinel website on January 31, 2018. It can be found here: https://www.sentinelinitiative.org/sites/default/files/Methods/CNDS_Design_Technical_Documentation.pdf
- Version 1.0 of the "Cross Network Directory Service Project User Documentation" was published on the Sentinel website on January 31, 2018. It can be found here: https://www.sentinelinitiative.org/sites/default/files/Methods/CNDS_User_Documentation.pdf
- "Cross Network Directory Service: A Socio-technical Platform to Enable Meaningful Collaboration across Organizations" was accepted as an oral presentation at the Health Care Systems Research Networks Conference in Minneapolis (April 12, 2018) and as a poster at the AMIA 2018 Informatics Summit (March 12-15, 2018)
- "Cross-Network Directory Service: Infrastructure to Enable Collaborations Across Distributed Research Networks" accepted for publication in the Learning Health Systems journal

Period of Performance:

7/15/15 – 6/30/18

Federal Point of Contact:

Michael Nguyen

Source Data Capture from Electronic Health Records: Using Standardized Clinical Research Data

The current information systems and data models that define both clinical care and clinical research are disparate in many ways. Thus, using data from EHRs and electronic data capture (EDC) systems necessitates a detailed understanding of each data source. Moreover, even data sources within the same database can have extensive differences in coding, schema, format, and usage.¹⁹ Given that clinical data are mostly collected through EHRs, a guideline or protocol for the use of health care data in clinical research and trials could be helpful to enable PCOR studies. In 2013, FDA published guidance on Electronic Source Data in Clinical Investigations,²⁰ which provides recommendations on the capture, review, and retention of electronic data in FDA-regulated clinical trials. To provide a working example of the FDA's recommended approach, the FDA has initiated the OneSource project, which will provide a real

world demonstration of their guidelines and commitment to EHR-to-EDC in clinical research environments.

Project Purpose and Goals:

This project will demonstrate an approach to collecting data for clinical trials that populates an EDC system directly from an EHR systems in an FDA-compliant way using the RFD standard, HL7 CCD, Integrated Health Enterprise (IHE), FHIR®, and Clinical Data Interchange Standards Consortium (CDISC) standards. The demonstration involves a collaboration with the University of California at San Francisco (UCSF) as part of a phase 3 trial in breast cancer entitled “I-SPY 3 TRIAL” (Investigation of Serial Studies to Predict Your Therapeutic Response with Imaging and moLecular 3). The I-SPY 3 study is a large-scale trial that will use adaptive approaches to simultaneously test several drugs for the treatment of breast cancer using biomarkers and involving the collection of electronic patient reported outcomes (ePRO). PCOR stakeholders will be provided with a cloud-based Health Insurance Portability and Accountability Act (HIPAA)-compliant and 21 C.F.R. Part 11-compliant tool to seamlessly integrate into any EHR and EDC systems. This tool will enable PCOR stakeholders (i.e., health care providers) to track patient data at the point of care in their EHRs²¹.

The project objectives are to:

- Demonstrate an end-to-end, EHR-to-EDC, standards-based technology solution, OneSource, for the capture and transmission of regulated clinical research data.
- Assess the utility of the standards-based technology solution processes to facilitate FDA activities related to inspection, reconstruction, and evaluation of the trial/clinical investigation.
- Assess the impact of the standards-based technology solution on current workflow processes at the sites, and on FDA’s remote monitoring of studies.
- Develop guidelines for future implementations in both health care and clinical research.
- Provide recommendations for the improvement of existing standards and implementation guides to standards development organizations.
- Develop a general framework (technologies, processes, policies, governance, and standards) for the capture and use of electronic structured data in regulated clinical trials.

Contributions to PCOR Data Infrastructure Functionalities:

- *Standardized Collection of Standardized Clinical Data.* The project supported standardization through: the development of a set of research CDEs in specific gaps areas and the development of a governance structure for CDEs; and the development of a core set of standards for the collection and integration of prevalent use cases of PPI for PCOR.
- *Use of Clinical Data for Research.* The project also supports the use of EHR and PRO through the implementation of standards for the collection and use of this data in FDA-regulated clinical trials.

Accomplishments and Deliverables:

The FDA has made significant progress towards demonstrating the EHR-to-EDR data capture:

- The RFD and CCD technical standards to capture data have been piloted.
- The RFD and CCD technical standards are maintained through updates to the technical standards, implementation guides, reference implementation documentation, and testing documentation.
- The project team has defined quality improvement metrics, including key metrics such as data fidelity and performance, in particular time and projected cost savings compared to the standard I-SPY 2 workflow requiring manual abstraction and entry into eCRFs.
- The project team has completed the evaluation of OpenClinica (an EDC vendor) as a future eSource Electronic Data Capture platform. Clinical Data Interchange Standards Consortium (CDISC), a standards developing organization, plans to leverage UCSF findings from OneSource in their development of oncology clinical trial standards.

Disseminated Products:

- A podium presentation at the 2018 AMIA Annual Symposium on eSource to OneSource: Process and Quality Improvements in the I-SPY 2 TRIAL
- A presentation to the Digital Data & Health Therapeutics Group for Production implementation.
- A presentation to the University of California Office of the President, demonstrating how OneSource could serve as a platform across the University of California System
- OneSource was presented as part of the University of California San Francisco (UCSF) Share Case IT Symposium “Data Integration Real World Impact & USCF – Integration Overview” on October 12, 2017.
- A presentation as part of the FDA panel, “Use of Electronic Health Records (EHRs) as eSource in Clinical Investigations”, at the 2017 annual DIA conference in Chicago, Illinois
- A presentation to the UCSF Orthopedics group as part of the Health Innovation & Technology in Ortho Committee (HITO) on June 14th, 2017
- At I-SPY consortium’s workshop/retreat (which focus on innovation in breast cancer clinical trials) with FDA participation (March 7-8, 2017)

Coordination with Other Federal Agencies:

The team collaborated on a report detailing recommendations in collaboration with CDISC for the improvement of existing standards. The team also provided project presentations to the CERSI and the Center for Drug Evaluation and Research (CDER) Health IT Board.

Period of Performance:

9/1/16 – 12/30/18

Federal Point of Contact:

Mitra Rocca

Standardization and Querying of Data Quality Metrics and Characteristics for Electronic Health Data

One of the challenges in using data from EHRs is that no standards exist for describing the quality and completeness of electronic health data. Understanding the characteristics of a data source is critical for investigators in their determination about whether the data are fit for a specific use. Effective use of the growing number of data sources and distributed networks will require adoption of a uniform approach to describing the quality characteristics of electronic health data, as well as the data capture characteristics at the institutional, provider, and health plan level and data domain level. FDA is supporting a project to develop, test, and implement a standards-based approach to describing data and presenting data quality metrics.

Project Purpose and Goals:

This project will create and implement a metadata standards data capture and querying system for: 1) data quality and characteristics; 2) data source and institutional characteristics; and 3) “fitness for use”.

The project objectives are to:

- Develop metadata standards and technical specifications for implementation
- Develop and test the metadata standards data capture and querying approach
- Implement the new metadata standards in at least two distributed networks
- Incorporate the standards in an open source software release

Contributions to PCOR Data Infrastructure Functionalities

- *Use of Clinical Data for Research.* Effective use of the growing number of electronic health data sources and distributed data networks—by researchers and across the federal government—will require adoption of a standardized approach to describing the quality and general characteristics of these data, as well as the information related to how such data are captured, stored, and maintained. This project supports the development and testing of metadata standards that describe data quality to determine fitness of use for particular research purposes.
- *Use of Enhanced Publicly-Funded Data Systems for Research.* The project will create an open source tool that can be implemented with any data model ensuring broad applicability. With this work, the FDA team aims to provide a harmonized approach to data characterization across multiple sources so in order to better understand candidate data sources before analyzing them. This will increase research planning efficiency through enhanced understanding of a data source before putting resources into conducting study, asking a question, or distributing complex analytic programs. Additionally, the work will help improve the interpretability of analytic results across agencies and research teams.

Accomplishments and Deliverables:

The project team made significant advances toward their objectives in 2018.

- The team documented the ideal end-state and reference implementation designs and developed an initial data model that will hold quality metrics. The model is organized around a central table that captures measurements (counts of patients, maximum or minimum values, frequency of

values, etc.) and which is surrounded by tables that identify for each measurement the source system, the context (patient, member, encounter, claim, etc.), any relevant stratifications (age ranges, e.g.), and other important qualifiers. The model is based on a list of use cases documented by the team, the quality assurance processes of Sentinel and PCORnet, and the Harmonized Data Quality framework by Kahn and colleagues.²² The team concluded work with the data modeler at the end of September 2018. The six distinct models that were developed were combined into one overall model for the project.

- A list of over 100 data quality metrics was documented, which includes metrics of interest for PCORnet, Sentinel, and other electronic health data sources. Metrics of interest include; 1) how many encounters have a discharge date before an admission date; 2) how many rows in the demographics table have no value for rate; and 3) number of prescriptions dispensed per patient year in the source system.
- Work in 2018 also included evaluating the existing FHIR[®] data quality standards, which resulted in the development of a data structure patterned after the FHIR[®] Measure resource. The Measure resource defines the computational structure of the associated clinical quality measure (CQM).²³ The team is exploring implementing a system that can accommodate FHIR[®]-formatted data to populate the data quality metrics database.
- The Discovery and Design phase of the project, which focused on describing the proposed metadata standards, technical specifications for implementing the standards, and a data dictionary, concluded in September 2018. Work products resulting from the Discovery & Design phase include the use cases and the query/analytic tool mock-ups that will be implemented as part of the overall implementation.
- Contracting for implementation work was finalized in October and implementation activities began in October 2018.
- To inform an environmental scan, the project team conducted a literature review on existing data quality methods and standards processes, resources, and standards.

Coordination with other Federal Agencies:

Because of the close relationship between this project and the CNDS project (another OS-PCORTF-funded project), the two share team members within the FDA to improve collaboration and cross-fertilization.

Period of Performance:

8/20/16 - 9/30/19

FDA Federal Point of Contact:

Michael Nguyen

Utilizing Data from Various Data Partners in a Distributed Manner

Information on patients' health care is captured across various data sources held across institutions. Health care information about the diagnosis, treatment, and management of disease and illness is now largely available in electronic format through the claims submitted by health care providers to insurance companies or government payers. Linkage of health care data sources provides more robust cross-sectional and longitudinal patient profiles, which enhances the supplemental uses of this data for research and analysis. However, privacy and security concerns exist regarding the disclosure of sensitive individual-level and institution-level information. This project seeks to develop the capability to conduct

rapid and secure Distributed Regression Analysis (DRA) that alleviates concerns through the use of privacy-protecting analytic methods deployed across a distributed research network.

In a distributed network, data for different people are held at different institutions (*horizontally partitioned*), or information about a single person is held at different institutions (*vertically partitioned*). DRA enables sites to maintain control of patient-level data while generating valid regression estimates across the network without the need to transfer protected health information or conduct meta-analysis. This kind of system provides a balance between analytic requirements, patient privacy and confidentiality, and proprietary considerations. This FDA project builds upon previous work on DRA, creating and demonstrating a new analytic capability to fully automate DRA in horizontally partitioned data in accordance with participating data partners' permissions. The demonstration of this new analytic capability uses PopMedNet™ to automate the multi-step interactive process.

Project Purpose and Goals:

This project will develop and test the functionality to conduct timely and secure DRA in distributed data networks. The identified use case for this pilot test is a comparison of bariatric surgery on several outcomes (e.g., body mass index change and hospitalization one year post-surgery). The project deliverables are an open source software application for use with PopMedNet, documented software testing and results, and the corresponding technical and user documentation to automate DRA in other distributed research networks.

The project objectives were to:

- Develop a new open-source software application that will use PopMedNet,™ that enables the creation, operation, and governance of distributed health networks, to automate DRA query workflow on horizontally-partitioned data; develop this software application so that it can be supported by PopMedNet™ and can be modified and adopted for non-PopMedNet™ applications.
- Test the new DRA in an actual distributed research network.
- Provide technical and user documentation to accompany the new software and allow for its widespread adoption. Explore the feasibility of conducting DRA in which data from the same people are held at different institutions (vertically-partitioned data).

This project produced the following deliverables for horizontally partitioned data: 1) developed and tested code for multiple regression analyses and 2) a technical report.

Contributions to PCOR Data Infrastructure Functionalities:

- *Use of Clinical Data for Research.* The ability to link clinical data across health care organizations' databases provides more robust cross-sectional or longitudinal patient profiles, enhancing secondary uses of electronic health care information for research purposes, and improving access to information that would not be present in a single data source (e.g., claims, registry, or EHR data).
- *Use of Enhanced Publicly-Funded Data Systems for Research.* By developing this solution to work with PopMedNet, DRA has the potential to affect the way research is conducted across disparate research networks (e.g., PCORnet, Sentinel), many of which use PopMedNet™ as their platform, to apply a consistent approach to these types of analytics.

Accomplishments and Deliverables:

The project focused on developing a stable, feasible approach to enable secure distributed linear, logistic, and cox regression analysis within a distributed data network while not requiring sharing of any patient-level datasets from participating data partners. The team found that it is possible to perform automatable, routine DRA in horizontally partitioned DDNs that employ PopMedNet™ as their distributed data-sharing platform. The team successfully implemented a pilot DRA query workflow with select data partners in the Sentinel System. Internal and external tests consistently produced statistically equivalent regression parameter and standard error estimates to those from the pooled individual-level data analysis. External tests with the three select Sentinel Data Partners showed that DRA could be completed in under 20 minutes, excluding the time required to assemble the analytic dataset at each Data Partner.

The project team developed the distributed regression analysis SAS code, which is now publicly available on FDA's Sentinel website. The DRA was fully integrated into the 2017 PopMedNet™ release and the new version of the software is available on www.popmednet.org and on Sentinel's website at <https://www.sentinelinitiative.org/sentinel/methods/utilizing-data-various-data-partners-distributed-manner>. In addition to source code and documentation for the algorithms, the Sentinel website also provides test data and sample reports for each regression model type.

The team also created two DRA SAS packages used to run DRA, one for data partners and one for the analysis center. The packages include all algorithms for linear, logistic, and Cox regression. This open source software utilizes PopMedNet™ and allows stakeholders to perform real-world distributed regression within actual PCOR distributed data networks on horizontally partitioned data.

Lastly, the team developed documentation of the DRA algorithms and set up a SAS-based DRA application for execution in a horizontally partitioned distributed data network. This documentation also includes information on how the software can be modified for use on platforms other than PopMedNet™, the types of research that can and cannot be answered by this new method, limitations of the new method, and testing done to validate this method.

The team also explored the feasibility of performing linear DRA within vertically partitioned DDNs, a data environment in which different databases include information about the same individuals, using a publicly available dataset. Vertical partitioning would allow for the combination of both insurance claims data and EHR records to be traced over time for a single patient. For the proof-of-concept vertical DRA, the team found the PopMedNet™ DRA query workflow could be used to conduct DRA within vertically partitioned data environments. However, additional enhancements are required to integrate vertical DRA algorithms into the PopMedNet™ DRA query workflow, and additional internal and external testing is required to assess operational performance.

Disseminated Products:

- Final Report: Her QL, Vilks Y, Young J, Zhang Z, Malenfant J, Malek S, Toh D. Distributed Regression Analysis In A Large Distributed Data Network: Sentinel. October 24, 2018. https://www.sentinelinitiative.org/sites/default/files/Methods/Distributed_Regression_Analysis_Final_Technical_Report.pdf
- Darren Toh, ScD presented at the 10th Sentinel Initiative Annual Public Workshop in Washington, DC.
- The distributed regression analysis SAS code is available here: <https://www.sentinelinitiative.org>

- Sample reports and data files (listed below) can be found at: <https://www.sentinelinitiative.org/sentinel/methods/utilizing-data-various-data-partners-distributed-manner>
 - **Linear DRA Sample Report:** Report generated for distributed linear regression analysis with the partitioned Boston Housing dataset
 - **Logistic DRA Sample Report:** Report generated for distributed logistic regression analysis with the partitioned Boston Housing dataset
 - **Cox DRA Sample Report 1:** Report generated for distributed Cox regression analysis with the partitioned Maryland convict dataset
 - **Cox DRA Sample Report 2:** Report generated for distributed stratified (Data Partner site identifier) Cox regression analysis with the partitioned Maryland convict dataset
 - **Boston Housing Data:** Zip file of the Boston Housing dataset, and the three partitioned datasets used for distributed linear and logistic regression analysis testing.
 - **Maryland State Prison Data:** Zip file of the Maryland State Prison dataset, and the three partitioned datasets used for distributed cox regression analysis testing.

Period of Performance:
7/15/15 – 6/30/18

Federal Point of Contact:
Michael Nguyen

VIII. National Institutes of Health (NIH)

NIH is administering six active OS-PCORTF-funded projects (Exhibit 7).

Exhibit 7. NIH Active Projects

NIH-Funded Projects
Creation of LOINC Equivalence Classes
Emergency Medicine Opioid Data Infrastructure: Key Venue to Address Opioid Morbidity and Mortality
Use of the ADAPTABLE Trial to Strengthen Methods to Collect and Integrate Patient-Reported Information with Other Data Sets and Assess Its Validity
Developing a Strategically Coordinated Registry Network to Support Research on Women’s Health Technologies*
Harmonization of Various Common Data Models and Open Standards for Evidence Generation*
Technologies for Donating Medicare Beneficiary Claims Data to Research Studies*

* Denotes a cross-agency funded project that involves more than one federal agency; these projects are described in the “Cross-Agency Funded Projects” section.

Creation of LOINC Equivalence Classes

Interoperability relies in part on common vocabulary standards that allow disparate systems to interpret and exchange information with the same clinical meanings. Logical Observation Identifiers Names and Codes (LOINC®), developed by the Regenstrief Institute, Inc., is a universal coding system for laboratory tests and other clinical observations. LOINC is a national standard with widespread adoption and

recognition of its utility; for example, LOINC is required under the CMS EHR Incentive Program, state and national public health reporting mandates, and in research networks like PCORnet.

Mapping local terms to LOINC codes can be complex and resource intensive. This complexity can lead to differences in mapping granularity and specificity that limit interoperability. Furthermore, not all important data uses require the full granularity of detailed LOINC names; for some purposes, “equivalencing” LOINC codes may be sufficient. “Equivalencing,” or rolling-up a hierarchy, groups similar LOINC codes together in clinically relevant and similar groups. This project’s efforts to create equivalencing classes addresses the important interoperability barriers created when clinical and laboratory terms are mapped to LOINC terms with different levels of granularity.

The LOINC Equivalence Class (LOINC release 2.64) contains 36 parent groups, containing 5,650 groups that organize 24,075 unique LOINC terms.

Project Purpose and Goals:

The goal of this project was to create a flexible, extensible, and computable mechanism for rolling LOINC codes into clinically relevant equivalence groups that enable more efficient processing and aggregation of data from diverse health IT systems. The primary focus of this work was on laboratory tests.

