Building the Data Capacity for Patient-Centered Outcomes Research:

The 2019 Annual Report

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

Suggested citation: Dullabh P, Dhopeshwarkar R, Heaney-Huls K, Sanders E, Hovey, L, Rajendran N, Moriarty E, Sidi M. Building the Data Capacity for Patient-Centered Outcomes Research: The 2019 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

List of Exhibits

I. Introduction

The U.S. health care system as a whole is making progress toward an interoperable data ecosystem that leverages data science methods and new technologies to make sense of the vast amounts of patient-generated and traditional health care services data. To answer complex research questions aimed at improving value and whole-person care requires the ability to link data sets and aggregate data from disparate sources. Advancements in data access and use are critical to supporting patient-centered research that produces new scientific evidence to inform the health care decisions of patients, caregivers, and their health care providers.

However, data infrastructure challenges remain for using the wealth of data generated by the health system to support patient-centered research. For example, poor data quality and lack of standardization present barriers to using clinical data for research. Addressing these data infrastructure challenges and enhancing foundational research capabilities to collect, link, and analyze data is the charge of the Office of the Secretary Patient-Centered Outcomes Research Trust Fund (OS-PCORTF) portfolio.

OS-PCORTF investments have resulted in a rich portfolio of projects meeting emerging policy priorities and filling gaps in data infrastructure needed for patient-centered comparative effectiveness research. The OS-PCORTF portfolio includes projects that are developing and testing standards that improve data interoperability, piloting novel approaches to patient-provided data collection, using real-world data in evidence generation, and addressing challenges to data linkages.

This annual report provides a synopsis of the 27 OS-PCORTF projects active in Fiscal Year (FY) 2019. The projects featured in this report demonstrate the range of data capacity building projects, and the breadth of the portfolio to address emerging policy areas including, substance abuse and mental health, value-based care, patient empowerment and interoperability, data innovation, and most recently a focus on expanding data capacity or infrastructure for current policy priorities including the opioids epidemic, childhood obesity, and emergency preparedness and response.

II. Background and Context

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] The Office of the Secretary of HHS delegated authority to Office of the Assistant Secretary for Planning and Evaluation (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 2010, the Office of Health Policy (HP) of ASPE has funded and supported a portfolio of approximately 70 individual projects aimed at building data infrastructure capacity for patient-centered outcomes research (PCOR). The goal of this report is to describe the 27 OS-PCORTF projects active in the past fiscal year.

The topics of these projects simultaneously respond to the Secretary's priorities, major federal legislation, and individual agency data strategies and evolving patient-centered research needs. In the sections that follow, we synthesize the work of five projects that concluded activities in FY 2019. For these five projects, the report describes how the project products contribute to PCOR data infrastructure needs and how the resulting project products can be used by end users. The report also describes how the portfolio is addressing key HHS Secretary priorities. Finally, the report provides individual project profiles for each active project. These project profiles describe the project goals, objectives, major accomplishments, and disseminated products and illustrate how the projects contribute to the five core PCOR data research functionalities.

III. 2019 Major Accomplishments

Across the portfolio, the OS-PCORTF projects have made meaningful contributions to the data infrastructure available to support patient-centered research. In keeping with ASPE's Strategic Framework for PCOR data infrastructure, the portfolio projects continue to make strides in linking data for research and developing standards and services that help data flow more freely and smoothly. The portfolio projects have also focused on increasing access to existing federal data sources and supporting the capture of patient-reported outcomes (PROs) that can enrich research and clinical care. Each year, the portfolio of projects produce tools and resources that are available to the research community.

To illustrate the impact and scope of the OS-PCORTF contributions to data infrastructure, the report highlights six projects that concluded in 2019 and whose activities offer meaningful, usable solutions for researchers and/or clinicians. These projects offer innovations in the following areas:

- **Tools for enhancing patient participation in research and access to PROs**
- **Tools for linking data across databases and research registries**
- **Tools to standardize data for analysis across networks**
- **Methods for converting narrative text into data that can be more easily analyzed for public health**
- **Tools for supporting research through patient-donated data**

Tools to Enhance Patient Participation in Research and Access to PROs

PROs are increasingly recognized as a key perspective in patient care, offering insights into health status, symptom burden, and other quality-of-life issues that are not always captured in clinical or administrative health records as part of routine care. To better integrate patient perspectives into health care research and care delivery, the data must be captured in standardized ways and incorporated into the patient's electronic health record (EHR). To address the issue, the Agency for Healthcare Research and Quality (AHRQ) and the Office of the National Coordinator for Health IT (ONC) conducted a joint project, ***Advancing the Collection and Use of ePROs through Health IT***. The project involved modifying an existing application used to collect PROs called the OBERD[3] app, and developing a new app, PRISM, to capture and integrate PRO data into EHR systems in a user-friendly way. The OBERD app allows clinicians to administer the Patient-Reported Outcomes Measurement Information System (PROMIS®) physical function measures via computer-adapted testing on a tablet; the results are integrated into EHRs in real-time. The PRISM app was developed through a challenge competition supported by AHRQ. It collects standardized PRO data on physical function and enables patients to discuss results with their health care providers.

Both patients and their providers could benefit from the apps. Both apps would allow patients to easily report PRO data outside of the health care setting. With appropriate IT architectures, providers would be able to view the patients' PRO data in their EHR systems alongside other clinical data, which would facilitate clinical discussions and shared decision-making.

As part of the project work, AHRQ and ONC produced a variety of resources to support interested users, including an implementation guide (IG) detailing how to capture and use of PRO data, and reports that describe lessons learned during the design and testing of the two PRO apps:

- **PRO Fast Healthcare Interoperability Resources (FHIR®) IG (and Pilot Test):** The report offers guidelines that can inform electronic capture and exchange of PRO data, in accordance with the FHIR standard. https://build.fhir.org/ig/HL7/patient-reported-outcomes/
- **AHRQ PRO Applications Pilot Test Report:** The report discusses the pilot test approach and lessons learned. [*Link pending*]
- **Challenge Competition-Winning App, PRISM:** The PROMIS Reporting and Insight System from Minnesota offers a range of functionalities (e.g., data collection, score trending, personalized recommendations, and educational materials) and leverages FHIR for efficient data integration with EHRs. https://apps.apple.com/us/app/prism-for-minnesota/id1454714605 and https://play.google.com/store/apps/details?id=com.perkmotivation.PrismForMedstar
- **PRISM App Open Source Code:** This underlying code for the PRISM app is available for developers who wish to use the code to build their own apps. [*Link pending*]
- **AHRQ Step Up App Challenge Report:** The report describes the end-to-end challenge design and operations process as well as key success metrics and outcomes. https://digital.ahrq.gov/sites/default/files/docs/citation/ahrq-step-up-app-challenge-summary-2019.pdf

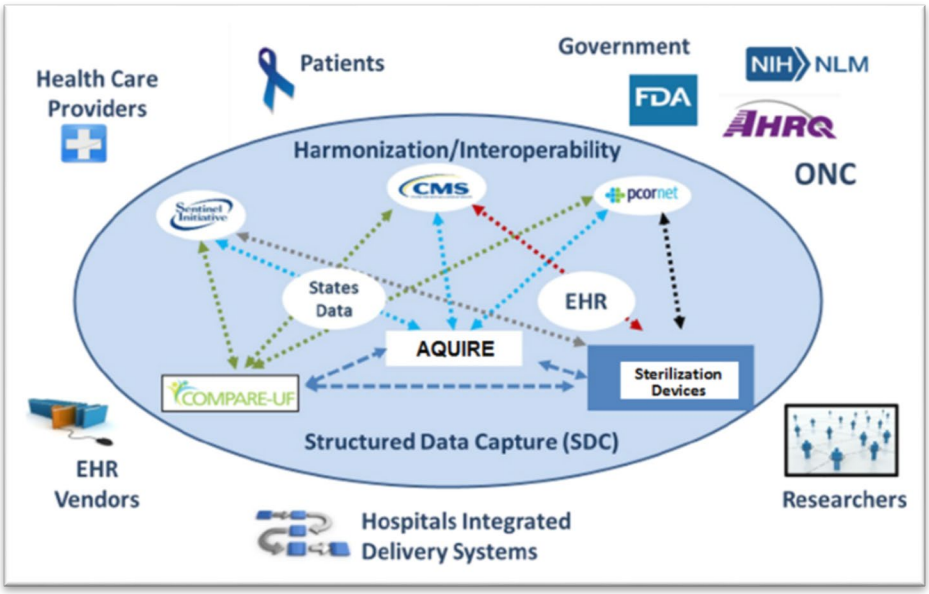
Tools for Linking Data across Databases and Registries

There is a recognized need for better data to support research on the unique issues affecting women's health, and specifically on the safety of medical devices designed to address these issues. As in many health disciplines, data related to women's health treatments and outcomes are captured in a variety of databases and registries whose data are not necessarily complete, standardized, and shareable among the larger research community. Lack of standardized data from diverse sources means that robust research is difficult and creates gaps in evidence-based guidance for women's health issues.

To address this issue, the Food and Drug Administration (FDA), National Institutes of Health (NIH)/National Library of Medicine (NLM), and ONC developed a national infrastructure to collect clinical data for studying health technologies uniquely affecting women's health. The ***Developing a Strategically Coordinated Registry Network (CRN) to Support Research on Women's Health Technologies*** project created the Women's Health Technologies Coordinated Registry Network (WHT-CRN). The WHT-CRN for health technologies are strategically partnered electronic health information systems that support: 1) the implementation of structured device identifiers, core minimum data elements, and definitions; and 2) the ability to share complementary data across information systems. WHT-CRN consists of four types of data domains (registries, claims, EHR data, and patient generated data) totaling over 550,000 records with detailed data and statistics for each, describing population demographics, disease presentation, device exposure, follow-up duration, and relevant clinical outcomes.4 The project identified and then harmonized data elements commonly captured across four women's health areas so registries can begin to capture these data in a standardized way. This will help researchers conduct

studies using data from multiple registries and answer new questions about the safety of women's health devices. The registry network is intended to improve post-market surveillance, increase possible evidence generation, support adverse event analyses, and complement existing surveillance tools. Exhibit 1 depicts the various stakeholders identified to contribute resources, knowledge, and experience with women's health devices for the WHT-CRN project.

Exhibit 1. WHT-CRN Stakeholder Engagement



The project team has produced several resources to support the use of the WTH-CRN to generate evidence around the safety and efficacy of medical devices:

- **Women's Health Technology Medical Device Epidemiology Network Website:** This website (hosted by MDEpiNet) is intended to raise awareness of the WHT-CRN project. http://mdepinet.org/womens-health-crn/
- **The WHT-CRN Common Core Data Set:** The common core data set consists of the four harmonized clinical core data sets that researchers can analyze to study patient care and outcomes. These common data elements can also be used by researchers and providers outside the CRN who want to standardize their own data for analysis. https://cde.nlm.nih.gov/
- **HL7 WHT-CRN FHIR IG:** The IG is a "recipe book" that provides general guidelines for how the implementation of the FHIR standard can be applied to support the capture and exchange of data among the registries. The knowledge and lessons learned can also be applied to other clinical areas, beyond women's health and medical devices. http://hl7.org/fhir/us/womens-health-registries/2018Sep/
- **Training Resources:** NIH/NLM produced training resources to enhance this effort, including training videos and written guides that highlight how to find and reuse the CDEs (https://cde.nlm.nih.gov/guides), and how to create new and advanced searches.
- **The Stress Urinary Incontinence (SUI) Surgery Module:** The SUI module is a new component of the AUGS Urogynecology Quality Registry (AQUIRE) and is aimed at collecting quality data on all types of SUI surgeries, including slings, Burch procedure, and periurethral bulking agents. https://www.augs.org/clinical-practice/aquire-sui-surgery-module/

Standardizing Data for Analysis across Networks

Two portfolio projects target the need to build bridges across networks and databases, so that information captured in each source can be combined and used for research. Using diverse data sources can be a boon for research, allowing studies to take in a greater volume of data that is more representative of a greater number of patients. However, researchers often find the data difficult to compare across sources because the information is collected and coded in different ways. The need for standardization is true across clinical areas, as well as across data sources (e.g., EHRs, claims, registries).

The first project, ***Harmonization of Various Common Data Models and Open Standards for Evidence Generation***, harmonized the data available in four major research networks to make it easier to conduct research across the networks. **The *Standardization and Querying of Data Quality Metrics and Characteristics for Electronic Health Data*** project created a system to help researchers harmonize data taken from different "real world data" (RWD) sources like EHRs.

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

Common data models (CDMs) are standardized structures used to organize data into a format that makes them comparable across databases. This allows researchers to more easily analyze data to answer patient-centered research questions. Networks currently utilize their own CDMs to ensure their own data is internally consistent; this multi-agency collaboration sought to harmonize the CDMs of four major networks to increase the ease of cross-network research: 1) the FDA Sentinel; 2) The Accrual to Clinical Trials (ACT) Network, which is housed alongside Informatics for Integrating Biology at the Bedside (i2b2); 3) the National Patient-Centered Outcomes Research Network (PCORnet); and 4) the Observational Medical Outcomes Partnership (OMOP).

The initial effort, led by FDA, focused on harmonizing cancer-related data elements, to assist in research on the safety and effectiveness of newly approved oncology drugs. But the effort has wider applications as well. By harmonizing CDMs, the research community will have an opportunity to aggregate more data than is currently possible, resulting in access to a larger sample size and additional demographics (e.g., elderly, pediatrics, non-U.S. population). The tools and programs developed by one network can be reused by other networks. This solution has the potential to support a diverse range of exploration beyond cancer—everything from basic PCOR queries, to data mining in order to generate hypotheses for future research, to large-scale sophisticated analysis, including randomized clinical trials.

The project produced several tools, modules, and mechanisms to support querying and use of data across CDMs:

- **Common Data Models Harmonization (CDMH) Common Data Elements:** The relevant data elements have been exported from the NIH/National Cancer Institute's (NCI) database, the cancer Data Standards Registry and Repository (caDSR), and have been imported into the NIHCDE Repository, which provides access to structured human and machine-readable definitions of data elements that have been recommended or required by NIH Institutes and Centers. NIH/NCI/PCORTF CDEs: https://cde.nlm.nih.gov/cde/search?selectedOrg=NCI&classification=PCORTF%20CDMH
- The **CDMH FHIR IG:** The IG will help researchers who want to use the project work to map and translate data into FHIR format. https://build.fhir.org/ig/HL7/cdmh/

- The **Study Data Tabulation Model (SDTM) Export Tool:** The SDTM Export Tool can help researchers who want to export record-level results from the databases in the Clinical Data Interchange Standards Consortium (CDISC) SDTM format to do analysis. https://www.cdisc.org/standards/foundational/sdtm
- **CDMs-to-Biomedical Research Integrated Domain Group (BRIDG) Mapping:** The mapping shows the alignment of data elements and the existing gaps. https://bridgmodel.nci.nih.gov/
- **BRIDG Model Updates:** The BRIDG Model has been updated with CDMH Data Elements promoting implementation strategies for use of BRIDG. https://bridgmodel.nci.nih.gov/download-model/bridg-releases
- **Data Governance Framework:** The data governance framework document details policies and practices for accessing to and using of RWD derived from data-sharing networks. https://cde.nlm.nih.gov/resources
- **Perceived Training and Data Science Support Needs for Use of Real World Data for Clinical Research**: NIH/NLM conducted a survey of NIH intramural researchers to understand their use of real world data (RWD) and need for support to utilize analytical tools. https://cde.nlm.nih.gov/resources

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

This project was designed to fill the gap in standards that describe the quality and completeness of electronic health data. Although the data are available through existing research networks, each network has its own data quality and validation processes. Further, there is not a central resource that researchers can use to understand the characteristics of the data and the sources of the data to determine if the data are fit-for-use. The project will help facilitates better use of existing health data networks to support evidence generation. The work was guided by a broad aim to create the infrastructure to curate and explore standardized data quality metrics across research networks. To do so, the project developed metadata standards and a prototype for researchers to assess data quality, ensure data has similar characteristics and formatting, captures information about the data source and institution, and helps researchers assess whether the data are appropriate for their research questions. Using common terms and capturing data in a standard format, researchers can determine if the data from disparate sources have the same meaning, and whether the data can be used harmoniously.

The project has produced several tools that can be used across research networks and data sources:

- **Data Quality Metrics Authoring and Querying Platform:** This cloud-based, open-source tool allows user to develop and author new metrics, capture data quality metric measures, and support the evaluation and visualization of supplied measures. https://github.com/PopMedNet-Team/DataQualityMetrics
- **Technical Documentation Report:** The technical report provides technical documentation appropriate for software developers and other technical users to facilitate their use of the DQM system. It is available in the GitHub repository for reference with the system source code. https://github.com/PopMedNet-Team/DataQualityMetrics

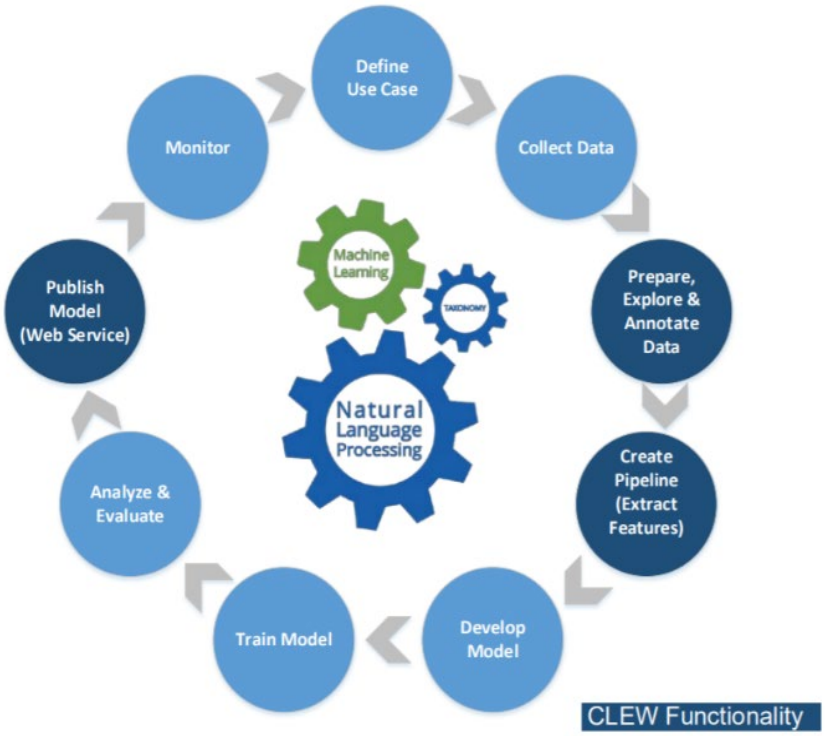
- **User Documentation:** The User Documentation report provides detailed user documentation related to the use of the web-based DQM system. The report is written to support researcher/investigator users of the system by describing all elements of the web-based system and providing instructional detail on use by an individual. https://github.com/PopMedNet-Team/DataQualityMetrics

Methods for Converting Narrative Text into Computable Data for Public Health

To maintain up-to-date information on cancer incidence, treatment, and outcomes, U.S. cancer registries collect data from multiple sources such as hospitals, laboratories, physician offices, and independent diagnostic and treatment centers. Much of this data arrives in unstructured formats, such as narrative notes, which contain critical information that is both time consuming and expensive to convert into analysis-ready data. Similarly, the FDA mandates the reporting of adverse events to monitor the safety of vaccines and other medical treatments, which arrives from diverse sources and contains a multitude of unstructured, highly valuable data.

To create better tools for handling unstructured data, the Centers for Disease Control and Prevention (CDC) and FDA launched a natural language processing (NLP) project to harness the wealth of data available in narrative text by standardizing it for use in research and analysis. The project's objective was to develop a web service that used NLP to help researchers convert unstructured clinical information into structured and standardized coded data that can be analyzed for research. To accomplish this goal, a platform was built as a resource for the research and engineering community called the Clinical Language Engineering Workbench (CLEW) platform. This platform provides free access to vetted, open-source NLP and machine learning tools to assist engineers who can use the tools and features to develop new NLP applications that address specific domain needs. Researcher end-users interested in using language models and algorithms to process and convert unstructured text to coded data also use the platform. Exhibit 2 illustrates the mechanism required to build a NLP machine learning model.

Exhibit 2. NLP Machine Learning Process



These NLP tools and services were piloted on two types of clinical information: extraction and coding of cancer data from cancer pathology reports (from the CDC) and safety surveillance data extracted from medical safety reports (from the FDA).

The NLP platform, including its services and tools, and the findings from the development process and pilots are now publicly available and ready to be applied to other types of health data:

- **CLEW Final Report:** The project team submitted a final report to ASPE about overall project accomplishments, challenges, and future activities. https://aspe.hhs.gov/system/files/pdf/259016/NLP-CLEW-FinalReport-508.pdf
- **CLEW Lessons Learned Document:** The report discusses lessons other researchers can draw from the project. https://aspe.hhs.gov/system/files/pdf/259016/NLP-CLEW-LessonsLearned-508.pdf
- **CLEW Workbench:** CLEW hosts NLP and machine learning tools, clinical NLP services, and the opportunity for tool development. https://www.cdc.gov/cancer/npcr/informatics/nlp-workbench/index.htm
- **CLEW Prototype Source Code and Documentation Software:** The publicly available software provides the code and installation instructions for the project tools. CDC GitHub: https://github.com/CDCgov/NLPWorkbench. FDA GitHub: https://github.com/FDA/
- **CLEW Technical Report:** This report describes the architecture design of CLEW and pilot use cases. https://aspe.hhs.gov/system/files/pdf/259016/NLP-Workbench-Web-Services-Technical-Report-508.pdf
- **CLEW User Guidance Document:** The guidance document is designed for NLP users who wish to install and use CLEW. https://aspe.hhs.gov/system/files/pdf/259016/NLP-CLEW-UserGuidanceDocument-508.pdf**Environmental Scan and Literature Review:** The team published an article, “Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review,” which provides a compilation of all identified open-source NLP and machine learning tools, frameworks, and systems. https://doi.org/10.1016/j.jbi.2017.07.012
- **Published Account of Adverse Events Reporting Tool:** The following article describes an annotated data set created for use in training NLP models: Foster M, Pandey A, Kreimeyer K, Botsis T. (2018). Generation of an annotated reference standard for vaccine adverse event reports. *Vaccine*, *36*(29), 4325-4330. https://doi.org/10.1016/j.vaccine.2018.05.079

Tools for Supporting Research through Patient-Donated Data

The *All of U*s Research Program is a national, large-scale research enterprise designed to support discoveries that increase our ability to better treat and prevent disease. Achieving this goal will involve enrolling 1 million or more volunteers who are willing to donate their health data for research and analysis.

The OS-PCORTF project, ***Technologies for Donating Medicare Beneficiary Claims Data to Research Studies***, aims to increase patient engagement in the *All of Us* Research Program and other research by creating an easy process for data donation. The project is a collaboration between the National Institutes of Health (NIH), the Centers for Medicare & Medicaid Services (CMS), and multiple EHR vendors, leveraging existing work on the Sync for Science™ (S4S) platform and Blue Button 2.0 program. The team built secure mechanisms (e.g., patient portals, apps, and application programming interfaces [APIs])

to allow Medicare beneficiaries to donate their medical and prescription drug claims data (i.e., Medicare Parts A, B, and D claims data) to research studies in which they would like to participate. Medicare beneficiaries could, for example, easily share their claims data from the MyMedicare.gov portal with national research efforts such as the *All of Us* Research Program using the S4S app.[5] To support this function, the team also produced an open source, data visualization web app for users to see data they contribute, from both clinical and beneficiary portals, in one consolidated form. The app supports additional features that can be activated by users, such as allowing patients to electronically annotate their data and information sharing between patients and health care providers. Exhibit 3 presents a visual illustration of patients accessing and donating their claims data to researchers and research institutions.

Exhibit 3. Accessing and Donating Claims Data

These data are highly valuable to the *All of Us* Research Program, and the lessons learned through this project could be applied to future efforts to enable individual patients to donate their health data to the larger medical research community to support a wide range of health care studies.

There are a variety of resources to support the donation and use of Medicare claims data for research:

- **Blue Button 2.0 API:** The interface allows beneficiaries to share their Medicare data with researchers. The Blue Button source code is continually updated and maintained on GitHub: https://github.com/CMSgov/bluebutton-web-server
- **Blue Button API Support Documentation:** This documentation provides instructions for using the API to connect patient health data to health apps and services, including support for S4S implementation. https://bluebutton.cms.gov/developers/
- **S4S API Webpage:** This dedicated page provides details to support the sharing of data via portals. http://syncfor.science/api-calls/
- **S4S Test Suite:** The test suite supports app developers who wish to use the platform for building and testing related apps. https://github.com/sync-for-science/test-suite/wiki/Inferno-Migration-Guide and https://github.com/sync-for-science/test-suite
- **Research App API Framework:** This framework provides researchers with the technical information needed to use S4S. https://github.com/sync-for-science/research-app-api

IV. Portfolio Contributions to Key HHS Policy Priorities

Since the formation of the OS-PCORTF in 2010, many new policy priorities have emerged within HHS in response to growing public health crises, such as the national opioid epidemic and a shifting health care landscape. The OS-PCORTF portfolio has evolved to support the data infrastructure needs for new HHS policy priorities through promoting inter-agency collaboration around linking data sets and making data more accessible for research, testing new tools and services for data linkage and de-duplication, and supporting the development of innovative data resources and analytic approaches.[6]

This report highlights four key policy areas in which current OS-PCORTF projects are working to develop solutions. These key HHS policy areas are:

- **Using real-world data for evidence generation**
- **Improving women and maternal health outcomes**
- **Addressing the national opioids epidemic**
- **Supporting value-based care transformation**

The featured OS-PCORTF projects within these areas represent examples from across the portfolio of active projects, including some of the newly funded FY 2019 projects whose work is just getting underway. Many of these projects address multiple priorities.

