



OSPCORTF
OFFICE OF THE SECRETARY PATIENT CENTERED
OUTCOMES RESEARCH TRUST FUND

BUILDING DATA CAPACITY FOR PATIENT-CENTERED OUTCOMES RESEARCH

Portfolio Highlights (2016 – 2019) Impact, Opportunities, and Case Studies

Full Report



ASPE
ASSISTANT SECRETARY FOR
PLANNING AND EVALUATION

OFFICE OF
HEALTH POLICY

Building Data Capacity for Patient-Centered Outcomes Research

Portfolio Highlights (2016 – 2019) Impact, Opportunities, and Case Studies

Full Report

Prepared for:

Office of the Assistant Secretary for Planning and Evaluation
Office of Health Policy
U.S. Department of Health and Human Services
200 Independence Avenue, SW
Washington, DC 20201

Prepared by:

Rina Dhopeswarkar^a
Prashila Dullabh^a
Rachel Dungan^b
Aldren Gonzales^c
Lauren Hovey^a
Allison Isaacson^b
Krysta Heaney-Huls^a
Susan Lumsden^c
Elizabeth Moriarty^a
Emily Sanders^a
Scott Smith^c
Marcos Trevino^c

^aNORC at the University of Chicago

^bAcademyHealth

^cOffice of the Assistant Secretary for Planning and Evaluation

This report was prepared under contract # HHS23320160020I. 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. The views expressed in this article are the authors and do not necessarily represent the views of the U.S. Department of Health and Human Services (HHS) or the United States Government.

Acknowledgements

The authors gratefully acknowledge contributions of the many federal agency partners who carried out the project activities of the OS-PCORTF portfolio which are described within this report.

Case Study Contributions

The content of the case studies presented in the third chapter of this report was informed by a webinar series which featured presentations of several individuals involved with OS-PCORTF project award activities. We acknowledge the important contributions of the following individuals who presented during the webinar series and reviewed drafts of the case studies:

Robert Ball, M.D., M.P.H., Sc.M.

Center for Drug Evaluation and Research
Food and Drug Administration

Richard Ballew, Ph.D.

Health, Research, Informatics and Technology
ICF

Jeffrey Brown, Ph.D.

Department of Population Medicine
Harvard Medical School

Carol DeFrances, Ph.D.

National Center for Health Statistics
Centers for Disease Control and Prevention

Stephanie Garcia, M.P.H.

Office of the National Coordinator for Health
Information Technology

Kenneth Gersing, M.D.

Division of Clinical Innovation
National Center for Translational Science

Gideon Scott Gordon, Ph.D.

Center for Drug Evaluation and Research
Food and Drug Administration

Janey Hsiao, Ph.D.

Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Wei Hua, Ph.D., M.H.S., M.S.

Division of Epidemiology
Food and Drug Administration

Sandy Jones, M.D.E.

Cancer Surveillance Branch, Division of
Cancer Prevention and Control
Centers for Disease Control and Prevention

David Kreda

Department of Biomedical Informatics
Harvard Medical School

Alexa Limeras, J.D.

Privacy and Confidentiality Unit
Centers for Disease Control and Prevention

Danica Marinac-Dabic, M.D., Ph.D., M.M.Sc., F.I.S.P.E.

Center for Devices and Radiological Health
Food and Drug Administration

Michael Nguyen, M.D.

Center for Drug Evaluation and Research
Food and Drug Administration

Emily O'Brien, Ph.D.

Department of Neurology
Duke University School of Medicine

Lori Pettebone-Koraganie, M.S.

Office of Enterprise Data & Analytics
Centers for Medicare & Medicaid Services

Mitra Rocca, Ph.D.

Center for Drug Evaluation and Research
Food and Drug Administration

Paul Sutton, Ph.D.

National Center for Health Care Statistics
Centers for Disease Control and Prevention

Robin Taylor, M.L.I.S.

National Library of Medicine
National Institutes of Health

Wendy Weber, N.D., Ph.D., M.P.H.

National Center for Complementary
and Integrative Health
National Institutes of Health

Technical Expert Panel

The authors sought to incorporate diverse perspectives of end-users of data infrastructure for patient-centered outcomes research and therefore convened a technical expert panel (TEP) of researcher and industry stakeholders. The TEP members were consulted throughout the development of this report over the course of two virtual and one in-person meeting and provided individual input as opposed to collective recommendations. The authors thank the following individuals for their insightful input:

Thomas Carton, Ph.D., M.S.
Louisiana Public Health Institute

Charles Friedman, Ph.D.
Department of Learning Health Sciences
University of Michigan

John Glaser, Ph.D.
Cerner Corporation

Michael Kahn, M.D., Ph.D.
Department of Pediatrics
University of Colorado Denver

Abel Kho, M.D.
Feinberg School of Medicine
Northwestern University

Julia Skapik, M.D., M.P.H.
National Association of Community Health Centers

Adam Wilcox, Ph.D.
Department of Biomedical Informatics & Medical
Education
University of Washington Medicine