The project objectives were to:

- Identify high priority and clinically relevant content subsets for representation in the new LOINC hierarchy. To support this work, the Regenstrief LOINC team developed an enhanced software tool that searches the LOINC database and stores subsets for later retrieval. Input on the high priority subsets was gathered from an expert advisory group and the global LOINC user community.
- Develop a clinically relevant roll-up hierarchy for LOINC terms. The Regenstrief LOINC team reviewed the candidate subsets to identify “equivalence groups” and defined the term attributes necessary to create the group. The goal was to create rule-based group definitions and save them so that they could be updated and re-executed over time and in subsequent LOINC releases.
- Disseminate the aggregation hierarchy within the main LOINC release distribution. The new hierarchy was published with each bi-annual LOINC release and is available publicly at no cost.

Contributions to PCOR Data Infrastructure Functionalities:

- *Standardized Collection of Standardized Clinical Data.* This project supports the standardized collection of data by developing a standard mechanism for assigning newly defined LOINC codes to represent clinically-relevant hierarchies. The application of these hierarchies by LOINC users will help address issues of variability in LOINC mappings and will allow for more consistent comparisons of data across institutions.

Accomplishments and Deliverables:

NIH successfully completed the deliverables of this project. This project created LOINC groups to provide a flexible, computable mechanism for rolling up LOINC terms to support standardized data aggregation and retrieval. The team created three releases of the LOINC Equivalence Class artifact. The first version, LOINC Group File Alpha 1 was released on June 23, 2017 (included in LOINC release 2.61). This release

contained 12 Parent Groups and 2,178 Groups that aggregate a total 6,438 unique LOINC terms. After iterative feedback, review, and expansion, the next version, LOINC Group File Alpha 2, was released on December 15, 2017. This version included the revised format and the inclusion of ancillary data such as molecular weight. It contained 21 Parent Groups, over 4,100 equivalence groups that aggregate more than 17,500 unique LOINC terms. The final version published under this award was released on June 15, 2018 (included in LOINC release 2.64) as LOINC Group File Beta 1. It contains 36 parent groups, containing 5,650 groups that organize 24,075 unique LOINC terms. It includes groups in several new areas, including radiology terms by region imaged, document ontology groups based on the setting, several groups based on a broader clinical concept, such as social determinants of health and physical activity. Inside the LOINC Group File are UsageNotes and Component-level molecular weights from the PubChem database for nearly 900 Groups.

To foster ongoing input from the LOINC community, the team developed multiple avenues to obtain feedback on the current version of the LOINC Groups file. One avenue is a portal for users to recommend and submit groups. Another avenue for feedback is a survey, which is sent to each person who downloads the current version of the LOINC groups file. The team has received substantial contributions of requested content that will be considered for incorporation into subsequent LOINC releases, including additional groups related to microbiology, drug testing, smoking, pregnancy, and pain terms.

In addition to defining and publishing the LOINC Group content in the standard release file format, the team made the latest LOINC Groups content available via an API based on the FHIR® standard specification. They deployed a publicly accessible FHIR®-based terminology server on the HAPI open-source FHIR® library in JAVA that makes all of the LOINC Groups available as FHIR® ValueSet resources. Making the LOINC Groups available in this format provides dynamic, queryable, and machine readable content via a standardized interface.

Within the project's period of performance, the project team made an active effort to disseminate information on the newly created LOINC groups to the LOINC user community, which has over 60,000 users, the Clinical Laboratory Improvement Advisory Committee (CLIA), and to health information technology developers at the FHIR® DevDays conference.

Disseminated Products:

- LOINC Group Beta 1 released in June 2018; the artifact was published along with LOINC version 2.64 and is available at: <https://loinc.org/groups/>
- LOINC Groups Alpha 2 released in December 2017; the artifact was published along with LOINC version 2.63 and is available at: <https://loinc.org/downloads/archive/>
- LOINC Groups Alpha 1 released in June 2017; the artifact was published on the LOINC website along with LOINC version 2.61 and is available at: <https://loinc.org/downloads/archive/>
- Presentation of third release of the equivalence class artifact and example groups at the public June 2018 Laboratory LOINC Conference. Handouts from the LOINC Committee Meeting are available at: <https://loinc.org/conference/summer-2018/meeting/>.
- Presentation by Dr. Vreeman at the FHIR® DevDays Conference about the new API capabilities for accessing LOINC content, including LOINC groups. The full educational session slides and a detailed tutorial of exercises is available at: <https://danielvreeman.com/fhir>.
- The project team presented the systems demonstration, 'LOINC Groups: A novel tool for data aggregation' at the AMIA 2018 Annual Symposium in San Francisco, California (November 3-7, 2018).

- The latest LOINC Groups content available via an API based on the FHIR® standard specification is available at: <https://loinc.org/fhir/>
- The publicly accessible FHIR®-based terminology server is available at: fhir.loinc.org
- The HAPI open-source FHIR® library is available at: <http://hapifhir.io>
- Presentation by Dr. Vreeman to Clinical Laboratory Improvement Advisory Committee (CLIAC), an advisory committee to HHS, about improving lab data interoperability including use of the LOINC equivalence groups. Slides from this presentation are available at: <https://danielvreeman.com/presentation/advancing-lab-data-interoperability-with-loinc/>.

Period of Performance:

8/31/16 –4/30/18

Federal Point of Contact:

H. Timothy Hsiao

Emergency Medicine Opioid Data Infrastructure: Key Venue to Address Opioid Morbidity and Mortality

EHR data are often used in research to improve patient outcomes. The quality of this research could be improved by including opioid relevant CDEs in clinical data registries and EHRs. EHR vendors have been slow to incorporate CDEs related to opioid use disorders (OUDs), and clinical data on OUDs are not collected in a uniform format. There are also inconsistencies in terminology used and types of information recorded about OUD patients. The American College of Emergency Physicians (ACEP) developed the first emergency medicine registry called the Clinical Emergency Data Registry (CEDR). CEDR has the functionality to measure emergency medicine outcomes, identify practice patterns and trends, improve the quality of acute care, exceed quality reporting standards, and eliminate and/or increase payer revenue.²⁴

Integration of patient-reported information and EHR-derived data has the potential to both enhance evaluation of outcomes that are meaningful to patients and to improve data quality and validity for PCOR.

This project will facilitate standardized measurement through use of OUD-specific CDEs in order to track and improve the quality of care at the point of contact with persons with OUD. In recent years, there has been an increased focus on PROs. PROs place patients at the center of health care research and ensure that research is of maximum value for both clinicians and patients. This project will also explore the feasibility of collecting PRO measures in ED setting and after an ED visit, and identify PROs most appropriate for inclusion in ED settings. Improved measurement and enhanced EHR infrastructure could provide benchmarking data such as how many providers provide naloxone for OUD, and can facilitate tracking of quality improvement efforts.

Improving clinical data infrastructure in ED settings addresses key areas of the HHS 5-point strategy to combat the opioids crisis by improving access to treatment and recovery services; promoting use of overdose-reversing drugs; providing support for cutting-edge research on addiction and pain; and advancing better practices for management of OUD and pain.²⁵

Project Purpose and Goals:

The goal of this project is to build clinical data research infrastructure that will begin to enhance capacity to use EHR data and PROs to conduct opioid related research in EDs. The project will improve

interoperability and linkages between EHRs, research networks, and registries for research relevant to the opioid epidemic.

The project objectives are to:

- Identify existing OUD CDEs relevant to the ED setting by conducting an environmental scan of current, publicly available data systems, data elements, and quality measures. A consensus data dictionary and data system guidance for OUD research in ED EHRs will be developed.
- Demonstrate that CDE from ED EHR test sites can be integrated into the ACEP CEDR by translating and mapping electronic OUD data elements to the NLM Value Set Authority Center (VSAC) and testing the validity and feasibility of the electronic OUD data dictionary in ACEP CEDR test sites. If successfully, researchers would be able to aggregate data from the CEDR from multiple sites and analyze the data.
- Explore feasibility of collecting electronic PROs such as PROMIS and other measures (e.g., pain intensity, substance use disorder treatment/status) at one or more of NIDA's Clinical Trial Network's (CTNs) sites.²⁶ The PRO tool will be implemented and integrated into routine ED clinical workflows. If successful, researchers will be able to aggregate data from the CEDR from multiple studies and analyze the data.

Contributions to PCOR Data Infrastructure Functionalities:

- *Standardized Collection of Standardized Clinical Data.* This project focuses on standardizing ED EHR data capture relevant to opioids for reporting to the ACEP CEDR. A steering committee of experts provides input to identify OUD data elements and definitions.
- *Collection of Patient-Provided Information.* The project will advance collection of PPI by identifying existing and potential CDEs for OUD that are relevant to the ED setting and by testing the feasibility of collecting PROs most appropriate for collection in ED setting and inclusion in ACEP CEDR.
- *Linking of Clinical and Other Data for Research.* This project focuses on demonstrating that CDE from ED EHR test sites can be transmitted or integrated into the ACEP CEDR.

Accomplishments and Deliverables:

Since the beginning of the award on June 1, 2018, the project has begun to lay the groundwork for the three project tasks and has made significant progress towards their achievement.

- The literature review and environmental scan of existing OUD CDEs is complete.
- Potential test sites have been identified for the demonstration. The project team has developed recruitment materials and will engage with potential demonstration sites soon to discuss collaboration on the CDE validity testing.
- A prototype of the application to collect PROs has been developed.

Period of Performance:
6/1/18- 5/30/20

Federal Point of Contact:
Kristen Huntley

Use of the ADAPTABLE Trial to Strengthen Methods to Collect and Integrate Patient-Reported Information with Other Data Sets and Assess Its Validity

The ADAPTABLE (Aspirin Dosing: A Patient-centric Trial Assessing Benefits and Long-Term Effectiveness) trial is the first major randomized comparative effectiveness trial to be conducted by PCORnet. This pragmatic clinical trial will compare the efficacy and safety of two different daily doses of aspirin widely used for patients with chronic cardiovascular disease.²⁷ As part of this new type of comparative effectiveness trial, the ADAPTABLE trial encompasses several key features, including enrollment of 20,000 patients across six large health care systems; an internet portal to consent patients and collect patient-reported information regarding risk factors, medications, and experiences; and reliance on existing EHR data sources for baseline characteristics and outcomes follow-up.

Integration of patient-reported information and EHR-derived data has the potential to both enhance evaluation of outcomes that are meaningful to patients and to improve data quality and validity for PCOR.

Project Purpose and Goals:

Because ADAPTABLE will rely on patients to report key information at baseline and throughout follow-up, it represents a unique opportunity to develop, pilot, and evaluate methods to validate and integrate patient-reported information with data obtained from the EHR. The project will generate tools and data standards that could be deployed in other PCOR studies beyond the ADAPTABLE trial.

The project objectives are to:

- Develop, test, and validate metadata standards for patient-reported information to describe the completeness, consistency, and fitness-for-use of patient-reported data in EHR research.
- Evaluate the validity of patient reported data through systemic comparison with EHR data. The project will develop a Patient-Reported Data Assessment Tool to quickly and efficiently evaluate concordance of patient-reported data and EHR data.
- Develop approaches to resolve inconsistencies between patient-reported data and EHR-derived data.

This project will inform future efforts to synthesize potentially inconsistent data from patient-reported and EHR sources to identify opportunities to streamline data capture, and to facilitate enrollment in study-specific target populations within larger health systems.

Contributions to the PCOR Data Infrastructure Functionalities:

- *Standardized Collection of Standardized Clinical Data.* This project focuses on methods to streamline and standardize data capture from the EHR. Baseline and follow-up data will be obtained from the PCORnet CDM, which serves as the foundation of the PCORnet distributed research network.
- *Collection of Participant-Provided Information.* This project utilizes the ADAPTABLE web portal to consent patients and collect patient-reported information regarding risk factors, medications, and experiences. Specifically, patients will provide information for four domains in the portal: 1) PROs

for general domains of health; 2) specific information about medications they take; 3) specific details about reasons for stopping aspirin when this occurs; and 4) hospitalizations.

- *Use of Clinical Data for Research.* The trial will leverage data from the EHR to identify the optimal dose of aspirin therapy for secondary prevention in atherosclerotic cardiovascular disease.

Accomplishments and Deliverables:

The project made significant progress toward their objectives.

- For the peer-reviewed manuscript, the team completed a literature review on data and metadata standards for patient-reported data in EHR-based trials to inform their development of a priority list of metadata standards. After discussing initial findings, the project leadership recommended that elements also be connected to the NIH CDE browser. The project team also finalized their concordance plan (statistical analysis plan) and explored smoking status categories in the CDM for mapping.
- As a result of the ADAPATABLE Supplement Roundtable Meeting held on September 14, 2017, the project team produced a jointly-authored white paper on available resources for best practices, key challenges, information gaps, and future research needs for promoting best practices in the use of patient-reported health data in pragmatic studies.
- Related to the Patient-Reported Data Assessment Tool developed on the PopMedNet™ platform: The team developed a statistical analysis plan that has been vetted internally with leadership, statisticians, and informaticists, as well as with ASPE and the sub awardees. The team has refined elements to be included in menu-driven query tool, refined the tool development plan/timeline and completed a preliminary assessment of patient-reported data.
- Through collaboration with its sub-awardees, the project team determined that patient-reported data elements would be submitted to LOINC. They were published on the PCORnet website and included in the NIH/NLM CDE Resource Portal and the REDCap shared library to support implementation of patient-reported data elements in future studies.
- The data and metadata standards report was finalized and posted to the NIH Collaboratory Living Textbook.
- The selection of appropriate concordance elements from the statistical analysis plan for inclusion in the menu-driven query tool was completed. Additionally, the Duke, Harvard and Vanderbilt teams collaborated on the abilities and the limitations of the Menu-Driven Query (MDQ) Tool, as well, as the needs of the study and future use. Finalization of the contracts between Duke and Harvard and Harvard and their programmers are in process.
- The patient-reported data elements were included in the June 2018 release of LOINC.
- The literature review of approaches to resolution of inconsistencies between patient-reported and EHR data has been completed.
- The ADAPATABLE Supplement Roundtable Meeting (2017) resulted in the development of two white papers on available resources, key challenges, existing information gaps, and future research needs for promoting best practices in the use of patient reported health data in pragmatic studies.

Disseminated Products:

- A summary report of the ADAPTABLE Supplement Roundtable Meeting (2017) was published in January 2018 and is available here: https://www.nihcollaboratory.org/Products/ADAPTABLE%20ROUNDTABLE%20SUMMARY_1_2_2.pdf
- The literature review of data and metadata standards for patient-reported data in EHR-based trials includes descriptions of the metadata standards and a dictionary describing each element was made available in order to reach the appropriate audience and serve the informatics needs of the research community. The report is posted to the NIH Collaboratory Living Textbook and available here: <http://rethinkingclinicaltrials.org/news/july-23-2018-new-report-summarizes-patient-reported-health-data-and-metadata-standards-from-the-adaptable-trial/>
<http://rethinkingclinicaltrials.org/news/july-23-2018-new-report-summarizes-patient-reported-health-data-and-metadata-standards-from-the-adaptable-trial/>
- The report "ADAPTABLE Supplement Report: Patient-Reported Health Data and Metadata Standards in the ADAPTABLE Study"²⁸ was published to the NIH Collaboratory Living Textbook website and is available here: <https://dcricollab.dcri.duke.edu/sites/NIHKR/KR/ADAPTABLE%20Supplement%20Patient-Reported%20Health%20Data%20Standards-2018-07-19.pdf>
- LOINC Patient-Reported Data Elements were published in the NLM CDE repository and REDCap shared library here: <https://github.com/ADAPTABLETRIAL/REFERENCE/wiki> and here: <http://rethinkingclinicaltrials.org/news/august-28-2018-adaptable-patient-reported-health-data-codes-now-available/>

Coordination with Other Federal Agencies:

The patient-reported data elements submitted for inclusion into LOINC will also be published on the PCORnet website and submitted for inclusion in the NLM CDE Resource Portal and the REDCap shared library to support implementation of patient-reported data elements in future studies conducted at REDCap partner institutions.

Period of Performance:

8/31/16 – 9/30/17; NCE: 8/31/18

Federal Point of Contact:

Wendy Weber

IX. Office of the National Coordinator for Health Information Technology (ONC)

ONC is administering six active OS-PCORTF-funded projects (Exhibit 8).

Exhibit 8. ONC Active Projects

ONC-Funded Projects
Conceptualization of a Data Infrastructure for the Capture and Use of Patient-Generated Health Data
Security and Privacy Standards for Patient Matching, Linking, and Aggregation
Advancing the Collection and use of Patient-Reported Outcomes through Health IT*
Developing a Strategically Coordinated Registry Network to Support Research on Women’s Health Technologies*
Harmonization of Various Common Data Models and Open Standards for Evidence Generation*
Privacy and Security Blueprint, Legal Analysis and Ethics Framework for Data Use, and Use of Technology for Privacy*

* Denotes a cross-agency funded project that involves more than one federal agency; these projects are described in the “Cross-Agency Funded Projects” section.

Conceptualizing a Data Infrastructure for the Capture and Use of Patient-Generated Health Data

PGHD are “health-related data created, recorded, or gathered by or from patients (or family members or other caregivers) to help address a health concern.”²⁹ PGHD are different from data gathered in clinical settings and through provider encounters in two ways: 1) patients are responsible for capturing the data; and 2) patients decide how to share or distribute the data with clinicians in other health care settings and others in their care team. Examples of PGHD include blood glucose readings using a home blood glucose monitors, data generated by mobile fitness tracking apps, and survey data collected from patient questionnaires in both paper and electronic formats. The use of PGHD offers a unique opportunity to fill in gaps in information and provide a more comprehensive picture of ongoing patient health for use during care, resulting in potential cost savings and improvements in health care quality and outcomes, care coordination, and patient safety.

Understanding the potential benefits and concerns surrounding PGHD, ONC began a series of activities in 2012 to increase the capture and use of PGHD to meet the needs of patients, researchers, and providers. Lessons learned from those activities informed this project, which aims to identify best practices, gaps, and opportunities for progress in these seven PGHD topic areas:

1. Patient recruitment for research studies and trials
2. Collection and validation of data and tools
3. Data donation
4. Ability to combine PGHD data with medical record data in multiple ways
5. Data interoperability
6. Big data analysis
7. Regulatory overview

Project Purpose and Goals:

This project developed a white paper that describes the current policy landscape, challenges and opportunities for the capture, use, and sharing of PGHD in research and care delivery through 2024. Additionally, the project conducted pilot demonstrations that tested the concepts discussed in the white paper, resulting in a practical guide. Two supporting infographics summarize the information in the white paper and the results of the pilot demonstrations.

Contributions to PCOR Data Infrastructure Functionalities:

- *Collection of Participant-Provided Information.* The white paper and practical guide support the sharing and use of PGHD. The increased use of tools and applications to collect PGHD (e.g., from mobile devices, PRO assessments, remote monitoring devices) offers an opportunity for patients to be more engaged in their health care and their use can facilitate joint patient-clinician decision-making.³⁰ The white paper provides useful insights that describe the current policy landscape, challenges, and opportunities that can be used by a variety of stakeholders to support the capture, use, and sharing of PGHD. The practical guide offers suggested practices and questions for stakeholders to consider when implementing the capture, use, and sharing of PGHD in the clinical and research settings.
- *Standardized Collection of Standardized Clinical Data.* The insights from the project deliverables can be used to inform future policies and guidance that promote standardized data collection of PGHD, increase its consistency and utility for research, clinical uses, and other analytic activities.