Increase the Use of Real-World Data for Evidence Generation

Using "real-world evidence" to bring new treatments to patients as part of the 21st Century Cures Act (the "Cures Act") is a key priority for HHS. Specifically, the Cures Act tasked the FDA to develop a framework and guidance for evaluating real-world evidence in the context of drug regulation.[7] Under the FDA's framework, real-world evidence is the clinical evidence regarding the usage and potential benefits or risks of a medical product derived from analysis of real-world data (RWD), which can come from a range of sources including EHRs, claims data, registries, and patient-generated health data (PGHD).[8] However, high-quality, real-time health data are often difficult to access because of lack of data standardization, cost, patient privacy, or other legal and intellectual property restrictions. Further, due to a variety of interoperability issues, it is often difficult to bring data together from different real-world data sources. As the FDA and other regulatory agencies consider how to leverage RWD to inform key decisions, standardized approaches for evaluating these data and curation processes to ensure their fitness for the intended use are necessary.[9] Several OS-PCORTF projects are addressing gaps in the accessibility and use of real-world sources.

The FDA's 2016 ***Standardization and Querying of Data Quality Metrics and Characteristics for Electronic Health Data*** project will help researchers incorporate RWD into their studies, and to do so with the confidence that the data are complete and high quality. As more RWD data become available, there is also a need for a reliable way to determine whether data are appropriate for a given study to avoid wasting time and project resources. To address these areas, the project has created an online platform with an open-source toolkit that contains visualization and analytic tools, and a data quality model to help researchers assess data sources, regardless of the different CDMs involved. These meta-data standards help create a fingerprint for each data source that gives researchers

> The Data Quality Metrics user guide and IG are publicly available to support interested researchers and developers on the project GitHub repository: https://github.com/PopMedNet-Team/DataQualityMetrics

vital information about the data, such as the setting (e.g., labs, hospitals) and the measures (e.g., how many responses fall outside of an expected range) that allow them to determine whether the data set is suitable for their own study. The platform and tools have been designed, tested, and are currently available to assist researchers looking for RWD to use in their research.

In 2018, the FDA completed the ***Source Data Capture from Electronic Health Records: Using Standardized Clinical Research Data*** project**,** creating tools and methods to automate the flow of EHR data into clinical trial data collection systems in a standard way. Researchers rely on robust clinical data to generate meaningful findings, and yet the information systems into which clinical care data and clinical research data are captured, organized, and stored can be highly disparate. To overcome these data infrastructure barriers, the FDA and its partners built OneSource, a single source to capture clinical health care data from EHRs and transfer the data in a standardized usable form to external data systems used in clinical research and trials. Removing manual processes improved the speed and efficiency of data transfer, reduced costs and burden for health care providers and research staff, and improved the data quality. OneSource focused on data elements related to large adaptive clinical trials for breast cancer, including biomarkers and PROs, but its tools are appropriate for broader use. OneSource created open-source forms, source code, and standards enhancement recommendations that have been released to the public and organizations interested in using EHR data for conducting clinical research.

> The final report, IG, and project products are publicly available to support widespread use of the OneSource platform and tools.

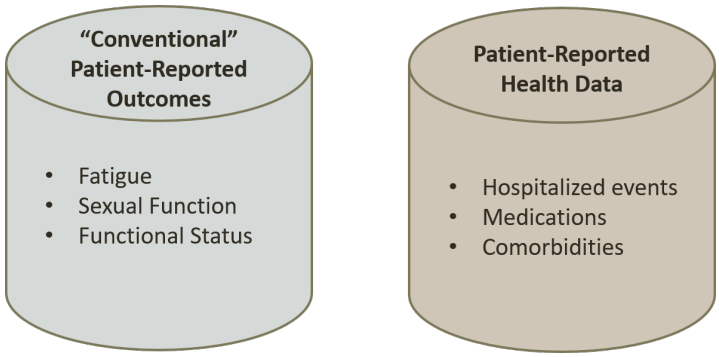
The FDA's 2019 project on ***Bridging the PCOR Infrastructure and Technology Innovation through Coordinated Registry Networks (CRN) Community of Practice (COP)*** focuses on capturing and sharing standardized data that can shed light on the safety of medical devices. This project builds upon the work of a previous OS-PCORTF project that created a CRN for sharing high quality, timely, and actionable evidence on women's health devices. The FDA is now strengthening the CRN by creating a minimum data set, including device identification numbers, adding patient-generated data, and linking the data to other data sources (i.e., clinical, claims, registries) through COPs. The COPs will span 12 clinical areas with the goal of enhancing data collection, sharing, and research across the participating CRNs to improve knowledge on clinical outcomes associated with specific women's health devices.

> The team has been refining the FHIR IG developed under the Women's Health Technologies CRN, which will be made publicly available upon completion. They have also made progress on a Core Minimum Data Set for some of the participating CRNs.

Under the NIH's 2016 project, ***Use of the ADAPTABLE Trial to Strengthen Methods to Collect and Integrate Patient-Reported Information with Other Data Sets and Assess Its Validity***, the team developed, piloted, and evaluated methods to validate and integrate patient-reported health data with EHR data for clinical trials research. Leveraging patient-report health (PRH) data to generate insight into patient-experienced outcomes in the real world has gained increased attention from scientists, industry, and regulators in recent years.[10] As shown in Exhibit 4 below, subjective patient-reported outcomes (PROs) include information such as symptoms, functional status, or fatigue; whereas PRH data includes patient- or caregiver-reported information on clinical events or treatments that are likely to exist in the EHR already, including hospitalizations, comorbid conditions, and medications their providers have recorded. The NIH project successfully demonstrated the completeness, consistency, and fitness of PRH data for use in EHR-based clinical trial research. The data helped to identify clinical events or encounters

of interest experienced by patients outside of the setting where the clinical trial is taking place as well as supplementing EHRs with important patient-reported and demographic information that is of use to researchers. These findings demonstrate the value of PRH to provide real-world evidence that enables robust and rigorous clinical trial activities.

Exhibit 4. Comparing "Conventional" Patient-Reported Outcomes and Patient-Reported Health Data



CDC's FY 2018 ***Childhood Obesity Data Initiative (CODI): Integrated Data for Patient-Centered Outcomes Research Project*** is also working to integrate RWD across sources. In 2017, the U.S. Preventive Services Task Force recommended that children should be screened for obesity and, if applicable, referred to intensive pediatric weight management interventions (PWMIs) to improve their physical and mental health. To date, RWD to study factors influencing the health outcomes and effectiveness of interventions used for childhood obesity prevention or treatment are often unavailable for research. This project is working to bridge the gap by developing and testing an expanded CDM including pediatric obesity-related information, patient linkage and de-duplication tools and data query services to bring together data stored across different organizations to create an individual-level, linked longitudinal record that includes individual and community-level risk factors, weight management interventions delivered in clinical and community settings, and clinical outcomes across health information systems.. The project is piloting tools and services (e.g., patient record linkage and de-duplication services) in the Colorado Health Observation Regional Data Service (CHORDS)—a PCORnet Clinical Data Research Network (CDRN)—in order to expand the availability of real-world data for childhood obesity research. This will support evidence to help health professionals in developing and tailoring interventions specific to the needs of children.

CODI will be piloted in the Colorado Health Observation Regional Data Service (CHORDS). CHORDS operates a PCORnet-compatible data model called the CHORDS Virtual Data Warehouse. Pilot participants include Denver Health, Children's Hospital Colorado, and Kaiser Permanente Colorado. Other potential clinical and community partners are being considered. The pilot is anticipated to go live in late 2020.

While these two projects are focused on building linkages between EHRs and other RWD sources, a FY 2019 project from the CDC, ***Making Electronic Health Record (EHR) Data More Available for Research and Public Health,*** is focused on building greater interoperability across EHR systems. Specifically this project is working to develop an application that can access data across multiple EHR systems using industry standards (e.g., FHIR and health information technology [health IT] certification requirements that underpin many EHRs and are required by the ONC, such as the Common Clinical Data Set and APIs). The application will be modeled on three diverse use cases (hepatitis, cancer, and health care surveys). After development, the project team will test the application on a hepatitis use case, to ensure it can extract data from multiple clinical organizations using multiple EHR platforms mapped to a CDM to provide RWD to researchers in order to generate real-world evidence about a given health problem.

Finally, emergency medicine is an area where real-time data is particularly needed to guide decision-making. AHRQ and the Office of the Assistant Secretary for Preparedness and Response (ASPR) are collaborating on a project around ***Assessing and Predicting Medical Needs in a Disaster, which*** began in FY 2018. This project is working to build a data platform to analyze an expanded Healthcare Cost and Utilization Project (HCUP) data set (used to identify, track, and analyze national trends in health care utilization, access, charges, quality, and outcomes) to include quarterly emergency department and inpatient data. This will enable evidence-based decision-making/responses to medically-related disasters and recovery interventions and inform efforts as well as support research queries that generate evidence around the effectiveness of different disaster response, recovery, and emergency preparedness interventions.

> The WHT-CRN FHIR Implementation Guide (IG) focuses on capturing and exchanging data related to women's health. The data that is captured will be made available to both providers and authorized researchers. While the CRN FHIR IG can be applied to multiple use cases, the current requirements have been drawn from PCORNet use cases and implementations.

Improve Women and Maternal Health Outcomes through Establishing and Enhancing the Data Infrastructure Needed to Support Women's Health Research

Women experience many unique health issues and concerns that necessitate tailored treatments and interventions to address. The need for better women's and maternal health care has been underscored by a high maternal mortality rate, particularly among black women. According to the CDC's Pregnancy Mortality Surveillance System, from 2007 to 2016, black mothers died at a rate of 3.2 times that of white mothers (*National Vital Statistics Report*, January 2020).[11] Better addressing women's health, and empowering women to be more engaged in their own health care, requires precise and robust data to monitor, understand, and address key outcomes that are unique to women. However, researchers have faced many difficulties with capturing and aggregating women's and maternal health outcomes data, which are often captured across multiple, disparate platforms using unique data elements that limit researchers' ability to view and analyze these data in aggregate. Our ability to address disparities related to women's health, however, relies on the standardization and aggregation of women's and maternal health data.[12] To meet this need, several OS-PCORTF projects are working to establish and enhance the data infrastructure needed to support best practices for women's health research via adequate data linkage, collection, and standardization.

The project ***Developing a Strategically Coordinated Registry Network (CRN) for Women's Health Technology*** (or WHT-CRN), a collaboration between the FDA, NIH/NLM, and ONC that began in FY 2017, developed an infrastructure to evaluate medical devices in three clinical areas unique to women's health: uterine fibroid treatments, pelvic organ prolapse treatments, and elective female sterilization therapies. The WHT-CRN addresses the need for an infrastructure to study health technologies in women by developing resources to support the use of diverse registry data sources and standardized data elements for post-market surveillance of therapies and evidence generation to the development of innovative therapies.

According to a 2018 CDC *Morbidity and Mortality Weekly Report*, rates of opioid use disorder (OUD) in pregnancy have increased fourfold in the last 20 years.[13] However, there is a lack of surveillance systems for collecting outcomes associated with different treatments clinically recommended for OUD during pregnancy. [14] The CDC, in partnership with NIH, the Substance Abuse and Mental Health Services

Administration (SAMHSA), and CMS, are tackling the challenge of a lack of national-level data on maternal health outcomes through a FY 2019 project establishing the ***MAT-LINK: MATernaL and Infant Network to Understand Outcomes Associated with Treatment for Opioid Use Disorder during Pregnancy*.** This network is working to enhance the data infrastructure for surveillance of maternal (as well as infant and child) health outcomes associated with medication-assisted treatment (MAT) for OUD during pregnancy. The project will establish a health outcomes surveillance network across multiple sites to rapidly collect data and monitor maternal, infant, and child health outcomes that can be analyzed, shared, and disseminated to inform patient-centered care for pregnant women with OUD as well as infants and children with prenatal opioid exposure.

Through their work to standardize data and develop data linkages to establish and enhance the infrastructure for women's health research, these two OS-PCORTF projects are contributing to the evidence base for more tailored and targeted research to inform and improve policies, clinical practice recommendations, and clinical decisions to improve women's health and health outcomes.

Address the Opioids Epidemic through Better Data and Data Infrastructure for Patient-Centered Outcomes Research

The U.S. is experiencing a historic opioid epidemic, characterized by a rapid increase in opioid overdose deaths within the past 20 years.[1,15] In 2018, Congress and the President enacted the bipartisan Substance Use-Disorder Prevention that Promotes Opioid Recovery and Treatment (SUPPORT) for Patients and Communities Act, which includes numerous provisions to prevent opioid misuse, increase access to treatment, and control the supply of illicit drugs.[16] The SUPPORT Act also directs government agencies to conduct studies on aspects of the opioid epidemic.[17] HHS has led the federal response to the opioid crisis, spending over $5 billion on opioids in FY 2018.[18] In 2017 HHS launched its broad 5-Point Strategy to Combat the Opioid Crisis.[19] This strategy includes the need for "better data" in order to "strengthen public health data reporting and collection to improve the timeliness and specificity of data, and to inform a real-time public health response as the epidemic evolves."[20]

Robust and reliable data that are available to researchers within and outside of the federal government are critical to understanding and tracking the epidemic, informing treatment and prevention efforts, and ultimately to reducing opioid-related morbidity and mortality. However, the data infrastructure (how data are *produced* and *maintained*) and the quality and timeliness of the data itself have been key challenges.[21] Currently, 10 OS-PCORTF projects across HHS agencies are working to expand data capacity or data infrastructure for patient-centered outcomes research on opioids. Two main areas of focus are:

- **Building linkages to address comorbidities:** According to a 2017 SAMHSA report, an estimated 1 in 8 U.S. adults (18 percent) have both severe mental illness and misuse opioids.[22] Two OS-PCORTF projects are addressing this issue through enhancing existing data sources to better ensure interoperability across systems and building data linkages. CDC, in partnership with SAMHSA, FDA, NIH/National Institute on Drug Abuse (NIDA), and the National Institute of Mental Health (NIMH), initiated a project in FY 2019around ***Identifying Co-Occurring Mental Health Disorders among Opioid Users Using Linked Hospital Care and Mortality Data.*** This project seeks to address the lack of data on health outcomes of opioid users with co-occurring substance use and mental health issues. The project, is working to build an enhanced, linked mortality file using mortality files on drug overdose deaths from the National Hospital Care Survey (NHCS), National Death Index (NDI) and the National Vital Statistics System and combining them with additional information on co-occurring disorders. The project will make this restricted data more

widely available to researchers through the National Center for Health Statistics (NCHS) restricted Research Data Center (RDC). This will enable greater understanding of the contributors to opioid-related mortality.

ASPE's and the Administration for Children and Families' (ACF's) ***Linking State Medicaid and Child Welfare Data for Outcomes Research on Treatment for Opioid Use Disorder (OUD)*** FY 2019 project is working to address the correlation between areas of the U.S. with high overdose death rates and high rates of children placed into foster care.[23] In partnership with several state agencies, the project is pilot-testing linkage between state Medicaid records and child welfare records in order to build a single, harmonized data set. These data sets will contain linked patient-level data, including Medicaid enrollment, patient diagnoses, services, and claims, along with child welfare outcomes (e.g., length of time in foster care, repeat maltreatment). This will allow researchers to identify parents with children in the child welfare system in order to better understand their needs for treatment of OUD, other substance use disorders (SUDs), and co-occurring mental health problems.

- **Building capacity for collection of patient-reported outcomes:** NIH/NIDA is working on two projects to enhance data collection and capacity to conduct PCOR on OUD. In FY 2018, NIH/NIDA initiated the ***Emergency Medicine Opioid Data Infrastructure: Key Venue to Address Opioid Morbidity and Mortality*** project, which seeks to standardize data to track patients presenting in the emergency department a main point of entry into the health care system for many OUD patients. Specifically the project is working to identify CDEs related to OUD and integrate these into the American College of Emergency Physicians Clinical Emergency Data Registry. Additionally, the project is exploring the feasibility of collecting electronic PROs (e.g., pain intensity, substance use disorder treatment/status) through EHR dashboards as well as a software that allows patients to consent for their data to be used in research as well as provides access to their data. They have partnered with Yale Medicine on a pilot to collect PRO data during and after emergency department visits using a mobile application with standardized opioid-use disorder-specific common data elements, created under the project, to track and improve quality of care at the point of contact.

  In FY 2019, NIH/NIDA initiated ***AMNET: An Addiction Medicine Network to Address the Opioid Epidemic in the United States***. Through this project, NIH/NIDA is working to establish a new practice-based research network and an electronic patient registry to gather standardized data on patients' characteristics, treatments, and outcomes for patients treated with buprenorphine and naltrexone in office-based practices. The project adapted an existing electronic registry (the American Psychiatric Association's PsychPRO) to collect data for practice-based research on OUDs and treatments and is developing CDEs for OUDs based on a variety of existing, validated sources. The resulting network will be used to provide near-real-time data to clinicians, researchers, and other stakeholders on OUD patients' mental health, pain, and SUDs and "real world" office-based routine treatment delivery.

Support Value-Based Care Transformation by Promoting Interoperability and Care Coordination

Value-based programs are a key part of HHS' quality strategy to reform health care delivered and paid. According to CMS, value-based programs support their objectives for improving care for individuals and populations as well as lowering the cost of care.[24] In April 2019, HHS and CMS announced the CMS Primary Cares Initiative, which aims to transform primary care through the delivery of value-based care models. Specifically, this Initiative seeks to reduce administrative burdens and empower providers to

spend more time caring for patients in order to deliver better value for patients throughout the health care system.[25] The transformation to a value-based health care system involves several significant clusters of activity, including measure and data development and building greater interoperability across data systems to maximize care coordination and enable organizational efficiencies. Critically, the delivery of value-based care requires bringing in the patient perspective in order to make continuity of care possible.[26] Several OS-PCORTF projects are working to support bringing in the patient voice, particularly among patient populations with multiple chronic conditions (MCCs) who have complex health needs handled by diverse providers and often undergo frequent care transitions (e.g., hospital to home, primary care to specialist, etc.). This makes data aggregation to inform quality and coordination of care particularly challenging, and as a result, their care is often fragmented, poorly coordinated, and inefficient.

- In FY 2019, AHRQ and the NIH/National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) initiated a project on **Data Capacity for Patient-Centered Outcomes Research through Creation of an Electronic Care Plan for People with Multiple Chronic Conditions**. The project team is working to build data capacity to improve care continuity and the flow of clinical and social information for people with MCCs by developing an interoperable, accessible electronic care (eCare) plan. The eCare plan extracts data from EHRs to facilitate aggregation and sharing of critical patient-centered data across home-, community-, clinic-, and research-based settings. Importantly, the eCare plan is the first patient-centered eCare plan that crosses multiple sectors, enabling it to serve as an overarching, longitudinal blueprint of the prioritized health concerns, goals, interventions, and health status of an individual patient. This ensures that the patient voice is brought in across all care settings where a patient with MCC receives care. The project is currently developing and will pilot an eCare Plan tool for use with patients who have chronic kidney disease, cardiovascular disease, diabetes, chronic pain, and/or OUD.
- Also in FY 2019, ASPE, AHRQ and CDC are collaborating on a project around ***Validating and Expanding Claims-based Algorithms of Frailty and Functional Disability for Value-based Care and Payment.*** This project seeks to refine and validate claims-based algorithms (e.g., within the Chronic Condition Warehouse [CCW] or within CMS claims data) around frailty and functional disabilities using EHR databases. Under value-based care programs, EHR data are frequently used for quality reporting through electronic clinical quality measures. The project began by working to identify patients with frailty and functional disabilities across payer and patient populations (Medicare, Medicaid, and dual eligible) and create/update the CMS CCW with indicators of frailty and functional disability. These indicators, along with the EHR-modified algorithms for integrating claims data that the project is also developing, can be used to assess quality of care in VBC programs to identify cases and/or conduct risk adjustment for comorbidities.

Making value-based care more patient-centered also requires addressing the social determinants of health (SDOH) that can influence patient care and preferences in order to increase the quality and lower the cost of care.[27,28] Multiple federal initiatives are integrating the electronic capture of SDOH into value-based payment programs and performance-based programs.[29,30] Many health care systems are already moving toward obtaining critical SDOH data to improve care coordination and the quality of health care services for vulnerable populations. Research has demonstrated that for many SDOH factors, small-area data (data at the community or sub-county level) may be necessary to conduct meaningful analyses. While researchers have spent substantial resources linking multiple data sets to create data files suitable for analyses, current data sets lack standardized metrics or estimates at the small-area level, which limits researchers' ability to examine SDOH in small geographic areas in order to target services.

- AHRQ and CDC are working to tackle this issue through the FY 2019 project ***Enhancing Patient-Centered Outcomes Research (PCOR): Creating a National Small-Area Social Determinants of Health Data Platform*****.** This project will develop a consolidated set of national standardized databases on valid and reliable SDOH factors at the small-area and other geographic levels, building on existing databases developed by federal agencies (e.g., AHRQ, the Health Resources and Services Administration (HRSA), CDC, ASPE, and NIH). The database will include SDOH data elements related income, employment, food, housing, environment, economics, education, safety, transportation, etc. in addition to health status and health care access and utilization. The project will make this longitudinal data set of SDOH variables for use by federal and non-federal researchers working to better understand the impact of value-based care at a small-area level.

V. Active OS-PCORTF Funded Projects

There were 27 projects active in FY 2019 (Exhibit 5). These 27 projects represent 40 individual agency awards.

The project profiles that follow were constructed based on a review of five sources of documentation reflecting project activities in the past fiscal 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 in the public domain. When possible, project deliverables submitted in tandem with the quarterly progress reports were also reviewed.

Exhibit 5. OS-PCORTF 27 Active Projects

Active Project	Project Description
Agency for Healthcare Research and Quality (AHRQ)	
Capstone for Outcomes Measures Harmonization Project	■ 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.
Enhancing Patient-Centered Outcomes Research (PCOR): Creating a National Small-Area Social Determinants of Health (SDOH) Data Platform	■ Develop a consolidated set of national standardized databases on valid and reliable SDOH factors at the small-area and other geographic levels to increase longitudinal data on SDOH available for patient-centered health services research.
Centers for Disease Control and Prevention (CDC)	
Augmenting the National Hospital Care Survey (NHCS) Data through Linkages with Administrative Records	■ Link the 2016 NHCS with Medicare Fee-for-Service claims data and federal housing assistance program data collected from the U.S. Department of Housing and Urban Development (HUD). The linked data will allow researchers to examine the role of federal social support programs in health outcomes and treatment efficacy for persons with stable housing, while also having the ability to focus on specific subpopulations, including persons with SUDs.

Active Project	Project Description
Childhood Obesity Data Initiative: Integrated Data for Patient-Centered Outcomes Research (CODI)	■ 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	■ Link three federal data sets—the NHCS, the NDI, and the NVSS-M-Drug-Involved Mortality (DIM)—to enhance the identification of opioid-related mortality.
Identifying Co-Occurring Disorders among Opioid Users Using Linked Hospital Care and Mortality Data	■ Develop a new set of algorithms using the linked NHCS, NDI, and NVSS-M-DIM to identify patients with an opioid event and co-occurring mental health disorders.
Making Electronic Health Record (EHR) Data More Available for Research and Public Health	■ Develop and test an application to support the real-time data extraction and exchange of data between EHRs and public health systems.
MAT-LINK: MATernaL and Infant NetworK to Understand Outcomes Associated with Treatment for Opioid Use Disorder during Pregnancy	■ Establish a surveillance network to monitor maternal, infant, and child health outcomes following treatment for OUD during pregnancy to inform patient-centered care for pregnant women with OUD and for infants and children with prenatal opioid exposure.
Strengthening the Data Infrastructure for Outcomes Research on Mortality Associated with Opioid Poisonings	■ Enhance the national vital statistics system through improved data collection and electronic data capture of opioid-related death data.
Food and Drug Administration (FDA)	
Bridging the PCOR Infrastructure and Technology Innovation through Coordinated Registry Networks (CRN) Community of Practice (COP)	■ Strengthen the CRNs as a real-world data source for high-quality, relevant, reliable, timely, and actionable evidence to improve patient outcomes of medical devices, specifically for technologies affecting women's health. The CRN COP will promote registries' maturity through harmonization of their minimum core data sets, commitment to incorporation of device identification, and patient engagement throughout and by facilitating the linking of registries to other data sources.
SHIELD—Standardization of Lab Data to Enhance Patient-Centered Outcomes Research and Value-Based Care	■ Improve the quality and semantic interoperability of laboratory data within and between institutions so that diagnostic information can be pulled from different sources or shared between institutions to help illuminate clinical management and understand health outcomes.
Standardization and Querying of Data Quality Metrics and Characteristics for Electronic Health Data	■ 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."
National Institutes of Health (NIH)	
NIH/NIDA's AMNET: An Addiction Medicine Network to Address the United States Opioid Crisis	■ Establish a new practice-based research network and electronic patient registry to improve the study of medication-assisted drug therapy in office-based practices.