Table of Contents

REPORT OVERVIEW	1
INTRODUCTION	1
Background.....	1
Strategic Framework.....	2
CHAPTER 1 PORTFOLIO ASSESSMENT	4
Chapter Executive Summary	4
Extent to Which The Portfolio Advances The Strategic Framework.....	4
Extent to Which The Portfolio Addresses Gaps Identified by The 2017 Evaluation	5
Extent to Which The Portfolio Has Evolved to Meet The Emerging Health Policy Landscape	5
Technical Expert Panel Feedback on Future Direction of the Portfolio	5
Introduction	6
Methods	6
Findings	9
How have the awards addressed the Strategic Framework?	9
Functionality 1: Use of Clinical Data for Research.....	9
Functionality 2: Standardized Collection of Standardized Clinical Data	10
Functionality 3: Linking Clinical and Other Data For Research	12
Functionality 4: Collection of Participant-Provided Information.....	13
Functionality 5: Use of Enhanced Publicly Funded Data Systems for Research	14
How Has the Portfolio Addressed Gap Areas Identified in the 2017 Evaluation?.....	16
How Has the Portfolio Evolved to Address the Data Infrastructure Needs of Emerging Policy Priorities?	20
Discussion: Areas for Future Work	23
Conclusion.....	25
CHAPTER 2 PORTFOLIO SYNERGIES AND OPPORTUNITIES WITH STRATEGIES	26
Chapter Executive Summary	26
Introduction	27
Methods	29
Findings	31
Functionality 1: Use of Clinical Data for Research.....	31
Functionality 2: Standardized Collection of Standardized Clinical Data	32
Functionality 3: Linking Clinical and Other Data for Research	32
Functionality 4: Collection of Participant-Provided Information (PPI)	32
Functionality 5: Use of Enhanced Publicly Funded Data Systems for Research	33
Discussion.....	33
Conclusions.....	39
CHAPTER 3 PORTFOLIO CASE STUDIES	40
Chapter Executive Summary	40
What Activities Can Be Pursued to Ensure That The Work Has Enduring Value?	41
What Non-Traditional Opportunities Or Platforms for Products, Outputs, Or Learnings Could Be Leveraged to Support Engagement between Researchers And End-Users?.....	42

Data Interoperability and Novel Approaches to Enhancing Researchers’ Access to Patient-Donated Data for Real-World Evidence.....	42
Case Study: Use of The ADAPTABLE Trial to Strengthen Methods to Collect, Validate, and Integrate Patient-Reported Information with Electronic Health Record Data	43
Case Study: Advancing the Collection and Use of Patient-Reported Outcomes through Health IT	46
Case Study: Technologies for Donating Medicare Beneficiary Claims Data to Research Studies	48
Data Standardization	51
Case Study: Development of a Natural Language Processing (NLP) Web Service for Public Health Use	52
Case Study: Standardization and Querying of Data Quality Metrics and Characteristics for Electronic Health Data.....	55
Case Study: Harmonization of Various Common Data Models and Open Standards for Evidence Generation	57
Enhancing National Death Index Data.....	60
Case Study: Enhancing Data Resources for Studying Patterns and Correlates of Mortality in Patient-Centered Outcomes Research: Project 4 — NDI Workshop and Strategy Paper.....	61
Case Study: Enhancing Data Resources for Studying Patterns and Correlates of Mortality in Patient-Centered Outcomes Research: Project 1—Adding Cause-Specific Mortality to NCHS’ National Hospital Care Survey by Linking to the National Death Index.....	62
Case Study: Enhancing Data Resources for Studying Patterns and Correlates of Mortality in Patient-Centered Outcomes Research: Project 2—Pilot Linkage of NDI+ to Commercially and Publicly Insured Populations	65
Supporting Data Asset Linkages	67
Case Study: Security and Privacy Standards for Patient Matching, Linking and Aggregation (PMAL).....	67
Case Study: Developing a Strategically Coordinated Registry Network (CRN) for Women’s Health Technologies	71
Case Study: PCOR: Privacy and Security Blueprint, Legal Analysis, and Ethics Framework for Data Use, & Use of Technology for Privacy	74
Discussion.....	77
What Activities Can Be Pursued to Ensure That The Work Has Enduring Value?.....	77
What Non-Traditional Opportunities Or Platforms for Products, Outputs, Or Learnings Could Be Leveraged to Support Engagement Between Researchers And End-Users?	78
Conclusion.....	80
APPENDICES.....	81
Appendix A. Strategic Framework Milestones by Functionality.....	82
Appendix B. Featured OS-PCORTF Awards	84
Appendix C. 2017 Evaluation Gap Area Criteria.....	86
Appendix D. Technical Expert Panel	88
Appendix E. HHS Data Strategy	89
Appendix F. High-Level Overview of Agency-Specific Frameworks and Data Strategies.....	91
Appendix G. Case Study Strategic Framework Functionalities	99
REFERENCES.....	100