Accomplishments and Deliverables:

The intent of this project is to advance patient engagement in research by increasing the capability to utilize PGHD in research and health care delivery. This project has successfully concluded and has developed important deliverables that can support future development of a PGHD policy framework including a white paper, a practical guide, and two supporting infographics.

With the increase of consumer health technologies (i.e., mobile applications, wearable devices), there has also been an increase in the frequency, amount, and types of PGHD available. Using these technologies, patients and their caregivers can now independently and seamlessly capture and share their health data electronically with clinicians. The white paper describes key opportunities and challenges and offers enabling actions that various stakeholder groups can take to further enhance PGHD capture, use, and sharing for health care delivery and research settings through 2024. This project aligns with several calls to action referencing PGHD in ONC's 10-year vision to achieve an interoperable health IT infrastructure by 2024. The white paper findings anticipate that digital health technologies will become more pervasive, offering more opportunities for patients to capture, use, and share their PGHD in support of health care delivery and research. The white paper was completed and shared publicly on HealthIT.gov on January 23, 2018, along with a supporting infographic depicting examples of PGHD, PGHD sources and the benefits of using PGHD.

To further validate and expand on the findings of the white paper, pilot demonstrations were conducted with two care delivery partners. In one pilot demonstration, Validic and Sutter Health partnered to test personalized care that leveraged remotely-collected PGHD across patients participating in a diabetes-focused research study. In the other pilot demonstration, TapCloud worked with AMITA Health to gather PGHD on several medical conditions, such as orthopedic surgery, stroke, behavioral health, and kidney transplant. The pilot demonstrations concluded at the end of 2017 and informed the development of the

practical guide, which offers suggested practices for implementing of the capture, use, and sharing of PGHD in clinical and research settings. Through the deployment of advanced clinical workflows leveraging PGHD, the pilot demonstration results showed the feasibility of PGHD use and informed necessary improvements for clinical care delivery and research models. The practical guide and a supporting infographic are publicly available and published on ONC's website.

The team concluded that the capture of PGHD alone is not sufficient to cause change within the health IT ecosystem. Joint action from across the ecosystem is necessary to overcome cultural, technical, and regulatory barriers. However, through collaboration, these barriers can be addressed, resulting in improved insights for clinicians and researchers and improved care for patients.

Disseminated Products:

- The final white paper is called Conceptualizing a Data Infrastructure for the Capture, Use, and Sharing of Patient-Generated Health Data in Care Delivery and Research through 2024: https://www.healthit.gov/sites/default/files/onc_pghd_final_white_paper.pdf
- The final white paper supporting Infographic is available here: https://www.healthit.gov/sites/default/files/onc_pghd_final_white_paper_infographic.pdf
- Funded by an FDA grant, the Duke-Margolis Center for Health Policy, convened a workgroup to create an action plan for accelerating the use of PGHD in research. The working group used the draft white paper to inform their discussions and was referred to in the Duke-Margolis Center action plan as a resource to readers, which is available here: https://healthpolicy.duke.edu/sites/default/files/atoms/files/mobilizing_mhealth_innovation_for_real-world_evidence_generation.pdf
- Conceptualizing a Data Infrastructure for the Capture, Use, and Sharing of Patient-Generated Health Data in Care Delivery and Research through 2024 (Practical Guide) was published in January 2018. The Practical Guide is available here: https://www.healthit.gov/sites/default/files/onc_pghd_practical_guide.pdf
- The Practical Guide infographic is available here: https://www.healthit.gov/sites/default/files/onc_pghd_pilot_demonstrations_infographic.pdf

Period of Performance:
6/16/15 – 6/30/18

Federal Point of Contact:
Michelle Murray

Security and Privacy Standards for Patient Matching, Linking, and Aggregation

Linking and aggregating patient data from disparate sources cannot be done without “matching” these data in a secure manner that protects patient privacy. Effective patient matching allows users of the data (e.g., patients, providers, or researchers) to draw correct inferences using the data that have been linked. Without accurate patient matching, stakeholders may inadvertently draw inaccurate conclusions that could significantly impair patient safety and privacy.^{31, 32} Challenges also arise when policies on data use and patient matching are in conflict.

Aggregating records requires strong and standardized methods (i.e., patient matching algorithms) to ensure that data on the same patient are kept together, and that data on different patients are kept separate. These patient matching algorithms rely on analyzing both

It is critical to have the ability to uniquely match patients for any infrastructure building around data sharing.

unique identifiers and personal characteristics of the patient record (i.e., patient attributes) to match records. The accuracy and reliability of patient matching results vary widely due to inconsistent data quality and the variability of algorithms used in matching. For instance, one may believe “John Smith” and the likely misspelled “Jon Smtih” to be the same person. However, even small differences introduce the chance that two patient records do not correspond to the same patient. Measuring this risk and developing risk thresholds is challenging and dependent on the end use of the linked datasets. These issues may be further exacerbated when information collected in the EHR is incomplete making it more difficult to uniquely identify patients across data sources. While there are data exchange standards in place to electronically exchange information across health IT systems, there are no widely adopted standards that consistently address patient identity, authorization, and consent.

Project Purpose and Goals:

This project focused on the identification and testing of standards to address the challenge of linking patients’ data across research, clinical, and claims data sets in order to support the PCOR data infrastructure. This project involves work along five distinct tracks.

The project objectives were to:

- Standardize attributes and improve algorithm match rates. Use cases, an environmental scan, a standards list, and a specifications/implementation guide were developed for this portion of the project. Patient matching standards (e.g., WEDI, IHE, S&I, and HL7) were reviewed and considered for optimizing a patient matching algorithm. The project also supported a Patient Matching Algorithm Challenge³³ to bring about greater transparency and data on the performance of existing patient matching algorithms, spur the adoption of performance metrics for patient data matching algorithm vendors, and positively impact other aspects of patient matching such as de-duplication and linking to clinical data.
- Create an open source visual tool for patient matching and aggregation. ONC repurposed an existing set of open source tools, developed for creating synthetic patient records for testing clinical quality measures, and added a Patient Match module (Patient Match Test Harness) to create a patient matching toolkit. The toolkit allows researchers to: 1) inspect patient match results; 2) quickly create test data for sharing; 3) incorporate results from clinical and claims feeds and PCORnet; and 4) model new patient attributes.
- Create a privacy and security API or PCOR infrastructure security “layer.” This privacy and security “layer” (i.e., authentication, authorization, consent, and data provenance) ensures that the data are being matched: 1) with the appropriate consents; 2) by a user who is authorized to do the matching; and 3) with authentication (i.e., validation that people are who they say they are) of those trying to access the data. The API work provided the specifications for this layer along with corresponding implementation guides to manage the authentication, authorization, and consent policies necessary for all participants (i.e., patients, researchers, providers) to safely retrieve and contribute patient data to the PCOR infrastructure.
- These algorithms, alongside the tools developed through this project (e.g., Patient Match Test Harness) enable researchers to more accurately link patient data from different sources, vastly increasing the availability of data for research.
- Include clinical data research networks in the piloting and testing of the proposed standards and services.

- Integrate the National Plan and Provider Enumeration System (NPPES) provider identification as an additional attribute to improve patient matching across sources.

Contributions to PCOR Data Infrastructure Functionalities:

- *Linking of Clinical and Other Data for Research.* PMAL supported the improvement of data quality by standardizing patient attributes and algorithms that can be used to reliably perform patient matching across clinical and claims data sets to improve algorithm match rates. These algorithms, alongside the tools developed through this project (e.g., Patient Match Test Harness) enable researchers to more accurately link patient data from different sources, vastly increasing the availability of data for research.
- *Use of Clinical Data for Research.* Addressing issues of data provenance, which has a role in nearly all data use and exchange activities, has the potential to open up new data sources for inclusion in PCOR.³⁴

Accomplishments and Deliverables:

This project encompassed a number of inter-related projects. Pilot projects were staged to address data quality, patient match rates, and promote attribute standardization for interoperability; a number of open source tools and resources have been developed under this project; and multiple dissemination efforts have been made.

ONC secured a commitment of participation from a Community Health Applied Research Network (CHARN) partner to pilot all aspects of the algorithm, data management, and the privacy and security layer at six sites. Pilot testing is now complete; lessons learned were documented in a series of key findings reports that will be released together in the spring of 2019. These reports describe the identification, testing, and optimization of a number of patient matching algorithms against data sets previously identified as gold standards for testing purposes.

Patient Match Algorithm Challenge

The goal of this prize challenge was to bring about greater transparency and data on the performance of existing patient matching algorithms, spur the adoption of performance metrics for patient data matching algorithm vendors, and positively impact other aspects of patient matching such as de-duplication and linking to clinical data.

The Patient Test Match Harness development work has also concluded. This toolkit allows researchers to inspect patient match results, create test data for sharing, model new patient attributes, and incorporate results for claims, clinical, and research network data sources.

Five final privacy and security profiles were delivered to the OpenID Foundation³⁵ standards organization promoting the use of OpenID authentication protocols, and submitted to the broader community for public review in the summer of 2017. Final voting for the approval to promote the profiles to formal implementers' Drafts status occurred on July 5, 2017.

ONC utilized the Health Relationship Trust (HEART) Profile, an authentication and authorization protocol to facilitate more consistent user access management. This was developed with health IT stakeholders to advance privacy and security specifications that enable an individual to control the authorization of access to health data. In parallel with this effort, ONC worked to align SMART on FHIR[®] profiles, geared toward provider workflows, with HEART profiles, focused on patient-mediated exchange.

The API development work occurred under two open vendor challenges, the Move Health Data Forward (MHDC) Challenge and the Health Data Provenance Challenge. The MHDF Challenge invited developers to use the HEART profiles to create an API to enable consumers to authorize the movement of their health data to destinations they choose. The winners of the challenge were announced on May 31 2017. The Health Data Provenance Challenge sought to address the challenges of data provenance and the lack of related meta-data standards. Phase 1 solicited white papers that identified problems for the conveyance of provenance and propose viable solutions. Phase 2 asked the four winners from Phase 1 to pilot those solutions. On February 13, 2018, ONC hosted a webinar that included live demonstration by the four finalists from this project. On March 13, 2018, ONC announced the winning solutions submitted by 1upHealth and RAIN Live Oak Technology. 1upHealth piloted the use of its partner's provider application to surface provenance information and help providers find aggregated data from various sources using the HL7[®] FHIR[®] standard along with improvements afforded by smart contracts on the blockchain's public ledger. RAIN Live Oak created a software toolkit enabling health information systems of any size to integrate provenance into their data-flow without disrupting existing practices or data repository requirements.

ONC has completed its deliverables for the NPPES work to support the provider directory integration of the NPPES. As a result, ONC and the Federal Health Architecture have formed a collaborative, the Healthcare Directory Technology Learning Community, to explore health care directory governance and sustainability issues, recognizing that technical solutions alone are insufficient for successful implementation of healthcare directories.

Disseminated Products:

- The Patient Match Test Harness is publicly available for download from GitHub here: <http://mitre.github.io/test-harness-interface/>.
- The Open Source Toolkit for Patient Matching and Aggregation is available on GitHub. Key resources from this toolkit include:
 - Test Harness interface: <http://mitre.github.io/test-harness-interface/>
 - Description of Patient Match Services: <https://github.com/mitre/ptmatch/blob/master/api/swagger.yaml>
 - Description of eCQM Services: <https://github.com/mitre/ecqm/blob/master/api/swagger.yaml>
- Tool Specifications Document: Design & Component Development here: <https://github.com/mitreid-connect/OpenID-Connect-Java-Spring-Server/>.
- API profile specifications are available here: <https://openid.bitbucket.io/HEART/>.
- The Health Data Provenance Challenge has ended. Additional information can be found here: <https://www.cccinnovationcenter.com/challenges/provenance-challenge/>.
- The Move Health Data Forward Challenge has ended. Additional information can be found here: <https://www.healthit.gov/buzz-blog/interoperability/announcing-winners-move-health-data-challenge/>

Period of Performance:
6/16/15 – 9/30/18

Federal Point of Contact:
Debbie Bucci

X. Cross-Agency Funded Projects

There are eight cross-agency funded active OS-PCORTF-funded projects (Exhibit 9).

Exhibit 9. Cross-Agency Funded Active Projects

Cross-Agency Funded Projects
Advancing the Collection and Use of Patient-Reported Outcomes through Health IT
Assessing and Predicting Medical Needs in a Disaster
Developing a Strategically Coordinated Registry Network to Support Research on Women’s Health Technologies
Development of a Natural Language Processing Web Service for Public Health Use
Enhancing Data Resources for Researching Patterns of Mortality in PCOR: Projects 1 – 4
Harmonization of Various Common Data Models and Open Standards for Evidence Generation
Privacy and Security Blueprint, Legal Analysis and Ethics Framework for Data Use, and Use of Technology for Privacy
Technologies of Donating Medicare Beneficiary Claims Data to Research Studies

Advancing the Collection and Use of Patient-Reported Outcomes through Health Information Technology

The patient perspective is essential to health care decision making and health management, and can provide important data to inform PCOR. A PRO is “any report of the status of a patient’s health condition that comes directly from the patient, without interpretation of the patient’s response by a clinician or anyone else.”³⁶ PROs provide complementary perspectives to those of providers and can provide insights into many health care facets, such as health status, symptom burden, health behaviors, and quality of life. Given the value of PROs, ONC and AHRQ are conducting a joint project focused Advancing the Collection and Use of PROs through Health IT.

The capacity of EHRs to capture PROs in standardized form is currently limited and underutilized. PCORI published a user guide on PRO and EHR integration, and detailed multiple challenges involved with integration options, from the technical and user perspectives (i.e., provider, patient, and researcher perspectives).³⁷ Providers/health systems wishing to integrate PROs directly into EHR systems may need to work with EHR vendors or IT teams to properly customize and present the PRO questions to patients. Additionally, the EHRs may not be optimized to leverage the PRO data for clinical, research, or administrative needs. Patients may find logging into an EHR portal at home to record their PROs to be more burdensome than reporting their outcomes on paper handouts. Collecting paper PRO assessments and scanning them into EHRs is an option, but it creates its own unique challenges, such as extra manual processes and workflows, and reduced opportunity for user-friendly displays and automated clinical decision support. A separate system could be used to collect PRO data and then integrate the data into the EHR. However, utilizing two systems may be expensive, require additional IT expertise, and be less secure. The lack of standards for collecting and integrating PRO represents an additional hurdle. This project aims to widen the availability of PRO data for PCOR through supporting standardization and promoting the development of PRO collection applications.

Project Purpose and Goals:

The purpose of this project is to develop user-friendly applications for standardized PRO data collection and integrate PRO data into an EHR or other health IT systems.

The PRO project objectives are:

- Refining and/or harmonizing health IT standards and implementation specifications that can be used to support sharing of PRO data through APIs and relevant health IT products for research;
- Supporting the development of user-friendly, PRO-collection applications that utilize the health IT standards; and
- Implementing private/public partnerships for pilot-testing these applications in a health system that supports both healthcare delivery and research.

Contributions to PCOR Data Infrastructure Functionalities:

- *Collection of Patient-Provided Information.* This project involves using standards which will enable easier sharing of PROs between patients, providers, and researchers. PRO data collection applications will be developed through a developer competition, and then pilot-tested in a health care system. The current state of PRO use will also be explored. Data capture standardization would improve the interoperable sharing of PRO data across providers.
- *Use of Clinical Data for Research.* This project will include developing implementation specifications that improve the standardization of the PRO data, thereby improving the potential usability of the data for research.

Accomplishments and Deliverables:

Since the project commenced in January 2017, the project has made progress toward its objectives.

- ONC drafted a summary of key informant interviews regarding current PRO use for functional status assessments, standards for electronic capture and use of PROs, and research programs and networks that use functional status to study patient outcomes. The final interview report was completed in March 2018.
- AHRQ launched the AHRQ Step Up App Challenge in August 2018. This is a three-phase competition that encourages participants to develop and present user-friendly apps that are able to collect standardized PRO data in primary and specialty care settings. The winner will have the opportunity to test their app in nine practice settings around the D.C., Maryland, and Virginia area. ONC is leading stakeholder efforts to develop the technical specifications for the competition.
- ONC finalized the report of the technical approach for standards refinement in April 2018.

Step Up App Challenge Update

- *Phase 1 received 56 submissions. Ten winners were chosen and announced in late October 2018. These 10 winners proceeded to Phase 2.*
- *Phase 2 participants use the ONC's technical specifications and other usability and functionality to develop the app. Participants are required to provide a video demonstration of their app, and submit pilot test implementation plan. This Phase will conclude in February 2019.*

- Pilot sprint 1 was completed in July 2018, along with a draft report for review.
- ONC is currently in the midst of the second sprint of the implementation guide development working with pilot participants REACHnet and pSCANNER. The second pilot sprint kicked off in August 2018.
- MedStar Health, AHRQ’s contractor for the pilot testing, is recruiting practices and preparing the health IT infrastructure for the pilot test. MedStar is also working with its sub-contractor to modify an existing app using ONC’s technical specifications. The modified app will also be pilot tested.

Coordination with Other Federal Agencies:

NIH, FDA, CMS, and HRSA are providing advice for this project. NIH is advising on work related to the Patient-Reported Outcomes Measurement Information System (PROMIS),³⁸ whose PRO measures are being tested for the standards refinement task. CMS and FDA will provide advice regarding how standards would be applied to their regulatory needs. HRSA is providing advice regarding the applicability of deliverables for the safety-net population.

AHRQ Period of Performance:

1/15/17 to 4/15/19

AHRQ Federal Point of Contact:

Janey Hsiao

ONC Period of Performance:

1/15/17 to 4/15/19

ONC Federal Point of Contact:

Margeaux Akazawa

Assessing and Predicting Medical Needs in a Disaster

HHS leads the US public health and medical response to disasters and emergencies. These disasters occur in all geographic regions, yet every geography has distinct disaster types and distinct medical needs. Often, researchers are unable to address geographic differences when designing studies, which leads to inaccurate and non-generalizable results. The Office of the Assistant Secretary for Preparedness and Response (ASPR) focuses specifically on preventing, preparing for, and responding to the adverse health effects of public health emergencies and disasters. They coordinate the National Disaster Medical System (NDMS), a group of professionals who supplement health systems and response capabilities in emergency settings. In the midst of Hurricane Harvey, ASPR and the AHRQ collaborated to explore the use of Healthcare Cost and Utilization Project (HCUP) data to predict medical needs by region. HCUP is a group of health care databases and software tools developed through an AHRQ sponsored Federal-State-Industry partnership. These databases bring together data collection efforts at multiple levels to create a resource with national encounter-level health care data. The overall objective of HCUP is to conduct and translate research to inform decision-making and improve health care delivery.³⁹

Examples of CER/PCOR Questions that Can be Answered by This Project

1. *Comparing patient outcomes for various interventions such as medical treatment received in shelters versus patient outcomes in hospitals*
2. *Assessing whether new approaches to providing dialysis services in the 2017 disasters improve patient outcomes*
3. *Determining if naloxone treatment in shelters is effective in overcoming opioid overdoses when the patient history is unknown.*

This project aims to close the gap in understanding how to tailor disaster medical response to the local level for each event. The project will focus on the creation of a data platform that can be used to conduct PCOR related to disaster response and recovery operations. The initial platform will consist of public facing statistical query pathways that provide access to statistical tables, maps, and graphic on patient health outcomes and a restricted access set of analytic files derived from the HCUP that can be accessed by researchers at AHRQ, ASPE, ASPR, and other federal partners. The restricted file can be integrated with supplemental data sources on disaster impacts and emergency interventions at the county level. In turn, analyses can reflect the health needs of specific populations, thus improving information to deploy appropriate medical expertise. This project will initially support CER/PCOR questions such as comparing differences between communities directly affected by a disaster versus other comparable communities for outcomes. Eventually, researchers can use these data to assess different interventions based on disaster type and population.