Active Project	Project Description
Emergency Medicine Opioid Data Infrastructure: Key Venue to Address Opioid Morbidity and Mortality	■ Identify, develop, and test clinical data elements relevant to OUD 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	■ 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 Assistant Secretary for Planning and Evaluation (ASPE)	
Linking State Medicaid and Child Welfare Data for Outcomes Research on Treatment for Opioid Use Disorder	■ Develop a linked data set of state Medicaid and child welfare data systems to study SUD diagnoses and treatment outcomes of parents with children in the welfare system.
Office of the National Coordinator for Health Information Technology (ONC)	
A Synthetic Health Data Generation Engine to Accelerate Patient-Centered Outcomes Research	■ Enhance Synthea by developing or updating data generation modules for opioid, pediatric, and complex care use cases to increase the number and diversity of synthetic patient health records to meet PCOR needs.
Cross-Agency Funded Projects	
Advancing the Collection and Use of Patient-Reported Outcomes through Health Information Technology	■ Develop technical tools for collecting and integrating standardized PRO data into EHRs or other health information technology systems. ■ *Agencies:* AHRQ, ONC
Assessing and Predicting Medical Needs in a Disaster	■ Create infrastructure to collect and query information to assess and predict medical utilization and need to aide in disaster response and recovery operations. ■ *Agencies:* AHRQ and ASPR
Data Capacity for Patient-Centered Outcomes Research through Creation of an Electronic Care Plan for People with Multiple Chronic Conditions	■ Develop and pilot an interoperable longitudinal electronic care (eCare) plan to facilitate data aggregation and exchange of critical patient-centered data across home-, community-, clinic- and research-based settings. The pilot eCare Plan tool will be designed for use with patients who have chronic kidney disease (CKD), cardiovascular disease (CVD), diabetes, chronic pain, and/or (OUD). ■ *Agencies:* AHRQ and NIH/National Institute of Diabetes and Digestive and Kidney Diseases(NIDDK)
Development of a Natural Language Processing Web Service for Public Health Use	■ 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. ■ *Agencies:* CDC, FDA

Active Project	Project Description
Developing a Strategically Coordinated Registry Network to Support Research on Women's Health Technologies (WHT-CRN)	■ Create a coordinated registry network for women's health technologies that will collect PROs and employ structured data capture from EHRs for data collection and exchange. ■ *Agencies:* FDA, NIH/NLM, ONC
Enhancing Data Resources for Researching Patterns of Mortality in Patient Centered Outcomes Research: Projects 1 and 2	■ Link health care claims and EHR data from the NHCS to death record information from the NCHS NDI; and separately link to administrative data from the CMS Master Beneficiary Summary File (MBSF) and Chronic Conditions Warehouse (CCW) (Project 1). Pilot the linkage of NDI+ to commercially and publicly insured populations in existing distributed research networks. ■ *Agencies:* CDC, FDA
Harmonization of Various Common Data Models and Open Standards for Evidence Generation	■ 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. ■ *Agencies:* FDA, NIH/NCI, NIH/National Center for Advancing Translational Sciences(NCATS)), NIH/NLM, ONC
Technologies for Donating Medicare Beneficiary Claims Data to Research Studies	■ 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. ■ *Agencies:* CMS, NIH
Training Data for Machine Learning to Enhance Patient-Centered Outcomes Research (PCOR) Data Infrastructure	■ Develop high-quality training data sets to develop, train, and improve machine learning algorithms. ■ *Agencies:* NIH/NLM, ONC
Validating and Expanding Claims-based Algorithms of Frailty and Functional Disability for Value-based Care and Payment	■ Validate Medicare and Medicaid claims-based algorithms used to identify a person's risk for physical and cognitive decline and impairment to support more robust-risk adjustments, case identification and assess quality in VBC programs. ■ *Agencies:* ASPE, AHRQ, CDC

VI. Agency for Healthcare Research and Quality (AHRQ)

AHRQ is administering a total of six active OS-PCORTF-funded projects including four cross-agency projects described later in Section XII (Exhibit 6).

Exhibit 6. AHRQ Active Projects

AHRQ-Funded Projects
Capstone for Outcomes Measures Harmonization Project
Enhancing Patient-Centered Outcomes Research (PCOR): Creating a National Small-Area Social Determinants of Health (SDOH) Data Platform

Capstone for Outcomes Measures Harmonization Project

Period of Performance
6/1/18 – 5/1/21

Federal Point of Contact
Elise Berliner

Since early 2018, the AHRQ Registry of Patient Registries has compiled almost 4,000 registries for research, quality improvement, public reporting, and post-market surveillance.[31] Even though these registries have considerable potential to be used for PCOR, analysis of data derived from those registries can be challenging as they lack standardized definitions of outcomes measures. An AHRQ OS-PCORTF-funded project to harmonize measures in five clinical areas, ***Harmonization of Clinical Data Element Definitions for Outcome Measures in Registries***, 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 unstructured 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 Capstone 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.*

This Capstone project will address these barriers by exploring the treatment of depression as a use case. Researchers will collect complex outcome measure information on depression from clinical sites and will transfer those data to existing patient registries. Using tools developed from an OS-PCORTF-funded project, the Capstone will implement those tools 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 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), which will then be transferred to electronic patient registries; and 3) using new data collection from structured data capture or through natural language processing of clinical notes.

This Capstone 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. Findings from this project will enable other registries to perform collection and use of data that are cost effective for sites and that reasonably fit into the 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

- ***Collection of Participant-Provided Information:*** The project will use a standards-based approach for collecting patient-provided data and will develop tools for integrating depression outcomes calculated from these data into EHRs and registries. Also, the project aims to address the clinical workflow and technical barriers to collecting patient-provided data and transmitting that data into registries for use by researchers.
- ***Use of Clinical Data for Research:*** AHRQ's work will support interoperability between EHRs and data registries. The ability to link EHR and registry data will provide a more comprehensive look into patient profiles and support analyses across populations that may not have been studied otherwise. Additionally, the Capstone intends to address the technical barriers to integrating data from EHR systems into registries by creating supportive tools and guidelines.

Accomplishments

The project is working toward completing a number of activities in an aim to facilitate and advance the use of registries and patient-centered outcomes research.

- The team has completed initial steps to convene the Stakeholders Panel. The stakeholders include representatives from patient and provider organizations, payers, research funders, and federal agencies. The aim of the panel is to provide feedback throughout implementation of the Capstone project.
- The requirements for the FHIR app were finalized to enable the adoption of AHRQ's core set of depression outcome measures in the EHRs.
- The team is continuing to recruit clinical sites to participate in the Capstone data collection infrastructure building activities.

Publications and Other Publicly Available Resources

- Project findings are presented in a paper, "Harmonized Outcome Measures for Use in Depression Patient Registries and Clinical Practice," published in the *Annals of Internal Medicine* and available here: https://www.acpjournals.org/doi/10.7326/M19-3818.

Coordination with Other Federal Agencies

AHRQ will work with NIH/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 health care databases.

Enhancing Patient-Centered Outcomes Research (PCOR): Creating a National Small-Area Social Determinants of Health (SDOH) Data Platform

Period of Performance
3/1/19 – 9/1/22

Federal Point of Contact
Patricia Keenan

Given that a great deal of effort has been put toward improving the health care system processes and emphasizing prevention efforts in the population, policymakers are looking at other stressors on the health care system, including a better understanding of the community context in which health services are provided. In particular, there is a growing demand for a data platform that integrates information about social determinants of health (SDOH), health service utilization, and systems of care. For many decades, researchers have emphasized the importance of SDOH and have developed many different conceptual models to explain the inter-relationship between individual, family, and societal factors on the health of an individual. Research has demonstrated that for many SDOH factors, small-area data (data at the community or sub-county level) may be necessary to conduct meaningful analyses, and that for other SDOH factors, data at other geographic levels are more meaningful.

Considerable evidence exists about the relationship between the inequities in SDOH factors and poor health outcomes (e.g., mortality, acute and chronic disease, disability). Some studies have shown a relationship between SDOH factors and health care utilization, and studies are just beginning to emerge on the effectiveness of the health care interventions that integrate patient and community SDOH information on patient and community health outcomes. Health care systems are already moving toward obtaining critical SDOH data to improve care coordination and the quality of health care services for vulnerable populations.

Researchers spend substantial resources linking multiple data sets to create data files suitable for analyses because, the current data sets lack standardized metrics or estimates at the small-area level. Those databases are often derived for the purposes of looking at the health of an individual or health of the community and include information about the neighborhood and built environment, health and health care, social and community context, education, economic stability. However, these databases are limited in their ability to examine the SDOH in the small geographic areas. There is no complete source of longitudinal information with uniformly formatted community-level data on SDOH readily available for health service research.

Project Purpose and Goals

The goal of the project is to develop a consolidated set of national standardized databases on valid and reliable SDOH factors at the small-area and other geographic levels, building on existing federal agencies databases maintained by agencies such as AHRQ, HRSA, CDC, ASPE, and NIH. Data elements to be included will span the SDOH landscape and include measures of income, employment, food, housing, environment, economics, education, safety, transportation, justice system, market structure, laws, federal, state and local intervention initiatives, social capital, and health status and health care access and utilization.

The overall objectives of the project are to:

- Conduct an environmental scan to identify a comprehensive set of HHS and other federal data sets with existing or analyzable small-area level and other geographic level data on SDOH variables.

- Design and create a publicly available data platform of valid and reliable standardized set of SDOH data sources at various geographic areas to include, but not limited to: census blocks, zip codes, counties, primary care service areas, market areas, and states.
- Coordinate and expand the data collection efforts on SDOH across HHS.
- Use the new data platform to conduct a minimum of two PCOR studies.
- Disseminate the SDOH data platform to end-users across the federal government, PCOR researchers, and health services researchers.
- Establish a sustainability and growth plan for the SDOH data platform.

Contributions to PCOR Data Infrastructure Functionalities

Linking of Clinical and Other Data for Research: The SDOH data platform will make available data files that can be linked to existing federal databases via geographic identifiers.

Use of Enhanced Publicly Funded Data Systems for Research: The project will utilize existing data platforms developed by HHS agencies (e.g., AHRQ, HRSA, CDC, ASPE, and NIH) in an effort to capture small-area data at the community-level to improve researchers' ability to study the effectiveness interventions aimed at delivering whole person care.

Accomplishments

As the project is fairly new, the team has completed a number of initial tasks.

- The team has reached out to other departments in an effort to prepare for the federal Technical Expert Panel (TEP). The TEP is meant to provide expertise on the federal databases and guidance on the project objectives.
- Since the start of the project, the team has worked to establish the contracting mechanism for implementation of the SDOH data and analytic platform and anticipates issuing the award during FY 2020.

Coordination with Other Federal Agencies

In order to build a comprehensive SDOH databases, AHRQ will work with numerous departmental and federal agencies to access and link their data sets. Coordination with those agencies will be critical to continuing to support health care services research.

VII. Centers for Disease Control and Prevention (CDC)

CDC is administering ten active OS-PCORTF-funded projects including three cross-agency-funded projects described later in Section XII (Exhibit 7).

Exhibit 7. CDC Active Projects

CDC-Funded Projects
Augmenting the National Hospital Care Survey (NHCS) Data through Linkages with Administrative Records: A Project
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
Identifying Co-Occurring Disorders among Opioid Users Using Linked Hospital Care and Mortality Data
Making Electronic Health Record (EHR) Data More Available for Research and Public Health
MAT-LINK: MATernaL and Infant NetworK to Understand Outcomes Associated with Treatment for Opioid Use Disorder during Pregnancy
Strengthening the Data Infrastructure for Outcomes Research on Mortality Associated with Opioid Poisonings

Augmenting the National Hospital Care Survey (NHCS) Data through Linkages with Administrative Records

Period of Performance
6/1/19 – 12/1/21

Federal Point of Contact
Lisa Mirel

The NHCS, conducted by the NCHS, is designed to provide accurate and reliable health care statistics, describing national patterns of health care delivery in hospital-based settings, including prevalence of conditions, health status of patients, and health services utilization. The NHCS collects patient-level identifiers, which enables linkage of patient episodes of care within hospital inpatient and emergency department settings to other administrative data sources, providing a more complete picture of patient care. Previously funded OS-PCORTF projects have linked the 2014 and 2016 NHCS to mortality data collected from the NDI, creating a new unique data resource to support the study of post-hospitalization mortality outcomes in more than 3.2 million patients. OS-PCORTF also supported the linkage of the 2014 NHCS to the 2014-2015 CMS MBSF, which links claims data from NHCS patients with Medicare coverage to Medicare program enrollment and summary cost and utilization data at the time of hospitalization and after.

This project will expand on previously funded OS-PCORTF projects that increased the capacity of the NCHS to support a wide range of patient-centered outcomes research questions. This project will link the 2016 NHCS with 2016-2017 CMS Medicare Fee-for-Service claims, Medicare Advantage encounters, and patient assessment data from long-term care facilities and home health providers and with federal housing assistance program data collected from the HUD. These linked data resources will make it possible to integrate information on mortality, health care service utilization, prescription drug use, provider health assessments, and receipt of federal housing with a given patient's hospital administrative claims and EHR data. The linkage between the NHCS to CMS and HUD data sources will expand data

capacity to support research studies focused on a wide range of patient health outcomes including initiatives targeting opioid use and mental health care services, efficacy of treatment protocols, medical interventions and prescription drugs, health outcomes associated with different types of post-acute care services utilization, and other health disparities. The 2016 NHCS and 2016-2017 HUD linked data sources will allow researchers to examine the role of federal social support programs in health outcomes and treatment efficacy for persons with stable housing, with the ability to focus on specific subpopulations, including persons with.

Project Purpose and Goals

The project focuses on the following four objectives:

- Conduct a patient-level record linkage of the 2016 NHCS hospital administrative claims and EHR data to 2016/2017 CMS Medicare claims encounter, and health assessment data.
- Conduct a patient-level record linkage of the 2016 NHCS hospital administrative claims and EHR data to the 2016/2017 HUD administrative records on federal housing program participation.
- Refine probabilistic matching algorithms and disseminate detailed statistical methodology report to support high-quality future data linkage activities within and beyond the patient-centered outcomes research community.
- Create research files and user guidance documents to support researchers in using the new NHCS-linked data resources. The linked data sets will be available through the NCHS Research Data Center, and documentation will be made available via NCHS website.
- Disseminate tools and analytic guidance to stimulate the broader use of this new data resources to expand the capacities of patient-centered outcomes researchers.

Contributions to PCOR Data Infrastructure Functionalities

- ***Linking of Clinical and Other Data for Research:*** This project will expand and enhance the data capacity of publicly funded data systems to support research studies focused on a wide range of patient health outcomes.

Accomplishments

Since work on the project began in June of 2019, the project team has made progress on the following:

- A request for proposal was released during Q3 FY 2019. On August 19, 2019, CDC awarded the contract to conduct the linkage of the NHCS data to the CMS and HUD administrative records.
- The project team has begun work on the 2016 NHCS-2016/2017 CMS Data linkage.

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

Period of Performance
5/30/18 – 9/30/21

Federal Points of Contact
Aly Goodman
Marissa Sucosky

Childhood obesity affects over 20 percent of children age 2–19 years old.[32] In 2017, the U.S. Preventive Services Task Force (USPSTF) issued a recommendation to screen children for obesity and refer children with obesity to intensive PWMIs to improve physical and mental health.[33] 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, those working in PCOR have prioritized 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. An expanded health information data set 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.[34]

Exhibit 8. Data Networks

Child health data, including clinical information, social health determinants, weight management programs, and geographic markers are maintained throughout communities in many separate information systems operated by hospitals, provider networks, and clinical and community-based programs (see Exhibit 8 from the CODI project team). Currently, distributed patient-centered research networks routinely gather data collected in health care settings and structure these data in a common way. 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 USPSTF in 2017. These data will help researchers answer questions such as whether all children are being screened appropriately for obesity

and whether disparities exist by demographic groups and will also assist in the identification of PWMIs that are most effective.

The project will pilot enhanced tools and services (e.g., patient record linkage and de-duplication 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 will expand the availability of data for childhood obesity research that help health professionals develop and tailor interventions that are specific to the needs of children.

This project will take place in three phases: 1) align efforts and resources to support collaborative efforts across participants; 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, PWMI data, and community (census) data from multiple Denver health care and public health institutions that have existing collaborations for childhood obesity PCOR.

CODI aims to address the following research questions as they highlight data gaps within childhood obesity research:

1. *When is obesity screening (measuring BMI) occurring? In what settings? What actions does it trigger, including comorbidity screening?*
2. *What "dose" & characteristics of weight management interventions are associated with effectiveness?*
3. *What is the cost and cost effectiveness of weight management intervention?*
4. *What is the prevalence of obesity among children age 2–19 years and trends over time within small geographic areas?*

Project Purpose and Goals

The purpose of this project is to link pediatric clinical EHR data, PWMI 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 PWMI data, as well as community-level census data by expanding to capture CODI clinical and intervention data elements.
- Identify and describe the current business and technical processes and tasks for capturing the required childhood obesity data within health care and PWMI 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 de-duplication data tools, services, and IG will be publicly available on CDC's cloud-based Surveillance Data Platform.
- Pilot the enhanced CODI technical services and IG to test the ability to link and query data in CHORDS to produce data sets 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 patient-level EHR and weight management program intervention data 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

Since the project began in 2018, the CDC project team has completed numerous activities in an effort to expand the availability of childhood obesity patient-centered outcomes data.

- The CDC team completed use case documentation for two use cases: surveillance and health services, and a technical environmental scan. Also, the team finalized the evaluation plan and Business Process Analysis report.
- The team finalized the CODI Data Models IG.
- Stakeholder's functional, non-functional, and technical requirements have been completed, as well as a landscape analysis of record linkage and de-duplication services and tools to inform CODI.
- The CDC Collaborative Work Group meets regularly. At the December 2018 in-person meeting, the Work Group focused on identifying and finalizing priority research questions.

Publications and Other Publicly Available Resources

- The infographic is available on the project website are at: https://www.cdc.gov/obesity/initiatives/codi/childhood-obesity-data-initiative.html
- A technical information fact sheet is also available at: https://www.cdc.gov/obesity/initiatives/codi/codi-technical-information-sheet.html

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

Period of Performance
4/15/18 – 4/14/21

Federal Point of Contact
Carol DeFrances

National-level statistics on opioid-related hospitalizations are limited and often incomplete. From 2005 to 2014, the rate of opioid emergency department visits increased by an estimated 99.4 percent. Opioid overdose deaths in the emergency department also increased 27.7 percent from 2015 to 2016.35 These estimates do not identify specific drugs causing the rise in emergency department visits and overdoses, but the EHR data do contain much more detail, such as clinical notes and laboratory results, allowing a wider perspective. These data will allow health policymakers, clinicians, and researchers to develop effective prevention strategies and improve patient care.

The 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: 1) the NHCS collects inpatient, emergency department, and outpatient claims and EHR data from over 500 participating hospitals; 2) the NDI includes all deaths occurring within the U.S., along with cause of death, and 3) the National Vital Statistics System (NVSS) restricted-use files on drug-involved mortality data (DIM) which includes information on specific drugs involved in overdose death. Each source 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, have addressed improved specificity within each source. Now, the combination of the three 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 for emergency department 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 research data set by linking data between hospital care and mortality by merging the NHCS, NDI, and NVSS-M-DIM. 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-DIM data set 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.

This project aims to enhance the ability to identify opioids in the hospital setting through 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

- ***Use of Enhanced Publicly Funded Data Systems for Research:*** The linkage of the NHCS, NDI, and NVSS-M-DIM data into one merged data set for analysis will enhance researchers' ability to use these data to identify opioid-specific hospital encounters and deaths.
- ***Use of Clinical Data for 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

Since work on the project began in April of 2018, the project team has made notable progress on several tasks in support of project objectives:

- Convened five TEP meetings:
  - The first TEP was held on September 10, 2018.The meeting provided an overview of the project and status of the work completed since the award was made. Attendees were federal government stakeholders, including staff from FDA, NIH/NIDA, SAMHSA and ASPE. Preliminary results from the linked NHCS and NDI data were also presented, including that the linked data allows researchers to pull important trend information like the most frequent hospital diagnoses for people that died of a drug overdose.
  - The second TEP was held on December 7, 2018. TEP members recommended the project team identify federal agencies and projects who have used NLP for research in the development of the enhanced opioid-identification algorithm.
  - On April 8, 2019, the third TEP meeting was held to discuss progress in the development of the opioid-involved hospital visit case definition.
  - On June 13, 2019, the project hosted its fourth TEP on the project's NLP methodology. Hospital Medication Codes, code-based algorithms, and NLP methodology were discussed, along with preliminary findings on opioid mentions calculated using two different methodologies.
  - On September 17, 2019, the project hosted its fifth TEP, presenting its enhanced diagnosis code algorithm with preliminary analysis and a comparison of the enhanced code list and the original code list. Additionally, updates on the NLP methodology and process were discussed. The NLP process methodology and process includes the development of an annotation guide and use of clinical annotators to train the machine classifier. The machine classifier will be developed to identify confirmed opioid-involvement hospital encounters.
- A pharmacist, health research analyst, and program manager/clinical subject matter expert (SME) joined the project team, in order to provide research and analytic support to the project's NLP subject matter expert, on developing, tuning, training, and testing the NLP classifier.
- On February 1, 2019, the DMI SAS program was modified to run on NHCS 2016 data, in order to produce baseline counts of encounters with at least one opioid-related DMI term.
  - The case definition meeting was held on March 1, 2019, with staff from federal agencies and researchers. The group was able to reach a consensus on the definition of what is an opioid/opioid-like substance, how opioid involvement should be defined, and what evidence should be used to define an opioid-involved hospital encounter or mortality.

- The project team applied an enhanced diagnosis code algorithm to identify opioids in the 2016 NHCS encounters.

Publications and Other Publicly Available Resources

- The 2014 NCHS and 2014–2015 NDI data was linked to the 2014 NVSS-M-DIM file, which is now available in the NCHS RDC. This linked data file includes information on hospital care, mortality post-hospital discharge, and specific drugs mentioned in the literal text on the death certificate. Analysts created documentation for the RDC about the merged file.
- Completed the analysis on the merged NHCS/NDI/NVSS-M-DIM file, which included the development of research questions to understand the numbers of opioids found in hospitals and death certificates. A report on the analysis can be found on the NCHS website: https://www.cdc.gov/nchs/data/nhcs/opioid_health_outcomes.pdf
- "Enhancing Identification of Opioid-Involved Health Outcomes Using Linked Hospital Care and Mortality Data: Summary Report on Task 1" white paper was published online by NCHS and is available at https://www.cdc.gov/nchs/data/nhcs/opioid_health_outcomes.pdf. This report provides a brief description of the data sources, the processes used to link the data sets, and the variables available in the linked data sets. The report also presents example analyses to demonstrate the value of the linked data.

Coordination with Other Federal Agencies

NCHS will collaborate with SAMHSA, FDA, NIH/NIDA, and ASPE in the development of identification algorithms and dissemination.

Identifying Co-Occurring Disorders among Opioid Users Using Linked Hospital Care and Mortality Data

Period of Performance
5/1/19 – 10/31/21

Federal Point of Contact
Carol DeFrances

According to the 2017 National Survey on Drug Use and Health (NSDUH), the number of adults with SUDs who had any mental illness was about 8.5 million, and the number with severe mental illness was about 3.1 million people.[36] Based on 2015 NSDUH data, approximately 1.5 million adults with severe mental illness had misused opioids in the past year, which is equivalent to a co-occurrence of opioid misuse and severe mental illness in an estimated 1 in 8 adults (13%).[37] It is important for the National Hospital Care Survey (NHCS) to monitor the role that co-occurring disorders plays in opioid-related morbidity and mortality outcomes.

A previously funded OS-PCORTF FY 2018 project at NCHS, *Enhancing Identification of Opioid-Involved Health Outcomes Using Linked Hospital Care and Mortality Data*, has provided enhanced methodology to accurately identify a hospital patient's use of opioids in any form (i.e., used as directed, misused in a manner contrary to provider instructions, used intentionally to become intoxicated or for the purpose of self-harm, taken accidentally, etc.). Additionally, the project identified the specific legal or illicit opioid agent taken. This OS-PCORTF FY 2019 Capstone project will build upon the FY 2018 project methodology to flag evidence of co-occurring mental health disorders. Both projects use algorithms that determine the occurrence of an event (the use of opioids, type of opioid agent taken, and presence of a substance use or mental health issue) by selecting combinations of coded items (diagnoses, procedures, lab results, etc.) and terms contained in free-text (clinical notes, cause of death literal text). Both projects

will result in the creation of linked files that combine three data sources to enable access to data that follows patients with an opioid event for one year following hospital discharge. This will allow for retrospective analysis of the extent to which specific opioid agents and the co-occurrence of mental health disorders were involved in hospital encounters preceding post-discharge deaths.

Project Purpose and Goals

The goal of this project is to improve public health surveillance and expand researchers' access to data on health outcomes of opioid users with co-occurring substance use and mental health issues by completion of the following objectives:

- Develop a new set of algorithms that uses the linked NHCS/NDI/NVSS-M-DIM files to identify hospital encounters and death records involving patients with co-occurring disorders using medical code-based algorithm and natural language processing.
- Conduct a study to validate algorithms from this project and the FY 2018 *Enhancing Identification* project to identify the use of opioids and the existence of co-occurring disorders.
- Apply the validated algorithm to identify prevalence of opioid-involved emergency department visits and co-occurring disorders among opioid users in the 2016 linked NHCS, NDI, and NVSS-MDIM files.
- Provide data on opioid use and co-occurring disorders and make that data available through: 1) the NCHS Research Data Center, and 2) a previously developed interactive web portal for NHCS participating hospitals.
- Disseminate research findings from the validation study and the application of the validated algorithm to calculate prevalence of co-occurring disorders among opioid users in the linked data between NHCS, NDI, and the NVSS-M-DIM files.