List of Exhibits

Exhibit 1.	Telling the Story of PCOR in HHS.....	1
Exhibit 2.	OS-PCORTF Strategic Framework.....	2
Exhibit 3.	Progress toward Using Clinical Data for Research	10
Exhibit 4.	Progress toward Standardized Data Collection.....	11
Exhibit 5.	Progress toward Linking Data	12
Exhibit 6.	Linked Data Files Produced by the OS-PCORTF Portfolio	13
Exhibit 7.	Progress toward the Collection of PPI	14
Exhibit 8.	Progress Toward Using Federal Datasets for Research	15
Exhibit 9.	Awards and Products that Address the Strategic Gap Areas Identified from 2017 Evaluation	18
Exhibit 10.	Opioid Topics Addressed by the Portfolio Award Products and Description	21
Exhibit 11.	Legislative Mandate Data Infrastructure Needs Supported by OS-PCORTF Award (Awards per Fiscal Year)	21
Exhibit 12.	Evolving Data Infrastructure Concepts of the Portfolio and Representative Awards from FY 2016–2019.....	22
Exhibit 13.	HHS Data Strategy	30
Exhibit 14.	The OS-PCORTF Strategic Framework Functionalities	30
Exhibit 15.	HHS Agency Data Strategies.....	31
Exhibit 16.	Reports and Proposed Rules	31
Exhibit 17.	S4S Patient-Mediated Data-Sharing.....	48
Exhibit 18.	Award Deliverables	49
Exhibit 19.	Overview of CDM Harmonization.....	58
Exhibit 20.	Percent of Mortality Records Transmitted to NCHS within 10 Days of the Date of Death	61
Exhibit 21.	Datasets to Be Linked by the Project.....	63
Exhibit 22.	Linkage Identifiers	63
Exhibit 23.	Short-term Implications of Project Work.....	66
Exhibit 24.	Implications and Impact.....	75
Exhibit 25.	Three Basic Questions When Deciding to Share Data	76
Exhibit F-1.	OS-PCORTF Strategic Framework for Building PCOR Data Infrastructure	91
Exhibit F-2.	National Institutes of Health (NIH) – Strategic Plan for Data Science.....	92

Exhibit F-3.	National Library of Medicine (NLM) – A Platform for Biomedical Discovery and Data-Powered Health: National Library of Medicine Strategic Plan 2017–2027.....	93
Exhibit F-4.	Food and Drug Administration (FDA) – Framework for FDA’s Real-World Evidence Program.....	93
Exhibit F-5.	Food and Drug Administration (FDA) – Sentinel System Five-Year Strategy 2019–2023	94
Exhibit F-6.	Food and Drug Administration (FDA) – Digital Health Innovation Action Plan.....	95
Exhibit F-7.	Centers for Disease Control and Prevention (CDC) – Surveillance Strategy	95
Exhibit F-8.	Centers for Medicare & Medicaid Services (CMS) – Interoperability and Patient Access Proposed Rule.....	96
Exhibit F-9.	Office of the National Coordinator of Health IT (ONC) – 21st Century Cures Act: Interoperability, Information Blocking, and the ONC Health IT Certification Program	97
Exhibit F-10.	Office of the Chief Technology Officer (CTO) – The State of Data Sharing at the U.S. Department of Health and Human Services.....	98

List of Tables

Table 1.	Percentage of Awards that Address Each of the Five Functionalities.....	4
Table 2.	Number of Funded Awards by Federal Fiscal Funding Years (2016-2019).....	7
Table 3.	Descriptive Table of Gap Analysis Results	17
Table 4.	Use of the ADAPTABLE Trial to Strengthen Methods to Collect, Validate, and Integrate Patient-reported Information with Electronic Health Record Data Award Products and Description	45
Table 5.	Advancing the Collection and Use of Patient-Reported Outcomes through Health IT Award Products and Description	47
Table 6.	Technologies for Donating Medicare Beneficiary Claims Data to Research Studies Award Products and Description	50
Table 7.	Development of a NLP Web Service for Public Health Use Award Products and Description.....	54
Table 8.	Standardization and Querying of Data Quality Metrics and Characteristics for Electronic Health Data Award Products and Description.....	56
Table 9.	Harmonization of Various Common Data Models and Open Standards for Evidence Generation Award Products and Description	59
Table 10.	Enhancing Data Resources for Studying Patterns and Correlates of Mortality in Patient-Centered Outcomes Research: Project 4 — NDI Workshop and Strategy Paper Award Products and Award Products and Description	62
Table 11.	Enhancing Data Resources for Studying Patterns and Correlates of Mortality in Patient-Centered Outcomes Research: Project 1—Adding Cause-Specific Mortality to NCHS’ National Hospital Care Survey by Linking to the National Death Index Award Products and Description	64