Project Purpose and Goals:

The purpose of this project is to develop a data platform to conduct PCOR related to medically related disaster response and recovery. ASPR and AHRQ will work on separate tasks to meet stated objectives. AHRQ will expand the HCUP database to include new quarterly emergency department and inpatient data from individual states. They will also compile data sources and create the platform with input from the project TEP. Finally, AHRQ will test the online query system and data analysis environment to ensure a useful and functional platform for end users. ASPR will developed an NDMS EHR application to capture near real-time data. It will collect health needs and treatments at HHS medical sites in a disaster. Through consulting with the TEP, ASPR will prioritize environmental hazard data sources according to data availability, quality, cost, value, and feasibility of incorporation into the existing ASPR mapping platform. ASPR will also convene a workshop to inform researchers of the new data available through this project, receive feedback, encourage future PCOR research, and learn about each disaster research center to facilitate future collaborations. Finally, ASPR will conduct an operational exercise using the newly-created data platform. Multiple data sets from multiple sources with specificity at the county-level will be combined to develop the data platform. One level of access is public and will contain query systems accessed through the AHRQ HCUP website while another level will include restricted-access files for federal health services researchers.

The overall objectives of the project are to:

- Create a disaster-relevant analytic platform available with two levels of access.
- Design a reporting system that can collect and share real-time reporting of medical encounters during a disaster response.
- Pilot test the database and platform via an operational disaster training exercise, and engage researchers to evaluate whether data can predict medical needs in a disaster.

Contributions to the PCOR Data Infrastructure Functionalities:

- *Use of Clinical Data for Current Research:* This project will use clinical data in real time to assess region-specific, urgent questions. These data will ensure that the medical response to disasters will consider PCOR.

AHRQ Accomplishments and Deliverables:

Since the project began in Quarter 4 of FY18, the project has made progress towards stated project aims.

- The team began drafting a project plan outlining the approach to: developing, testing, and disseminating the analytic files and underlying data through public-facing statistical query systems. The focus of the initial analytic platform will be on hurricanes.
- Initiated updates to HCUP quarterly data processing programs in preparation for the processing of quarterly outpatient data files.
- Established priorities for data recruitment and processing based on data availability for all HCUP State Partners.

Researchers at AHRQ conducted analyses to help ASPR refine their emergency response plans to Hurricane Florence. Using historical data from the HCUP State Emergency Department Databases (SEDD) and State Inpatient Databases (SID) from the periods of Hurricane Irene (August 2011) and Hurricane Matthew (October 2016), AHRQ consulted with ASPR in projecting estimated changes in hospital inpatient stays and emergency department visits in the first four weeks after Hurricane Florence.

ASPR Accomplishments and Deliverables:

Since the project began in Quarter 4 of FY18, the project has made progress towards stated project aims.

- The project team has planned and recruited members of a Technical Experts Panel to be convened in FY19.

Disseminated Products:

- A project team member gave the opening keynote presentation “How Good Data Leads to Good Response” at the Public Health in Disasters Conference in Utah (September 17-18, 2018).
- Two presentations were given at the Natural Hazards Center Workshop (July 8-11, 2018): “Better Outcomes: Strategies to Improve Health and Well-Being” and “Interactive Listening Session – Better Results: Data and Partnership Needs to Improve Disaster Response”

AHRQ Period of Performance:
6/15/18-9/14/20

AHRQ Federal Point of Contact:
Bill Freeman

ASPR Period of Performance:
6/15/18-9/14/20

ASPR Federal Point of Contact:
Leremy Colf

Developing a Strategically Coordinated Registry Network to Support Research on Women’s Health Technologies

The Developing a Strategically Coordinated Registry Network (CRN) to Support Research on Women’s Health Technologies is a joint project between the FDA, NIH/NLM, and ONC, to facilitate PCOR focused on women’s health by connecting three existing registries and several other federal data sources (e.g., Sentinel, claims data, PCORnet). The Women’s Health Technologies CRN (WHT-CRN) was initiated to improve clinical evidence generation and to better answer clinical questions on medical device technologies in clinical areas unique to women. Initially, the WHT-CRN focuses on three clinical areas: sterilization and long-acting contraception, uterine fibroid treatment, pelvic floor disorders and stress urinary incontinence.

Registries provide critical infrastructure that can be used for a variety of analyses related to patient care and outcomes. While single-purpose registries can meet the demand for data on real-world patient care, they can be expensive to maintain, use proprietary data formats, and often focus on a single therapy, when routine care can involve combinations of devices or therapies in multiple care settings. A network of registries, or CRN approach, presents an opportunity to address these challenges. The project will leverage structured data capture standards to extract relevant clinical information from EHRs to populate the registry. WHT-CRN registries will also use standard data elements and measure definitions to harmonize the data available for research. Finally, this data will be shared across registries using standard API interfaces, which can support more complex study designs evaluating the effect of combinations of devices or therapies.

Project Purpose and Goals:

The project's primary goals are to establish a CRN for research on women's health technologies, and to develop and test tools for the collection of standardized data and evaluate the completeness and flexibility of the HL7 FHIR® exchange messaging standard to support the evaluation of medical devices in clinical areas unique to women.

While working collaboratively to accomplish these goals, lead agency FDA, NIH/NLM, and ONC have a number of additional objectives:

- Convene a community of stakeholders (e.g., patients, providers, manufacturers, EHR vendors, SDOs, researchers, etc.) interested in improving data infrastructure in the area of women's health device safety (FDA leads).
- Establish a multi-stakeholder data governance model (FDA/NIH collaboration).
- Define clinically-meaningful patient-centered outcome measures for each device area using CDEs and value sets specified using standard vocabularies and codes sets (i.e., NLM CDE Repository, the NLM Value Set Authority Center, ONC EHR certification criteria, etc.) (FDA leads).
- Pilot and test FHIR® profiles developed by other OS-PCORTF projects (SMART on FHIR® platform and the ONC SDC initiative to extract clinical data from EHRs into the CRN (FDA/ONC collaboration)).
- Create a harmonized, interoperable platform and reusable tools (i.e., a data sharing framework) that will link the WHT-CRN registries to each other (ONC leads).
- Develop and ballot the HL7 FHIR® profiles incorporating the WHT-CRN data elements; conduct a pilot to demonstrate the capture and exchange of data within the CRN (ONC/FDA collaboration).
- Evaluate the completeness of the piloted HL7 FHIR® resources for meeting the project's research goals (FDA/NIH/ONC collaboration).

WHT-CRN Registries

- **COMPARE UF Registry.** A nationwide patient-centered outcomes registry of women with uterine fibroids funded by AHRQ and PCORI.
- **National Pelvic Floor Disorder Registry.** A private-public partnership for FDA mandated post-market surveillance of uro-gynecological mesh devices.
- **Collaborative Registry of Sterilization Therapies.** Registry infrastructure development supported through the CRN project. The registry will capture elective female sterilization therapies.

Contributions to PCOR Data Infrastructure Functionalities:

This project addresses standards and services across a number of the functionalities needed to build research data capacity.

- *Standardized Collection of Standardized Data.* The use of a data exchange standard like HL7 FHIR® provides an opportunity to collect structured, standardized, analysis-ready patient data at the point of care. Capturing standardized data will streamline data collection and support exchange of data across networks. The resulting data will be not only more consistent across organizations, but also more reflective of real-world evidence, such as supporting the inclusion of both the medication (using the clinical research standards and controlled terminologies) and the implantable device data (available through links to unique device identifier (UDI) data and meta-data available in the Global UDI Database). This standardization supports the increased use of clinical data captured and will allow researchers to collect longitudinal patient information and to link data sets with other relevant information for research (e.g., other research networks such as Sentinel, PCORnet, and available state data infrastructure).

Accomplishments and Deliverables:

The project continues to focus on engaging with end-users and has a robust data governance framework for securing buy-in from the end-user community. A critical component of this project is the collaborative governance structure established between the clinical and informatics teams, designed to ensure the final deliverables retain their clinical relevance and utility for clinical care, quality improvement, and device surveillance, as well as for research. The team is also developing a sustainability plan to ensure longevity of the WHT-CRN.

- Three clinical working groups were established to address each of the three device areas. The working groups are using a formal Delphi consensus process to create the core data sets from list of CDEs identified earlier in the project. The three working groups —Pelvic Organ Prolapse, Uterine Fibroids, and Sterilization/Long-acting Reversible Contraceptives—concluded the consensus review (Round 2 Delphi) and minimum core datasets for all three device areas using these CDEs have been finalized and are ready for review.
- A working group of informatics leaders representing the FDA, NIH, and ONC was formed to begin the process of translating the clinical data sets, definitions, and value sets into standard vocabularies used in data capture and exchange. The Informatics Working Group has moved onto standardizing and harmonizing the clinical data elements identified by each clinical working group. To support this next phase in the project, the Informatics Working Group has created a repository on MAX.Gov to help coordinate the data element standardization and harmonization efforts between the clinical and informative working groups. The repository includes a landscape analysis that examines the current state of women's health CRNs and reviews how the registries are able to capture, utilize, and share data. This helps to inform the feasibility of pilot studies.
- As part of the HL7 ballot process, the informatics team completed the HL7 Project Scope statement around the technical standards and process for capturing and exchanging the CRN core dataset. The HL7 balloting efforts included obtaining approval from working groups including Biomedical Research and Regulation (BR&R) in June 2018; Orders and Observation (O&O) in May 2018; Clinical Interoperability Council (CIC) in April 2018; and US Realm Steering Committee (USR-SC) in June 2018. The team developed and submitted HL7 FHIR® Implementation Guide: Women's Health Technology Coordinated Registry Network (CRN), Release 1 to HL7 as a "Comment Only" ballot. Balloting is complete, and the sponsoring HL7

workgroup BR&R passed a motion to accept all changes. Enhancement to the NIH CDE Repository capabilities to in support of needs for the HL7 Implementation Guide were developed, testing, and implemented. In August 2018, the team conducted a detailed walkthrough of the CRN implementation guide with the CRN community. The team completed the identification of comparable concepts across the different clinical working groups for the purpose of creating a harmonized set of CRN data elements. A form was created in the NIH CDE Repository to store the harmonized data elements.

- The team completed and posted the ONC/CRN Environmental Assessment and Pilot Analysis on the MAX.gov site. This assessment considers how the registries and working groups are able to use and share data. It also provides an update on the state of women's health CRNs.
- The team was successful in soliciting federal partners and women's health technology stakeholders to provide anecdotes of real-world applications of women's health technologies to inform the development of an HL7 FHIR® profile that relates to women's health technology issues.
- A patient partner was selected at the end of July to contribute the patient perspective to the project.

Disseminated Products:

- The most recent version of the HL7 FHIR® CRN Implementation Guide is publicly available here: <http://hl7.org/fhir/us/womens-health-registries/2018Sep/>
- Established a CRN project page under the HL7 BR&R Confluence site to track all ballot artifacts: <http://confluence.hl7.org/display/BRR/Biomedical+Research+and+Regulation>
- The team submitted a CRN Implementation Guide FHIR proposal: http://wiki.hl7.org/index.php?title=CRN_FHIR_IG_Proposal
- In collaboration with the Medical Device Epidemiology Network (MDEpiNet),⁴⁰ a public-private partnership to improve the infrastructure for medical device evaluation, the WHT-CRN project team developed a website (hosted by MDEpiNet) to raise awareness of the WHT-CRN project. The website is accessible at: <http://mdepinet.org/womens-health-crn/> and links to relevant working group activities. Additionally, the FDA project lead presented at the 2017 MDEpiNet Annual Meeting.
- MDEpinet, a collaboration of representatives from Cornell, ONC, NLM, and FDA, presented at ISPOR: International Society for Pharmacoeconomics and Outcomes Research in May 2018. A poster was on display for the WHT-CRN.
- NLM presented to the NIH Common Data Element Task Force on an overview of the project methodology for collaborating between informatics experts and clinical experts. Presenters also highlighted on-going project activities and accomplishments, which included leveraging the technical capabilities of the NIH CDE Repository.
- The core minimum dataset was prepared in REDCap to be presented at the WHT-CRN Annual Meeting, which took place on September 7, 2018.
- The team presented a CRN poster at the MDEpiNet Annual Meeting on October 22 and 23, 2018.

Coordination with Other Federal Agencies:

This multi-agency initiative has direct representation from other HHS agencies including AHRQ, CDC, and the Office of Women’s Health, on its working groups. Notably, there is cross-agency collaboration with AHRQ’s Outcome Measures Framework—a conceptual model for developing standard outcome measures. The WHT-CRN team has begun discussions on how to potentially leverage the findings and best practices from this project to inform AHRQ’s Outcome Measure Framework. Upon the project’s conclusion, NIH/NLM aims to include the data element resources within the NLM CDE Library, promoting the uptake of the standardize CDEs in future research projects.

FDA Period of Performance:

2/15/17 – 5/15/19

FDA Federal Point of Contact:

Danica Marinac-Dabic
Benjamin Eloff

NIH/NLM Period of Performance:

2/15/17 – 5/15/19

NIH/NLM Federal Point of Contact:

Lisa Lang

ONC Period of Performance:

2/15/17 – 5/15/19

ONC Federal Point of Contact:

Mera Choi
Farrah Darbouze

Development of a Natural Language Processing Web Service for Public Health Use

The Development of a Natural Language Processing (NLP) Web Service for Public Health Use is a joint project between the CDC and the FDA. NLP takes unstructured free text and codes clinical concepts into structured content that can be analyzed and used more readily.⁴¹ The CDC and FDA are leveraging a recently developed NLP system to translate unstructured, free-text data submitted to existing agency surveillance systems into standardized, structured form.

In the US, central cancer registries collect, manage, and analyze longitudinal data about cancer cases and cancer deaths. Cancer data are collected from multiple sources such as hospitals, laboratories, physician offices, and independent diagnostic and treatment centers. Hospital reporting of cancer cases has been standardized for over a decade; however, as the provision of cancer care has shifted away from the hospital, registries have had to expand their data collection efforts to include data from non-standard systems that contain large amounts of unstructured data. The CMS EHR Incentive Program, established under HITECH Act and since renamed the CMS Promoting Interoperability Program, established support for widespread adoption of EHRs and data and transmission standards to facilitate interoperable data exchange. The public health measures under Meaningful Use included specific cancer registry reporting requirements that facilitated the collection of structured cancer data from provider EHR systems. Additionally, over 90 percent of cancer cases are diagnosed using methods that generate pathology reports that are often documented in the form of unstructured text. The process of abstracting these crucial cancer data is very labor intensive and expensive.

Unstructured data limits the ability of researchers to analyze the information without manual review. Similarly, a considerable amount of clinical information submitted to the FDA Spontaneous Reporting Systems is unstructured. One of the FDA’s major responsibilities is the post-marketing safety surveillance through the review of spontaneous reports submitted to the Vaccine Adverse Event Reporting Systems (VAERS) and the FDA Adverse Event Report System (FAERS). The adverse events description in

VAERS and FAERS data are coded to the Medical Dictionary for Regulatory Activities (MedDRA) terms—a library of standardized medical terminology to facilitate sharing of regulatory information (e.g., registration, documentation, and safety monitoring) for medical products used by humans both before and after the product has been approved for use.⁴² However, a considerable amount of clinical information in both systems is either not coded (e.g., medical and family history) or is not linked to the MedDRA codes (e.g., exact time information for each symptom). Additionally, there may be duplicate entries for the same event, a phenomenon that impacts the surveillance process, requiring manual review of submitted reports to trace the adverse event.

Project Purpose and Goals:

This project proposes to develop an NLP web services environment that will be made accessible and publicly available to researchers to support conversion of unstructured clinical information to structured and standardized coded data. It will be made available on the Public Health Community Platform (PHCP), a cooperative platform for sharing interoperable technologies to address public health priority areas aimed at improving population health outcomes and health equity.⁴³ The NLP web services environment will contain NLP architectures and tools process spontaneous report narratives, extract clinical and temporal information from the text, formats the data for presentation, and map unstructured medical concepts (e.g., cancer data and safety surveillance data) into structured data (i.e., International Classification of Diseases 10th Edition Clinical Modification (ICD-10-CM), LOINC, SNOMED, and Medical MedDRA).

As a joint project, some project deliverables will be coordinated by a lead organization; as noted below.

The project objectives are to:

- Conduct an “as-is” environmental scan and literature review of all existing NLP algorithms, methods, and tools for possible inclusion in the NLP web service to receive unstructured clinical information and return standardized data needed for CDC cancer surveillance and FDA safety surveillance domains. The assessment also took into consideration possible requirements of other federal agencies, public health agencies, and/or PCORnet participant focus areas.
- Design and pilot the NLP Web Service technical requirements (CDC leads, FDA contributes).
- Build structured datasets using CDC and FDA resources to capture data and evaluate the performance of the pilot version of the NLP Web Service (CDC/FDA collaboration).
- Evaluate the pilot; and release the final NLP Web Service (CDC/FDA collaboration).
- Update the NLP Web Service and release the final version on the PHCP (CDC/FDA collaboration).

Contributions to PCOR Data Infrastructure Functionalities:

- *Use of Clinical Data for Research.* This project supports the use of narrative text data captured in electronic health systems for research by converting unstructured clinical information (e.g., cancer and safety surveillance data) into structured and standardized coded data. This is broadly applicable across health domains and will improve data quality for research.
- *Use of Publicly Funded Data Systems for Research.* This project uses publicly-available repositories by working with cancer registries and safety surveillance reporting systems. Cancer registries contain valuable information on cancer cases and deaths sent from various sources.

Many parts of medical records, laboratory reports, and other clinical reports submitted to these registries remain in narrative text format. Similarly, the FDA Spontaneous Reporting Systems (SRS) receives a substantial amount of clinical information on vaccine, blood products, and drugs that are not coded or linked to the appropriate MedDRA terms. This project aims to use the NLP Web Service to receive unstructured clinical information and return standardized data needed for CDC cancer surveillance and FDA safety surveillance domains. It will also provide federal agencies, public health agencies, and PCORnet participants with access to advanced NLP tools and pipelines to assist in their own research.