Contributions to PCOR Data Infrastructure Functionalities

- ***Use of Enhanced Publicly Funded Data Systems for Research:*** This project will use the linked NHCS/NDI/ NVSS-M-DIM files to identify hospital encounters and death records involving patients with co-occurring disorders.
- ***Use of Clinical Data for Research:*** This project will use clinical data gathered from hospital encounters and mortality data in order to study how co-occurring disorders, such as substance use disorders, may contribute to opioid-related morbidity and death.
- ***Linking of Clinical and Other Data for Research:*** This project will link data from the NCHS/NDI NVSS-M-DIM files in order to identify hospital encounters or death records involving opioid-users with co-occurring mental health issues using NLP and medical code algorithms. This linking of data will expand researchers' access to information that can aid them in calculating the prevalence of co-occurring disorders among opioid-users.

Accomplishments

Since work on the project began in May of 2019, the project team has made notable progress on several tasks:

- An initial list of standard medical codes and terms was created that will be refined by SMEs and used to identify suspected cases of co-occurring disorders.

- The methodology of the combined FY 2018 and FY 2019 OS-PCORTF project's annotation and associated guidance were developed. The annotation of the NHCS data to develop the FY 2018 and FY 2019 algorithms will happen simultaneously to enhance efficiency of the annotation process.
- The NLP methodology and process were developed and refined. The NLP process methodology and process includes the development of an annotation guide and use of clinical annotators to train the machine classifier. The machine classifier will be developed to identify confirmed co-occurring opioid-involvement hospital encounters and SUDs or mental health issues.
- On July 3, 2019, a quarterly TEP meeting was held to discuss the project's case definitions. The group was able to reach a consensus on the definition of what is a co-occurring disorder, how substance use and mental health disorder involvement should be defined, and what evidence should be used to define a co-occurring hospital encounter or mortality.
- On September 25, 2019, a special meeting of the TEP was held to discuss the development of the annotation form that will be used to capture information to be used in the training of the machine classifier.

Making Electronic Health Record (EHR) Data More Available for Research and Public Health

Period of Performance
4/15/19 – 10/18/23

Federal Point of Contact
Maria Michaels

Interoperability of patient data remains challenging to achieve in real-world applications, especially those that do not involve direct patient care or payment. Real-time data exchange between health systems and research and public health is inconsistent and insufficient. Many patient-centered outcomes researchers and public health surveillance programs share a common challenge: they rely on clinical data that are frequently inaccessible and under-reported. As a result, they may be unable to answer critical questions that could lead to better, more patient-centered care or to leverage patient-level data for public health action.

Similarly, lack of access to EHR data can preclude innovative partnerships between providers and public health to advance patient outcomes. Patient-centered outcomes researchers and public health professionals need better ways to access data from different EHR systems without posing additional burden on health care providers. In recent years, the maturation of standards such as FHIR and the ONC requirements for certification of health IT that underpin many EHRs—such as the U.S. Core Data for Interoperability (USCDI) and APIs—have created an environment that is ripe for developing scalable and extensible solutions to overcome interoperability challenges. The purpose of this project is to develop a reference architecture, processes, and an application to address some of the identified challenges exchanging data between health care providers and public health. The project aims to use industry standards, supported by health IT certification requirements, designed for scalability and extensibility and licensed as open source software, to address these challenges.

Project Purpose and Goals

The project has selected three diverse use cases (hepatitis C [HCV], cancer reporting, and health care surveys) to model and inform the design of the application. After development, the project team will fully test the application on an HCV use case to ensure it can extract data from multiple clinical organizations using different EHR platforms. The goal is to leverage a common reporting framework to provide the data

needed by researchers and public health officials to generate scientific evidence about a given health problem.

The project will focus on the following objectives:

- Define the use cases and research questions relevant to end-users.
- Use existing health data and exchange standards (e.g., FHIR) to develop an application for real-time data exchange between EHRs and research and public health systems.
- Establish the infrastructure to create a collaborative network of clinical, research, and public health organizations and health IT vendors as a ready testbed for testing developed tools.
- Implement the application for at least one use case (HCV) in both clinical research and public health surveillance contexts.
- Evaluate the HCV use case implementation to measure improvements in the timeliness and completeness of data reported for research and public health.
- Develop a plan for broad use and long-term sustainability of the application, including publishing the application as open source software.

Contributions to PCOR Data Infrastructure Functionalities

- ***Using Clinical Data for Research:*** Optimizing data for research by improving access, enhancing quality, and promoting interoperability of clinical data across multiple sources.
- ***Standardized Collection of Standardized Clinical Data:*** Better defining and standardizing key data terms and concepts (i.e., common data elements, or CDEs) to more effectively and efficiently share, link, and aggregate across data sources.
- ***Use of Enhanced Publicly Funded Data Systems for Research:*** Enhancing federal and state-level data systems to enable greater access, use, linkages, and analysis of publicly funded data for research.

Accomplishments

The project made progress in achieving the aforementioned objectives:

- The project completed a landscape analysis to evaluate implementation of FHIR and other standards among the EHR vendors.
- CDC hosted a 100-member TEP kick-off meeting on October 15, 2019, to gather information on the scientific, technical, and practical aspects of the application development. The TEP consists of subject matter experts in the three use cases.

Coordination with Other Federal Agencies

- The project established regular collaborators meetings with NIH/NCATS, FDA, NIH/NCI, and ONC to leverage other agencies expertise related to specific project objectives. For example, FDA is working with NIH/NCATS to harmonize the PCORnet CDM to FHIR, particularly for the FHIR mappings of the HCV use case.

MAT-LINK: MATernaL and Infant Network to Understand Outcomes Associated with Treatment for Opioid Use Disorder during Pregnancy

Period of Performance
3/4/19 – 12/31/22

Federal Point of Contact
Shin Kim

From 1999–2014, the prevalence of OUD among pregnant women in the U.S. quadrupled from 1.5 to 6.5 per 1,000 delivery hospitalizations.[38] Opioid use during pregnancy increases the risk of an infant being born with neonatal abstinence syndrome (NAS). Recent evidence suggests that children born with NAS may experience developmental delays;[39,40] however, the developmental trajectory of these children has not been systematically studied.

The American College of Obstetricians and Gynecologists recommends that pregnant women with OUD be offered an opioid agonist (methadone, buprenorphine) to treat OUD during pregnancy. In some clinical settings, pregnant women may also be offered medically supervised withdrawal (opioid detoxification).[41] There is limited information, however, comparing maternal, infant, and child health outcomes associated with these treatment regimens during pregnancy.

In 2016, CDC established a surveillance system called the U.S. Zika Pregnancy and Infant Registry (USZPIR) to monitor pregnant women and their infants with laboratory evidence of Zika virus infection. This system captures timely data, which are used to inform clinical and public health practice recommendations. The lessons learned and tools used to establish the USZPIR will be adapted and leveraged for the MAT-LINK, a surveillance system to monitor the maternal, infant, and child health outcomes following treatment for OUD during pregnancy.

Results from MAT-LINK will be used to improve understanding of the spectrum of maternal, infant, and child health outcomes following treatment for OUD during pregnancy and the role of mediating and moderating factors on maternal and infant outcomes, including exposure to multiple substances, maternal comorbidities, and other psychosocial factors. MAT-LINK will also be used to improve policies, clinical practice recommendations, and clinical decision-making. In addition, this project will develop and pilot a data platform to collect and link maternal, infant, and child data across clinical sites and which can be modified to collect linked data on other exposures during pregnancy.

Project Purpose and Goals

MAT-LINK will establish a surveillance network, consisting of three to five clinical sites, to collect data on maternal, infant, and child health outcomes associated with treatments for OUD during pregnancy. The overall objectives of the project are to:

- Establish an organizational structure, which would include federal agency representatives, clinical and public health partners, and the CDC Steering Committee, to provide critical input on linked maternal and infant health data collection approaches and analytical priorities.
- Develop a data platform to collect linked maternal and infant data among women treated for OUD during pregnancy.
- Analyze and disseminate preliminary results to inform patient-centered care for pregnant women with OUD and for infants and children with prenatal opioid exposure.

Contributions to PCOR Data Infrastructure Functionalities

- ***Standardized Collection of Standardized Clinical Data:*** This project will create a core set of variables to standardize surveillance of OUD pregnancies.
- ***Collection of Participant-Provided Information (PPI):*** This project will include extracting and abstracting data from patient's medical records. Lessons learned will inform best practices for collecting and utilizing medical record information.

Accomplishments

In the first few months of the project, the CDC team accomplished the following tasks:

- Completed a project objective by establishing an organizational structure, including a CDC Steering Committee and Partners Group comprised of federal, clinical and public health partners. After an inaugural meeting, the Steering Committee convened to discuss status updates and define clinical site inclusion criteria.
- Completed a project deliverable by publishing a Notice of Funding Opportunity to identify an implementation partner. After reviewing applications, CDC selected the Public Health Informatics Institute (PHII).
- PHII released the request for proposals (RFP) for the pilot clinical sites and is preparing to award four clinical sites to participate in the collection of maternal and infant data. Following the release of the RFP, PHII and CDC received 25 Letters of Intent from clinical sites located across the continental U.S. PHII posted the RFP on their website and continues to post MAT-LINK updates on their public-facing webpage: https://phii.org/blog/MAT-LINK.
- A manuscript has been drafted to provide the background and justification for MAT-LINK, a list of maternal, infant, and child health variables topics, and propose potential key questions to be addressed by MAT-LINK, with a tentative publication date of late spring/summer 2020.

Publications and Other Publicly Available Resources

- A public-facing CDC project webpage was created to post background information about their work, implementation partner, external partners, goals, example variables, and link to their implementation partner's website. The website can be accessed here: https://www.cdc.gov/ncbddd/aboutus/mat-link.html

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

Period of Performance
5/1/18 – 5/14/20

Federal Points of Contact
Steven Schwartz
Kate Brett

Cause of death information from death certificate data is often used by researchers and those in public health for programmatic, policy, and outcomes research needs. This project, will redesign and enhance the Vital Statistics Rapid Release (VSRR) and Medical Mortality Data System (MMDS), both subsets of the NVSS, 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 U.S... Developed in the 1980s, it uses algorithms to assign underlying cause of death, multiple causes 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 a physician completes the certificate using their EHR.[42] Medical examiners and coroners typically certify death in the case of drug overdose, so integrating the case management systems used by coroners with state electronic death registration system (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 FDA will be used as an enhanced prototype to strengthen the MMDS coding process through collaboration with NIH/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 the 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 using machine learning and NLP techniques.
- 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 an 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

The project team made a great deal of progress in setting up the systems, contracts, and processes needed to achieve the project goals:

- NCHS established a new development server to support MedCoder, the modernized cause-of-death coding system. The server was then tested to validate that the output of the new system with the current system. Continued testing of the capacity of the new MedCoder system at production capacity is ongoing, and the results of the testing are driving system improvements.
- The project team, with support from NIH/NLM, completed mapping a process for creating and maintaining a drug database, beginning with the use of RxNorm for commercially manufactured drugs and expanded with illicit drug data obtained from the Department of Justice. NIH/NLM has created the initial set of data, which is ready to populate in the database for the Division of Vital Statistics (DVS).
- Contracts were awarded to six vital registration jurisdictions (CA, FL, GA, MI, NH, and 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.
- In February, representatives from the six implementers' community vital records staff, as well as software developers from several data systems along the mortality data pathway, demonstrated the use of an HL7 FHIR-enabled API to transmit data from one reference implementation of an EDRS to NCHS as well as to another reference EDRS. The project successfully transmitted data from EDRS to NCHS and one jurisdictional EDRS to another, and coded data from EDRS to a state cancer registry.
- Organizational representatives tested the Vital Records Death Reporting IG attended at the two-day HL7 FHIR Connectathon in Atlanta, Georgia on September 14–15, 2019. Many successes were achieved, including the transmission of data from EDRS to NCHS and back again, one jurisdictional EDRS to another, and coded data from EDRS to state cancer registry.
- NCHS contracted with three states who have added neonatal abstinence syndrome (NAS) as a reportable condition on their birth certificates to share data files linking hospital discharge reports with birth certificates.
- A National Health Statistics Report on a feasibility study demonstrating the types of research questions that can be answered will be publicly available in June 2020.

Publications and Other Publicly Available Resources

- The Vital Records Mortality and Morbidity Reporting FHIR IG was brought to HL7 ballot in May 2019 and was approved with comments for trail use. The Vital Records Mortality and Morbidity Reporting FHIR IG (v0.1.0: STU 1 Ballot 1) is available here: http://hl7.org/fhir/us/vrdr/2019May/

- NCHS released monthly provisional drug overdose death counts. The provisional data presented include: 1) the reported and predicted provisional counts of deaths due to drug overdose occurring nationally and in each jurisdiction; 2) 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 3) the reported and predicted provisional counts of drug overdose deaths involving specific drugs or drug classes occurring nationally and in selected jurisdictions. The data are accessible here: https://www.cdc.gov/nchs/nvss/vsrr/drug-overdose-data.htm
- NCHS published the following special report on fentanyl deaths: Spencer MR, Warner M, Bastian BA, Trinidad JP, Hedegaard H. (2019). Drug overdose deaths involving fentanyl, 2011–2016. *National Vital Statistics Reports*, *68*(3). The report can be found on CDC's website: https://www.cdc.gov/nchs/data/nvsr/nvsr68/nvsr68_03-508.pdf

Coordination with Other Federal Agencies

CDC is working closely with NIH/NLM and DEA on this project. NIH/NLM holds valuable data for pharmaceutically manufactured drugs. Currently, there is no comparable reference library or standardized terminology for the illicitly manufactured drugs. DEA has agreed to make a reference list of illicitly manufactured drugs, which will help the NIH/NLM project team move forward with creating a supplemental drug file, which will include all substances, both illicit and pharmaceutically manufactured, for accurate death certificate reporting.

VIII. Food and Drug Administration (FDA)

FDA is administering seven active OS-PCORTF-funded projects including four cross-agency-funded projects described later in Section XII (Exhibit 9).

Exhibit 9. FDA Active Projects

FDA-Funded Projects
Bridging the PCOR Infrastructure and Technology Innovation through Coordinated Registry Networks (CRN) Community of Practice (COP)
SHIELD—Standardization of Lab Data to Enhance Patient-Centered Outcomes Research and Value-based Care
Standardization and Querying of Data Quality Metrics and Characteristics for Electronic Health Data

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

Bridging the PCOR Infrastructure and Technology Innovation through Coordinated Registry Networks (CRN) Community of Practice (COP)

Period of Performance
4/16/19 – 9/30/22

Federal Point of Contact
Danica Marinac-Dabic

Existing comparative effectiveness research often relies on data captured at the point of care, re-entered into clinical research systems, and then consolidated and transformed for the analysis and research purposes. The process is viewed as complex, labor-intensive, and expensive, as it requires duplicate data entry, extensive data validation and normalization to assure accurate and effective evaluation. As a result, both study designs and study infrastructure for generating and appraising real-world evidence are often limited. To improve the capacity to study medical devices in real-world settings, the FDA Center for Devices and Radiological Health (CDRH) launched a series of strategic efforts to improve national capacity to study medical devices.

FDA and its partners have invested significant resources into standing up a strategic CRN infrastructure designed to facilitate the evolution of traditional single-purpose registries into CRNs.

Capturing standardized data will streamline data collection and support exchange of data across therapies, patient populations and episodes of care. For example, by including unique device identifiers (UDIs) in the CRN minimum core data set, the CRNs become suitable for routinely studying device-specific questions; by including patient-generated data in the CRN, the patient-centered research community can link the patient experiences to clinical data sources, and regulators can identify how devices are performing for patients and identify patient-centric endpoints for future studies. Lastly, by linking registry data to additional data sources (such as claims submitted by providers to bill for medical services), patients and providers can have a more comprehensive understanding of long-term outcomes and adverse events associated with health technologies. Such strengthened infrastructure would lead to better evidence generation in general and be better suited to address the gender/sex specific differences in health technologies.

The goal of this project is to strengthen the CRNs as a real-world data source for high-quality, relevant, reliable, timely, and actionable evidence to improve patient outcomes of medical devices—specifically for technologies affecting women's health. This will be done by advancing the CRNs' ability to capture standardized data, through harmonization of their minimum core data sets, commitment to incorporation of device identification, include patient-generated data, and link their registry data to additional data sources through the learning community (the CRN Community of Practice). The CRN Community of Practice (CRN-COP) consists of CRNs across 12 clinical areas. This collaborative initiative will provide an opportunity to collect structured, standardized, analysis-ready patient data at the point of care. Strengthening the CRN-COP and each individual CRN will result in a more strategic approach to addressing the needs of the entire PCOR stakeholder community via harmonized and interoperable infrastructure and potentially allowing for more complex study designs.

Project Purpose and Goals

The project's goal is to strengthen the CRNs as a national infrastructure for technology affecting women's health through the following five objectives:

- Advance the CRNs capacity for patient-centered outcomes research across 12 clinical areas through their development in 7 areas (attributes): patient engagement, unique device

identification, data quality, efficiency, governance, sustainability, and fitness for use during the total product life cycle.

- Pilot-test and refine the existing device-specific FHIR profiles (produced as part of the FY 2017 project *Developing a Strategically CRN for Women's Health Technologies*) in an expanded set of three to five CRNs to demonstrate the capture and exchange of CRN data using FHIR.
- Pilot-test and refine the existing instrument for capturing PROs to capture patient preferences in the End Stage Renal Disease registry to evaluate scientifically valid data regarding patient uncertainty in accepting a variety of benefit/risk tradeoffs within a CRN.
- Advance CRN capacity to produce linked data sets and combine heterogeneous data and develop machine learning techniques to validate the linked data sets.
- Develop a gender- and sex-specific outcome measure framework for devices and test it in the most mature CRNs.

Contributions to PCOR Data Infrastructure Functionalities

This project addresses all three components, services, standards, and policies and governance, and all five functionalities of the HHS Strategic Framework:

- ***Linking of Clinical and Other Data for Research:*** The project will build upon the registries established in the strategically CRN infrastructure and link the registries to additional data sources, such as claims data, EHR data, and patient generated data. This will allow researchers and regulators to have a more complete picture of patient outcomes and experiences and access to analysis-ready data for surveillance and evidence generation.
- ***Standardized Collection of Standardized Clinical Data:*** This project will promote the Standardized Collection of Standardized Clinical Data by including UDIs in the CRN minimum core data set. This will allow for consistent, reliable, and streamlined study of device-specific questions.
- ***Collection of Participant-Provided Information:*** In collaboration with the above functionality, the project plans to collect standardized patient data at the point of care. This patient-generated data will then be incorporated into the CRN, expanding the volume and depth of patient-provided information accessible to researchers.
- ***Use of Clinical Data for Research:*** Clinical data sources will serve as a key source of linkage to other data and registries, such as patient-generated information. This will promote interoperability across multiple data sources and ultimately expand the CRN's ability to capture patient preference and connect it with their health record and outcomes.

Accomplishments

The project began working toward its project goals in the middle of FY 2019:

- A funding award was issued to the MDEpiNET Coordinating Center.
- The draft CRN Governance Management Plan was developed establishing the governance approach and processes, decision-making authorities, and project monitoring objectives, metrics, and deliverables.

- The Core Minimum Data Set was identified for a few CRNs, including Women's Health, Study of Prostate Ablation Evidence Development (SPARED-CRN), Orthopedic Devices Coordinated, and International Consortium of Orthopaedic Registries (ICOR).
- A paper to describe the assessment for each CRN related to each of the seven attributes is underway. This will provide background on the applied methodology and implementation roadmap for the assessment tool.
- Using the FHIR IG developed under the first project, *Developing a Strategically Coordinated Registry Network (CRN) for Women's Health Technologies*, FDA has begun refining the guide for one of the CRNs. This will help inform and facilitate the exchange of clinical and administrative data to support device evaluation.

SHIELD—Standardization of Lab Data to Enhance Patient-Centered Outcomes Research and Value-based Care

Period of Performance
4/1/19 – 10/21/21

Federal Point of Contact
Michael Waters

Every day, physicians must draw conclusions about a patient's health status based on a snapshot in time during a brief visit, conversation with the patient, and limited historical health information. A simple and meaningful representation of a patient's diagnostic history can be invaluable in contextualizing their personal health care challenges at any single point in time. Laboratory data, however, is commonly excluded from many research efforts due to challenges with data quality and interoperability. For example, the absence of semantic interoperability—or, the ability of two or more systems to exchange, use, and analyze information the same way—such as, variation in the coding of *in vitro* diagnostic (IVD) tests—has been frequently cited as a significant impediment to continuity of patient care, research, and public health care in general. While IVD data are largely digitized, which is a benefit, IVD data should be represented in a standard way and yet is often coded different between institutions (or even within an institution) leading to ambiguity and reducing data utility in research and health care.

To address this critical need to improve laboratory data infrastructure, public workshops were held in 2015 and 2016, leading to the formation of the SHIELD (Systemic Harmonization and Interoperability Enhancement for Laboratory Data) collaborative. The SHIELD collaborative is a multi-agency/stakeholder network consisting of FDA, CDC, NIH, ONC, CMS, U.S. Department of Veterans Affairs (VA), IVD manufacturers, EHR vendors, laboratories, College of American Pathologists, standards developers, Pew Charitable Trusts, National Evaluation System for Health Technology (NEST), and academia.

This project will expand the collaborative efforts of FDA and other stakeholders involved in the SHIELD initiative to help improve laboratory information interoperability by ensuring that critical laboratory results (e.g., was an infection detected?) are recorded the same way across different EHR systems. To realize this goal, this project will develop manuals to consistently map the same LOINC codes (Logical Observations Identifiers Names and Codes) to the same type of IVD. This is important because without the specific guidance, laboratories and registries often assign different (and frequently incorrect) LOINC codes for the same type of IVD test. This project will also pilot implementation of SHIELD standard digital formats, which will include the seamless distribution of LOINC and SNOMED-CT (Systematized Nomenclature of Medicine—Clinical Terms) coding to update the infrastructure among those provider institutions and registries.

Project Purpose and Goals

This project aims to improve the quality, interoperability and portability of laboratory data within and between institutions so that diagnostic information can be pulled from different sources or shared between institutions to help illuminate clinical management and understand health outcomes. The overall objectives of the project are to:

- Use the process created by the Regenstrief Institute (i.e., the owner, developer, and curator of LOINC codes) and used for Microbiology, to develop LOINC code mapping manuals for the remainder of IVD domains: Chemistry, Drug/Toxicology, Allergy, Serology/Hematology, Cell Markers, and Molecular Pathology.
- Conduct pilot implementations and testing of an interoperability upgrade to existing laboratory information systems and registries by incorporating SHIELD-approved, high-quality, industry-defined and supported format to facilitate the publication and exchange of LOINC codes for vendor IVD test results.
- Assess the interoperability and value of the systems and tools tested in six pilot laboratory settings and EHR systems pre- and post-implementation.

Contributions to the PCOR Data Infrastructure Functionalities

The project addresses the five functionalities of the HHS Strategic Framework for PCOR data infrastructure by proposing changes that will enable:

- ***Standardized Collection of Standardized Clinical Data:*** This project will help standardize clinical data by harmonizing the LOINC codes being applied to diagnostic testing and the SNOMED codes related to results, which will in turn improve data sharing and analysis by reducing inconsistencies and redundancies in data.
- ***Linking Clinical and Other Data for Research:*** SHIELD will work to standardize data so that it can be accurately aggregated from multiple labs and/or registries and be accurately identified as either duplicative or non-duplicative of results from other data streams.
- ***Use of Clinical Data for Research:*** This project will enhance the use of clinical data for research by improving the standardization of lab data, which will increase both its accuracy and its utility.
- ***Use of Enhanced Publicly Funded Data Systems for Research:*** The project will leverage existing federal data sources (i.e., the FDA's Unique Device Identification [UID] System), in addition to private clinical and laboratory data, to enhance the compatibility of lab results from different systems.

Accomplishments

The project has recently commenced activities and therefore its activities are in progress.

- At present, two of the six IVD LOINC mapping manuals have been drafted.
- Many of the tests intended for implementation have been identified, mapped, and coded in LIVD (for LOINC to vendor IVD)—the standard specification for publishing and exchanging LOINC codes for vendor IVD test results.[43] The project team has also established the clinical information models and version control necessary for the pilot-test implementation within up to six health care institutions.

- The project team has met with the Clinical Laboratory Standards Institute (CLSI) to develop a document to describe to laboratories how to consistently implement the laboratory informatics standards developed through the project.

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

Period of Performance
8/20/16 – 9/30/19

FDA Federal Point of Contact
Jamila Mwidau

Currently, limited standards exist for describing and characterizing the quality and completeness of electronic health data. In order to best utilize this data, it is imperative that researchers and investigators understand the characteristics of the data sources in order to determine whether the data are fit-for-use. Effective use of the growing number of data sources and distributed networks will require adoption of a uniform approach to describing the characteristics of electronic health data, as well as the data capture characteristics at the institutional, provider, and health plan level and data domain level. This project supports the development and testing of metadata standards that will help researchers assess data quality to determine fitness of use for particular research purposes.

Project Purpose and Goals

This project developed and implemented 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 team leveraged the work of a prior APSE FY 2015 funded project—the Cross Network Directory Service (CNDS), which focused on helping researchers discover appropriate data for their studies. Data Quality Metrics (DQM) extends the work of the CNDS in two ways: first, by leveraging CNDS governance and access control capabilities, and second, by helping researchers investigate the characteristics of the data sources and the quality of specific data elements and domains.