Table 12.	Enhancing Data Resources for Studying Patterns and Correlates of Mortality in Patient-Centered Outcomes Research: Project 2—Pilot Linkage of NDI+ to Commercially and Publicly Insured Populations Award Products and Description	66
Table 13.	Security and Privacy Standards for Patient Matching, Linking and Aggregation (PMAL) Award Products and Description	69
Table 14.	Developing a Strategically Coordinated Registry Network (CRN) for Women’s Health Technologies Award Products and Description	73
Table 15.	PCOR: Privacy and Security Blueprint, Legal Analysis, and Ethics Framework for Data Use, & Use of Technology for Privacy Award Products and Description	77
Table A-1.	Strategic Framework Milestones by Functionality.....	82
Table B-1.	Awards Included in the Impact Report.....	84
Table D-1.	Technical Expert Panel Members	88
Table E-1.	HHS Data Strategy	89
Table G-1.	Functionalities Addressed by Each Case Study.....	99

Report Overview

The Assistant Secretary for Planning and Evaluation (ASPE) coordinates a portfolio of intradepartmental projects that build data capacity for conducting patient-centered outcomes research (PCOR). The portfolio supports the Department of Health and Human Services' (HHS) patient-centered research programs designed to produce new scientific evidence that informs and supports the health care decisions of patients, families, and their health care providers.¹ This report describes the impact of a subset of the portfolio awards funded from Fiscal Year (FY) 2016 through FY 2019 and considers opportunities for continued advancement of a robust data infrastructure for patient-centered research.

Progress toward Data Infrastructure Capacity. ASPE developed the *HHS Strategic Roadmap for Building Data Capacity for Clinical Comparative Effectiveness Research* to inform investments in the portfolio. This roadmap includes a Strategic Framework for the Office of the Secretary Patient-Centered Outcomes Research Trust Fund (OS-PCORTF) portfolio. The framework describes the core research functionalities needed to support the creation, enhancement, and usability of data for patient-centered research, as well as a series of milestones for tracking their progress.

This report is organized in three chapters. Each chapter assesses the impact of the portfolio toward building and expanding data infrastructure capacity for PCOR. First, the contributions of the OS-PCORTF awards to support the Strategic Framework are assessed. Second, cross-cutting themes between HHS data infrastructure strategies and the Strategic Framework are identified to highlight cross-agency synergies and opportunities for the future. Finally, 12 awards that have helped build data infrastructure for patient-centered research while also advancing HHS strategic priorities are described.

Opportunities for Future Work. Overall, results show that progress has been made to build data infrastructure for patient-centered research. Looking forward, there are opportunities to continue the work of the portfolio in the following areas:

- Refine the Strategic Framework to reflect the broader ecosystem of PCOR and comparative effectiveness research (CER) and address the evolving data infrastructure needs across HHS agencies, including non-technical factors (e.g., financial and policy drivers, incentives, governance).
- Identify and prioritize areas of collaboration with HHS agencies that leverage synergies across agencies' data infrastructure projects.
- Identify metrics to facilitate portfolio-wide assessment, award-specific assessment, and dissemination and translation to ensure that award goals are achieved.
- Build on existing dissemination efforts to increase the OS-PCORTF portfolio products and tools uptake more broadly, increasing visibility for the OS-PCORTF both inside and outside of HHS.
- Collaborate with fellow HHS agencies, engaging in cross-agency activities that further support awareness of PCOR, sharing of collective data and knowledge resources, and the building of data infrastructure.
- Expand work to collaborate with non-federal research networks to strengthen the utility and longevity of the data infrastructure that has been built and expand the potential end-user base.
- Focus on sustainability planning, for example, via integration into the award planning process to encourage more proactive identification of end-users who will utilize the products upon project completion.

INTRODUCTION

Since FY 2010, HHS under the OS-PCORTF has funded a portfolio of diverse and large-scale data infrastructure projects. This report describes the impact of a subset of 43 awards funded from FY 2016 through FY 2019. The goals of the report are to:

- Assess the progress of the awards toward building data capacity for patient-centered outcomes research
- Identify products available for dissemination and approaches for increasing adoption and use
- Identify potential areas for future work

The report is organized in three parts. Chapter 1 builds off of a previous formative evaluation that assessed the contributions of the portfolio awards funded between FY 2012 to FY 2016 (the “2017 Evaluation”). Specifically, Chapter 1 assesses the extent to which 43 individual awards funded between FY 2016 and FY 2019, and the resulting products, expand data capacity for patient-centered outcomes research. Chapter 2 identifies cross-cutting themes between HHS data infrastructure strategies and the OS-PCORTF Strategic Framework. Chapter 3 features

12 case studies that have helped to build data infrastructure for PCOR while also advancing HHS strategic priorities. The report concludes with suggestions for future areas of focus to build data capacity. Exhibit 1 presents a visual for how this report furthers the work of the 2017 Evaluation to understand the impact of the portfolio and develop a vision for future work.

Background

HHS agencies routinely collect, link, and analyze data that can be used to generate new scientific knowledge about federal programs and the patient populations these programs serve. These data are foundational to research that expands knowledge about the outcomes and effectiveness of health care treatments and interventions. As a consumer, producer, and regulator of key national health data, HHS is uniquely positioned to coordinate its programs to build national data capacity in support of the mission, statutory authorities, and annual priorities of each HHS agency and the Department as a whole.