Accomplishments and Deliverables:

Since work on the project began in June 2016, the CDC and FDA project teams have completed multiple deliverables in support of the NLP web service or workbench, called The Clinical Language Engineering Workbench (CLEW).

- The FDA project team completed an environmental scan of existing NLP tools (systems, components, applications, and algorithms), which included a literature review and multi-channel review that identified existing clinical NLP systems that generate structured information from unstructured free text. The environmental scan provided a long list of tools that are potentially useful in building pipelines for various clinical NLP domains, including the FDA Event-based Text-mining of Health Electronic Records (ETHER) system for clinical text, the CDC Electronic Mapping, Reporting, and Coding (eMaRC) Plus text mining functionality, and the NLM MetaMap tool for mapping text to terminologies in the Unified Medical Language System (UMLS). This work is described in a "Report of the NLP environmental scan results, including a complete review and inventory of existing NLP algorithms" and a published report on "Natural Language Processing Systems for Capturing and Standardizing Unstructured Clinical Information."⁴⁴
- The FDA project team also initiated and completed the generation of an annotated corpus with 1,000 vaccine reports. The annotated corpus supports not only the training of language models and development efforts in the project is publicly available to the research community on GitHub here: <http://github.com/fda/VAERS-Annotations>.
- In preparation for the development of CLEW, the NLP workbench web service, the FDA team finalized the selection of the safety surveillance use cases that will be incorporated in the NLP workbench pilot (clinical and temporal entity recognition, identification of the temporal relations between the entities, and case summarization), and identified existing solutions to support these uses cases, and determined the required software development. The FDA built the NLP workbench prototype in the FDA's environment and shared knowledge and code with the CDC to support the development of the web-based NLP Workbench being developed by the CDC team. The "Core NLP Approach and Corresponding Report" describes the development efforts in detail and the NLP Workbench local prototype developed at the FDA.
- In November 2018, the FDA submitted its fourth deliverable on the CLEW. The report presents a detailed description of the technical specifications of the core NLP approach of the CLEW system. This includes the specific, motivating use cases for NLP across a range of clinical domains, as well as the tools that are currently available as part of the CLEW. Several of these tools are also separated into their constituent processing components, so that pipelines combining pieces of multiple tools can be constructed with the ultimate goal of identifying the best-performing components for each task or task category.

The CDC project team, with assistance from the FDA, developed the high-level conceptual architectural design of the CLEW. The CDC completed the design of the web portal page content for the CLEW, integrated the Language Application (LAPPS) Grid framework developed with funding from the National Science Foundation (NSF). The team installed the CDC/FDA tools, and developed a technical web services report that describes the architecture and functionality. The LAPPS Grid provides access to and deployment of language resources and processing functions available from servers around the globe. The team completed development of CLEW/LAPPS Grid prototype webpages with content and vetted with the CDC/FDA teams, and is developing CLEW/LAPPS Grid website for demonstration and use.

- The eMaRC Plus application is used by central cancer registries to receive and process cancer pathology and biomarker data that are received in unstructured narrative text format. The CDC further developed the eMaRC Plus application to interface with CLEW web services to process unstructured pathology data. The CDC completed a functional requirements document detailing enhancements of eMaRC Plus to interface with the CLEW web service titled “eMaRC Plus Functional Requirements.” The team completed development of the eMaRC Plus enhancements to call the CLEW pathology SER and pathology coding services to analyze the narrative pathology reports and return five key data elements for diagnosed cancer cases, including histology, primary site, laterality, behavior, and grade and codes the data using the International Classification of Diseases for Oncology (ICD-O) standard.
- The CDC team received all of the de-identified pathology reports for their pathology pilot project, giving them a total of 2,000 pathology reports to train the pathology language model in the CLEW. The team completed development of three language models that utilize the Stanford NLP, OpenNLP, and Gate tools to process narrative anatomic pathology laboratory reports and compared the output from each pipeline to identify the best model and tools. After developing language models and pipelines for analyzing pathology reports using NLP and machine learning techniques, the FDA completed annotation of 1,000 post-market vaccine reports. The team then installed the cTakes (clinical Text Analysis and Knowledge Extraction) tool onto the CLEW/LAPPS Grid to create a fourth language model and pipeline to process narrative pathology reports and compare outputs with the other three models and pipelines.

The CDC has collected de-identified pathology data for four primary cancer sites (breast, lung, prostate, and colorectal) from three national labs and the Cancer Canada. They have used this data to create a language model which will be used to train the NLP web service to process pathology data.

The final report presents the requirements, architecture, and implementation of the core NLP approach for the CLEW. The approach was designed to be flexible for adapting and incorporating NLP tools for many clinical activities across different domains, identified through the comprehensive environmental scan. A subset of these tools, chosen for their applicability to specific use cases for safety surveillance and cancer pathology processing, have been tested and made available through a CLEW interface, which allows for the construction of multiple NLP pipelines that can combine components of separate tools. All of the current tools are also accessible through straightforward Java APIs that allow them to be easily programmatically accessed by developers of other software tools and applications.

Disseminated Products:

- All project systems, code, and documentation have been loaded onto the CDC public GitHub at <https://github.com/CDCgov/NLPWorkbench>
- The Conceptual Architectural Design of the NLP Workbench Web Services has been developed and posted on the NLP Workbench Web Service URL at <https://www.cdc.gov/cancer/npcr/informatics/nlp-workbench/index.htm>. This document provides an overview of the Workbench, which includes user interfaces, web service components, NLP applications, and offline components. The CDC and FDA have updated the Technical Report that describes the architectural design and functionality of the CLEW that will integrate much of the LAPPS Grid infrastructure. An updated version of the Technical Report is available for review, but there may be revisions based on completion of model refinements and lessons learned from pilot implementations. The paper "[Generation of an annotated reference standard for vaccine adverse event reports](#)" was electronically published in a peer reviewed medical journal, *Vaccine* on June 5, 2018.⁴⁵
- The FDA and CDC, in collaboration with NIH/NCI, hosted a Clinical Natural Language Processing Workshop on December 18-19, 2017 in Atlanta, Georgia. The workshop focused on structured data extraction, classification, and consolidation from heterogeneous free text documents.
- The environmental scan resulted in literature review titled, "[Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review](#)" and was published in the September 2017 issue of the *Journal of Biomedical Informatics*.⁴⁶
- NLP Workbench Web Services was presented at the North American Association of Central Cancer Registries (NAACCR) June 2017 Annual Conference: [Development of Natural Language Processing \(NLP\) Workbench Web Services](#).⁴⁷
- Presented at the "Natural Language Processing and Machine Learning Workshop" organized by NIH/NCI in December 2016 and at an ASPE-organized webinar in April 2017.

Coordination with Other Federal Agencies:

The FDA and CDC continue their collaboration on the design and architecture of the NLP Workbench, as well as related dissemination activities. In April 2017, the CDC and FDA held a webinar, which was attended by over 60 participants from Federal agencies, state health departments, and universities who have expertise and/or an interest in NLP and machine learning. The insight gained during this webinar was used to further develop the design of the NLP Workbench Web Services. The FDA and CDC leads also met with the National Science Foundation to learn about their open source platform, LAPPS Grid, which is an interoperable web service platform for NLP research and development. They also discussed a possible collaboration with the National Science Foundation (NSF) to further expand the LAPPS Grid to incorporate clinical care and public health use cases. The CDC and FDA met with NIH/NCI several times throughout 2017 to continue their collaboration on this project.

CDC Period of Performance:

6/1/16 – 5/31/18

CDC Federal Point of Contact:

Sandy Jones

FDA Period of Performance:

6/1/16 – 5/31/21

FDA Federal Point of Contact:

Mark Walderhaug

Enhancing Data Resources for Researching Patterns of Mortality in Patient Centered Outcomes Research: Projects 1 - 4

An important objective of the OS-PCORTF is to build data capacity for PCOR in order to collect, link, and analyze data on outcomes and effectiveness from multiple sources. This group of projects consists of four independently-led but supportive components designed to enhance data resources for researching patterns of mortality in PCOR. They are being undertaken by the CDC, FDA, and CMS, in consultation with each other and ASPE. Mortality is an important outcomes in PCOR and efforts to better harmonize, connect, and enrich the federal mortality data through the four projects described below will accelerate its availability and utility for PCOR.

Project 1 (CDC) – Adding Cause-Specific Mortality to National Center for Health Statistics’

National Hospital Care Survey by Linking to the National Death Index. This project will leverage data from CDC’s NCHS and CMS to create new data infrastructures for PCOR to advance studies on mortality following hospital care by making “cause of death” available.

This project will provide the first-ever data linkage of EHR data from a nationally representative US sample to the NDI.

Together these linkages will enable the development of national estimates of cause-specific death rates following ED visits and/or hospital inpatient stays for specific conditions. In addition, this project will provide the first-ever data linkage of EHR data from a nationally representative U.S. sample to the NDI, enable evaluations of EHR and claims data on their quality and

complementarity, and create new approaches to optimize linkage by using the personally identifiable information available in claims and EHR data.

Project 2 (FDA) – Pilot Linkage of NDI+ to Commercially and Publicly Insured Populations.

Information on death and cause of death is often needed in PCOR, but administrative claims and EHR databases generally do not have complete capture of information on fact, cause, and manner of death. This project will develop a standard, repeatable, and efficient process for linking a distributed data network of commercial and public health plans with the NDI+. The capability to link distributed data networks like Sentinel or PCORnet to the NDI+ will enable many types of PCOR, including adverse event surveillance, predictive risk modeling, and comparative effectiveness research.

Project 3 (CMS) – Enhancing Data Resources for Researching Patterns of Mortality in Patient

Centered Outcomes Research. Currently, Medicare and Medicaid research data contain information on fact of death, but only certain years of Medicare data (1999-2008) and Medicaid data (1999-2007) contain information on the cause and manner of death. CMS will link the NDI data on cause and manner of death with claims data to produce updated research files with Medicare descendants data from 2008-2016 and Medicaid decedents data from 2007-2015. Linking health care claims data, including diagnoses with cause and manner of death enables many types of PCOR research, including descriptive epidemiology, predictive modeling to identify high value intervention targets, and comparative effectiveness. The benefit of linkage is particularly high for recent years of data given changes to the programs, for example, Medicare prescription drug benefits starting in 2006, and Medicaid expansion in many states starting in 2012.

Project 4 (CDC) – NDI Workshop and Strategy Paper.

NCHS will work together with FDA, ASPE, CMS, NIH, CDC Centers, states, and others to develop a long-term national strategy on access and use of NDI data particularly to support PCOR. NCHS will hire a Service Fellow expert in vital statistics and contract with the National Association for Public Health Statistics and Information Systems (NAPHSIS) to review/evaluate the outcomes of the MITRE work (Project 2) and to expand on that work in assessing the non-economic barriers limiting access to and use of the NDI data for PCOR. The NAPHSIS will also

conduct a state-level assessment of the laws/policies/preferences and perceptions of its members in restricting the use of the death data by the Federal Government and, given those restrictions, explore opportunities for minimizing the barriers to the use of NDI data.

The NDI is the only central data source containing information on both fact and cause of death for all deaths occurring within the U.S. It is heavily used by researchers in medical and health studies, for mortality ascertainment to assess the causes and risk factors of diseases and conditions and for the effectiveness of a wide range of interventions and drug therapies. Together, these projects attempt to build capabilities for systematic and more complete ascertainment of death information that is linkable, shareable across health systems, and more useful to PCOR researchers.

Project Purpose and Goals:

Project 1 (CDC) – This project will link health care claims and EHR data from the NHCS to death record information from the NCHS NDI; and separately to administrative data from the CMS Master Beneficiary Summary File (MBSF) and Chronic Conditions Warehouse (CCW). A data set that includes the 2014 NHCS claims and 2014/2015 NDI linked data will be available to the NCHS RDC.

Specifically, the objectives of this project are to:

- Link the 2014 NHCS inpatient and ED claims data to the 2014 and 2015 NDI (to ensure at least 12 months of post-discharge follow up for each hospital event).
- Link the 2014 CMS MBSF and CCW to 2014 and 2015 NDI.
- Link the 2016 NHCS inpatient and ED claims and EHR data to the 2016 and 2017 NDI.

Project 2 (FDA) – The project will create the capability to link distributed data networks to NDI+ with the goal of enabling many types of PCOR, and will address the logistical challenges of data linkage across multiple health plans through its primary objectives:

- To develop standard, repeatable, and efficient technical solutions for linking the NDI's death and cause of death data to large, publicly insured populations.
- To demonstrate the feasibility of linkage by using use cases to assess associations between select medications and death or cause of death as an outcome.

Project 3 (CMS) – The primary objectives were to:

- Develop standard, repeatable, and efficient technical solutions for linking the NDI's death and cause of death data to large, commercially and publicly insured populations.
- Demonstrate the feasibility of linkage by using use cases to assess associations between select medications and death or cause of death as an outcome.

Project 4 (CDC) – The primary objective of Project 4 is to examine the barriers to the access and use of NDI for PCOR (economics, legal, statutory, technical, operational, policy, etc.) and then, using a national workshop approach, attempt to develop a long-term strategy for NDI acceptable to the federal researchers and the states. At a minimum, these meetings and workshop will examine and make recommendations for:

- Altering the economic model used to support NDI
- Identifying and minimizing the non-economic barriers to accessing and using NDI

- Re-using cause of death data (e.g., sharing the data with other research proposals; multiple use of approved data for other research studies)
- Improving the efficiency of the administrative aspects of linkages
- Improving the timeliness and quality of the NDI data
- Improving the timeliness of the NDI approval process

The outcome of this project will be a report on the long-term strategy for NDI.

Contributions to PCOR Data Infrastructure Functionalities:

Project 1 (CDC):

- *Use of Clinical Data for Research/Linking of Clinical and Other Data for Research.* This project aims to link multiple types of data including EHR data, claims data and NDI data which will support research related to the inpatient setting as well as post-discharge care and outcomes.
- *Use of Enhanced Publicly-Funded Data Systems for Research.* The linkages described in Project 1 will enhance the value of NDI data as well as administrative data sets including the CMS MBSF and CCW.

Project 2 (FDA):

- *Linking of Clinical and Other Data for Research.* In the immediate future, the data created by the linkages from this project will be available to researchers in participating health plans under contract to FDA according to the data use agreements required by the NDI Program. This addresses a gap identified in the portfolio-wide evaluation of “Developing technical services and standards for services that allow patient data to be securely linked to other data sources.”

Project 3 (CMS):

- *Linking of Clinical and Other Data for Research.* Project 3 supports the further development of key federally-initiated data systems for research. It also supports the enhancement of strategic publicly-funded data systems (including CMS data) to facilitate their access and use, and ease retrieval of data for research purposes.

Project 4 (CDC):

- *Use of Enhanced Publically-Funded Data Systems for Research.* This project supports the enhancement of strategic publicly-funded data systems (including CMS data) to facilitate their access and use, and ease retrieval of data for research purposes, and in particular a coordinated, multi-agency long term strategy for this pursuit.

Accomplishments and Deliverables:

Since work on the projects began in February 2017, the project teams have made progress on several early deliverables, and the Project 3 team has concluded its work.

Project 1 (CDC):

The CDC team has made significant progress in linking NHCS data to NDI data. The team has used multiple linkage methodologies to link the 2014 NHCS data collected from inpatient and emergency

department settings to the 2014 and 2015 NDI data. The team is currently evaluating the different matching methodologies and has proposed changes to both streamline and increase the match rate. They are developing an alternative methodological approach for partial agreement weight scenarios.

- The linked 2014 NHCS claims – 2014/2015 NDI data is currently available at: https://www.cdc.gov/nchs/data-linkage/nhcs_linkage.htm
- The report on 2014 NHCS claims – 2014/2015 NDI data linkage is available here: <https://www.cdc.gov/nchs/data-linkage/nhcs-ndi.htm>
- The team has also begun the linkage process and report on the 2016 NHCS – 2016/2017 NDI linkage.
- National Center for Health Statistics. Office of Analysis and Epidemiology. The Linkage of the 2014 National Hospital Care Survey to the 2014/2015 National Death Index: Methodology Overview and Analytic Considerations, April 2018. Hyattsville, Maryland. Available at the following address: https://www.cdc.gov/nchs/data/datalinkage/NHCS14_NDI14_15_Methodology_Analytic_Consider.pdf

Project 2 (FDA):

Administrative progress: The FDA made a contract award to Harvard Pilgrim Health Care who has begun the process to initiate subcontracts with health plans and scientific collaborators and the process to initiate conversations with NDI to ensure that the single IRB process is accepted by NDI as part of the application development. In the past 3 months, the contractor has continued to work with study sites on the specific requirements necessary to obtain approval for a central IRB. The contractor has also worked with NDI on multiple issues in preparation for an application.

Scientific progress: Protocol development is underway for the proposed technical approach (methods and processes) for linkage and the study team has discussed the selection of use cases from regulatory and scientific perspectives. FDA and the contractor reached consensus on various technical issues for protocol development and the final draft protocol was submitted to FDA on 6/28/2018.

Project 3 (CMS):

CMS has successfully linked the NDI's death and cause of death data with the MBSF and MESF files.

- CMS has completed validating and updating the MBSF and MESF files with the NDI segment for Medicare descendants data from 2008-2016 and Medicaid decedents data from 2007-2015. All these documentation and changes were updated on the ResDAC and CCW websites. The data created by the linkages will be made accessible to researchers in participating health plans under agreement with FDA. A final report will be delivered to ASPE. CMS hopes to create an annual report that details how many times each file was requested and by which organizations.
- The files were publicly posted on the ResDAC and CCW websites in April 2018. The ResDAC files are available here: <https://www.resdac.org/cms-data/files/mbsf> and the CCW files can be found here: <https://www.cwdata.org/web/guest/data-dictionaries>.
- As of the end of FY18, CMS has accepted and approved 13 requests for the updated NDI data and the linked data has been provided to researchers. Some example studies and topics that were approved for the data include:
 - Surveillance of Cancer and Blood Disorders in the General Medicare Population

- Medicare Co-Insurance and Risk of Death From Colorectal Cancer
- State Innovations Model All-Payer Operations
- Surveillance of Diabetes, Prediabetes, and Prevention Efforts
- Development and Validation of a Risk Stratification Index Based Suite of Predictive Algorithms to Analyze Health Care Outcomes and Costs
- Enhancing the Measurement of Health Care Quality to Improve the Value of Care in Medicare and Medicaid

Project 4 (CDC):

The CDC reports progress on several important contracting, planning, and information gathering activities: NCHS awarded a contract to the NAPHSIS to conduct an assessment of the barriers to NDI and the rationale for those barriers from the perspective of the states. NAPHSIS has fielded the survey of the perception of state vital registration officials of the laws and statutes governing access and use of the mortality data from the state perspective, garnering an 88 percent (50 out of 57 jurisdictions) response rate.