The project objectives were to:

- Develop, test, and implement a standards-based approach to describing data quality and presenting data quality metrics, focusing on data held in EHRs and claims systems.
- Implement the new metadata standards in at least two distributed research networks.
- Incorporate the standards in open source software tools (web application, flexible data model, visualization templates)

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.
- ***Use of Enhanced Publicly Funded Data Systems for Research:*** The project will create an open source tool that can be implemented with any common data model ensuring broad applicability. With this work, the FDA team aims to provide a harmonized approach to data characterization across multiple sources 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

Since work on the project began in August 2016, the project team has completed several activities in support of their goals and objectives. The project's work occurred in three phases: 1) discovery and design; 2) development and testing; and 3) implementation and release.

During the first phase, discovery and design, the project team evaluated existing data quality frameworks and processes, and developed a data quality data model to enable the exploration of DQM in a way that is flexible and agnostic to the CDM. The team began by conducting an environmental scan and literature review of publications about data quality methods, standards processes, resources, and standards. Through exploration of existing practices and published work, the team developed an initial data quality data model. The model is organized around a central table that captures measurements (counts of patients, maximum or minimum values, frequency of values, etc.) and that is surrounded by tables that identify for each measurement the source system, the context (patient, member, encounter, claim, etc.), any relevant stratifications (e.g., age ranges), 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.[44] Along with the data quality data model, the team delivered the proposed set of standards and 78 use cases in the Discover and Design Report – which also includes the technical specifications for implementing the standards and a dictionary describing each metadata element.

During the second phase, development and testing, the project team developed a web-based system and accompanying data quality data model database to store data quality information. The team began with developing a prototype of the system, which was then shared with stakeholders for feedback. The team incorporated the feedback from the stakeholders as part of the final software development phase. There were over 25 participants that attended the stakeholder sessions.

For the third and final phase of the project, implementation and release, the project team documented all the project activities and made them publically available to the stakeholder community. Some of the deliverables include: 1) the technical system documentation targeted toward software developers and other technical audiences; and 2) user documentation targeted towards system end-users (e.g. investigators). Both the technical system documentation and user documentation can be found on the DQM public GitHub repository. In addition, the project team published the software code for the web-based system.

Throughout the project, the team documented several lessons learned with two themes centered on governance and requirements for contributors. Regarding governance, the team acknowledged that operationalizing the DQM system will require designated funding and a 'Coordinating Center' to operate the system. In addition, although the team developed a beta-version of the system and provided the code and technical documentation on the DQM public GitHub repository, the project team recognizes that they need a production-level version. The production-level version of the system with web hosting would enable any individuals interested in the system to utilize it. Regarding potential requirements for contributors, the project team discussed the incentives for and barriers to participating with multiple stakeholders. One proposed approach is to have one or multiple networks support the implementation of the DQM system, and have their network data sources join to create a mass of users, that may influence further participation. Another approach discussed by the project team is to implement a production

version of the system with a Coordinating Center within FDA to enable the data quality of data sources in the Sentinel system and more broadly.

Publications and Other Publicly Available Resources

- The Data Quality Metrics Authoring and Querying Platform is a cloud-based, open-source tool that allows user to develop and author new metrics, capture data quality metric measures, and support the evaluation and visualization of supplied measures. https://github.com/PopMedNet-Team/DataQualityMetrics

- The Technical Documentation report provides technical documentation appropriate for software developers and other technical users to facilitate their use of the DQM system. It is available in the GitHub repository for reference along with the open source code for the DQM System.

- The User Documentation report provides detailed user documentation related to the use of the web-based DQM system. The report is written to support researcher/investigator users of the system by describing all elements of the web-based system and providing instructional detail on use by an individual. It is available in the GitHub repository for reference

- The project team presented and demonstrated the system during four stakeholder sessions in the month of September. All four sessions were recorded and posted on the site at the following link: https://dataquality.healthdatacollaboration.net/resources

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.

IX. National Institutes of Health (NIH)

NIH is administering eight active OS-PCORTF-funded projects including five cross-agency-funded projects described later in Section XII (Exhibit 10).

Exhibit 10. NIH Active Projects

NIH-Funded Projects
Emergency Medicine Opioid Data Infrastructure: Key Venue to Address Opioid Morbidity and Mortality
NIH/NIDA's AMNET: An Addiction Medicine Network to Address the United States Opioid Crisis
Use of the ADAPTABLE Trial to Strengthen Methods to Collect and Integrate Patient-Reported Information with Other Data Sets and Assess Its Validity

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

Period of Performance
6/1/18 – 5/31/20

Federal Point of Contact
Kristen Huntley

To conduct impactful PCOR, merging patient-reported information and EHR-derived data can help develop meaningful data sets that support enhanced evaluation of patient outcomes. Emergency departments are critical points of entry for OUD patients, and therefore they present an opportunity to organize OUD data and conduct OUD-related research. The American College of Emergency Physicians (ACEP) developed the first emergency medicine registry, called the Clinical Emergency Data Registry (CEDR) 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.[45]

This project will assess use of OUD-specific CDEs that could help improve the quality of care at the point of contact. In recent years, PROs have become recognized as capturing important information that is often excluded from or not accurately captured in EHRs, such as pain intensity and SUD treatments. This project will also explore the feasibility of collecting PRO measures in emergency department settings and after an emergency department visit and will identify PROs most appropriate for inclusion in emergency department 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 emergency department settings addresses key areas of the HHS 5-point strategy to combat the opioids overdose 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.[46]

Project Purpose and Goals

The goal of this project is to study emergency department clinical data research infrastructure to enhance capacity to conduct opioid-related research in emergency departments through PROs and registries.

The project objectives are to:

- Identify existing OUD CDEs relevant to the emergency department 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 emergency department EHRs will be developed.
- Demonstrate that CDE from emergency department EHR test sites can be integrated into the ACEP CEDR by translating and mapping electronic OUD data elements to the NIH/NLM Value Set Authority Center (VSAC) and testing the validity and feasibility of aggregating and analyzing data from the CEDR test sites.
- Explore the feasibility of collecting electronic PROs, such as PROMIS and other measures (e.g., pain intensity, SUD treatment/status) at NIH/NIDA's Clinical Trial Network (CTNs) sites.[47] The PRO tool will be implemented and integrated into routine emergency department clinical workflow to study feasibility of collection and use.

Contributions to PCOR Data Infrastructure Functionalities

- ***Standardized Collection of Standardized Clinical Data:*** The project focuses on identifying existing and potential CDEs for OUD that are relevant to the emergency department setting to facilitate standardized reporting from the EHR 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 patient-provided information by testing the feasibility of collecting PROs most appropriate for collection in emergency department settings.
- ***Linking of Clinical and Other Data for Research:*** The project focuses on demonstrating that CDEs from emergency department EHR test sites can be transmitted or integrated into the ACEP CEDR.

Accomplishments

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 toward their achievement.

- The literature review and environmental scan were completed during Q4 FY 2019. The purpose of those deliverables are to examine current publicly available data systems, data elements, and quality measures to identify OUD data elements suitable for capture in the EHR or use in the emergency department setting of existing OUD CDEs.
- A technical report identifying existing CDEs for OUD relevant to the emergency department setting was adapted for a manuscript, approved by the CTN Publications review committee, and accepted for publication in *Addiction Science & Clinical Practice*.
- Four potential test sites have been identified and selected to assess whether CDEs from the EHRs can be transmitted into the ACEP CEDR. The project team plans to finalize a chart review guide and data collection form for those test sites.
- A prototype of the application to PROs and test data element validity has been developed and is being tested.

NIDA's AMNET: An Addiction Medicine Network to Address the United States Opioid Crisis

Period of Performance
4/15/19 – 10/15/2022

Federal Point of Contact
Carlos Blanco

Only about one-quarter of the over 2 million Americans with OUD receive treatment.[48] The three FDA-approved medications for treatment of OUD—buprenorphine, methadone, and extended-release naltrexone—are provided in many office-based and community medical practices throughout the U.S.[49] Office-based practices, which provide buprenorphine and naltrexone, play a key role in the treatment response to the epidemic because they have the capacity to treat many more patients than can be accommodated in the limited number of opioid treatment programs (OTPs), including rural and other underserved communities. However, little is known about outcomes for the patients treated with buprenorphine and naltrexone in office-based practices. These practices typically do not collect standardized data on patients' characteristics, treatments, and outcomes and have not been harnessed to conduct patient-centered outcomes research. By connecting these settings to a practice-based research

network, this project will enhance addiction-related data collection and capacity to conduct PCOR focused on treatment of OUD.

Project Purpose and Goals

This project will establish a new practice-based research network and an electronic patient registry named the Addiction Medicine Network (AMNet). AMNet will be designed to collect data for clinical and health services research related to addiction. The AMNet registry will serve as a platform for research on OUD and its treatments. The foundation for AMNet will be the American Psychiatric Association's (APA) clinical data registry (PsychPRO) and the over 40,000 members of the APA and the American Society of Addiction Medicine.

AMNet's addiction medicine practitioners will collect standardized treatment and outcomes data, including clinical and patient-reported data relevant to the opioid epidemic from routine clinical practice (including in understudied populations). Participation in AMNet will improve clinical decision-making among addiction treatment providers and support research by AMNet participants and extramural scientists. With a focus on practice-based research, AMNet will develop the foundations of a research network for community-based clinical trials (e.g., comparative effectiveness of extended-release formulations of OUD medications). AMNet will provide real-time data on patient characteristics, care delivery, and recovery service utilization that can be used to improve patient outcomes through performance improvement efforts.

The specific objectives of this project are:

- Establish AMNet, an addiction medicine practice-based research network.
- Adapt PsychPRO to support data collected for AMNet.
- Perform feasibility and validity testing of AMNet measures and OUD CDEs.
- Expand addiction medicine research capacity and outreach through AMNet participant training and related dissemination activities.
- Develop business requirements for linking AMNet to other databases and registries. The first data linkage project will link AMNet data for providers participating in CMS' Merit-Based Incentive Program (MIPS).

Contributions to PCOR Data Infrastructure Functionalities

- ***Standardized Collection of Standardized Clinical Data:*** In order to ensure interoperability, AMNet will standardized the collection of treatment and outcomes data by establishing OUD Common Data Elements to effectively share data across the research network.
- ***Collection of Participant-Provided Information:*** The AMNet registry will incorporate patients-reported data to more fully capture the patient's experience through treatment and recovery to better evaluate outcomes.
- ***Use of Clinical Data for Research:*** The project will leverage APA PsychPRO's existing process for implementing secure patient and clinician portals to develop an effective registry that helps combat the current opioid epidemic. Mainly, the clinical data registry will enable AMNet to collect standardized treatment and outcomes data, including patient-reported data.

- ***Linking Clinical and Other Data for Research:*** Linkage between AMNet and Researched, Abuse, Diversion, and Addiction-Related Surveillance System (RADARS) will be established to improve identification and tracking of patients with OUD.

Accomplishments

As the project is still in its launch phase, the team has completed some initial activities to establish the groundwork needed to establish the research network.

- The team has selected external Steering Committee members to provide guidance on numerous items, including assembling inclusion criteria for assessment measures and CDEs, identification of research questions, and defining publication policies.
- Environmental scan of screening and assessment measures has been drafted.
- Data use agreement policy has been reviewed.

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

Period of Performance
8/31/16 – 8/31/19

Federal Point of Contact
Wendy Weber

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.[50] 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 patient-centered research.

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 comparative effectiveness 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

- ***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. The project will support patient-reported data standardization by submitting for inclusion into LOINC patient-reported data elements.
- ***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

The team compiled 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. The report included recommendations for researchers on how to merge PRH and EHR data, proposed an evaluation to investigate building knowledge around data standards, and provided guidance to inform future PRH data use for research. The team has developed a Patient-Reported Data Assessment tool on the PopMedNet™ to enable investigators to compare PRH data and EHR information using a menu-driven query tool. When tested, the tool was found to support the conduct of pragmatic clinical trials, yet, additional research is needed to understand investigators' awareness and uptake of the tool. A drafted technical report and use documentation have been developed to support utilization of the tool.

In collaboration with other sub-awardees, the team identified 68 patient-reported data elements for submission to LOINC, and published these on the PCORnet website. The data elements were also included in the REDCap shared library to support implementation of 50 patient-reported data elements in future studies. The patient-reported data elements were included in the June 2018 release of LOINC. In terms of dissemination products, the team has published a joint white paper on available resources for best practices, key challenges, information gaps, and future research needs in the use of patient-reported health data in pragmatic studies. Also, a report highlighting recommended approaches to resolving inconsistences between patient-reported and EHR data has been completed and submitted for publication to the *American Heart Journal*.

Publications and Other Publicly Available Resources

- The team produced a Patient Data Assessment Tool using PopMedNet that allows investigators to compare PRH and EHR information using a menu-driven query tool, and is available on GitHub here: https://github.com/PopMedNet-Team/ADAPTABLE

- The user documentation is published here: https://dcricollab.dcri.duke.edu/sites/NIHKR/KR/ADAPTABLE%20Supplement%20MDQ%20Documentation_v1.0.pdf. Both can be accessed here: https://github.com/PopMedNet-Team/ADAPTABLE
- A report summarizing the development of the menu-driven query tool is available here: https://dcricollab.dcri.duke.edu/sites/NIHKR/KR/ADAPTABLE%20Supplement%20MDQ%20Testing%20Summary_v1.0.pdf

- The project team jointly authored a report of the ADAPTABLE Supplement Roundtable Meeting (2017) to explore recommendations for best practices, key challenges, information gaps, and future research needs for promoting best practices in the use of PRH data in pragmatic studies (2018). The summary report can be found here: https://dcricollab.dcri.duke.edu/sites/NIHKR/KR/ADAPTABLE%20ROUNDTABLE%20SUMMARY_1_22.pdf
  - A consensus statement emerging from this work was published in the *Journal of American Medical Informatics Association,* and is available ahead of print here: rhttps://dcricollab.dcri.duke.edu/sites/NIHKR/KR/ADAPTABLE%20ROUNDTABLE%20SUMMARY_1_22.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 to 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/
- LOINC Patient-Reported Data Elements were published in the NIH/NLM CDE repository and REDCap shared library to support implementation of 50 patient-reported data elements in future studies and can be found here: https://cde.nlm.nih.gov/cde/search?selectedOrg=External%20Forms&classification=ADAPTABLE
- The addition of 50 ADAPTABLE patient-reported data elements were included in the June 2018 release of LOINC and can be found here: https://r.details.loinc.org/LOINC/89070-7.html

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 NIH/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.

X. Office of the Assistant Secretary for Planning and Evaluation (ASPE)

ASPE is administering two active OS-PCORTF-funded projects including one cross-agency-funded project described later in Section XII (Exhibit 11).

Exhibit 11. ASPE Active Projects

ASPE-Funded Projects
Linking State Medicaid and Child Welfare Data for Outcomes Research on Treatment for Opioid Use Disorder

Linking State Medicaid and Child Welfare Data for Outcomes Research on Treatment for Opioid Use Disorder

Period of Performance
6/1/19 – 12/31/22

Federal Point of Contact
Robin Ghertner

Research has demonstrated that parental substance use has negative impacts on a child's health and other outcomes later in life, including child maltreatment,[51] medical, and behavioral health issues[52,53]. SUDs affect parents' ability to effectively carry out parental responsibilities, and as a result is a leading determinant of child maltreatment and foster care placement.[54] Child welfare-involved parents with SUDs are an especially vulnerable population facing additional barriers to receiving SUD treatment that are important to understand in order to provide successful treatment and sustain recovery. Some of these barriers include: 1) ability to pay for treatment; 2) adherence to treatment and the inability of courts to monitor adherence effectively; and 3) inadequate recovery supports for individuals receiving treatments and their families. A family's involvement in the child welfare system can be an opportunity to get connected to integrated, evidenced-based treatment and services to support their path to recovery.

Understanding the correlation between treatments, additional supports provided by the child welfare system, and a return to successful daily functioning and family stability, is central to assessing positive patient outcomes for parents receiving treatment. State child welfare agencies generally track outcomes of children involved in their systems, including permanency outcomes (such as reunification with family or adoption/guardianship), and regularly monitor outcomes as youth transition to adulthood. However, health outcomes for child-parent dyads involved with the child welfare system are rarely tracked regularly and consistently, and this impedes the ability to understand how a parent's SUD affects child-welfare involved youth.

End-Users of the Data Include

- Researchers in public health, substance use, and child welfare fields
- State Medicaid, child welfare, and SUD treatment agencies
- National and state health policy groups, government agencies, including HHS, as well as nongovernmental entities
- Child welfare advocacy groups, agencies, and legislators

Linking Medicaid and child welfare case records for children in the foster care system is required for new Comprehensive Child Welfare Information Systems (CCWIS)—the case management information systems for child welfare programs. CCWIS are currently

under development in most states, as a result of federal regulations published in 2016; however, states do not typically link parent Medicaid records to the CCWIS for research. Little is known about use of Medicaid for treatment of SUD or mental health within the population of parents who have children in the child welfare system (this can mean a child protective services investigation for maltreatment or placement in foster care), or how that treatment may be associated with child welfare outcomes (including repeated maltreatment, family reunification, and later health and wellbeing outcomes). There is also a gap in knowledge around how services provided by child welfare agencies align with those paid for by Medicaid. Few if any states have the capacity to monitor SUD treatment outcomes of parents with children in child welfare systems, as adequate data on parents are rarely collected.

Two major barriers to developing the evidence are: 1) a lack of comprehensive data relating SUD diagnoses and treatment to parents involved with the child welfare systems and, 2) a lack of funding to implement SUD treatment programs in child welfare systems. The latter problem is being addressed by recent federal legislation, the Family First Prevention Services Act (FFPSA), signed into law in February 2018. This project will address the former challenge, and subsequently aid in implementing FFPSA.

Project Purpose and Goals

The purposes of this project is to enhance data infrastructure and increase data availability for research on parents that have children in the child welfare system and are in need of treatment for OUD, other SUDs, or behavioral health issues.

The overall objectives of the project are to:

- Develop data sets that link records from State Medicaid and from child welfare systems in two to four states. These data sets will contain linked patient-level data including Medicaid enrollment, patient diagnoses, services, and claims, along with child welfare outcomes.
- Prepare a single, harmonized research use data set and develop a process for external researchers to access the data.
- Develop an approach for states to have an ongoing link between Medicaid and child welfare data systems, including integrating Medicaid eligibility, enrollment, and claims data with the new CCWIS. In addition, create a roadmap for other states to follow that documents the process as well as lends insight into lessons learned, challenges and successes.
- Design, conduct, and encourage analyses on the research use linked data set.

Contributions to PCOR Data Infrastructure Functionalities

- ***Use of Enhanced Publicly Funded Data Systems for Research:*** This project will use data from the states' Medicaid records and child welfare case records in order to study the negative impact of SUD among parents on their child's health.
- ***Linking of Clinical and Other Data for Research:*** This project will link records from State Medicaid and child welfare systems in order to develop a single data set. Researchers will leverage this data set to analyze and identify parents who have children in the child welfare system who may be experiencing SUD and are in need of medical treatment or recovery services and to support future effectiveness research on new interventions.

Accomplishments

Since work on the project began in June of 2019, the project team has made notable progress on several tasks.

- Published a solicitation for a contractor to perform the data linkages.
- Finalized a short list of states to participate in the project based on selection criteria that included: 1) state interest and buy-in from leadership; 2) existing infrastructure and capacity; 3) policies around data sharing; and 4) research partners.
- Convened the first meeting of the TEP charged with advising on the key project activities and best practices related to linking and harmonizing Medicaid and child welfare data.
- Engaged with key stakeholders to generate interest, receive guidance, and recruit outside entities that will help the sustainability of the project

Coordination with Other Federal Agencies

This is a collaborative project between ASPE's Office of Human Services and Policy and the ACF Office of Planning Research and Evaluation.

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

ONC is administering five active OS-PCORTF-funded projects including four cross-agency-funded projects described later in Section XII (Exhibit 12).

Exhibit 12. ONC Active Projects

ONC-Funded Projects
A Synthetic Health Data Generation Engine to Accelerate Patient-Centered Outcomes Research

* Denotes a cross-agency-funded project that involves more than one federal agency; these projects are described in

A Synthetic Health Data Generation Engine to Accelerate Patient-Centered Outcomes Research

Period of Performance
4/1/19 – 4/30/22

Federal Point of Contact
Stephanie Garcia

High-quality health care-related data are often difficult to access because of cost, patient privacy, or other legal and intellectual property restrictions. To protect patient privacy, researchers and developers often depend on anonymized data to test theories, data models, algorithms, or prototype innovations. However, the risk of re-identification of anonymized data is high and has been impossible to completely eliminate, especially with rare conditions. Further, due to a variety of interoperability issues, it is often difficult to bring data together from different resources for the purpose of robustly testing analysis models, algorithms, or assisting in the development of software applications. Synthetic data can be used to initiate, refine, or test innovative research approaches more quickly. This project proposes to address the

need for research-quality synthetic data by increasing the amount and type of realistic, synthetic data that the Synthea™ software program can generate. Synthea™ is an open-source, free, and publicly available software program that uses standards such as FHIR to create high-quality, clinically realistic, synthetic patient health records in large volumes.

Project Purpose and Goals

A synthetic data engine is a potentially important piece of the greater patient-centered outcomes research data infrastructure because it provides researchers with a low-risk, readily available synthetic data source complementing their use of real clinical data and enhancing their ability to conduct rigorous analyses and generate relevant findings that can inform health care and treatment decisions.

The project will address the following objectives:

- Enhance Synthea by developing five to seven priority use cases for new or updated data generation modules in the following topics: opioids, pediatrics, and complex care. Administer a prize competition (a "challenge") to encourage researchers and developers to validate the realism of the generated synthetic health records.
- Support awareness and use of Synthea, including its updated modules, module builder, and the generated synthetic data through various dissemination mechanisms.

Synthea™ is a synthetic patient data generator that models the medical history of patients. The output are high-quality synthetic, realistic but not real patient data and associated health records in HL7 FHIR and HL7 C-CDA formats. Currently Synthea™ data models include:

- *Primary care, emergency department, and symptom-driven encounters.*
- *Conditions, allergies, medications, vaccinations, observations/vitals, lab, procedures, CarePlans*
- *Birth-to-death lifecycle*
- *Configuration-based statistics and demographics*

Source: https://synthea.mitre.org/about

Contributions to PCOR Data Infrastructure Functionalities

- ***Use of Enhanced Publicly Funded Systems for Research:*** Synthea exclusively uses publicly available health statistics, combining them with clinical guidelines and medical coding dictionaries to increase the number and diversity of clinically relevant synthetic patient health records. The three identified topics areas address key HHS policy priorities in which access to data for outcomes research can be difficult, challenging researchers' ability to study new treatment options and interventions.

Accomplishments

Since the project launch in the spring of 2019, the project team has made progress toward several activities in support of the project's objectives:

- ONC issued the contract award for performance of the work and held the project kick-off meeting on September 24, 2019.The inaugural meeting of the TEP took place during the ONC Annual Meeting. The TEP is comprised of a group of diverse stakeholders representing viewpoints external to the federal government to provide advice on the use case selection.

Publications and Other Publicly Available Resources

- A public-facing website was created on ONC's HealthIT.gov. The website will be updated throughout the course of the project and is accessible here: https://www.healthit.gov/topic/research-evaluation/synthetic-health-data-generation-accelerate-patient-centered-outcomes

XII. Cross-Agency Funded Projects

There are 11 cross-agency funded active OS-PCORTF-funded projects (Exhibit 13).

Exhibit 13. 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
Data Capacity for Patient-Centered Outcomes Research through Creation of an Electronic Care Plan for People with Multiple Chronic Conditions
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 and 4
Enhancing Patient-Centered Outcomes Research (PCOR): Creating a National Small-Area Social Determinants of Health (SDOH) Data Platform
Harmonization of Various Common Data Models and Open Standards for Evidence Generation
Technologies for Donating Medicare Beneficiary Claims Data to Research Studies
Training Data for Machine Learning to Enhance Patient-Centered Outcomes Research (PCOR) Data Infrastructure
Validating and Expanding Claims-based Algorithms of Frailty and Functional Disability for Value-based Care and Payment

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

AHRQ Period of Performance
1/15/17 – 4/15/19

AHRQ Federal Point of Contact
Janey Hsiao

ONC Period of Performance
1/15/17 – 4/15/19

ONC Federal Point of Contact
Stephanie Garcia

The patient perspective is essential to health care decision-making and health management, and it can inform evidence generation regarding treatments and outcomes and allow investigators to better answer questions that are important to patients and their families. 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."[55] 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 conducted a joint project, *Advancing the Collection and Use of PROs through Health IT*.

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).[56] Providers wishing to integrate PROs directly into EHR systems may need to work with EHR vendors or IT teams to customize and present the PRO questions to patients. EHRs may not be optimized to leverage the PRO data for clinical, research, or administrative needs. Patients may find that logging into an EHR portal at home to record PROs is 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 aimed to widen the availability of PRO data for patient-centered outcomes research by increasing PRO standardization and promoting the development of PRO collection apps.

Project Purpose and Goals

The purpose of this project was 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 were to:

- Refine and/or harmonize 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.
- Support the development of user-friendly PRO-collection applications that utilize the health IT standards.
- Pilot-test these applications in a health system that supports both health care delivery and research.

Contributions to PCOR Data Infrastructure Functionalities

- ***Collection of Patient-Provided Information:*** This project involved using standards that simplify the PRO sharing process between patients, providers, and researchers. An existing PRO application was modified to incorporate the standards. Through a developer competition, new PRO data collection applications were developed using the standards. Both the modified application and the competition winning application were then pilot-tested in a health care system. The current state of PRO use was also explored. Improvements in the methods and standardization of PRO capture can improve the interoperable sharing of this data.
- ***Use of Clinical Data for Research:*** This project developed the implementation specifications needed to help standardize PRO data, thereby improving the usability of the PRO data for research.