Research data is foundational to several major HHS initiatives like using “real-world evidence” to bring new treatments to patients as part of the 21st Century Cures Act (the “Cures Act”), the historic National Institutes of Health (NIH) *All of Us* Research Program that will gather data from 1 million or more Americans to accelerate research and improve health, and the many efforts to use data to address the opioid crisis in the United States.

Coordination Role of ASPE. Since 2011 ASPE’s Office of Health Policy has convened HHS agency leaders to oversee the development and approval of awards to address Department priorities that build data capacity

Exhibit 1. Telling the Story of PCOR in HHS



for patient-centered outcomes research. A major goal of building data capacity is to support HHS research programs that generate scientific evidence that informs decisions about patient health outcomes. The common interest in building data capacity for patient-centered research brings together the expertise of HHS agency leaders, informaticians, technologists, and researchers to identify priorities, share expertise and resources, and collaborate on projects. The activities to build data capacity involve coordination with agency leaders and researchers; development of a Strategic Framework and shared vision for data infrastructure; annual setting of funding priorities to meet strategic goals; implementation of new projects; and periodic performance evaluations.

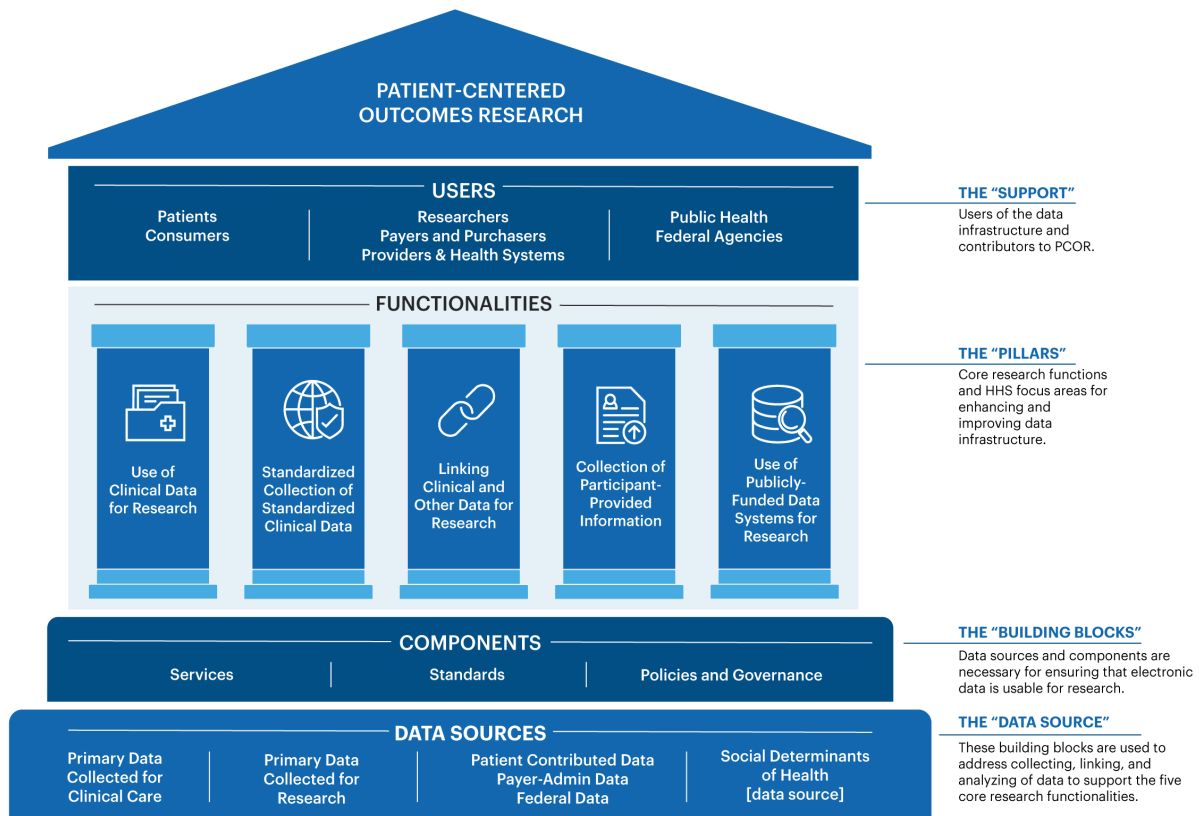
End-Users. The end-users of patient-centered research data are primarily researchers, research networks, and research programs that study patient-centered questions, particularly comparing different interventions. These groups are not mutually exclusive and are often interconnected with one another by

statute, regulation, mission, professional interests, funding, or in other ways. Given the OS-PCORTF statutory charge to “coordinate federal programs to build data capacity ... including the development and use of clinical registries and health outcomes research networks,” the portfolio focuses on the needs of research networks that use data to conduct health outcomes studies that address patient-centered questions or concerns. Together, these groups represent key partners for identifying and carrying out projects that are supported by OS-PCORTF funding.