NCHS held a national NDI Workshop on March 14 and 15 at the NCHS facility in Hyattsville, Maryland with state registration officials, Federal researchers, and academic researchers. The workshop was facilitated by staff from the Illinois Public Health Institute and focused on the barriers to accessing and using the NDI. Prior to the workshop, NCHS held a webinar discussing the access and use issues identified in the NAPHSIS survey and the NCHS stakeholder interviews. To inform NDI improvement activities, NCHS conducted a series of interviews with federal researchers and the NDI Advisory Committee about their experiences with NDI and the issues from their perspectives. CDC also convened the national NDI Workshop from March 14 – March 15, 2018. Findings from the assessment, interviews, and Workshop are being used to develop a draft strategic plan to outline the issues identified by all the stakeholders and the strategies to resolve the identified issues. The CDC published its final report on Project #4 in June 2018, identifying key issues with the NDI and proposed mitigation strategies. These issues relate to the NDI application process, a range of data issues, regulations and policies, and the cost of NDI services.

Following the publication of this report, the CDC convened a series of meetings with agency partners to begin improving the NDI, including forming and convening a Federal Interagency Workgroup on Mortality Data System Improvement; meeting with NIH leadership to discuss the NDI Strategic Plan and the possibility of a joint initiative; meeting with the NIH All of Us Program to talk about their project and how the use of NDI may support their needs; securing funding from NCHS to begin development of an NDI Portal to improve/streamline the application process for NDI matching services; and securing NIH buy-in for NCHS-NIH enhancement of the NDI.

Disseminated Products:

- CDC posted a methodological report describing the linkage of the 2014 NHCS claims to the 2014/2015 NDI. The report includes a brief overview of the data sources, a description of the methods used for linkage, and analytic guidance to assist researchers while using the files. This report and resources are available here:
 - Report: <https://www.cdc.gov/nchs/data-linkage/nhcs-ndi.htm>

- Methodology: https://www.cdc.gov/nchs/data/datalinkage/NHCS14_NDI14_15_Methodology_Analytic_Consider.pdf.
- Linked Data Files: <https://www.cdc.gov/rdc/index.htm>
- A codebook for the 2014 NHCS claims data linked to the 2014 and 2015 NDI file is available on the NCHS RDC, and is available here: https://www.cdc.gov/nchs/data/datalinkage/LMF2015_DataDictionary.pdf
- The linkage of the 2014 NHCS data to death certificate records from the 2014 and 2015 NDI was featured in the April 27, 2018 OS-PCORTF program update email. The update also mentioned the file can be accessed through the NCHS RDC network.

CDC Project 1 Period of Performance:
2/1/17 - 1/31/21

CDC Project 1 Federal Point of Contact:
Delton Atkinson

FDA Project 2 Period of Performance:
2/1/17 - 7/31/19

FDA Project 2 Federal Point of Contact:
Greg Pappas; Robert Ball

CMS Project 3 Period of Performance:
3/1/17 - 4/29/18

CMS Project 3 Federal Point of Contact:
Andrew Shatto

CDC Project 4 Period of Performance:
2/1/17 - 1/31/21

CDC Project 4 Federal Point of Contact:
Delton Atkinson

Harmonization of Various Common Data Models and Open Standards for Evidence Generation

In order to achieve a sustainable data network infrastructure, promote interoperability and foster the creation of a Learning Health System as laid out in the ONC Connecting Interoperability Roadmap,⁴⁸ there is a need to map and transform data across various Common Data Models (CDMs) and leverage open-source standards. The Harmonization of Various CDMs and Open Standards for Evidence Generation project is a collaborative project among the FDA, NIH/NCI, NIH/NCATS, NIH/NLM, and ONC. This project aims to leverage existing OS-PCORTF investments, including DAF and the NIH CDE Repository, as well as existing infrastructure (e.g., the NCI cancer Data Standards Registry and Repository (caDSR) and the NLM Value Set Authority), which can help with building reusable data mapping and transformation services. By mapping various CDM data elements and leveraging existing OS-PCORTF investments, it is feasible to reuse the data, methods, and other resources from existing networks (i.e., PCORnet, Sentinel, the Observational Health Data Sciences and Informatics (OHDSI), and i2b2).

After mapping the CDMs, the identified observational data will be utilized to evaluate factors associated with the safety and effectiveness of newly approved oncology drugs (i.e., immune checkpoint inhibitors, PD1 and PDL1) in combination with other immunotherapy agents. Oncology as a field has recently been transformed by the emergence of immunotherapy in cancer care. As immunotherapy represents an entirely new modality of therapy which has not been commonly used until now, our information about the experience of treating cancer patients with these drugs is largely limited to the clinical trials setting. Observational data is lacking, and in the real world, patients have a wide array of comorbid conditions that may affect the safety and efficacy of these drugs and may not be identified in clinical trials.

By harmonizing various CDMs, the research community will have an opportunity to aggregate more data, resulting in access to a larger sample size and additional demographics (e.g., elderly, pediatrics, non-US population, etc.). For example, PCORnet has a network of approximately 70 million patients from mainly EHR sources. Sentinel has access to 350 million patients from mainly administrative claims plans. By harmonizing the CDMs in these two networks, the researchers will have access to not only EHR data, but to administrative claims data. Researchers will have a unified access tool to multiple networks of observational data. The much larger and representative patient cohorts also have a greater breadth of information that is useful for rare events.

In addition, tools and programs developed by one network of observational data can be re-used by other networks. This solution has the potential to support a diverse range of exploration from basic PCOR queries, to data mining to generate hypotheses for future research, to large-scale sophisticated analysis, including randomized clinical trials.

Project Purpose and Goals:

This project will harmonize CDMs developed by four different networks: Sentinel, PCORnet, OHDSI and i2b2-ACT. The process will involve architecture design and software development, creation of queries, CDM mapping, etc. Ultimately, this harmonization will enable researchers in federal agencies or academia to have access to data from a larger network of patients.

Although FDA serves as the lead for the project, the scope of the project is shared by a number of agencies working with FDA. The FDA will initially focus on the data elements required to assess the safety of newly approved oncology drugs in combination with other immunotherapy agents. The team at NIH/NCATS will serve as co-leads on this project and will provide informatics and implementation strategy. The SMEs at NIH/NCI will serve as oncology researchers, who will evaluate the proposed solution, identify the data elements needed to assess drug safety, and test it against the oncology use case. NIH/NLM, as the designated central coordinating body for clinical terminology standards within HHS, will provide its expertise, tools and assistance to assure the general applicability and utility of the project products for the PCOR community. ONC will serve as advisors in data standards and proposed approach. Once initial data model mapping are complete from the chosen data models to FHIR®/DAF, ONC will convene discussions among data model owners to ensure that these mappings are reflected accurately.

Project deliverables include a mapping tool to serve as the common data architecture; a web-based portal accessible to researchers to map their source data into the target data; and analytics and visualization tools for researchers will be developed as part of this project.

The overall project objectives are to:

- Develop common data architecture as the intermediary between the four research network CDMs
- Develop a flexible data model that can be used to create outbound data in multiple formats for multiple purposes
- Test the common data architecture by using it to study factors associated with the safety and effectiveness of newly approved oncology drugs that boost patients' immune response to cancer
- Establish methods and develop processes, policies and governance for ongoing curation, maintenance and sustainability of the common data architecture, building upon existing resources, standards and tools

Contributions to the PCOR Data Infrastructure Functionalities:

- *Use of Clinical Data for Research.* This project will ultimately provide PCOR researchers with access to larger and more diverse types of observational data, or data derived from the delivery of health care in routine clinical settings (with appropriate data partner permissions). The enhanced data infrastructure created through this project will have the capacity to support evidence generation on patient-centered outcomes that can inform regulatory and clinical decision making within federal programs.
- *Standardized Collection of Standardized Clinical Data.* A key dimension of this project is to develop standards that support secure, electronic query of structured data across clinical research and delivery systems, including standards for open-source access.
- *Linking Clinical and Other Data for Research.* This project will enable the harmonization of several existing CDMs in order to support research and analyses across multiple data networks. The aim is to advance the utility of data and its interoperability across networks to facilitate PCOR.

Accomplishments and Deliverables:

The project team has made good progress on the aims and associated tasks.

FDA

As an initial step to developing the CDM Mapping Tool, the FDA, in collaboration with source CDMs, NIH/NCATS, and other experts, documented existing CDM artifacts (environmental scan) and their mappings into an aggregate resource. ONC completed an evaluation report of these activities that addressed: 1) the feasibility of using an intermediary data model for data extraction (the Federal Health Architecture's Federal Health Information Model) and 2) the feasibility of using Model Driven Health Tools in Model-Driven Message Interoperability as tooling capabilities for model mapping and transformations.

- FDA and NHI/NCI oncologists finalized the protocol-based assessment for the safety of immunology products evaluation use case.
- The FDA is actively pursuing approval from HL7 for their harmonization standards and is developing new HL7 FHIR® extensions.

NIH/NCATS

- The project team, led by NIH/NCATS, surveyed the market for an existing open source extract, transform and load (ETL) software tool to automate the data mapping process and documented their efforts via an environmental scan. After completing an evaluation of existing ETL tools, the team opted to partner with a software and data integration company, to advance this work. This partner will also provide the analytic and visualization package.
- They have built a variety of tools, including a Query Builder that allows researchers to build a query and search for data; a Query Translator that translates a research query into the syntax necessary to communicate with each CDM partner; Query Transformation tool protocols for data return and testing of responses; integration and transformation of the data; and a Query Results Viewer. All of these components are underway and at various stages of completion.

NIH/NCI

- The CDMs were mapped to the selected transition model, the Biomedical Research Integrated Domain Group (BRIDG) model, and are being reviewed by the team. NCI will use this mapping to load the model data elements into caDSR. Information from that test will also help customize the ETL tool. The team completed concept annotation for BRIDG 5.1 (balloted version of BRIDG that addresses CDMH requirements).
- Key components were identified and procured: an ETL solution was identified, purchased and installed; analytic and visualization package was identified, installed, and tested; and a software version control tool was implemented. The ETL was customized to support data mappings and transformations, the mappings are now complete and have been tested.
- The web-based portal and user interface has been designed and made fully functional. Testing is ongoing and will not be considered complete until the project ends.
- The team completed metadata curation of all 4 CDMs: Sentinel, PCORV4.0, OMOP, I2b2ACT, which will allow them to complete the registrations of the CDMs and BRIDG in caDSR and then with NLM. The CDMs have been incorporated into caDSR, which will be a lasting product, available into the future.

NIH/NLM

- NIH/NLM conducted an initial assessment and exploration of adjustments needed to the NIH CDE Repository to support use of the standards as envisioned by the CDM harmonization, and collaborated on the initial strategic plan for needed educational resources to support the CDM tool in the context of other data and terminology resources needed by the NIH clinical research community. This led to the selection of the collaboration suite, population, and ETL tools.
- In the most recent quarter (Q3), NIH/NLM completed an initial draft of documentation about the CDMs in order to understand the data access rules, retention, privacy, and other matters needed for users to interact and use the data. NLM also completed an initial draft test strategy for user acceptance testing in order to understand how users interact with and what they want for the system and incorporated it into the Draft Test Plan.
- NIH/NLM has also been participated in technical discussions with ONC and technical experts about the need for and possible formats of a FHIR® export from the query tool; it was determined that the FHIR® export model would not be mature enough due to need for extensions. The outcome of discussions resulted in the development of a requirement for NLM to create a CSV export from Adeptia which would have inherent capabilities to export data in multiple formats including CSV and JSON. NLM also continued enhancements to the NIH CDE Repository to support the ingestion of caDSR/NCI modeled data elements and support the use of FHIR®.
- NLM continued planning for the development of educational resources through holding stakeholder reviews of the planned assessment of the data analytic and terminology services needed by NIH and FDA intramural researchers. Based on stakeholder feedback the assessment questions were finalized, input into Survey Monkey, a communication plan was established, and survey participants were identified.

ONC

- NIH/NLM initiated an assessment and exploration of adjustments needed to the NIH CDE Repository to support use of the standards as envisioned by the CDM harmonization, with assistance from ONC. ONC and NIH completed a draft mapping between 4 data models and the

NIH conceptual model Biomedical Research Integrated Domain Group (BRIDG) conceptual model and from BRIDG to FHIR®. Extensions to both BRIDG model and FHIR® resources have been submitted through respective governance systems for approval. Finally, ONC is involved in the selection of the ETL tool and configuration underway to support research query formation.

Disseminated Products:

- FDA: The Harmonization of Various CDMs project was presented at the Drug Information Association (DIA) 2017 Annual Conference (June 2017).

Cross-Agency Collaboration:

The project team has engaged with the FDA Cross-Network Directory Service project to explore the re-use of that investment to further this CDM project and better understand the complexities of mapping query requests across research networks. The project team has also presented within their respective agencies to raise awareness of the work (e.g., FDA CDER Real Work Evidence Workgroup and NCI Genomic Data Commons). NCATS have been very involved in leveraging components from ONC US Core/DAF project and has regular meetings with PCORI to discuss and share materials.

FDA Period of Performance:

2/8/17 – 5/7/19

FDA Federal Point of Contact:

Mitra Rocca

NIH/NCATS Period of Performance:

2/8/17 – 5/7/19

NIH/NCATS Federal Point of Contact:

Kenneth Gersing

NIH/NCI Period of Performance:

2/8/17 – 5/7/19

NIH/NCI Federal Points of Contact:

Denise Warzel and Elad Sharon

NIH/NLM Period of Performance

2/8/17 – 5/7/19

NIH/NLM Federal Point of Contact:

Lisa Lang

ONC Period of Performance:

2/8/17 – 5/7/19

ONC Federal Point of Contact:

Albert Taylor

Privacy and Security Blueprint, Legal Analysis and Ethics Framework for Data Use, and Use of Technology for Privacy

The Privacy and Security Blueprint, Legal Analysis and Ethics Framework for Data Use and Use of Technology for Privacy project is a joint project between the CDC and ONC. Patient-level data are essential to understanding and improving health outcomes. These data must be made available to researchers in a way that ensures the protection of patient privacy while providing sufficient granularity to allow meaningful research questions to be assessed. However, current laws and policies around the use of patient level data are nuanced and sometimes conflicting, creating confusion for researchers, providers, and patients.

Making PCOR data-sharing a reality requires addressing many privacy and security-related policy issues. This project addressed these issues by: 1) conceptualizing and developing a privacy and security data infrastructure blueprint; 2) conceptualizing and developing the legal analysis and ethical frameworks needed to balance individual privacy rights with data use, sharing, and disclosure for PCOR; and 3)

identifying, refining, harmonizing, validating, recommending, and piloting standards that support individual consent and preferences for research.

Project Purpose and Goals:

ONC developed a privacy and security data infrastructure blueprint, legal analysis, and ethical framework to address the many legal and privacy and security-related policy issues that affect the use of data for various types of PCOR. The CDC lead an analysis of the public health legal and ethical analysis for use broadly, which serves as a supplement to the overall privacy and security data infrastructure blueprint developed by ONC. The CDC analysis explored the agency's traditional data collection activities in the context of supporting individually identifiable data for PCOR.

The CDC project objectives were to:

- Conduct an “as-is” analysis of public health laws and ethical considerations that relate to the release and use of CDC public health and health surveillance data. These data include public health data generated or collected by the CDC and generated or collected by other agencies or organizations using CDC funds. The “as-is” analysis includes an environmental scan, analysis of findings, and gap analysis, each of which will result in a report.
- Convene a work group of public health legal and ethics experts to inform the “as-is” analysis and provide a “view” of agency wide data use and report of findings.
- Prepare a final white paper of findings and recommendations for best practice practices for data release beyond CDC, and provide guidance to researchers interested in using CDC data for PCOR.

ONC project objectives were to:

- Develop a *Privacy and Security Data Infrastructure Blueprint*: This portion of the project identified common PCOR use cases and diagramed user scenarios/stories. These use cases facilitated analysis of the privacy and security requirements for PCOR infrastructure based on the identified user stories. ONC engaged with a broad spectrum of PCOR stakeholders to identify and diagram these use cases.
- Develop a *Legal Analysis and Ethical Framework*: The purpose of this framework is to enable PCOR researchers to analyze and evaluate their data needs against federal privacy, public health, and security laws. This means providing researchers with a consistent understanding of the legal obligations related to privacy, confidentiality, and security of data to determine if and how PCOR can support those obligations. This framework also includes a legal and regulatory gap analysis of privacy and security protections, models of data flows mapping to privacy and security legal requirements, identification and definition of updates to policies and minimum requirements for de-identification and re-identification, and ethical implications (CDC lead) for PCOR data use.
- Conduct an electronic consent technical functionality *landscape assessment and gap analysis* for the purpose of identifying and piloting standards to capture, manage, and communicate consent preferences for health information research.
- Develop and *pilot standards* for capturing and sharing individual patient consent preferences for PCOR.

Contributions to PCOR Data Infrastructure Functionalities:

- *Standardized Collection of Standardized Clinical and Claims Data:* The project explored use cases focused on the collection and use of clinical and administrative data.
- *Collection of Participant-Provided Information.* The project explored several use cases focused on the collection and use of PPI for PCOR.
- *Linking and Use of Clinical and Other Data for Research.* The project explored use cases focused on the linking of clinical data with other data sets, held by different business entities.
- *Use of Clinical Data for Research.* The project explored use cases that highlight the privacy and ethical considerations of using clinical data for research. These are all gap areas identified in the OS-PCORTF portfolio evaluation.

CDC Accomplishments and Deliverables:

The CDC portion of the project has produced several important deliverables which are described in more detail below.

The legal and ethical environmental scan and analysis of findings reports are complete. The key take-aways from these reports focused on data de-identification as a useful tool to further PCOR and the need for clarity surrounding the use of a single IRB versus multiple IRBs. CDC also completed a gap analysis to determine whether the current state of laws and ethical considerations that govern the release and use of CDC public health data address specific public health or health care research needs. Finally, the work concluded with CDC creating a framework that analyzes hypothetical scenarios of legal and ethical uses around its data and provides assistive questions to guide readers through these considerations. This framework can be found in the CDC's published White Paper, the *Legal and Ethical Framework to Use Centers for Disease Control and Prevention Data for Patient-Centered Outcomes Research*. The framework is summarized below.

The framework attempts to identify the tools CDC needs in order to determine how its data can be optimized to serve the public and therefore achieve the agency's overall mission. Specifically, the report describes various data use scenarios, putting the application of laws and ethics into practice.

The framework provides a systematic approach to review data sharing policies, build PCOR data infrastructure, and share data for PCOR proposals. It seeks to address three main questions:

1. "Can It?" Does CDC have the legal authority to share data as requested?
2. "Must It?" Is CDC required to share data as requested?
3. "Should It?" If the CDC has legal discretion to share data, should it do so? In this regard, what are the ethical considerations that support or weigh against providing the requested data?^[1]

Overall, this paper addresses PCOR governance infrastructure needs by evaluating ethical and legal principles that PCOR data users must consider when working with CDC data. As PCOR research gains more attention, it will be critical to have established practices in place to ensure the legality and ethicality around the use of data for research.