Accomplishments

The goals of the project were to develop technical infrastructure to better collect and use PRO data for patient-centered outcomes research.

The first aim of the project focused on refining and harmonizing health IT standards and implementation specifications to enhance the sharing of PRO data through APIs and other relevant health IT products for research. ONC conducted a series of key informant interviews and prepared an internal findings report on: 1) PROs used for functional status assessments; 2) standards for electronic capture and use of PROs; and 3) research programs and networks that use functional status to study patient outcomes.

In early 2018, ONC prepared a technical approach report that delineated the technical specifications needed to standardize the collection of PROs and to easily integrate PRO data into an EHR. The report detailed the technical approach for pilot sites, which was used to pilot-test the PRO collection app. ONC also developed and pilot-tested an HL7 FHIR-based implementation specification in two pilot organizations: REACHnet and pSCANNER. As a result of the pilot-testing, ONC developed the PRO FHIR IG, which was made publicly available on HL7's website. The IG has gone through several rounds of ballot comments through HL7, and ONC continues to reconcile ballot comments.

Step Up App Challenge Winners

- *PRISM was announced as the Step Up App Challenge grand prize winner. PRISM is an application that enhances the quality of clinical discussion between health care providers and patients by allowing for continued patient engagement outside of the clinical setting.*
- *PRISM also received the second place award at 2019 AMIA Annual Symposium FHIR App Showcase.*
- *PEER Technologies' app "Back Pain Tracker" was announced as the second-place winner and Cliexa-EASE mobile app as the third-place winner.*

The second aim of the project focused on supporting the development of user-friendly, PRO-collection applications that utilize health IT standards and implementation specifications identified by ONC. New PRO applications were developed through a prize competition supported by AHRQ, the "Step Up App Challenge," which launched in August 2018. This three-phase competition encouraged participants to develop and present user-friendly apps able to collect standardized PRO data in primary and specialty care settings. PRISM, the PROMIS Reporting and Insight System from Minnesota, was announced as the Step Up Challenge Grand Prize Winner. AHRQ also worked with MedStar to modify an existing app (OBERD app) by incorporating ONC's PRO FHIR IG to demonstrate that developers could utilize the PRO FHIR IG on existing apps.

The final aim of the project focused on implementing private/public partnerships for pilot-testing these PRO-collection apps in a health system. AHRQ pilot-tested two apps: PRISM and OBERD. AHRQ contracted with MedStar Health to recruit practices and prepare the health IT infrastructure for pilot-testing. The team identified a number of lessons learned: 1) EHR vendors are moving in the direction of adopting the FHIR standards, as an efficient and secure way to integrate PRO data; 2) understanding the needs of the industry, and the needs of individual sites that plan to collect PRO data, is essential for securing organizational commitment; 3) leveraging existing resources that organizations have invested in and using PRO instruments that align with organizational priorities will help mitigate compatibility issues. The project team plans to publish a report discussing the approach to pilot testing and lessons learned, along with the source code for the PRISM app.

The pilot tests provided early evidence of successful uptake and use of apps—using publicly available and pilot-tested FHIR IG. The IG will continue to be tested. It will increase the likelihood that developers continue to use it to capture different domains of PROs.

Publications and Other Publicly Available Resources

- In September 2019, AHRQ released the report on the App Challenge that details the winners and how the app can be applied in PCOR research: https://healthit.ahrq.gov/sites/default/files/docs/citation/ahrq-step-up-app-challenge-summary-2019.pdf
- The PRO FHIR IG:
  - Last balloted version: http://hl7.org/fhir/us/patient-reported-outcomes/2019May/
  - Continuous-build version (updated as comments are reconciled): http://build.fhir.org/ig/HL7/patient-reported-outcomes/
- ONC PRO Project and IG Webinar Presentation: https://www.healthit.gov/sites/default/files/page/2019-09/ONC_PRO_Webinar_Slides_FINAL.PDF
- Challenge competition winning app—PRISM: PROMIS Reporting and Insight System from Minnesota: https://apps.apple.com/us/app/prism-for-minnesota/id1454714605 and https://play.google.com/store/apps/details?id=com.perkmotivation.PrismForMedstar

Coordination with Other Federal Agencies

NIH, FDA, CMS, and HRSA provided advice for this project. NIH advised on work related to the PROMIS,[57] whose PRO measures were used as a use case in the project. . CMS and FDA will provide advice regarding how standards would be applied to their regulatory needs. HRSA provided advice regarding the applicability of deliverables for the safety-net population.

Assessing and Predicting Medical Needs in a Disaster

AHRQ Period of Performance
6/15/18 – 9/14/20

ASPR Period of Performance
6/15/18 – 9/14/20

AHRQ Federal Point of Contact
Pam Owens

ASPR Federal Point of Contact
Leremy Colf

HHS leads the U.S. 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. ASPR focuses specifically on preventing, preparing for, and responding to the adverse health effects of public health emergencies and disasters. In the midst of Hurricane Harvey, ASPR and 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.

Examples of Real-World Decision-Making Informed by Platform Analysis

Researchers at ASPR conducted analysis to contribute to real-world decision-making:

- *In response to Hurricane Dorian, ASPR conducted a quick-turnaround analysis of historical hurricanes that traveled up the Eastern Seaboard and caused substantial flooding.*
- *In response to the Kincade fires, ASPR presented a smoke inhalation analysis to California public health partners, emergency management, researchers, and health care professionals, which resulted in modified resource allocation.*

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 patient-centered outcomes research related to disaster response and recovery operations. The initial platform will consist of public-facing statistical query pathways (e.g., HCUPnet and/or FastStats) that provide access to statistical tables, maps, and graphics on patient health outcomes, and a restricted access set of analytic files derived from 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 comparative effectiveness research questions, such as which emergency management interventions at the county level were successful. 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 outcomes research 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. 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.

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

- ***Linking Clinical and Other Data for Research:*** AHRQ will focus on creating an analytic platform that enables integration of emergency department and inpatient outcome data captured during disasters. Once linked, ASPR will test the capacity of the platform to prepare emergency respondents and support patient-centered outcomes research.
- ***Use of Enhanced Publicly Funded Data Systems for Research*****:** The project will utilize HCUPnet and HCUP Fast Stats, which are tools that access the national and state-level health statistics system, particularly emergency department and ambulatory settings data.

AHRQ Accomplishments

Since the project began, the AHRQ team has made progress toward developing a national-level health care encounters platform.

- The team has revised the HCUP Hurricane Data Resource analytic files, a series of hurricane-specific HCUP and supplemental external data files for HHS researchers.
- AHRQ continues to acquire new quarterly inpatient and outpatient data. The project has completed processing four new quarterly outpatient data files from three states.
- The team has developed and tested hurricane impact estimate programs and, separately, conducted an analysis of historical hurricanes that resulted in flooding in the Eastern Seaboard.
- User documentation for the HCUP Hurricane Data Resource analysis from the California wildfires has been developed and is currently under internal review.
- The team is working to finalize an HCUP Statistical Brief highlighting emergency department visits after a hurricane with respect to injuries.

Publications and Other Publicly Available Resources

- The team also published the HCUP Fast Stats Hurricane Path, detailing inpatient and emergency department usage before and after 11 hurricane events. Announcement and additional details can be visited here: https://hcup-us.ahrq.gov/news/announcements/faststats2019Dec.jsp

ASPR Accomplishments

The ASPR team have completed significant activities to further engage with stakeholders and inform the research community.

- After conducting a feasibility study, National Disaster Medical System (NDMS) withdrew from the project due to insufficient Health Information Repository data quality. With input from the TEP, the team began utilizing new data sources, such as the Red Cross National Shelter System data, to capture stays in shelters, medical needs at those shelters, and facilities operating during disasters.

Publications and Other Publicly Available Resources

- After consulting TEP members, the team finalized a report detailing the new functional design layer of ASPR's platform, GeoHEALTH (https://geohealth.hhs.gov/arcgis/home/)

Data Capacity for Patient-Centered Outcomes Research through Creation of an Electronic Care Plan for People with Multiple Chronic Conditions

AHRQ Period of Performance
4/15/19 – 5/15/23

AHRQ Federal Points of Contact
Arlene Bierman
Janey Hsiao
Steve Bernstein

NIH/NIDDK Period of Performance
4/8/19 – 9/30/23

NIH/NIDDK Federal Points of Contact
Jenna Norton
Kevin Abbott

Interest in pragmatic research among researchers and key funding organizations is growing due to increased recognition that findings from traditional randomized trials may not apply to real-world situations. Pragmatic trials represent a cost-effective and efficient research approach, in which health IT systems such as EHRs facilitate use of point-of-care data to enhance our understanding of the effectiveness of health interventions in real-world practice. However, lack of interoperability and exchange of data across EHRs creates barriers to pragmatic patient-centered outcomes research, as essential data on patient-centered outcomes, as well as health risk and promoting factors, are frequently missing, inconsistent, or difficult to compile across settings and conditions.

Data aggregation is particularly important and challenging for people with multiple chronic conditions, who undergo frequent care transitions (e.g., hospital to home, primary care to specialist, etc.). These individuals have complex health needs handled by diverse providers, across multiple settings of care. As a result, their care is often fragmented, poorly coordinated, and inefficient. These challenges will increasingly strain the U.S. health system with the aging of the U.S. population. Research is needed to better understand optimal care for these complex patients, yet comprehensive data enabling the study of factors influencing outcomes across multiple conditions and disease states in real-world settings are largely unavailable.

Project Purpose and Goals

This project will build data capacity to conduct pragmatic patient-centered outcomes research by developing an interoperable electronic care (eCare) plan to facilitate aggregation and sharing of critical patient-centered data across home-, community-, clinic-, and research-based settings by extracting EHR data and exchanging that data across settings. The eCare plan will be an overarching, longitudinal

blueprint of the prioritized health concerns, goals, interventions, and health status of an individual patient across care settings and health care team members. The pilot eCare plan tool developed for this project will be designed for use with patients who have chronic kidney disease (CKD), cardiovascular disease (CVD), diabetes, chronic pain, and/or OUD.

The overall objectives of the project are to:

- Expand an existing data element and standards set focused on CKD to the following comorbid health conditions: CVD, chronic pain, OUD, and diabetes.
- Develop an open-source eCare plan application for people with MCCs and pilot it in patient populations with CKD.
- Establish an eCare plan repository and application development collaborative to support the project's development, testing, piloting, and implementation efforts and provide an open source repository available to enable search, access, sharing, and exchange of eCare plans.
- Disseminate all project products through free, open source channels (e.g., federal government websites, open-source software exchanges such as GitHub).

Contributions to the PCOR Data Infrastructure Functionalities

- ***Standardized Collection of Standardized Clinical Data:*** The project will build on existing eCare plan data elements and standards set for CKD that aligns with CMS and ONC-certified EHR technology requirements as part of the Promoting Interoperability programs, adding novel data elements and standards from common clinical terminologies for diabetes, CVD, pain and OUD.. The project team is also leveraging clinical information models, FHIR profiles, and FHIR APIs to facilitate data access and exchange.

AHRQ Accomplishments

The project team continues to make significant progress in an effort to improve coordinated care for chronically ill patients.

- AHRQ has awarded a contract to pilot and evaluate the eCare plan application.
- AHRQ completed an environmental scan to examine currently available eCare plans across different diseases and different sectors.

Publications and Other Publicly Available Resources

- A collaboration website was developed to allow for management of tasks, sharing of documents, and group discussions. Visit https://ecareplan.ahrq.gov/.

NIH/NIDDK Accomplishments

Since the project began in April of 2019, the project team has made notable progress on several tasks.

- In October 2019, NIH/NIDDK and AHRQ hosted an in-person TEP kick-off meeting to begin discussing the development of the expanded data element and standard set. Since then, TEP subgroups have been meeting on a monthly basis.
- The NIH/NIDDK project team has finalized the scope of work and materials for the development of the clinical information models, eCare Plan software application, and corresponding IG. The

contract was awarded to perform this work, and NIH/NIDDK held its kick-off meeting with the contractor on October 2019.

Coordination with Other Federal Agencies

The project has convened a Federal Stakeholder panel including diverse Federal agencies and organizations with interest in the eCare plan to acquire additional input and guidance on the project. Members of this panel include: Agency for Community Living, CMS, HRSA, Indian Health Service, numerous other NIH Institutes and Centers, ONC, PCORI, and the Veteran's Health Administration.

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

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

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.[58] 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 U.S., 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 nonstandard systems that contain large amounts of unstructured data. 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) to report adverse events.[59] 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 codes that provide key information like exact time 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 publicly available to researchers to help them convert unstructured clinical information into structured and standardized coded data. It will be made available on the Public Health Community Platform (PHCP), a cooperative platform for sharing interoperable health IT that addresses public health priority areas, such as improving

population health outcomes and health equity.[60] The NLP web services environment will contain NLP architectures and tools process spontaneous report narratives, help researchers extract clinical and temporal information from the text, provide 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 between CDC and FDA, 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 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 data sets 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. 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

Since work on the project began in June 2016, CDC and FDA project teams have completed multiple deliverables in support of the NLP web service or workbench called CLEW. The project's work has occurred in three phases involving: 1) an "as-is" environmental scan and literature review, identifying existing NLP architectural methods and tools; 2) design of the CLEW platform environment; and 3) pilot projects to test the use of the CLEW, focusing on cancer pathology and safety surveillance data.

In the first phase, the FDA project team completed the environmental scan, which included a literature review and multi-channel review that identified 54 existing open-source tools that are potentially useful in building pipelines for clinical NLP domains. Tools of potential interest to the project were then selected on

the basis of availability and relevance. 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."[61]

After the completion of the environmental scan, the project focused on the design of a CLEW platform environment to provide open-source NLP and machine learning tools to develop, experiment with, and refine clinical NLP models. The FDA project team selected safety surveillance use cases to include in the NLP workbench pilot, identified existing solutions to support these use cases, and determined the required software development. The FDA built the NLP workbench prototype and the CDC project team designed the CLEW platform as a user-friendly interface for NLP engineers to create pipelines from an expansive, preexisting catalogue of NLP tools and to transform data between multiple frameworks and tools.[62]

The CLEW was tested via two pilot projects: one led by CDC and the other led by the FDA. The CDC chose to focus on cancer pathology for the pilot project domain. Information from five key data elements (histology, primary site, behavior, laterality, and grade) were extracted from data sets provided by four national laboratories. In addition, the CDC team created a service to code the data for these five elements and mapped them to the national ICD oncology standard coding system. The project team saw highly promising results in this pilot project domain; data-matching results from the four models created by the project team were comparatively close in specificity and sensitivity. As a result of the pilot project, the state-based cancer registry system, eMaRC Plus, was expanded to use cancer pathology NLP machine learning models that were developed in this project and hosted on CLEW.

The FDA pilot project domain was the Safety Surveillance program, which tracks adverse events through post-market report-monitoring for medical products. The project team found that its machine learning model increased the rule-based system's sensitivity and specificity accuracy. Sensitivity rates were higher with the hybrid model (a rule-based model with a machine learning model built on top of it) than with either model alone. As a result of this pilot project, the safety surveillance NLP application was incorporated into CLEW for other NLP experts to use. The FDA team also created an annotated data set for training NLP models and uploaded the solution to GitHub for broader use.[63]

The project team learned several lessons while developing the NLP Web Service. For example, effective NLP model training requires a broad training set from all sources. To effectively utilize analytic methods and data science, department-level infrastructure is needed to enhance analytic capabilities and maximize the usefulness of data. The project team agrees that a well-trained model to address a specific issue will have much better sensitivity and specificity than a generalized model. More work is needed to fully implement cTAKES components onto CLEW/LAPPS Grid Platform, and a cloud environment is necessary for cross-agency development and joint sharing. The next step for the NLP field is the creation of a primary home for NLP Web Services to encourage the collaboration and sharing of NLP machine learning solutions across federal agencies, public health entities, and PCORI researchers.

Publications and Other Publicly Available Resources

- CLEW, the cloud-based, open-source NLP Workbench Web Service is available at: https://www.cdc.gov/cancer/npcr/informatics/nlp-workbench/index.htm
- All code and documentation have been uploaded to the CDC public GitHub at https://github.com/CDCgov/NLPWorkbench and the FDA public GitHub at https://github.com/FDA/

- CDC and FDA developed a CLEW User Guidance document which explains how to install and use the CLEW, as well as the products developed in the CDC and FDA pilots. The report will be made available on ASPE's OS-PCORTF webpage here: https://aspe.hhs.gov/system/files/pdf/259016/NLP-CLEW-UserGuidanceDocument-508.pdf.
- CDC and FDA developed a NLP CLEW Final Report which describes in detail the project's goals, major accomplishments including a description of the environmental scan, the CLEW platform design, pilot findings, lessons learned, and dissemination deliverables. The report is available on ASPE's OS-PCORTF webpage here: https://aspe.hhs.gov/system/files/pdf/259016/NLP-CLEW-FinalReport-508.pdf.
- CDC and FDA developed the NLP CLEW Workbench Web Service Technical Report, which presents a detailed technical description of the core NLP approach of the prototype version of the Workbench and two pilot applications developed using the Workbench. The report isavailable on ASPE's OS-PCORTF webpage here: https://aspe.hhs.gov/system/files/pdf/259016/NLP-Workbench-Web-Services-Technical-Report-508.pdf
- The project teams compiled a Lessons Learned Report. In this report, the teams summarize the key observations, and findings that inform future tools, systems development, and testing, and NLP and machine learning pipeline and model development. The report is available here: https://aspe.hhs.gov/system/files/pdf/259016/NLP-CLEW-LessonsLearned-508.pdf
- The results of the environmental scan and literature review, "Natural Language Processing Systems for Capturing and Standardizing Unstructured Clinical Information: A Systematic Review" were published in the September 2017 issue of the *Journal of Biomedical Informatics*.[64] https://www.ncbi.nlm.nih.gov/pubmed/28729030
- The FDA project team initiated and completed the generation of an annotated corpus to support training and development efforts of language models. The complete clinical and temporal annotations for the 1,000 Vaccine Adverse Event Reporting System (VAERS) reports are publicly available to the research community on GitHub here: http://github.com/fda/VAERS-Annotations.
- The team also published the final corpus in a paper describing the methodology used to create it so that researchers can assess the utility of the corpus to their own work. The paper titled "Generation of an annotated reference standard for vaccine adverse event reports" was published in *Vaccine* in 2019. https://www.ncbi.nlm.nih.gov/pubmed/29880244

Coordination with Other Federal Agencies

The FDA and CDC are continuing their collaboration on the design and architecture of the NLP Workbench and 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 (NSF) 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 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.

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

FDA Period of Performance
2/15/17 – 5/15/19

FDA Federal Points of Contact
Danica Marinac-Dabic
Benjamin Eloff

NIH/NLM Period of Performance
2/15/17 – 5/15/19

NIH/NLM Federal Points of Contact
Lisa Lang
Robin Taylor

ONC Period of Performance
2/15/17 – 5/15/19

ONC Federal Points of Contact
Kate Brett
Steven Schwartz

The Developing a Strategically Coordinated Registry Network to Support Research on Women's Health Technologies (WHT-CRN) 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 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 focused on three clinical areas: sterilization and long-acting contraception, uterine fibroid treatment, and 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 relevent 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.

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.*

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 e.g., the NIH CDE Repository, the NIH/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 Structured Data Capture (SDC) initiative) to extract clinical data from EHRs into the CRN (FDA/ONC collaboration).
- Create a harmonized, interoperable platform and reusable tools (e.g., 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).

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

This project created a CRN to address challenges related to evidence generation for women's health technologies. This was achieved through three phases.

The first phase focused on establishing a partnership structure composed of four clinical working groups. The four clinical groups were organized around four clinical topic areas including sterilization/long-acting reversible contraceptives, pelvic organ prolapse, stress urinary incontinence, and uterine fibroids. Additionally, a working group of informatics leaders representing the FDA, NIH, and ONC was formed to complete the process of translating the clinical data sets, definitions, and value sets into standard vocabularies to be used in data capture and exchange. The Informatics Working Group was successful in standardizing and harmonizing the clinical data elements identified by each clinical working group. Four

data sets, including data dictionaries, were developed for each of the four clinical areas. The Informatics Working Group also developed an Excel template of registry domain data elements (definitions, permissible values, and context) and a claims library. One of the key outputs of this phase was the Partnership Framework Report. This report provides lessons learned for building a community of practice, promoting cross-clinical collaboration, and building trust with stakeholders.

The second phase of the project focused on developing data infrastructure. Each clinical working group used a six-month Delphi process to arrive at a consensus on a core minimum data set for each of the four conditions, and ultimately successfully identified a set of CDEs. The CDEs were harmonized and standardized, and the clinical working groups were continuously engaged to make improvements. The final recommendation for the harmonized set of data elements was presented to the IG group to be included in the HL7 FHIR IG submission. Further, the Informatics Working Group 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 included a landscape analysis that examined the current state of women's health CRNs and reviewed how the registries are able to capture, utilize, and share data. This helped to inform the feasibility of pilot studies. Several pilot studies and demonstration projects were conducted, as described in the WHT final report. One such pilot site used existing registries to improve the recommendations in the WHT-CRN FHIR IG for use in a test or production environment. Another pilot integrated the uterine fibroids core minimum data set into the Ultra registry (UCSF), to be used for research, education, and quality purposes.

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 data set. 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 U.S. Realm Steering Committee (USR-SC) in June 2018. The team developed and submitted HL7 FHIR® IG: 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. Enhancements to the NIH CDE Repository capabilities in support of needs for the HL7 IG were developed, tested, and implemented. 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. In August 2018, the team conducted a detailed walkthrough of the CRN IG 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 final phase involved sustainability planning to ensure the value and longevity of this work, as well as considerations for using the established infrastructure for research. This includes pilot-testing of two new modules.The first is the Stress Urinary Incontinence and Pelvic Organ Prolapse module, which is included in the American Urogynecologic Society AQUIRE registry. AQUIRE is an open, national urogynecology-focused registry that is designed to measure and report health care quality and patient outcomes. A patient-facing SMART on FHIR mobile application was designed by ONC to collect PROs and is currently undergoing pilot testing under the Stress Urinary Incontinence Module of the AQUIRE registry. The second module will pilot-test the core minimum data elements for Pelvic Organ Prolapse and support the implementation and refinement of specifications in the WHT-CRN IG in a test environment.

The project team is preparing several reports to disseminate lessons learned and best practices from this project. The team has developed a Partnership Report that will include a comprehensive sustainability

plan for the work. The team also prepared an Infrastructure Report, which details population demographics, disease presentation, device exposure, follow-up duration, and relevant clinical outcomes for each CRN through metadata and descriptive statistics. Finally, the team developed a Data and Tools Report that describes the overall process of creating the harmonized set of data elements and the IG as well as a Structured Framework Document, which is a framework for data-sharing and interoperability among participating data sources and clinical sites participating in pilot-testing.

The project team reports that the project has already proven to be useful in to end-users. Two manufacturers of devices of stress urinary incontinence are in the process of leveraging the AQUIRE-SUI family of registries for evidence generation. Additionally, the academic partners involved in this project have begun exploring the use of WHT-CRN for future proposals to NIH and PCORI, which demonstrates the potential for sustaining continued work on this project beyond the OS-PCORTF funding period.

Publications and Other Publicly Available Resources

- The most recent version of the HL7 FHIR® CRN Implementation Guide is publicly available here: http://hl7.org/fhir/us/womens-health-registries/2019May/
- 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),[65] 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.
- The Stress Urinary Incontinence (SUI) Surgery Module, in conjunction with the American Urogynecologic Society AQUIRE registry, can be found here: https://www.augs.org/clinical-practice/aquire-sui-surgery-module/.
- The WHT-CRN common core data set, inclusive of the four core clinical data sets, can be found here, as part of the larger NIH CDE Repository: https://cde.nlm.nih.gov/
- The initial set of minimum core harmonized data elements, also incorporated in the NIH CDE repository, is available here: https://drive.google.com/file/d/1GGRZOJcLBAW_p_czcfaSe7xAirs8KPw6/view?usp=sharing
- The form containing the initial core minimum set of harmonized data elements in the NIH CDE Repository is available here: https://cde.nlm.nih.gov/formView?tinyId=XJwYSNJ4I
- Training videos on how to use the NIH CDE Repository and create new forms and CDEs are available here: https://cde.nlm.nih.gov/videos. and on YouTube:
  - *NIH CDE Repository Overview*: https://www.youtube.com/watch?v=tBHLNhX2nK8
  - *How to Create a Common Data Element (CDE)*: https://www.youtube.com/watch?v=oJKuG4FnyRc
  - *Creating Forms in the CDE Repository*: https://www.youtube.com/watch?v=LowLJh29-4M
- Written guides on how to use advanced features of the CDE Repository are available here: https://cde.nlm.nih.gov/guides

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 ARHQ's Outcome Measure Framework. Upon the project's conclusion, NIH/NLM aims to include the data element resources within the NIH CDE Repository, promoting the uptake of the standardize CDEs in future research projects.

Enhancing Data Resources for Researching Patterns of Mortality in Patient Centered Outcomes Research: Projects 1 & 2

CDC Project 1 Period of Performance
2/1/17 – 1/31/21

CDC Project 1 Federal Point of Contact
Carol DeFrances

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

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

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. Mortality is an important outcome in PCOR and efforts to better harmonize, connect, and enrich the federal mortality data through the two projects described below will accelerate its availability and utility for PCOR. 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.

This group of projects consisted of four independently led but supportive components designed to enhance data resources for researching patterns of mortality in PCOR. Two of the projects, led by CMS (Project 3) and CDC (Project 4), concluded in FY 2018 and are described in the 2018 Portfolio Report. Two of the projects remain active: CDC (Project 1) and FDA (Project 2).