Strategic Framework

In partnership with the OS-PCORTF Leadership Council,² ASPE developed the *HHS Strategic Roadmap for Building Data Capacity for Clinical Comparative Effectiveness Research*. The council includes representatives from the: Administration of Children and Families, Agency for Healthcare Research and Quality (AHRQ), Assistant Secretary for Preparedness and Response (ASPR), Centers for Disease Control and

Exhibit 2. OS-PCORTF Strategic Framework



Prevention (CDC), Centers for Medicare & Medicaid Services (CMS), Food and Drug Administration (FDA), Health Resources and Services Administration (HRSA), Office of the National Coordinator for Health Information Technology (ONC), NIH, Substance Abuse and Mental Health Services Administration (SAMHSA), and the HHS Chief Technology Officer (CTO). The Roadmap includes a Strategic Framework for OS-PCORTF investments as well as a series of milestones for tracking their progress.

The Strategic Framework is analogous to a three-level structure where each level supports the creation, enhancement, and usability of data for PCOR.³ As shown in Exhibit 2 above, the core components and data sources are the “building blocks” that support the standardized collection, linking, and data analysis for patient-centered and CER. Users and contributors of the data (e.g., consumers, providers, researchers, payers, public health agencies, and federal policymakers) provide additional support for PCOR. These users represent the key partners for implementing OS-PCORTF-funded awards.⁴

The “building blocks,” or the data sources and components, ensure that electronic data is usable for patient-centered research. Data sources represent a wide variety of primary and secondary data sources from across the continuum of health care and include the clinical, environmental, and socio-behavioral factors that impact health outcomes. Three main components guide the collection, linkage, and analysis of data from these data sources, specifically: 1) standards (e.g., specifications for capturing, storing, representing, linking, and exchanging data); 2) services (e.g., resources and tools for capturing, storing, linking, analyzing, and sharing data); and 3) policies and governance structures (e.g., federal rules or guidelines that ensure standards and services are followed and applied consistently).⁵

The “pillars” in the middle of this structure represent the core research “functionalities” that are the key focus areas for the OS-PCORTF portfolio to enable more robust data for end-users. These five functionalities are:

Using Clinical Data for Research – optimizing data for research by improving access, enhancing quality, and promoting interoperability of clinical data across multiple sources.

Standardizing Data Collection – 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.

Linking Data – linking clinical data (e.g., electronic health record—EHR—data, clinical registries) with other data types (e.g., claims data, program data, participant-provided information) in order to track patients across the continuum of care and/or capture a range of health-related outcomes.

Collecting Participant-Provided Information (PPI) – developing and using new standards and technologies to collect PPI so that participants can participate more fully in clinical research.

Using Federal Databases for Research – enhancing federal and state-level data systems to enable greater access, use, linkages, and analysis of publicly funded data for research.

Eighteen milestones have been specified across the five functionalities to serve as benchmarks of progress toward achieving the functionalities. Appendix A displays the full list of milestones aligned with the five functionalities. Together, the functionalities and milestones provide a method for assessing the impacts of the OS-PCORTF portfolio and identifying areas where additional work can be done.

CHAPTER 1

PORTFOLIO ASSESSMENT

Chapter Executive Summary

The goals of the OS-PCORTF portfolio are to build data capacity and data infrastructure for patient-centered research. The portfolio assessment consisted of three distinct analyses: 1) an assessment of the extent to which these awards are advancing the Strategic Framework functionalities; 2) a gap analysis to determine how well the portfolio addresses the gap areas identified by the 2017 Evaluation; and 3) a review of the portfolio’s contribution to emerging HHS policy priorities.

Extent to Which The Portfolio Advances The Strategic Framework

- Almost 75 percent of awards focus on optimizing the use clinical data for research. This includes developing tools and services that support researchers in accessing electronic clinical data across multiple sources such as EHRs, claims, registries, and patient portals; and improving the quality of these data. For example, a CDC award is making clinical data more accessible for research and public health surveillance by developing an application that will leverage existing health data and exchange standards to support real-time data exchange between EHRs and public health systems.
- Just over 50 percent of the awards are working to enhance the use of federal datasets for research by enabling greater access to the data, supporting linkages of the data, or facilitating the querying and analysis of the data. An example includes the awards to AHRQ and ASPR, who are collaborating to build a data platform to analyze an expanded Healthcare Cost and Utilization Project (HCUP) dataset to inform emergency preparedness and response. This platform will support research queries that assess the effectiveness of different disaster response and recovery interventions.
- Roughly 50 percent of the awards focused on standardizing data collection. This included defining CDEs so that data can be collected, linked, and analyzed across multiple data sources. For example, AHRQ and the NIH’s National Institute of Diabetes and Digestive and Kidney Diseases are developing an electronic care plan to collect data from across the continuum of care using new and expanded data elements to capture information on patients with multiple chronic conditions for use in pragmatic trials.