Disseminated Products:

- The final white paper was completed with input from the CDC Workgroup. The report is available publicly on ASPE's website:
https://aspe.hhs.gov/system/files/pdf/259016/PCOR_Legal_508_2.pdf
 - Centers for Disease Control and Prevention. (Prepared by Chrysler D, Girish Pai S, Limeres A, Homes C, et al.) Legal and Ethical Framework to Use Centers for Disease Control and Prevention Data for Patient-Centered Outcomes Research.
- Presented on the Ethical Framework to an internal CDC audience on November 17, 2017 as the second speaker as part of an ethics seminar.
- Presented to external audience at the PRIM&R Advancing Ethical Research Conference in San Diego on November 8, 2017. Audience included members from IRB's and other research institutions, as well as federal agencies. Title of panel presentation: The Boundaries of Privacy and Public Health Concerns.

ONC Accomplishments and Deliverables:

With this project, ONC undertook multi-pronged advancements to both understanding the legal and ethical constraints on data sharing for PCOR, and enhancing PCOR data infrastructure through the development and piloting of standards for consent. Phase 1 focused on the legal and ethical issues that researchers encounter when conducting PCOR. The privacy and security analysis and legal gaps analysis resulted in a practical user-facing guide to help researchers navigate this landscape. The technical portion (Phases 2 and 3) consisted of an analysis of the technical gaps and requisite functionalities involved in patient consent. Standards were developed to collect and share patient consent for research and successfully piloted at two sites. The technical team produced an implementation guide and white paper to guide others in using the consent standards.

The development of the Legal and Ethical Architecture involved a multi-phase process, informed by subject matter expert input and discussions on the development of PCOR data use scenarios, and a policy and legal and ethical gap assessment. The Architecture is a collection of tools and resources to help researchers navigate the complex legal and regulatory environment involved with collecting, aggregating, and using patient data for research.

For the technology-related components of the project, or The Patient Choice Technical Project, in Phase 1 ONC completed the environmental scan and gap analysis of consent management technologies for research. Informed by its environmental scan, ONC completed the 'Basic Consent' (Treatment, Payment, and Operations) pilot testing in cooperation with the Veterans Administration. This tested the functional requirements and scenarios described in the use case development phase, allowing veterans to proactively manage their privacy and exercise control over the sharing of their data. Based on these results, ONC developed a best practices implementation white paper describing the lessons learned from the pilot efforts, including policy, business, technical, and standards challenges and gaps for ONC consideration.

During Phase 2: Basic Choice for Research consisted of supporting the mapping of informed consent to the FHIR® Standard. REACHnet completed building its backend infrastructure mapping a consent template for all appropriate pediatrics cases and demonstrated the use of the standard in January 2018. In September 2018, the Basic Choice for Research Consent pilot concluded. During the pilot, the REACHnet development team gained experience working with FHIR® and integrating the standard into REACHnet's infrastructure.

For Phase 3: Granular Choice, the FHIR® standard was implemented for capture and exchange of consent and health information. MiHIN served as the pilot organization and completed the project and demonstration in June 2018. Patients' granular consent was clearly demonstrated using the contract resource, DSTU2 model, as the standard. Some of the highlights from this phase include patients being able to specify and then update consent preferences via a patient portal with which information they would like shared (e.g., transition of care information, medical visit information, and medication information). They were also able to revoke consent for sharing this information, which was then transmitted to other connected organizations. The white paper describing the standard specifications and the pilot findings is currently under review by ONC. The intent of the paper is to guide others in using the consent standards within a health information exchange environment. ONC has continued to map the informed consent to the FHIR® standard.

Disseminated Products:

- The ONC Privacy and Security Blueprint, Legal Analysis and Ethics Framework for Data Use, and Use of Technology for Privacy is available here: <https://www.healthit.gov/sites/default/files/page/2018-02/Legal-and-Ethical-Architecture-for-PCOR-Data-Final.pdf>
- ONC published the “Legal and Ethical Architecture for Patient-Centered Outcomes Research (PCOR) Data (“Architecture”) available here: <https://beta.healthit.gov/topic/legal-and-ethical-architecture-patient-centered-outcomes-research-pcor-data-architecture>
- A Live Community Pilot of the Standards Solution was presented at HL7 Educational Session on June 29, 2017. This presentation reviewed the Phase 1: Basic Choice for TPO piloting efforts and outcomes as related to HL7 FHIR®.
- Patient Choice Demo at HIMSS 2017
- Hosted HL7 FHIR® Connectathon Consumer Centered Data Exchange demonstration in collaboration with project partners. Project members presented with team members from the Basic Choice for Research and Granular Choice pilots two presentations at the Connectathon: “ONC Patient Choice Pilot: HloH Kids” and “eConsent Connectathon Summary” respectively.

Coordination with Other Federal Agencies:

The team held regular meetings with the ONC Patient Choice Technical Project, which was analyzing and developing standards to fulfill the technical capability for implementing and sharing individual consent (basic and granular choice) for health information exchange in healthcare and research settings.

CDC Period of Performance:

7/10/15 – 9/30/17

CDC Federal Point of Contact:

Carissa Holmes

ONC Period of Performance:

6/16/15 – 9/30/18

ONC Federal Point of Contact:

Rose-Marie Nsahlai, Stephanie Garcia

Technologies for Donating Medicare Beneficiary Claims Data to Research Studies

This project is a collaboration between NIH and the CMS to leverage the Sync for Science™ (S4S) and Blue Button 2.0 programs to enable Medicare beneficiaries to donate their medical and prescription drug claims data for scientific research studies using the S4S data donation workflow.

The *All of Us* Research Program (formerly known as the Precision Medicine Initiative Cohort Program) is a national, large-scale research enterprise designed to support discoveries that increase our ability to better treat and prevent disease by enrolling one million or more volunteers. S4S is beginning a provider pilot, using HL7's FHIR® standards and CMS' Blue Button 2.0 API to significantly improve the ease of data donation from Medicare beneficiaries to *All of Us*. The FHIR® framework translates data into a standardized, structured format that can be accepted by a wide range of applications and the Blue Button 2.0 API simplifies the data transfer.

In practice, the implementation of S4S FHIR® and the Blue Button 2.0 API will enable beneficiaries to directly donate their Medicare Parts A, B, and D claims data to research studies in which they choose to participate. These data would be highly valuable to the *All of Us* Research Program, and the lessons learned through this pilot phase could be applied to future efforts that enable individuals to donate their Medicare claims data to the larger medical research community to support a wide range of healthcare studies.

Project Purpose and Goals:

This project aims to provide a safe and secure mechanism for Medicare beneficiaries to donate at least three years of their individual Medicare claims data to scientific research studies. NIH is the lead agency, developing the technical guidelines and CMS will leverage its previous work (i.e., a FHIR®-based API, launched as Blue Button 2.0, to enable beneficiaries to connect their Medicare claims to third-party applications and services they trust) to allow data donation to S4S/*All of Us*.

The two teams will work in concert to achieve the following specific deliverables:

- Finalize a CMS-defined profile for the FHIR® financial resources, including (at least) Explanation Of Benefit (EOB) in FHIR® Standard for Trial Use version 3 (STU3) format.
- Expand S4S resource definitions to include this CMS profile on the FHIR® EOB resource.
- Ensure that CMS's Blue Button API supports the S4S permissions and API calls to retrieve EOB and patient resources.
- Ensure that CMS's Blue Button API is compatible with the S4S authorization specifications (SMART on FHIR®).
- Deploy S4S support into the CMS Blue Button 2.0 API for a pilot patient population.
- Conduct an S4S pilot in which CMS Medicare beneficiaries donate their claims data to *All of Us* via the CMS Blue Button 2.0 API.

Contributions to PCOR Data Infrastructure Functionalities:

This project addresses the services and standards components and two of the five functionalities of the HHS Strategic Framework:

- *Linking of Clinical and Other Data for Research.* This project will allow researchers to collect longitudinal patient information from Medicare and to link data sets with other relevant information for NIH-led research.
- *Use of Enhanced Publically-Funded Data Systems for Research.* This project leverage current investments in federal data infrastructure to inform future infrastructure development—combining advances in Blue Button on FHIR® (Blue Button 2.0) and S4S to enhance data collection by the *All of Us* initiative.

Accomplishments and Deliverables:

The project team has made significant progress on completion of each of the deliverables:

- CMS launched the Blue Button 2.0 production API on March 6, 2018. This means that app developers can now request production access to the API, which will then allow Medicare beneficiaries to share their claims data with third-party applications they choose to trust. S4S/*All of Us* will be able to connect to this production API in this manner.
- The Blue Button API source code documentation for the APIs, user experience guides, and reference tools are all publicly posted on the Blue Button API site: <https://bluebutton.cms.gov/>.
- The Blue Button API code has been completely modified to meet the S4S specifications. With the modifications complete, the CMS portion of the project has concluded.
- NIH has prepared updated S4S online documentation with a notebook-based tutorial for app developers: <https://github.com/sync-for-science/sync-for-science.github.io/blob/master/proxy-api-calls/SMART.ipynb>
- Modifications to the S4S reference implementation open source test suite for Medicare beneficiary claims data to work with the CMS Blue Button sandbox (automatic sign in and approval steps) are complete. Deployment of the test suite with EOB and Coverage resource validation is in production, and continued CMS sandbox compatibility enhancements are under development.
- CMS added 30,000 synthetic beneficiaries to the production environment for production application testing.

Disseminated Products:

- Source code published on GitHub, which is maintained continuously for the Blue Button Platform: <https://github.com/CMSgov/bluebutton-web-server>The developer preview documentation is available at: <https://cmsgov.github.io/bluebutton-developer-help/>
- Google group to support developers: <https://groups.google.com/forum/#!forum/Developer-group-for-cms-blue-button-api>
- Blue Button 2.0 Implementation Guide, published June 2018: <https://bluebutton.cms.gov/developers/#blue-button-implementation-guide>
- Documentation detailing the developer/application validation process to support operations: <https://bluebutton.cms.gov/developers/#production-api-access>
- HIMSS 2018, Office Hours: Donating Data for Research with Blue Button API & Sync for Science, Josh Mandel, David Kreda, Carly Medosch

- Presented S4S with context of Financial Data to developers at the HL7 Working Group Meeting May 2018, FHIR® Dev Days Boston 2018

NIH Period of Performance:

9/7/17 – 4/29/19

NIH Federal Point of Contact:

Ed Ramos

CMS Period of Performance:

9/7/17 – 4/30/18

CMS Federal Point of Contact:

Lori Pettebone-Koraganie

XI. Conclusion

This report describes the 25 OS-PCORTF projects' and their contributions to data infrastructure, in terms of their goals, objectives, and rationale and the utility of their products to real world problems. Collectively, these projects have contributed multiple frameworks, standards, services, tools and networks to advance the field of PCOR. With these solutions, they have targeted a number of pressing health concerns like the opioid crisis and policy issues and priorities like value-based care, health insurance reform, drug pricing and improving patient access to their health information both for personal use and data donation to research.

To further emphasize their practical applications of the OS-PCORTF projects, the report also profiled the tangible accomplishments of the 11 projects that concluded in 2018. Their findings and products have direct relevance to key issues in the field, such as the growing use of APIs, the need for PCOR-supportive data governance structures and continued efforts to standardize data, and solutions to challenging technical issues like patient matching, record linkages, and facilitating research across distributed network. The 11 concluding projects have produced usable, publicly available products that can be deployed to improve data access and use.

As both a funder and coordinator for the OS-PCORTF projects, ASPE has continued to show its commitment to supporting high impact projects and the uptake of the resulting knowledge and products. ASPE has actively pursued their dissemination at conferences and via white papers, and their demonstration via regular webinars and learning opportunities, and outreach to potential end users via the 2018 OS-PCORTF Workshop, led by ASPE and the Resource Center.

In 2019, ASPE and the Resource Center will continue identifying opportunities to apply the OS-PCORTF products to known gaps in knowledge and technical solutions, and to highlight key achievements, products, and applications to support dissemination and the long-term impact of the portfolio.

Appendix A. OS-PCORTF Project Portfolio

Table A1. Active OS-PCORTF Projects

Funded Agency	Project Title
Agency for Healthcare Research and Quality	Advancing the Collection and Use of Patient-Reported Outcomes through Health Information Technology (IT)*
	Assessing and Predicting Medical Needs in a Disaster*
	Capstone for Outcomes Measures Harmonization Project
	Harmonization of Clinical Data Element Definitions for Outcome Measures in Registries
Assistant Secretary for Preparedness and Response	Assessing and Predicting Medical Needs in a Disaster*
Centers for Disease Control and Prevention	Childhood Obesity Data Initiative: Integrated Data for Patient-Centered Outcomes Research Project
	Development of a Natural Language Processing Web Service for Public Health Use*
	Enhancing Data Resources for Researching Patterns of Mortality in Patient Centered Outcomes Research* Project 1 – Adding Cause - Specific Mortality to NCHS’s National Hospital Care Survey by Linking to the National Death Index
	Enhancing Data Resources for Researching Patterns of Mortality in Patient Centered Outcomes Research* Project 4 – NDI Workshop and Strategy Paper
	Enhancing Identification of Opioid-Involved Health Outcomes Using Linked Hospital Care and Mortality
	Improving the Mortality Data Infrastructure for Patient-Centered Outcomes Research
	Privacy and Security Blueprint, Legal Analysis and Ethics Framework for Data Use, and Use of Technology for Privacy*
	Strengthening the Data Infrastructure for Outcomes Research on Mortality Associated with Opioid Poisonings
	Centers for Medicare and Medicaid Services
Improving Beneficiary Access to their Health Information through an Enhanced Blue Button Service	
Technologies for Donating Medicare Beneficiary Claims Data to Research Studies*	
Food and Drug Administration	Collection of Patient-Provided Information through a Mobile Device Application for Use in Comparative Effectiveness and Drug Safety Research

Funded Agency	Project Title
	Cross-Network Directory Service
	Development of a Natural Language Processing Web Service for Public Health Use*
	Developing a Strategically Coordinated Registry Network to Support Research on Women's Health Technologies*
	Enhancing Data Resources for Researching Patterns of Mortality in Patient Centered Outcomes Research* Project 2 – Pilot Linkage of NDI+ to Commercially and Publicly Insured Populations
	Harmonization of Various Common Data Models and Open Standards for Evidence Generation*
	Source Data Capture from Electronic Health Records: Using Standardized Clinical Research Data
	Standardization and Querying of Data Quality Metrics and Characteristics for Electronic Health Data
	Utilizing Data from Various Data Partners in a Distributed Manner
National Institutes of Health	Creation of LOINC Equivalence Classes
	Developing a Strategically Coordinated Registry Network to Support Research on Women's Health Technologies*†
	Emergency Medicine Opioid Data Infrastructure: Key Venue to Address Opioid Morbidity and Mortality
	Harmonization of Various Common Data Models and Open Standards for Evidence Generation*†
	Technologies for Donating Medicare Beneficiary Claims Data to Research Studies*
	Use of the ADAPTABLE Trial to Strengthen Methods to Collect and Integrate Patient-Reported Information with Other Data Sets and Assess Its Validity
Office of the National Coordinator for Health Information Technology	Advancing the Collection and Use of Patient-Reported Outcomes through Health Information Technology (IT)*
	Conceptualizing a Data Infrastructure for the Capture and Use of Patient-Generated Health Data
	Developing a Strategically Coordinated Registry Network to Support Research on Women's Health Technologies*
	Harmonization of Various Common Data Models and Open Standards for Evidence Generation*
	Privacy and Security Blueprint, Legal Analysis and Ethics Framework for Data Use, and Use of Technology for Privacy*
	Security and Privacy Standards for Patient Matching, Linking and Aggregation

* OS-PCORTF project funding awarded to multiple agencies.

† OS-PCORTF project funding awarded within NIH

Table A2. Completed OS-PCORTF Projects

Funded Agency	Project Title
Assistant Secretary for Planning and Evaluation	Beta Testing of the Multi-Payer Claims Data [^]
	Comparative Effectiveness Research Inventory
Centers for Disease Control and Prevention	Expanding Data Collection Infrastructure of the National Program of Cancer Registries for Comparative Effectiveness Research
Centers for Medicare and Medicaid Services	Maintenance and Support of the Chronic Conditions Warehouse for Comparative Effectiveness Research
Health Resources and Services Administration	Strengthening and Expanding the Community Health Applied Research Network (CHARN) Registry to Conduct Patient-Centered Outcomes Research
National Libraries of Medicine	Development of Data Infrastructure for Use of EHRs in Comparative Effectiveness Research*
Office of the National Coordinator for Health Information Technology	Creating the Foundational Blocks for the Learning Health Care System: Data Access Framework
	Creating the Foundational Blocks for the Learning Health Care System: Structured Data Capture
	Development of Data Infrastructure for Use of EHRs in Comparative Effectiveness Research*
	Strategic Opportunities for Building Data Infrastructure for Patient-Centered Outcomes Research

[^] The Multi-Payer Claims Data (MPCD) project was a \$16 million CMS project with a contract period of performance of 09/14/2010 to 09/15/2013. On 09/24/2012, the contract was modified with ASPE-provided OS PCORTF funding to conduct a Beta Test. ASPE was responsible for leadership oversight of the Beta Testing of MPCD.

* OS-PCORTF project funding awarded to multiple agencies.