Project 1 (CDC)—Adding Cause-Specific Mortality to National Center for Health Statistics' National Hospital Care Survey by Linking to the National Death Index (NDI) and Medicare program enrollment and summary costs and utilization by Linking to the CMS Master Beneficiary Summary File (MBSF). This project will leverage data from CDC's NHCS, NDI, and CMS to create new data infrastructures to advance studies on mortality and post-acute care utilization following hospital care by linking inpatient and emergency department patient records collected in the 2014 and 2016 NHCS with death certificate records from the NDI and summary costs and utilization from the MBSF. Together these linkages will expand the capability of patient centered outcomes researchers to examine mortality outcomes and post-acute care utilization following emergency department visits and/or hospital inpatient stays for specific conditions and/or health care treatments and procedures. In addition, this project will provide the first-ever data linkage of EHR data from a national provider survey to the NDI, enable evaluations of EHR and

> *This project will provide the first-ever data linkage of EHR data from national provider survey to the NDI.*

claims data on their quality and complementarity, and create new approaches to optimize patient level 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 patient-centered outcomes research, 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 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 Medicare administrative data from the CMS MBSF. Restricted-use linked data files for the 2014 NHCS linkage to 2014/2015 NDI and 2014/2015 CMS MBSF, as well as the 2016 NHCS linkage to 2016/2017 NDI will be available for researcher use through the NCHS Research Data Center.

Specifically, the objectives of this project are to:

- Link the 2014 NHCS inpatient and emergency department 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 NHCS inpatient and emergency department claims data to the 2014 and 2015.
- Link the 2016 NHCS inpatient and emergency department 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.

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 and claims data, with NDI and CMS 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 NHCS survey data as well as administrative data sets including the NDI and CMS MBSF.

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 1 (CDC) Accomplishments

At the start of the project in February 2017, the project teams have made significant progress on three major data linkages. The linkage of the 2016 NCHS inpatient, emergency department claims, and EHR records to the 2016/2017 NDI, the 2014 NHCS inpatient and emergency department claims to the 2014/2015 NDI, and 2014 NHCS inpatient and emergency department claims to the 2014/2015 CMS MBSF data files has been completed and the linked data files and reports are currently available for research use.

Publications and Other Publicly Available Resources

- The linkage of the 2014 NHCS inpatient and emergency department claims to the 2014/2015 NDI has been completed and the linked data files are currently available for research use. More information is available on the NCHS Data Linkage website: https://www.cdc.gov/nchs/data-linkage/nhcs-ndi.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
- NCHS has published a report on The Linkage of the 2014 National Hospital Care Survey to the 2014/2015 National Death Index: Methodology Overview and Analytic Considerations: https://www.cdc.gov/nchs/data/datalinkage/NHCS14_NDI14_15_Methodology_Analytic_Consider.pdf
- The linkage of the 2014 NHCS inpatient and emergency department claims to the 2014/2015 CMS MBSF data files has been completed and the linked data files are currently available for research use. More information is available on the NCHS Data Linkage website here: https://www.cdc.gov/nchs/data-linkage/CMS-Medicare-Restricted.htm
- NCHS has published a report describing the methods used for linkage and analytic considerations for the 2014 NHCS linkage to the 2014-2015 CMS MBSF at: https://www.cdc.gov/nchs/data/datalinkage/NHCS-CMS-Medicare-Llinkage-Methods-and-Analytic-Considerations.pdf
- The linkage of the 2016 NCHS inpatient and emergency department claims and EHR records to the 2016/2017 NDI has been completed and the linked data files are currently available for research use. More information is available on the NCHS Data Linkage website https://www.cdc.gov/nchs/data-linkage/nhcs-ndi.htm
- NCHS has published a report on The Linkage of the 2016 National Hospital Care Survey to the 2016/2017 National Death Index: Methodology Overview and Analytic Considerations at:https://www.cdc.gov/nchs/data/datalinkage/NHCS16_NDI16_17_Methodology_Analytic_Consider.pdf
- Bercovitz A, Jamoom E, Lau DT. (2018). National Hospital Care Survey Demonstration Projects: Characteristics of Inpatient and Emergency Department Encounters among Patients with any

Listed Diagnosis of Alzheimer disease. National Health Statistics Reports; no 121. Hyattsville, MD: National Center for Health Statistics Reports. The report is available here: https://www.cdc.gov/nchs/data/nhsr/nhsr121-508.pdf

Project 2 (FDA) Accomplishments

Since the end of 2017, the project has made considerable progress toward developing new data linkage capabilities between commercial health plan data and the NDI+.

- The project has finalized the administrative workflow deliverable for IRB and NDI approvals, which will guide this multi-site research effort. The project utilized a central IRB protocol, which helped to minimize administrative burden. Additionally, templates and recommendations were developed throughout the NDI approval process, which will be used to support and expedite other research efforts.
- The team had completed the development and testing of the distributed linkage processes for data exchange between the health plans and the NDI.
- The project worked closely with the health plans to prepare logistics for exchanging files with NDI. Ultimately, all health plans have all submitted data files to the NDI and received matched results from the NDI. All health plans are currently testing the program package to link the matches provided by the NDI to their data.
- Efforts to link multiple study sites to the NDI are currently underway.

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

FDA Period of Performance
2/8/17 – 7/31/20

FDA Federal Point of Contact
Scott Gordon

NIH/NCATS Period of Performance
2/8/17 – 9/30/20

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 and Robin Taylor

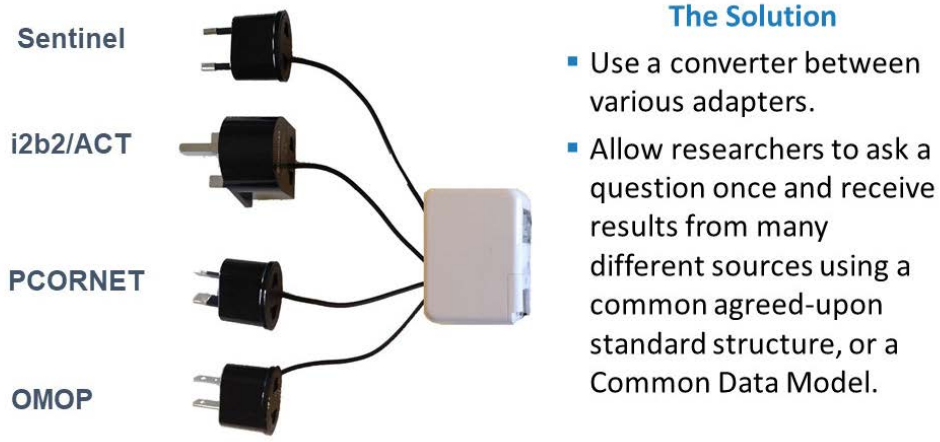
ONC Period of Performance
2/8/17 – 5/7/19

ONC Federal Point of Contact
Albert Taylor

To build a sustainable data infrastructure and promote interoperability, there is a need to map and harmonize data across various CDMs using open-source standards. This project is a collaboration among the FDA, NIH/NCI, NIH/NCATS, NIH/NLM, and ONC. The project is enabling the use of data among four major research networks—Sentinel, PCORnet, OMOP, and i2b2-ACT—to support research across a range of health issues.

By harmonizing various CDMs, the research community will have an opportunity to leverage large amounts of RWD from across the networks and will allow researchers to investigate trends related to patient demographics (e.g., elderly, pediatrics, non-US population, etc.). In turn, this will help answer research questions on the safety of different cancer treatment options, and generate RWE that supports patient and provider decision-making. By harmonizing the CDMs in these 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 reused by other networks. The harmonization of CDMs across these research networks would benefit researchers in a diverse range of patient-centered research inquiries, from basic PCOR queries, to data mining to generate hypotheses for future research, to large-scale sophisticated analysis, including randomized clinical trials.

Exhibit 14. The Four Major Data Networks and CDM Solution



Project Purpose and Goals

This project will harmonize CDMs developed by four different networks: Sentinel, PCORnet, OMOP, 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.

The FDA serves as the lead for the project, with multiple other agencies contributing. 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 project co-leads and will develop the informatics and implementation strategy. The oncology research at NIH/NCI will evaluate the proposed harmonization 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 the proposed approach.

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 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 support evidence generation on the effectiveness of oncology drugs on patient-centered outcomes, the results of which 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 harmonizes 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

Over the course of the project, numerous tools, modules, and mechanisms were developed and tested across the various participating agencies. These accomplishments created the foundation of multiple extension projects, including but not limited to: CDC using the work for enhanced surveillance; FDA leveraging the mappings for RWD submission in a regulated clinical trial; and the NIH using the tools and information to harmonize RWD for research across the 58 academic centers that make up the CTSA. These mappings are also the basis for moving research from proprietary data models to interoperable HL7 FHIR data standards, which facilitates sharing and harmonization of not only EHR data but registries and mobile devices.

Notably, along with leading an environmental scan of existing CDM artifacts, the FDA project team also developed the oncology use case for the PCORnet 3.1 and 4.0 CDMs. The NIH/NCATS team surveyed the market for an existing open source extract, transform, and load (ETL) software tool to automate the data mapping process, and prepared a report summarizing the selection process for the intermediary model. The NIH/NCATS team also created a "Query Builder," which serves as the front-end interface that offers researchers a simple way to construct and issue their research questions. The "Query Transformation" module transforms the original query into a version of the question that is compatible with each CDM. The CDM Harmonization Results Database and Viewer receives and analyzes the results of a query in one or more of the CDM formats. To process these results, the team created the SDTM export tool that exports record level results in the CDISC STDM format.

The NIH/NCI team completed the metadata curation of all four CDMs (Sentinel, PCORV4.0, OMOP, I2b2 ACT) which will allow them to complete the registrations of the CDMs and BRIDG in caDSR and then with NIH/NLM. The CDMs have been incorporated into caDSR, which will be a lasting product and available

into the future. Also, all the final CDM packages of the CDEs were sent to NIH/NLM to be reloaded and a visualization of the mappings across the CDEs was made available on an NIH/NCI website. The four CDM CDEs have since been uploaded to the NIH CDE Repository and the related report is being finalized. ONC and NIH teams completed the mapping of the four data models, the NIH BRIDG conceptual model, and from BRIDG to FHIR. FHIR® resource extensions and CDMH IG are in HL7 ballot reconciliation, and the ONC team anticipates publication within Q3 or Q4 of calendar year 2019. Also, pilot-testing of the CDMH IG has been completed and the package has been passed for the first-round of balloting and is in reconciliation. Lastly, NIH/NLM developed a governance framework document that outlines suggested policies and practices for access to and use of the RWD that are derived from data-sharing networks that connect CDMs.

Publications and Other Publicly Available Resources:

- The IG for utilizing CDMH as represented in HL7 FHIR® is now available online at HL7.org. All documentation regarding the use of these extensions and profiles for implementing the CDMH elements is presented. https://build.fhir.org/ig/HL7/cdmh/
- The SDTM Export Tool which exports record level results in CDISC STDM format is available at: https://www.cdisc.org/standards/foundational/sdtm
- The CDM-to-BRIDG mappings, which aligns the data models and shows the alignment of data elements and the existing gaps, can be used to transform data to various output versions. The mappings are available here: https://bridgmodel.nci.nih.gov/
- The CDMH/BRIDG-to-SDTM mapping provides the rules to export results in support of submissions to FDA. The mapping is available here: https://github.com/cdmhproject/cdmh
- The CDMH/BRIDG-to-FHIR mapping proves the alignment and gaps of the CDMs to existing HL7 FHIR resources, each are available here:
    - I2B2 CDE Links: https://cdebrowser.nci.nih.gov/cdebrowserClient/cdeBrowser.html#/search?programArea=0&contextId=6CB969CC-DD4B-1016-E053-F662850A40C7&classificationSchemeId=66589E50-F300-4B2E-E053-F662850A5342
    - OMOP: https://cdebrowser.nci.nih.gov/cdebrowserClient/cdeBrowser.html#/search?programArea=0&contextId=6CB969CC-DD4B-1016-E053-F662850A40C7&classificationSchemeId=339F8634-199C-3A8A-E050-BB89AD431025
    - PCORNet: https://cdebrowser.nci.nih.gov/cdebrowserClient/cdeBrowser.html#/search?programArea=0&contextId=6CB969CC-DD4B-1016-E053-F662850A40C7&classificationSchemeId=66589E50-F2A6-4B2E-E053-F662850A5342
    - Sentinel: https://cdebrowser.nci.nih.gov/cdebrowserClient/cdeBrowser.html#/search?programArea=0&contextId=6CB969CC-DD4B-1016-E053-F662850A40C7&classificationSchemeId=665A47EA-F6D2-2F9F-E053-F662850A1DCB
- The BRIDG model was updated to include concepts of the mapped CDMs, promoting implementation strategies for use of BRIDG with the various CDMs allowing use of the BRIDG model in CDM projects as well as linking the CDMs to HL7 FHIR and CDISC SDTM via BRIDG, https://bridgmodel.nci.nih.gov/download-model/bridg-releases

- The PCOR CDMH Program Area can be browsed. https://cdebrowser.nci.nih.gov/cdebrowserClient/cdeBrowser.html#/search?programArea=0&contextId=6CB969CC-DD4B-1016-E053-F662850A40C7
- The four CDM CDE are available on the NIH CDE Repository – CDMH Common Data Elements: https://cde.nlm.nih.gov/cde/search?selectedOrg=NCI&classification=PCORTF%20CDMH. The site provides access to structured human and machine-readable definitions of data elements that have been recommended or required by NIH Institutes and Centers and other organizations for use in research and for other purposes.
- A visualization site has been established at NIH/NCI depicting the BRIDG and cross-model mappings. https://vis-review-si.nci.nih.gov/
- The data governance framework document details policies and practices for accessing to and using of RWD derived from data-sharing networks. https://cde.nlm.nih.gov/resources
- NIH/NLM conducted a survey of NIH intramural researchers to understand their use of real world data (RWD) and need for support to utilize analytical tools. https://cde.nlm.nih.gov/resources

Cross-Agency Collaboration

The project team has engaged with the FDA Cross-Network Directory Service project to explore the reuse 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 NIH/NCI Genomic Data Commons). NIH/NCATS have been very involved in leveraging components from ONC U.S. Core/DAF project and has regular meetings with PCORI to discuss and share materials.

Training Data for Machine Learning to Enhance Patient-Centered Outcomes Research (PCOR) Data Infrastructure

NIH/NLM Period of Performance
8/1/19 – 2/28/22

ONC Period of Performance
8/1/19 – 2/28/22

NIH/NLM Federal Point of Contact
Stefan Jaeger

ONC Federal Point of Contact
Stephanie Garcia

Artificial Intelligence (AI) and associated innovative technologies like machine learning have the power to consume large amounts of data in varied, complex formats to more quickly identify effective treatments potentially accelerating clinical innovation by speeding up the research lifecycle and the application of evidence in clinical settings. Machine learning relies on training data sets to effectively “learn” how to identify patterns with little human intervention. Industry experts have acknowledged that large amounts of high-quality training data are a critical part of the foundation that will support researchers’ use of machine learning in patient-centered outcomes research.

Use Case Selection

The incidence of kidney disease and drug-resistance TB are growing. However, kidney disease and TB represent two conditions with limited evidence-based treatments. The project use cases were selected to address: 1) lack of definitive treatments for kidney disease, and 2) limited efficacy of TB treatment.

High-quality training data sets leverage CDMs and CDEs annotated by domain experts. These training data sets can combine previously unconnected data resources and can be used to train algorithms to elucidate knowledge and extract relevant data points for research to accelerate the discovery of novel disease-outcome correlations and associations, and inform the design of prevention and treatment studies. This project will curate high-quality training data sets using quality clinical research data that are collected by HHS and combine them with other kidney and tuberculosis (TB) related clinical data. These training data sets will then be used to develop, and train AI models for predication. The project will develop and disseminate papers and a final report that discusses the current strengths and limitations of AI for PCOR, industry, and HHS. It will also include a forward-looking section that will provide an initial high-level blueprint identifying the potential for HHS to use AI in discovery and safety surveillance and to address key issues facing the intended federal end-users (e.g., Medicare, Medicaid).

Project Purpose and Goals

This project will enhance the capacity of researchers to use machine learning by developing and disseminating a number of resources that will present not only training data and methods but also lessons learned. Evidence generated from this application of AI will support multiple federal and HHS initiatives in precision medicine in kidney and TB research programs so providers and public health practitioners can match patients to the best treatments based on their specific health conditions, life-experiences, and genetic/phenotypic profiles.

This project will address the following objectives:

- Develop high-quality training data sets and capture lessons learned from best practices in data annotation and curation and compile insights on the data quantity and quality requirements for machine learning applied in patient-centered outcomes research.
- Develop machine learning algorithms that will be used provisionally to test the training data sets and validate approaches in evaluating their performance using conventional metrics.
- Develop an IG detailing each method used and the generic aspects of the kidney disease and TB use case that each method leverages, with detail sufficient to facilitate its application to a wider array of use cases.
- Disseminate tools and training data and lessons learned to stimulate the application of these methods to a wider array of use cases by researchers. Project products will be made accessible to researchers through existing NIH/NLM and ONC repositories and data centers/enclaves.

Contributions to PCOR Data Infrastructure Functionalities

- ***Use of Clinical Data for Research:*** This project will use a diverse range of clinical data, from clinical images such as X-rays and CT scans to genomic and other registry data sets to improve machine learning algorithms to more quickly identify effective treatments and potentially accelerating clinical innovation by speeding up the research lifecycle and the application of evidence in clinical settings. The project seeks to apply this research to national large-scale cohort studies like the All of Us Research Program.

NIH/NLM Accomplishments

The NIH/NLM portion of the project effectively launched in August of 2019 and has focused its activities on initiating partner engagement.

- In order to obtain new training data related to the TB use case, NIH/NLM prepared the necessary paperwork to access the NIAID TB portal, and is working with external health care providers to acquire relevant data.
- NIH/NLM has been collaborating with partners in China to convene a workshop titled "Using Artificial Intelligence to Identify Multi-Drug Resistant TB and AIDS-TB Co-infections in Radiographs," which is aimed at raising awareness about the project and identifying other potential sources of training images for the algorithms. The workshop was held in Shenzhen, China, in parallel with the Medical Image Computing and Computer Assisted Intervention Society conference in October 2019.

ONC Accomplishments

Since the project launched in the spring of 2019, the project teams have made progress toward several activities.

- ONC issued a contract award for performance of the work. The project kick-off meeting was held on September 24, 2019.
- Planning is underway to convene a multi-stakeholder TEP to advice on the direction of the project.

Technologies for Donating Medicare Beneficiary Claims Data to Research Studies

CMS Period of Performance
9/7/17 – 4/30/18

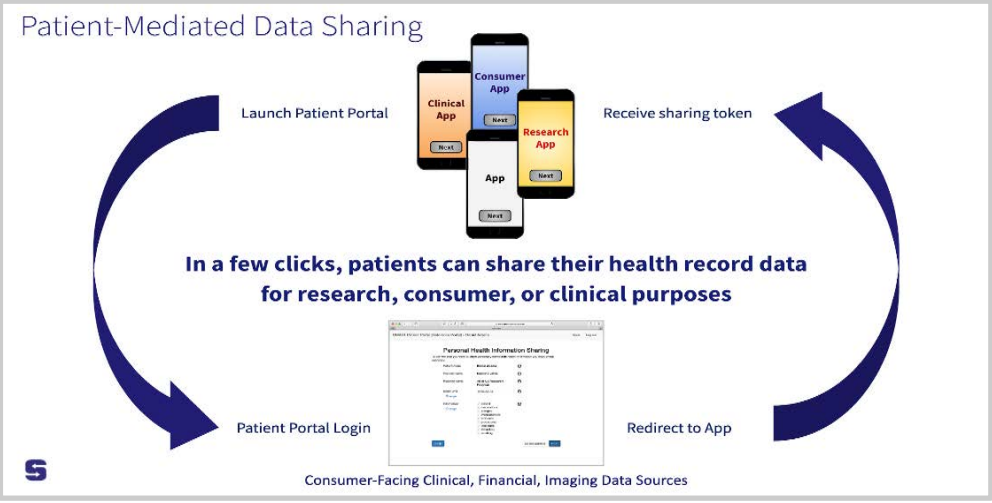
CMS Federal Point of Contact
Lori Pettebone-Koraganie
Katie Cox

NIH Period of Performance
9/7/17 – 4/29/19

NIH Federal Point of Contact
James McClain

This project is a collaboration between NIH and 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.

Exhibit 15. Patient-Mediated Data Sharing



The All of Us Research Program is a national, large-scale research enterprise designed to support discoveries that increase our ability to better treat and prevent disease by enrolling 1 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, via a few clicks using an app and patient portal (see Exhibit 15). These data will 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 health care 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 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' Blue Button API supports the S4S permissions and API calls to retrieve EOB and patient resources.
- Ensure that CMS' 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 Publicly Funded Data Systems for Research:*** This project leverages 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.

CMS Accomplishments

CMS' portion of the project is complete. In the section that follows, we summarize the project accomplishments for CMS' contribution.

A previous CMS OS-PCORTF project launched the Blue Button 2.0 API on March 6, 2018. Blue Button 2.0 enables app developers to request access to the API to design apps, allow Medicare beneficiaries to share their claims data with third-party applications they choose. The objective of the CMS Blue Button 2.0 project was to add technical enhancements to Blue Button by converting Medicare claims to the HL7 FHIR data standards.

This CMS project, in collaboration with NIH, further enhances Blue Button 2.0 functionality by adapting the Blue Button 2.0 API to allow beneficiaries to submit their Medicare claims data directly to the research studies of their choosing by piloting the use of the enhanced Blue Button 2.0 API to facilitate data-sharing with the *All of Us* Research Program through S4S.

CMS successfully modified the API code to meet the S4S specifications. Modifications were made to the S4S reference implementation open source test suite for Medicare beneficiary claims data to work with the CMS Blue Button 2.0 developer sandbox. Additionally, deployment of the test suite with EOB and Coverage FHIR resource validation was completed. In 2019, CMS finished incorporating native mobile app support to the OAuth2.0 authorization protocol. Additionally, CMS added 30,000 synthetic beneficiaries to the production environment for production application testing and increased Part D data updates from monthly to weekly collection.

To disseminate their work, CMS developed source code documentation, user experience guides, and reference tools, all available for public consumption on the Blue Button website (https://bluebutton.cms.gov/). The project's publicly available documentation benefits developers and other stakeholders.

While the project period of performance has concluded, the CMS team continues to work to enhance the system and support NIH's work. Specifically, CMS will on-board the *All of Us* Research Program S4S sites with Blue Button 2.0 once NIH's portion is complete.

Publications and Publicly Available Resources

- Source code published on GitHub, which is maintained continuously for the Blue Button Platform: https://github.com/CMSgov
- Instructions for understanding and using the CMS Blue Button 2.0 API to connect patient health data is available at: https://cmsgov.github.io/bluebutton-developers
- Google group to support developers: https://groups.google.com/forum/#!forum/Developer-group-for-cms-blue-button-api
- Blue Button 2.0 IG, published June 2018: https://bluebutton.cms.gov/developers/#blue-button-implementation-guide
- A listing of Blue Button 2.0 Production applications can be found at https://go.cms.gov/bluebuttonapps
- The Blue Button 2.0 Sandbox is available at: https://sandbox.bluebutton.cms.gov
- Provides an Extract-Transform-Load data pipeline, which moves data from CMS' Chronic Conditions data warehouse into the CMS Blue Button FHIR Server: https://github.com/CMSgov/bluebutton-data-pipeline

- Documentation detailing the developer/application validation process to support operations: https://bluebutton.cms.gov/developers/#production-api-access

NIH Accomplishments

The NIH project team has made significant progress on completion of its deliverables:

- NIH conducted pilot testing on the S4S App, which was successful in validating the technology and workflow.
- NIH developed the S4S Discovery App that supports viewing both clinical and CMS Blue Button patient data made available by the S4S Procure App.
- The NIH-developed S4S Procure App allows developers to obtain data from the CMS Blue Button API.

Publications and Other Publicly Available Resources

- S4S API. The FHIR API links to support the sharing of data via portals, including use cases in clinical data, financial data, and imaging data:
  - http://syncfor.science/api-calls/
  - http://syncfor.science/use-case/clinical/
  - http://syncfor.science/use-case/financial/
  - http://syncfor.science/use-case/imaging/
- S4S Test Suite.
  - The framework for testing S4S API implementations can be found here: https://github.com/sync-for-science/test-suite
  - The Inferno Testing Tool for testing server compatibility with S4S can be found here: https://github.com/sync-for-science/test-suite/wiki/Inferno-Migration-Guide
- Research App API. The list of endpoints available to any clients working with the Research App API can be found here: https://github.com/sync-for-science/research-app-api
- 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
- Shared design work with CARIN alliance to support FHIR R4 profiles for claims data, including an IG with suggested resources for payers to provide to their consumers: http://build.fhir.org/ig/HL7/carin-bb/

Validating and Expanding Claims-based Algorithms of Frailty and Functional Disability for Value-based Care and Payment

AHRQ Period of Performance
5/28/19 – 3/15/23

AHRQ Federal Point of Contact
Arlene Bierman

ASPE Period of Performance
5/28/19 – 4/28/23

ASPE Federal Point of Contact
LokWong Samson

CDC Period of Performance
6/24/19 – 6/23/23

CDC Federal Point of Contact
Violanda Grigorescu

Functional risk factors such as frailty and functional disabilities (e.g., vision impairment, deaf or hard of hearing, mental health conditions, etc.) play an important role in risk-adjustment of research studies as well as payment and value-based care programs. ASPE has conducted preliminary research on four categories of functional risk factors: frailty, predictors of functional dependence, mental and behavioral health disorders, and existing indicators in the CCW of other potentially disabling conditions. ASPE found that these indicators can improve the predictive/explanatory power of risk-adjustment models of Medicare outcomes measures used in value-based care programs, and they explain poorer outcomes in dually enrolled beneficiaries. Validation of these indicators would ensure broader use and acceptance for use in payment programs and for robust research studies. These indicators and algorithms of frailty and functional disabilities can be used to conduct robust PCOR and to assist case identification and risk-adjustment based on patient characteristics.