Table 1. Percentage of Awards that Address Each of the Five Functionalities

Functionality	Percentage of Awards that Address the Functionality
Use of Clinical Data	72%
Use of Federal Datasets	53%
Standardized Data Collection	49%
Use of PPI	49%
Linking Data	28%

- Likewise, nearly 50 percent of awards are developing tools to collect PPI to support patient participation in clinical research. For example, an NIH award is developing methods to validate and integrate patient-reported health data with EHR data for use in a pragmatic clinical trial that compares the effectiveness of two different doses of aspirin.
- Only 25 percent of awards focused on linking data; however, these awards are iteratively enhancing federal datasets that are highly valued by health care researchers by linking them together to support expanded research inquiries. For example, a CDC project is developing new algorithms that use linked National Hospital Care Survey (NHCS) data and National Death Index (NDI) data to improve identification of opioid-related deaths.
- Table 1 consolidates the statistics presented in the bullets above, showing the percentage of awards that address each Strategic Framework functionality.

Extent to Which The Portfolio Addresses Gaps Identified by The 2017 Evaluation

- The most frequently addressed gap area was disseminating research findings. Nearly every award described activities to publicly disseminate OS-PCORTF products to end-users, such as posting software code to open source platforms like GitHub.
- The portfolio also made progress toward closing gaps around implementing standards and improving data quality. For example, multiple projects are improving data quality and enhancing interoperability by developing, piloting, and refining implementation guides through industry standards developing organizations like Health Level Seven (HL7). Advancing the development of standards efforts such as Fast Healthcare Interoperability Resources (FHIR®) is critical to data standardization and improving data access and exchange for clinical and research purposes.

Extent to Which The Portfolio Has Evolved to Meet The Emerging Health Policy Landscape

- The work of the portfolio was more focused in the early years on data linkages, increased access to federal datasets, and improvements in clinical registries and research networks. For example, several awards are working to enhance the utility of federal datasets such as the NDI and Sentinel. Several projects are supporting the interoperability objectives of the Cures Act to enable the secure exchange of EHI without special effort by pilot testing application programming interfaces (APIs).
- The portfolio has since expanded to advance next-generation interoperability of health information, improving patient access to health information, and accelerating growth in advanced data science techniques. For example, one FY 2019 award is creating research-quality synthetic datasets to complement the use of real-world data and allow for more rigorous analysis.

Technical Expert Panel Feedback on Future Direction of the Portfolio

The TEP provided feedback in three areas: 1) potential areas of refinement to the Strategic Framework; 2) PCOR data infrastructure needs that should be prioritized in future portfolio activities and 3) metrics to assess portfolio impact going forward.

Areas of refinement to the Strategic Framework that were suggested included:

- Integrating external factors which can influence the data sources, data infrastructure, and types of patient-centered research inquiries that are carried out such as financial and policy drivers that influence provider documentation and the social determinants that impact health outcomes.

- Identifying and incorporating cross-cutting barriers which can impede progress in patient-centered research.
- Emphasizing the role of data provenance as a key component that influences data usability for research.
- Establishing and consulting with working groups for each of the five functionalities who can advise on relevant federal initiatives that could be leveraged by the portfolio, metrics for assessing progress of the functionality, and the evolving needs related to that functionality.

Priorities identified for future work include:

- Addressing the non-technical barriers that impede the use and sharing of data for research including data governance and privacy and security protections.
- Providing targeted support and guidance to projects teams with products that are ready for translation of their work for broader uptake.
- Engaging end-users during the planning phases of project awards to identify areas of greatest need, support translation of products, and inform sustainability efforts.

The TEP identified three categories of metrics for tracking progress and impact of the portfolio:

- Portfolio-wide metrics including metrics that track and quantify artifacts that are used by other project awards and metrics derived from a more prescriptive strategic roadmap that can assess progress along the roadmap;
- Award-specific metrics to assess whether each individual award achieved its articulated objectives;
- Dissemination and translation metrics that track the number and type of dissemination products, use of website analytics (e.g., measuring website traffic and counting the number of downloads of posted material), and tracking the ways in which other research initiatives leverage award outputs.

Introduction

NORC conducted a three-pronged analysis of the contributions of 43 OS-PCORTF awards to collect, link, and analyze data to enhance and expand data capacity for PCOR. Specifically, the areas of analyses include:

- Area 1: Describe how the awards address the key milestones of the five Strategic Framework research functionalities.
- Area 2: Highlight how the awards have addressed gaps in building data capacity identified by the 2017 Evaluation.
- Area 3: Describe how the awards align with current and emerging HHS priorities.

This chapter concludes with a discussion of remaining gaps and suggestions for areas of future focus that are relevant to building capacity for PCOR.

Methods

The portfolio assessment focuses on 43 awards funded through the OS-PCORTF between FY 2016 and 2019 (Table 2). These awards represent a subset of the total 69 awards that have been funded since FY 2010. ASPE has awarded several projects with multi-agency awards. This assessment focuses on the individual agency awards. The titles of all of the awards included in this report are in Appendix B.