Appendix B. Abbreviations

Acronym	Description
ACEP	American College of Emergency Physicians
AHRQ	Agency for Healthcare Research and Quality
API	Application Programming Interface
ASPE	Office of the Assistant Secretary for Planning and Evaluation
ASPR	Office of the Assistant Secretary for Preparedness and Response
caDSR	cancer Data Standards Registry and Repository
CDA	Clinical Document Architecture
CDC	Centers for Disease Control and Prevention
CDE	Clinical Data Element
CDM	Common Data Model
CEDR	Clinical Emergency Data Registry
CER	Clinical Effectiveness Research
CMS	Centers for Medicare and Medicaid Services
CNDS	Cross-Network Directory Service
CODI	Childhood Obesity Data Initiative
CRN	Coordinated Registry Network
CTO	Chief Technology Officer
DAF	Data Access Framework
DQ	Data Quality
DVS	Division of Vital Statistics
eCRF	Electronic case report form
EDC	Electronic Data Capture
EDRS	Electronic Death Registration Systems
EHR	Electronic Health Record
ETL	Extract-Translate and Load
FAERS	Food and Drug Administration Adverse Event Report System
FDA	U.S. Food and Drug Administration
FHIR®	Fast Healthcare Interoperability Resources
FISMA	Federal Information Security Management Act
HEART	Health Relationship Trust Profile
HCUP	Healthcare Cost and Utilization Project
HHS	U.S. Department of Health and Human Services
HITECH	Health Information Technology for Economic and Clinical Health Act

Acronym	Description
HL7	Health Level Seven International
HP	Office of Health Policy
ICD-10	International Classification of Diseases 10 th Edition
IG	Implementation Guide
IHE	Integrating the Healthcare Enterprise
IRB	Institutional Review Board
IT	Information Technology
LOINC	Logical Observation Identifiers Names and Codes
MACRA	Medicare Access and CHIP Reauthorization Act
MedDRA	Medical Dictionary for Regulatory Activities
MMDS	Medical Mortality Data System
NCHS	National Center for Health Statistics
NCATS	National Center for Advancing Translational Science
NCI	National Cancer Institute
NDI	National Death Index
NHCS	National Hospital Care Survey
NIDA	National Institute on Drug Abuse
NIH	National Institutes of Health
NLM	National Library of Medicine
NLP	Natural Language Processing
NORC	NORC at the University of Chicago
NVSS	National Vital Statistics System
OMB	Office of Management and Budget
ONC	Office of the National Coordinator for Health Information Technology
OS	Office of the Secretary
ODD	Opioid Use Disorders
PCOR	Patient-Centered Outcomes Research
PCORI	Patient-Centered Outcomes Research Institute
PCORnet	PCORI's National Patient-Centered Clinical Research Network
PCORTF	Patient-Centered Outcomes Research Trust Fund
PGHD	Patient-generated health data
PHCP	Public Health Community Platform
PMN	PopMedNet™
PMAL	Patient Matching, Linking, and Aggregation

Acronym	Description
PPI	Patient-Provided Information
PRO	Patient-Reported Outcome
PROMIS	Patient-Reported Outcomes Measurement Information System
SAMHSA	Substance Abuse and Mental Health Services Administration
S4S	Sync for Science™
SDC	Structured Data Capture
SDO	Standards Development Organization
SMART on FHIR®	Substitutable Medical Apps, Reusable Technology on Fast Healthcare Interoperability Resources
SNOMED	Systematized Nomenclature of Medicine
USPSTF	U.S. Preventative Services Task Force
VAERS	Vaccine Adverse Event Reporting Systems
VBP	Value-Based Payment
VSRR	Vital Statistics Rapid Release
WCP	Web Configuration Portal
WHT-CRN	Women's Health Technologies Coordinated Registry Network
WMP	Weight Management Program

Appendix C. Glossary

Key Terms	Description
Blue Button	A standard that makes patients the custodians of their data by allowing them to share and access it.
Clinical Element Models (CEM)	An approach to representing detailed clinical data models and the instances of data which conform to these models.
Clinical Data Research Networks (CDRN)	System-based networks (such as hospital systems) that have the potential to become an ideal electronic network, without structural impediments.
Common Data Elements (CDE)	Data elements shared between multiple data sets. ⁴⁹
Common Data Models (CDM)	An aggregated or centralized data model copies data from original sources and brings and standardizes these data in a centralized place. The copied data can then be queried and analyzed.
Data Governance	The process by which stewardship responsibilities are conceptualized and carried out, that is, the policies and approaches that enable stewardship.
Distributed Research Network (DRN)	A DRN is an approach in which data holders maintain control over their protected data and its uses. A DRN features a central portal that performs network functions, such as operations (e.g., workflow, policy rules, auditing, query formation, distribution) and security (e.g., authentication, authorization) and distributed data marts that remain under the control of the data holders.
Electronic Clinical Data (ECD)	Data in electronic form that may include clinical data from sources such as long term care facilities, pharmacies, laboratories, genetic tests, emergency medical services, and other clinical settings.
Electronic Health Record (EHR)	An electronic record of health-related information for a patient that contains information captured in clinical visits, lab and imaging studies, and other information important to the patient's medical past.
eMaRC	Used by central cancer registries to receive and process cancer pathology and biomarker data that are received in unstructured narrative text format.
Fast Healthcare Interoperability Resources (FHIR®)	A standard for translating health information data into a structured that can be accepted by a wide range of applications.
GitHub	A web-based service for developers to build software.
Health Information Technology for Economic and Clinical Health (HITECH) Act	A law passed by Congress in 2009 that authorizes expenditures of approximately \$20 billion over five years to promote the adoption and use of electronic health record technologies that will be connected through a national health information network.
International Statistical Classification of Diseases and Related Health Problems, 10th revision (ICD-10)	ICD-10 is the diagnostic classification standard for all clinical and research purposes.
Interoperability	The ability of health information technology (health IT) systems from different vendors to communicate and share information.

Key Terms	Description
Institutional Review Board (IRB)	A group that follows federal regulations, state laws, and institutional policy to review, monitor, and approve research in order to protect the ethical rights and privacy of the subjects involved.
Learning Health Care System	A health care system in which knowledge generation for research, science, quality assessment, outcomes, and safety standards are aligned for improvement and innovation.
Logical Observation Identifiers Names and Codes (LOINC)	A universal coding system for laboratory tests and other clinical observations. It is a national and international standard with widespread adoption and recognition of its utility.
Meaningful Use	Objectives regarding electronic health record use that eligible professionals and hospitals must achieve to qualify for Centers for Medicare and Medicaid Services (CMS) Incentive Programs. ⁵⁰
Medicare Access and CHIP Reauthorization Act (MACRA)	The Medicare Access and CHIP Reauthorization Act of 2015 ended the Sustainable Growth Rate formula and consolidated three existing quality programs into two new tracks under the Quality Payment Program.
Metadata	The term metadata refers to “data about data”. The term is ambiguous, as it is used to describe two fundamentally different concepts. Structural metadata concerns the design and specification of data structures and is more properly called “data about the containers of data”; descriptive metadata, on the other hand, concerns individual instances of application data, that is, the data content.
Natural Language Processing (NLP)	A computational model that analyzes texts using several linguistics approaches, such as syntax, semantics, and pragmatics, for the purpose of achieving human-like language results.
Patient-Centered Outcomes Research (PCOR)	Patient-Centered Outcomes Research helps people make informed health care decisions and allows their voices to be heard in assessing the value of health care options. It answers four patient-focused questions: “Given my personal characteristics, conditions and preferences, what should I expect will happen to me?” “What are my options and what are the benefits and harms of those options?” “What can I do to improve the outcomes that are most important to me?” “How can the health care system improve my chances of achieving the outcomes I prefer?”
Patient-Generated Health Data (PGHD)	Patient-generated health data is health-related data created, recorded, gathered, or inferred by or from patients or their designees to help address a health concerns.
PCORI’s National Patient-Centered Clinical Research Network (PCORNet)	A “network of networks” that brings together Clinical Data Research Networks and Patient-Powered Research Networks to support patient-centered outcomes research. ⁵¹
PopMedNet™	PopMedNet™ is an open source software application that enables the creation, operation, and governance of distributed health data networks through a no cost license.

Key Terms	Description
RxNorm	RxNorm is a normalized naming system for clinical drugs (both generic and brand name) and a tool for supporting semantic interoperability between drug terminologies and pharmacy knowledge data bases. ⁵² RxNorm is maintained by NLM.
Semantic Interoperability	The ability of computer systems to transmit data with unambiguous, shared meaning.
Sentinel	A distributed research network, using existing electronic health care data from multiple sources to support monitoring FDA regulated medical products and devices.
SMART on FHIR®	SMART on FHIR® is a set of profiles specified in the Argonaut Implementation Guide that builds upon FHIR® by defining an authorization model for apps based on the OAuth standard. ^{53, 54, 55}
Systematized Nomenclature of Medicine (SNOMED)	A standard for the electronic exchange of clinical health information that has been designated for use by U.S. Federal Government systems. ⁵⁶
Value-Based Purchasing (VBP)	VBP programs reward health care providers with incentive payments for quality and safety improvements in patient-centric care.

References

- ¹ U.S. Department of Health and Human Services. Office of the Assistant Secretary for Planning and Evaluation. Patient-Centered Outcomes Research Trust Fund FAQs. Retrieved from: <https://aspe.hhs.gov/patient-centered-outcomes-research-trust-fund-faqs>.
- ² Public Health Services Act Section 937(7).
- ³ U.S. Department of Health and Human Services. Secretary Priorities. Retrieved from: <https://www.hhs.gov/about/leadership/secretary/priorities/index.html>.
- ⁴ U.S. Department of Health and Human Services. Office of the Assistant Secretary for Planning and Evaluation. OS PCORTF Strategic Framework. Available from: <https://aspe.hhs.gov/os-pcortf-strategic-framework>.
- ⁵ Novak V. et al. Troponin Criteria for Myocardial Infarction after Percutaneous Coronary Intervention. *Arch Intern Med* 2012; 172(6):502-208.
- ⁶ Gliklich RE, Leavy MB, Karl J, Campion DM, Levy D. Outcome Measure Framework Design Document. Effective Health Care Program Research Report No. 43. (Prepared by the Outcome DEcide Center under Contract No. 290-2005-0035-1). AHRQ Publication No. 14-EHC019-EF. Rockville, MD: Agency for Healthcare Research and Quality. May 2014. Retrieved from www.effectivehealthcare.ahrq.gov/reports/final.cfm.
- ⁷ Agency for Healthcare Research and Quality. (n.d.). RoPR - The Registry of Patient Registries. Retrieved from <https://patientregistry.ahrq.gov/>
- ⁸ Hales, C. M. (2017). *Prevalence of obesity among adults and youth: United States, 2015-2016*. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics.
- ⁹ US Preventive Services Task Force. Screening for Obesity in Children and Adolescents: US Preventive Services Task Force Recommendation Statement. *JAMA*. 2017;317(23):2417–2426. doi:10.1001/jama.2017.6803
- ¹⁰ Dietz W, et al. *National Academy of Medicine, Washington, DC, 2017*
- ¹¹ Weiss, A. J., Elixhauser, A., Barrett, M. L., Steiner, C. A., Bailey, M. K., & O'Malley, L. Opioid-Related Inpatient Stays and Emergency Department Visits by State, 2009–2014. HCUP Statistical Brief# 219. December 2016. *Agency for Healthcare Research and Quality, Rockville, MD. Agency for Healthcare Research and Quality. Rockville, MD.*
- ¹² Vivolo-Kantor AM, Seth P, Gladden RM, et al. Vital Signs: Trends in Emergency Department Visits for Suspected Opioid Overdoses — United States, July 2016–September 2017. *MMWR Morb Mortal Wkly Rep* 2018;67:279–285. DOI: Retrieved from <http://dx.doi.org/10.15585/mmwr.mm6709e1>.
- ¹³ Merianne Spencer, "Using Linked Data and Natural Language Processing to Support Patient-Centered Outcomes Research on Opioids." 2018 Academy Health Annual Research Meeting
- ¹⁴ Centers for Disease Control and Prevention National Center for Health Statistics National Hospital Care Survey [website] <https://www.cdc.gov/nchs/nhcs/>.
- ¹⁵ Weiskopf, N. G., & Weng, C. (2013). Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *Journal of the American Medical Informatics Association*, 20(1), 144-151.

-
- ¹⁶ SentinelInitiative.org. Sentinel System's Story for the Public. Retrieved from <https://www.sentinelinitiative.org/public>.
- ¹⁷ PCORnet® The National Patient-Centered Clinical Research Network. About PCORnet. Retrieved from <https://pcorner.org/about-pcorner/>.
- ¹⁸ Lavalley, Danielle C., et al. "Incorporating Patient-Reported Outcomes Into Health Care To Engage Patients And Enhance Care." *Health Affairs*, vol. 35, no. 4, 2016, pp. 575–582. doi:10.1377/hlthaff.2015.1362.
- ¹⁹ Voss, Erica A., Rupa Makadia, Amy Matcho, Qianli Ma, Chris Knoll, Martijn Schuemie, Frank J. DeFalco, Ajit Londhe, Vivienne Zhu, and Patrick B. Ryan. "Feasibility and utility of applications of the common data model to multiple, disparate observational health databases." *Journal of the American Medical Informatics Association* 22, no. 3 (2015): 553-564.
- ²⁰ Food and Drug Administration Guidance of Industry Electronic Source Data in Clinical Investigations. Retrieved from <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM328691.pdf>.
- ²¹ Food and Drug Administration Advancing Regulatory Science – OneSource Pathology Checklist. Retrieved from <https://www.fda.gov/ScienceResearch/SpecialTopics/RegulatoryScience/ucm507090.htm>
- ²² Kahn, M. G., Callahan, T. J., Barnard, J., Bauck, A. E., Brown, J., Davidson, B. N., Estiri, H., Goerg, C., Holve, E., Johnson, S. G., Liaw, S. T., Hamilton-Lopez, M., Meeker, D., Ong, T. C., Ryan, P., Shang, N., Weiskopf, N. G., Weng, C., Zozus, M. N., ... Schilling, L. (2016). A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. EGEMS (Washington, DC), 4(1), 1244. doi:10.13063/2327-9214.1244
- ²³ 14.6 Resource Measure - Content. (2016, August). Retrieved from <http://hl7.org/fhir/2016sep/measure.html>
- ²⁴ American College of Emergency Physicians. (n.d.). Clinical Emergency Data Registry. Retrieved from <https://www.acep.org/administration/quality/cedr/cedr-home/#sm.0014sf1ap1dc4dnotc215pdp4942r>
- ²⁵ Assistant Secretary for Public Affairs (ASPA). (2018, July 23). Secretary Priorities. Retrieved from <https://www.hhs.gov/about/leadership/secretary/priorities/index.html#opioids>
- ²⁶ Ilkonis, P. A., Yu, L., Dodds, N. E., Johnston, K. L., Lawrence, S. M., Hilton, T. F., Mccarty, D. (2016). An Item Bank for Abuse of Prescription Pain Medication from the Patient-Reported Outcomes Measurement Information System (PROMIS®). *Pain Medicine*. doi:10.1093/pm/pnw233
- ²⁷ Adaptable, the Aspirin Study: A Patient-Centered Trial . Retrieved from <http://theaspirinstudy.org/>.
- ²⁸ Yang, L; Tenenbaum, J., "ADAPTABLE Supplement Report: Patient-Reported Health Data and Metadata Standards in the ADAPTABLE Study", NIH Collaboratory Living Textbook. 2018. Retrieved from: <https://dcricollab.dcri.duke.edu/sites/NIHKR/KR/ADAPTABLE%20Supplement%20Patient-Reported%20Health%20Data%20Standards-2018-07-19.pdf>.
- ²⁹ HealthIT.gov Consumer eHealth Patient-Generated Health Data [website]. <https://www.healthit.gov/policy-researchers-implementers/patient-generated-health-data>.
- ³⁰ HealthIT.gov Consumer eHealth Patient-Generated Health Data [website]. <https://www.healthit.gov/policy-researchers-implementers/patient-generated-health-data>.
- ³¹ Shah, G.H., Lertwachara, K., and Ayanso, A. Record linkage in healthcare: applications, opportunities, and challenges for public health. *International Journal of Healthcare Delivery Reform Initiatives*, 2010. 2(3): 29-47.

-
- ³² Dimitropoulos, L. L. Privacy and Security Solutions for Interoperable Health Information Exchange; Perspectives on Patient Matching: Approaches, Findings, and Challenges. Report written for Jodi Daniel and Steven Posnack at the Office of the National Coordinator for Health IT and P. Jon White at the Agency for Healthcare Research and Quality. June 2009.
- ³³ HHS. Patient Match Algorithm Challenge Award Winners. Retrieved from: <https://www.hhs.gov/about/news/2017/11/08/hhs-names-patient-matching-algorithm-challenge-winners.html>
- ³⁴ Posnack, S. "Oh, the Places Data Goes: ONC Announces Data Provenance Challenge. Stephen Posnack. Office of the National Coordinator for Health Information Exchange Health IT Buzz. April 16, 2017. Retrieved from: <https://www.healthit.gov/buzz-blog/interoperability/places-data-onc-announces-data-provenance-challenge/>.
- ³⁵ OpenID Foundation. Retrieved from: <http://openid.net/foundation/>.
- ³⁶ U.S. FOOD AND DRUG ADMINISTRATION. Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Federal Register 2009;74(35):65132-133.
- ³⁷ Patient-Centered Outcomes Research Institute. Users' Guide to Integrating Patient-Reported Outcomes in Electronic Health Records. 2017. <https://www.pcori.org/sites/default/files/PCORI-JHU-Users-Guide-To-Integrating-Patient-Reported-Outcomes-in-Electronic-Health-Records.pdf>
- ³⁸ National Institutes of Health. Patient-Reported Outcomes Measurement Information System (PROMIS). Retrieved from <https://commonfund.nih.gov/promis/index>
- ³⁹ HCUP Overview. Healthcare Cost and Utilization Project (HCUP). January 2018. Agency for Healthcare Research and Quality, Rockville, MD. Retrieved from www.hcup-us.ahrq.gov/overview.jsp.
- ⁴⁰ Medical Device Epidemiology Network. Retrieved from <http://mdepinet.org/about-us/>.
- ⁴¹ Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? J Biomed Inform. 2009 Oct;42(5):760-72.
- ⁴² International Conference on Harmonization of Technical requirements for Registration of Pharmaceuticals for Human Use (ICH). Retrieved from <http://www.meddra.org/how-to-use/support-documentation/english>.
- ⁴³ The Public Health Community Platform Retrieved from <http://www.thephcp.org/home>
- ⁴⁴ Kreimeyer K, Foster M, Pandey A, Arya N, Halford G, Jones SF, Forshee R, Walderhaug M, Botsis T. Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review. J Biomed Inform. 2017 Sep;73:14-29.
- ⁴⁵ Foster M, Pandey A, Kreimeyer K, Botsis T. Generation of an annotated reference standard for vaccine adverse event reports. Vaccine. 2018 Jul 5;36(29):4325-4330.
- ⁴⁶ Kreimeyer K, Foster M, Pandey A, Arya N, Halford G, Jones SF, Forshee R, Walderhaug M, Botsis T. Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review. J Biomed Inform. 2017 Sep;73:14-29.
- ⁴⁷ Jones S. Development of Natural Language Processing (NLP) Workbench Web Services. NAACCR 2017 Annual Conference. 2017 June. <https://20tqt36s1la18rvn82wcmprn-wpengine.netdna-ssl.com/wp-content/uploads/2017/07/Development-of-A-Natural-Language-Processing-Workbench-Web-Services-Jones.pdf>

⁴⁸ Office of the National Coordinator for Health Information Technology. Connecting Health and Care for the Nation. A Shared Nationwide Interoperability Roadmap. Version 1.0. Retrieved from: <https://www.healthit.gov/sites/default/files/hie-interoperability/nationwide-interoperability-roadmap-final-version-1.0.pdf>.

⁴⁹ NIH U.S. National Library of Medicine. Common Data Element (CDE) Resource Portal Glossary. Retrieved from: <https://www.nlm.nih.gov/cde/glossary.html>.

⁵⁰ Health IT.gov. EHR Incentives & Certification. Retrieved from: <https://www.healthit.gov/providers-professionals/meaningful-use-definition-objectives>.

⁵¹ Patient-Centered Outcomes Research Institute. PCORnet: The National Patient-Centered Clinical Research Network. Retrieved from: <https://www.pcori.org/research-results/pcornet-national-patient-centered-clinical-research-network>

⁵² NIH U.S. National Library of Medicine. RxNorm Overview. Retrieved from: <https://www.nlm.nih.gov/research/umls/rxnorm/overview.html>.

⁵³ SMART®. Retrieved from: <https://smarthealthit.org/an-app-platform-for-healthcare/about/>.

⁵⁴ SMART®. SMART of FHIR Profiles. Retrieved from: <http://docs.smarthealthit.org/data/profiles.html>.

⁵⁵ Health Level Seven FHIR Argonaut Project. Argonaut Data Query Implementation Guide Version 1.0.0. Retrieved from: <http://www.fhir.org/guides/argonaut/r2/>.

⁵⁶ NIH U.S. National Library of Medicine. SNOMED-CT. Retrieved from: <https://www.nlm.nih.gov/healthit/snomedct/>.