ASPE is partnering with CMS to update and validate the current list of potentially disabling conditions in the CCW and potentially add these additional indicators in both Medicare and the Transformed Medicaid Statistical Information System (T-MSIS) data. ASPE is also partnering with the CDC and AHRQ to develop and test EHR-modified algorithms that identify patients with frailty and functional disabilities across patient populations (i.e., Medicare, Medicaid, and dual-eligible beneficiaries). The algorithm development and testing process will begin with a close look at descriptive statistics, including the demographics available, and types of EHR data fields available or used to capture patient function or functional disabilities, led by ASPE (claims-based algorithms) and CDC (EHR-modified algorithms). This process will also involve a direct comparison of EHR-based algorithms and claims-based algorithms in detecting disability-related fields, and iterative refinement.

Project Purpose and Goals

The purpose of this project is to build data capacity to conduct robust PCOR by refining and validating claims-based algorithms of frailty, functional disabilities, and potentially disabling conditions. This project will modify the claims-based algorithms to develop EHR-based algorithms for use in different payer populations, including commercial payers. The project will then use the claims-based algorithms that rely on diagnoses and further incorporate data fields available in EHR systems to create modified versions of the algorithms that can be used to mine EHR data. The EHR-based algorithms will include data elements such as the results of patient disability screening tools, data fields on functional status and disability where available, and potentially physician notes.

Specifically, the objectives of this project are to develop:

- A set of validated and refined claims-based algorithms using Medicare and/or Medicaid claims that identify patients' functional risk and frailty, to be made available to the public through the CCW.

- A set of draft EHR-modified algorithms that have been tested in a health system.
- A final IG to support users in implementing the EHR algorithms for research or quality-related risk adjustment, including researchers, health systems, and payers. The IG may identify areas where users may need to modify the algorithms to adapt their EHR system. It may also include summaries of the EHR Learning Network meetings and background meeting materials for reference.

Contributions to PCOR Data Infrastructure Functionalities

Standardized Collection of Standardized Clinical Data: The project will start with claims-based algorithms that rely on diagnoses and further incorporate functional risk factor data fields from EHR systems in order to create modified versions of the algorithms. The modified algorithms will include data elements such as the results of patient disability screening tools, data fields on functional status and disability where available, and, potentially, physician notes with the goal of facilitating patient-centered analysis of disability risk and health outcomes.

Linking of Clinical and Other Data for Research: This project will validate claims-based and EHR-based algorithms. The algorithms will help identify and extract information for patients with functional risk factors and facilitate analysis across these data sets, which can be put to use in both value-based care programs and outcomes research.

Use of Enhanced Publicly Funded Data Systems for Research: This project will support the enhancement of strategic publicly funded data systems (including CMS data) by updating and validating the current list of potentially disabling conditions in the CCW. Validation of these indicators would ensure broader use and acceptance for use in payment programs and for robust research studies that evaluate quality of care, patient health risks and comorbidities, etc.

Accomplishments

The project will be led by ASPE, in collaboration with AHRQ, CDC and CMS. As the project commences, major accomplishments ASPE has achieved include finalizing the scope of work and starting the contract bid selection process. CDC drafted a report presenting findings from the descriptive analysis of patients in the ambulatory EHR data, which included a list of variables used to capture patient function and functional disability.

XIII. Conclusion

The OS-PCORTF projects featured in this report demonstrate ASPE's commitment to enhancing data capacity for patient-centered outcomes research. Individual agency products and deliverables continue to improve the ability of researchers to collect, link, and analyze data to generate evidence that furthers the transformation of the health care system toward more patient-centered care delivery and decision-making.

The work of the six projects that were completed in FY 2019 highlight the breadth of the contributions made in building data capacity. From work that increases the collection of patient-provided data for clinical decision-making and comparative effectiveness research, to tools that improve researchers ability to conduct population studies and safety surveillance using data derived from multiple distributed research networks, to machine learning techniques that increase data usability, these projects leverage

cutting edge standards and technologies that resulted in tools now available to the researcher community to advance the field of PCOR.

Guided by the *HHS Strategic Roadmap for Building Data Capacity for Clinical Comparative Effectiveness Research,* the portfolio, as demonstrated by the 13 new FY 2019 projects, also continues to evolve to meet emerging HHS policy priorities. This report highlights four key policy areas in which the current OS-PCORTF projects are working to develop solutions. With the recent reauthorization of the PCOR Trust Fund, ASPE will build upon its successful collaboration with agency leadership to identify the work that addresses emerging policy issues such as those enumerated in the reauthorization i.e., intellectual disabilities, maternal mortality, and an expanded understanding of the types of outcomes, preference and values that patient use when making health care decisions.

In FY 2020 and beyond, ASPE 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.

Key Products for Researchers

- An HL7 FHIR IG for the electronic capture and exchange of PRO data: PRO FHIR IG
- An HL7 FHIR IG and governance model for establishing a network of registries that resulted in the formation for the first CRN focused on women health technologies: WHT-CRN FHIR IG
- A tool and HL7 FHIR IG that harmonizes data across different research network CDMs: CDM FHIR IG
- A platform for capturing and sharing data quality metrics and assessing fitness-for-use of data across research networks: DQM platform
- A tool that uses NLP to structure unstructured surveillance data for analysis: CLEW Workbench
- An API that allows patients to donate their Medicare Blue Button data to researchers: Sync for Science API

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
	Data Capacity for Patient-Centered Outcomes Research through Creation of an Electronic Care Plan for People with Multiple Chronic Conditions*
	Enhancing Patient-Centered Outcomes Research (PCOR): Creating a National Small-Area Social Determinants of Health (SDOH) Data Platform
	Validating and Expanding Claims-based Algorithms of Frailty and Functional Disability for Value-based Care and Payment*
Assistant Secretary for Preparedness and Response	Assessing and Predicting Medical Needs in a Disaster*
Centers for Disease Control and Prevention	Augmenting the National Hospital Care Survey (NHCS) Data through Linkages with Administrative Records
	Childhood Obesity Data Initiative: Integrated Data for Patient-Centered Outcomes Research
	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 Identification of Opioid-Involved Health Outcomes Using Linked Hospital Care and Mortality
	Identifying Co-Occurring Disorders among Opioid Users Using Linked Hospital Care and Mortality Data
	Making Electronic Health Record (EHR) Data More Available for Research and Public Health
	MAT-LINK: MATernaL and Infant Network to Understand Outcomes Associated with Treatment for Opioid Use Disorder during Pregnancy
	Strengthening the Data Infrastructure for Outcomes Research on Mortality Associated with Opioid Poisonings
	Validating and Expanding Claims-based Algorithms of Frailty and Functional Disability for Value-based Care and Payment
Centers for Medicare and Medicaid Services	Technologies for Donating Medicare Beneficiary Claims Data to Research Studies*

Funded Agency	Project Title
Food and Drug Administration	Bridging the PCOR Infrastructure and Technology Innovation through Coordinated Registry Networks (CRN) Community of Practice (COP)
	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*
	SHIELD - Standardization of Lab Data to Enhance Patient-Centered Outcomes Research and Value-based Care
	Standardization and Querying of Data Quality Metrics and Characteristics for Electronic Health Data
National Institutes of Health	Data Capacity for Patient-Centered Outcomes Research through Creation of an Electronic Care Plan for People with Multiple Chronic Conditions*
	Developing a Strategically Coordinated Registry Network (CRN) for 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*┼
	NIDA's AMNET: An Addiction Medicine Network to Address the United States Opioid Crisis
	Technologies for Donating Medicare Beneficiary Claims Data to Research Studies*
	Training Data for Machine Learning to Enhance Patient-Centered Outcomes Research (PCOR) Data Infrastructure┼
	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 Assistant Secretary for Planning and Evaluation	Linking State Medicaid and Child Welfare Data for Outcomes Research on Treatment for Opioid Use Disorder
	Validating and Expanding Claims-based Algorithms of Frailty and Functional Disability for Value-based Care and Payment*
Office of the National Coordinator for Health Information Technology	A Synthetic Health Data Generation Engine to Accelerate Patient-Centered Outcomes Research
	Advancing the Collection and Use of Patient-Reported Outcomes through Health Information Technology (IT)*

Funded Agency	Project Title
	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*
	Training Data for Machine Learning to Enhance Patient-Centered Outcomes Research (PCOR) Data Infrastructure*

* 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
Agency for Healthcare Research and Quality	Harmonization of Clinical Data Element Definitions for Outcome Measures in Registries
Assistant Secretary for Planning and Evaluation	Beta Testing of the Multi-Payer Claims Data^
	Comparative Effectiveness Research Inventory
Centers for Disease Control and Prevention	Enhancing Data Resources for Studying Patterns and Correlates of Mortality in Patient-Centered Outcomes Research: Project 4 – NDI Workshop and Strategy Paper*
	Expanding Data Collection Infrastructure of the National Program of Cancer Registries for Comparative Effectiveness Research
	Improving the Mortality Data Infrastructure for Patient-Centered Outcomes
	Improving Beneficiary Access to their Health Information through an Enhanced BlueButton Service
	Privacy and Security Blueprint, Legal Analysis and Ethics Framework for Data Use, & Use of Technology for Privacy*
Centers for Medicare & Medicaid Services	Enhancing Data Resources for Researching Patterns of Mortality in Patient Centered Outcomes Research: Project 3*
	Maintenance and Support of the Chronic Conditions Warehouse for Comparative Effectiveness Research
Food and Drug Administration	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 EHRs: Using Standardized Clinical Research Data
	Utilizing Data from Various Data Partners in a Distributed Manner
Health Resources and Services Administration	Strengthening and Expanding the Community Health Applied Research Network (CHARN) Registry to Conduct Patient-Centered Outcomes Research

Funded Agency	Project Title
National Institutes of Health	Development of Data Infrastructure for Use of EHRs in Comparative Effectiveness Research*†
	Creation of LOINC Equivalence Classes
Office of the National Coordinator for Health Information Technology	Conceptualizing a Data Infrastructure for the Capture and Use of Patient-Generated Health Data
	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*
	Privacy and Security Blueprint, Legal Analysis and Ethics Framework for Data Use, & Use of Technology for Privacy*
	Security and Privacy Standards for Patient Matching, Linking and Aggregation
	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.

† OS-PCORTF project funding awarded within NIH

Appendix B. Abbreviations

Acronym	Description
ACEP	American College of Emergency Physicians
ACF	Administration for Children and Families
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
CDISC	Clinical Data Interchange Standards Consortium
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

Acronym	Description
HHS	U.S. Department of Health and Human Services
HITECH	Health Information Technology for Economic and Clinical Health Act
HL7	Health Level Seven International
HP	Office of Health Policy
HRSA	Health Resources and Services Administration
ICD-10	International Classification of Diseases 10th Edition
IG	Implementation Guide
IHE	Integrating the Healthcare Enterprise
IRB	Institutional Review Board
IT	Information Technology
LOINC	Logical Observation Identifiers Names and Codes
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
NIDDK	National Institute of Diabetes and Digestive and Kidney Diseases
NIMH	National Institute of Mental Health
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
OMOP	Observational Medical Outcomes Partnership
ONC	Office of the National Coordinator for Health Information Technology
OS-PCORTF	Office of the Secretary Patient-Centered Outcomes Research Trust Fund
OUD	Opioid Use Disorders
PCOR	Patient-Centered Outcomes Research
PCORI	Patient-Centered Outcomes Research Institute
PCORnet	PCORI's National Patient-Centered Clinical Research Network

Acronym	Description
PGHD	Patient-generated health data
PHCP	Public Health Community Platform
PMN	PopMedNet™
PMAL	Patient Matching, Linking, and Aggregation
PPI	Patient-provided Information
PRH data	Patient-reported health data
PRO	Patient-Reported Outcome
PROMIS	Patient-Reported Outcomes Measurement Information System
PWMI	Pediatric Weight Management Interventions
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
SUD	Substance Use Disorder
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

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 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.[66]
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 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.
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.
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.
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 recognition of its utility.

Key Terms	Description
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.[67]
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.
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.[68] RxNorm is maintained by NIH/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 IG that builds upon FHIR® by defining an authorization model for apps based on the OAuth standard.[69,70,71]
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.[72]

References

[1] U.S. Department of Health and Human Services. Office of the Assistant Secretary for Planning and Evaluation (ASPE). Patient-Centered Outcomes Research Trust Fund FAQs. https://aspe.hhs.gov/patient-centered-outcomes-research-trust-fund-faqs.

[2] Public Health Services Act Section 937(7).

[3] OBERD. Smart Outcomes Data Collection. https://new.oberd.com/

[4] HL7 FHIR. Women's Health Technologies (WHT) Coordinated Registry Network (CRN) FHIR Implementation Guide. http://hl7.org/fhir/us/womens-health-registries/2019May/

[5] CMS Blue Button 2.0. Blue Button 2.0. https://bluebutton.cms.gov/

[6] Dhopeshwarkar, R., Dullabh, P., Dungan, R., Gonzales, A., Hovey, L., Isaacson, A., Heaney-Huls, K., Lumsden, S., Moriarty, E., Sanders, E., Smith, S., & Trevino, M. (2019). Building Data Capacity for Patient-Centered Outcomes Research Portfolio Highlights (2016–2019): Impact, Opportunities, and Case Studies.

[7] Caffrey, M. Getting Ready for the Use of Real-World Evidence. (February 4, 2019). Blog. https://www.ajmc.com/focus-of-the-week/getting-ready-for-the-use-of-realworld-evidence

[8] FDA. (December 2018). Framework for FDA's Real-World Evidence Program. www.fda.gov/media/120060/download

[9] Carino T and McClellan MB. (September 19, 2019). Accelerating the Shift Toward Person-Generated Real-World Evidence. Health Affairs Blog. https://www.healthaffairs.org/do/10.1377/hblog20190918.730735/full/

[10] McDonald L, Malcolm B, Ramagopalan S, Syrad H. (2019). Real-world Data and the Patient Perspective: the PROmise of Social Media? *BMC Medicine*, *17*(1), 11. doi:10.1186/s12916-018-1247-8

[11] Hoyert, DL, Uddin, SF Miniño, AM. (January 30, 2020). Evaluation of the Pregnancy Status Checkbox on the Identification of Maternal Deaths. *National Vital Statistics Report*. Available at https://www.cdc.gov/nchs/data/nvsr/nvsr69/nvsr69_01-508.pdf

[12] Mayer, R, Dingwall, A, Simon-Thomas, J, Sheikhnureldin, A, Lewis, K. The United States Maternal Mortality Rate Will Continue To Increase Without Access To Data. Blog. https://www.healthaffairs.org/do/10.1377/hblog20190130.92512/full/

[13] Haight SC, Ko JY, Tong VT et al. Opioid Use Disorder Documented at Delivery Hospitalization — United States, 1999–2014. *Weekly* / August 10, 2018 / 67(31);845–849. https://www.cdc.gov/mmwr/volumes/67/wr/mm6731a1.htm

[14] Mayer, R, Dingwall, A, Simon-Thomas, J, Sheikhnureldin, A, Lewis, K. The United States Maternal Mortality Rate Will Continue To Increase Without Access To Data. Blog. https://www.healthaffairs.org/do/10.1377/hblog20190130.92512/full/

[15] NIDA. (January 2019). Opioid Overdose Crisis. https://www.drugabuse.gov/drugs-abuse/opioids/opioid-overdose-crisis

[16] H.R.6 - SUPPORT for Patients and Communities Act. https://www.congress.gov/bill/115th-congress/house-bill/6

[17] H.R.6 - SUPPORT for Patients and Communities Act. https://www.congress.gov/bill/115th-congress/house-bill/6

[18] Tracking Federal Funding to Combat the Opioid Crisis. (March 2019). https://bipartisanpolicy.org/library/tracking-federal-funding-to-combat-the-opioid-crisis/

[19] U.S. Department of Health and Human Services (HHS). 5-Point Strategy to Combat the Opioid Crisis. https://www.hhs.gov/opioids/about-the-epidemic/hhs-response/index.html

[20] HHS. Strategy to Combat the Opioid Abuse, Misuse, and Overdose: A Framework Based on the Five Point Strategy. https://www.hhs.gov/opioids/sites/default/files/2018-09/opioid-fivepoint-strategy-20180917-508compliant.pdf

[21] National Science and Technology Council. (October 2018). Health Research and Development to Stem the Opioid Crisis: A National Roadmap. https://www.nih.gov/sites/default/files/Health-RD-to-Stem-Opioid-Crisis-2018-Roadmap-for-Public-Comment.pdf

[22] Substance Abuse and Mental Health Services Administration. (2017). 1.5 Million Adults Have Serious Mental Illness and Misused Opioids in the Past Year. The CBHSQ Report, The National Survey on Drug Use and Health. Retrieved from https://www.samhsa.gov/data/sites/default/files/report_2734/Spotlight-2734.pdf

[23] Ghertner, R, Baldwin, M, Crouse, G, et al. (March 9, 2018). The Relationship Between Substance Use Indicators and Child Welfare Caseloads. Office of the Assistant Secretary for Planning and Evaluation. https://aspe.hhs.gov/system/files/pdf/258831/SubstanceUseCWCaseloads.pdf

[24] Centers for Medicare & Medicaid Services. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/Value-Based-Programs

[25] CMS. (April 22, 2019). HHS NEWS: HHS To Deliver Value-Based Transformation in Primary Care. https://www.cms.gov/newsroom/press-releases/hhs-news-hhs-deliver-value-based-transformation-primary-care

[26] Heath, S. (October 29, 2018). As Value-Based Care Models Improve, Patients Must Take Center Stage. https://patientengagementhit.com/news/as-value-based-care-models-improve-patients-must-take-center-stage

[27] Centers for Medicare and Medicaid Services. https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Readmissions-Reduction-Program.html

[28] Agency for Healthcare Research and Quality. https://www.ahrq.gov/hai/hac/tools.html

[29] Health Payer Intelligence. (February 2018). 80% of Payers Aim to Address Social Determinants of Health. https://healthpayerintelligence.com/news/80-of-payers-aim-to-address-social-determinants-of-health

[30] Hirsch R. (August 2018). New Z Codes Include Code for Millennials. RACMonitor. https://www.racmonitor.com/new-Z codes-include-code-for-millennials

[31] Agency for Healthcare Research and Quality (AHRQ). RoPR—The Registry of Patient Registries. https://patientregistry.ahrq.gov/

[32] 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.

[33] US Preventive Services Task Force. (2017). Screening for Obesity in Children and Adolescents: US Preventive Services Task Force Recommendation Statement. *JAMA, 317*(23), 2417–2426. doi:10.1001/jama.2017.6803

[34] Dietz W, et al. (2017). National Academy of Medicine, Washington, DC.

[35] Vivolo-Kantor AM, Seth P, Gladden RM, et al. (2018). Vital Signs: Trends in Emergency Department Visits for Suspected Opioid Overdoses — United States, July 2016–September 2017. *Morbidity and Mortality Weekly Report*, *67*, 279–285. http://dx.doi.org/10.15585/mmwr.mm6709e1.

[36] SAMHSA. NSDUH 2017: http://www.samhsa.gov/data/release/2017-national-survey-drug-use-and-health-nsduh-releases

[37] SAMHSA. NSDUH 2015: http://www.samhsa.gov/data/release/2015-national-survey-drug-use-and-health-nsduh-releases

[38] Haight SC, Ko JY, Tong VT, Bohm MK, Callaghan WM. (2018). Opioid use disorder documented at delivery hospitalization – United States, 1999–2014. *Morbidity and Mortality Weekly Report*, *67*, 845-849.

[39] Fill MA, Miller AM, Wilkinson RH, Warren MD, Dunn JR, Schaffner W, et al. (2018). Educational Disabilities among Children Born with Neonatal Abstinence Syndrome. *Pediatrics*, *124*(3).

[40] Oei JL, Melhuish E, Uebel H, Azzam N, Breen C, Burns L, et al. (2017). Neonatal Abstinence Syndrome and High School Performance. *Pediatrics*, *139*(2).

[41] CDC. Substance Use During Pregnancy. https://www.cdc.gov/reproductivehealth/maternalinfanthealth/substance-abuse/substance-abuse-during-pregnancy.htm

[42] 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.

[43] LOINC® from Regenstrief. (June 1, 2017.) New specification available for IVD vendors to link their test codes to LOINC. https://loinc.org/news/new-specification-available-for-ivd-vendors-to-link-their-test-codes-to-loinc/

[44] Kahn, M. G., Callahan, T. J., Barnard, J., Bauck, A. E., Brown, J., Davidson, B. N., Estiri, H., et al. (2016). A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. *EGEMS*, *4*(1), 1244. doi:10.13063/2327-9214.1244

[45] American College of Emergency Physicians. Clinical Emergency Data Registry. https://www.acep.org/cedr/

[46] Assistant Secretary for Public Affairs (ASPA). (July 23, 2018). Secretary Priorities. https://www.hhs.gov/about/leadership/secretary/priorities/index.html

[47] 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

[48] National Institute on Drug Abuse. (June 2018). Medications to Treat Opioid Use Disorder. https://www.drugabuse.gov/publications/research-reports/medications-to-treat-opioid-addiction/overview

[49] U.S. Food and Drug Administration. (February 14, 2019). Information about Medication-Assisted Treatment (MAT). https://www.fda.gov/drugs/information-drug-class/information-about-medication-assisted-treatment-mat

[50] ADAPTABLE, the Aspirin Study: A Patient-Centered Trial. http://theaspirinstudy.org/.

[51] Kepple, N. J. (2018). Does Parental Substance Use Always Engender Risk for Children? Comparing Incidence Rate Ratios of Abusive and Neglectful Behaviors across Substance Use Behavior Patterns. *Child Abuse & Neglect*, *76*, 44–55. https://doi.org/10.1016/j.chiabu.2017.09.015

[52] Smith, V. C., & Wilson, C. R. (2016). Families Affected by Parental Substance Use. *Pediatrics*, *138*(2), e20161575–e20161575. https://doi.org/10.1542/peds.2016-1575

[53] Peleg-Oren, N., & Teichman, M. (2006). Young Children of Parents with Substance Use Disorders (SUD): A Review of the Literature and Implications for Social Work Practice. *Journal of Social Work Practice in the Addictions, 6*(1-2), 49-61. http://dx.doi.org/10.1300/J160v06n01_03

[54] Kepple, N. J. (2018). Does parental substance use always engender risk for children? Comparing incidence rate ratios of abusive and neglectful behaviors across substance use behavior patterns. *Child Abuse & Neglect*, *76*, 44–55. https://doi.org/10.1016/j.chiabu.2017.09.015

[55] U.S. Food and Drug Administration. (2009). Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. *Federal Register 2009*, *74*(35), 65132-133.

[56] Patient-Centered Outcomes Research Institute. (2017). Users' Guide to Integrating Patient-Reported Outcomes in Electronic Health Records. https://www.pcori.org/sites/default/files/PCORI-JHU-Users-Guide-To-Integrating-Patient-Reported-Outcomes-in-Electronic-Health-Records.pdf

[57] National Institutes of Health. Patient-Reported Outcomes Measurement Information System (PROMIS). Retrieved fromhttps://commonfund.nih.gov/promis/index

[58] Demner-Fushman D, Chapman WW, McDonald CJ. (October 2009). What can natural language processing do for clinical decision support? *Journal of Biomedical Informatics*, *42*(5), 760-72.

[59] International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). http://www.meddra.org/how-to-use/support-documentation/english.

[60] The Public Health Community Platform. https://sites.google.com/site/publichealthcommunityplatform/

[61] Kreimeyer K, Foster M, Pandey A, Arya N, Halford G, Jones SF, Forshee R, Walderhaug M, Botsis T. (September 2017). Natural Language Processing Systems for Capturing and Standardizing Unstructured Clinical Information: A Systematic Review. *Journal of Biomedical Informatics, 73*:, 14-29. https://www.sciencedirect.com/science/article/pii/S1532046417301685?via%3Dihub

[62] LAPPS Grid. The Language Application Grid. https://www.lappsgrid.org/

[63] Foster M., Pandey A., Kreimeyer K., Botsis T. (2018). Generation of an Annotated Reference Standard for Vaccine Adverse Event Reports. *Vaccine*, 36 (29), 4325-4330.

[64] Kreimeyer K, Foster M, Pandey A, Arya N, Halford G, Jones SF, Forshee R, Walderhaug M, Botsis T. (September 2017). Natural Language Processing Systems for Capturing and Standardizing Unstructured Clinical Information: A Systematic Review. *Journal of Biomedical Informatics, 73*, 14-29.

[65] Medical Device Epidemiology Network. http://mdepinet.org/about-us/.

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

[67] 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

[68] NIH U.S. National Library of Medicine. RxNorm Overview. Retreived from: https://www.nlm.nih.gov/research/umls/rxnorm/overview.html.

[69] SMART®. Retrieved from: https://smarthealthit.org/an-app-platform-for-healthcare/about/.

[70] SMART®. SMART of FHIR Profiles. Retrieved from: http://docs.smarthealthit.org/data/profiles.html.

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

[72] NIH U.S. National Library of Medicine. SNOMED-CT. Retrieved from: https://www.nlm.nih.gov/healthit/snomedct/.