While this portfolio assessment picks up where the 2017 Evaluation left off, the methods of analyses are not meant to replicate those of the 2017 Evaluation. Rather, this assessment uses award-monitoring data to summarize progress since the last evaluation.

Table 2. Number of Funded Awards by Federal Fiscal Funding Years (2016-2019)

Federal Fiscal Funding Year	Number of Funded Awards
FY 2019	12
FY 2018	7
FY 2017	16
FY 2016	8
Total	43

The methods for the portfolio assessment focused on measuring progress in three areas: 1) how the awards addressed the Strategic Framework; 2) how the awards addressed the gaps identified in the 2017 Evaluation; and 3) how the portfolio has addressed emerging health priorities over time. The specific analytic approaches used for each area are described below.

Area 1 Analytic Approach. To quantitatively assess progress on the Strategic Framework, there are two sources of data: 1) self-reported data from quarterly progress reports submitted by active FY 2016 through FY 2018 awards (N=31); and 2) statements of works for FY 2019 awards that had not yet submitted their first quarterly progress reports (N=12).

As part of their quarterly progress reporting activities submitted to ASPE, awards funded between FY 2016 and FY 2018 were presented the list of the 18 functionality milestones (as shown in Appendix A) and were asked to select those milestones that they felt their award activities most closely aligned with. These data were analyzed to determine how awards aligned with each of the five functionalities and associated milestones.

For the 12 FY 2019 awards that did not yet have self-reported quarterly progress reports available, two researchers independently reviewed award descriptions from the statements of work and then coded this information to milestones under each Strategic Framework functionality. The coded milestone data for the FY 2019 awards was combined with the self-reported milestone data from the quarterly progress reports (for the FY 2016–FY 2018

awards) to produce a comprehensive accounting of milestones achieved across the 43 awards.

Area 2 Analytic Approach. The 2017 Evaluation evaluated 20 OS-PCORTF awards that were active or completed between FY 2012 and FY 2016. The goal of this evaluation was to assess the progress of these awards in building the data capacity for PCOR and identify areas in which ASPE could focus future portfolio work.⁶ This evaluation identified five gap areas to further advancing data infrastructure capacity: 1) implementing standards; 2) enhancing data governance; 3) improving data quality; 4) balancing access with enhancing privacy and security; and 5) disseminating research findings.

To assess progress toward each of these five gap areas, the research team reviewed the qualitative data from interviews with stakeholders, which was used to identify the gap areas in the 2017 Evaluation. These results describe stakeholders' view on the comprehensiveness of the Strategic Framework to address gaps in data capacity for patient-centered research, how award products contribute to the functionalities, and how awards contribute to and address end-users' research needs. The research team used this information to more specifically define the type of awards that could address each gap area. For example, the team incorporated stakeholder recommendations around improving data completeness and determining fitness-of-use of clinical data for research purposes to identify the type of awards that could contribute toward the gap area of improving data quality. Sample classification criteria for each gap area are provided below (see Appendix C for a complete list of the criteria):

1. **Implementing Standards** – Awards that develop and/or pilot data content and exchange standards and implementation guides.
2. **Enhancing Data Governance** – Awards that develop data access policies or frameworks; awards that develop and/or pilot data provenance standards.
3. **Improving Data Quality** – Awards that improve data completeness, validity, and reliability; awards that develop and/or pilot approaches to analyze unstructured data; awards that develop data linkages.

4. **Balancing Access with Enhancing Privacy and Security** – Awards that develop and/or pilot security, access, and data donation standards.
5. **Disseminating Research Findings** – Awards with plans to publish results; awards that support researcher education.

Two researchers independently coded the 43 awards using the classification criteria to assess the extent to which these awards addressed the gap areas.

Area 3 Analytic Approach. In recent years, HHS has responded to emerging policy priorities including value-based care, the opioid crisis, real-world evidence generation, and patient access. OS-PCORTF data infrastructure awards have supported HHS efforts by building data capacity across these areas.

In addition to assessing the impact of the portfolio and how the portfolio addressed the gaps raised in the 2017 evaluation, the team also reviewed how the portfolio has supported the Department’s current and emerging priorities. The team reviewed portfolio program and strategic planning documents to identify key legislative and regulatory activities related to PCOR. Specifically, the portfolio and strategic planning documents included: 1) ASPE’s statutory charge⁷ to coordinate and invest OS-PCORTF funds to build data capacity for CER; 2) the 2017 Evaluation; 3) the 2017 Annual Portfolio;⁸ 4) the 2018 Planning Guide for OS-PCORTF Portfolio Participants;⁹ and 5) HHS Secretary Alex Azar’s priorities for HHS.¹⁰

The following Acts are included in this analysis:

- The Medicare Access & CHIP Reauthorization Act of 2015 (MACRA)
- The Foundation of Evidence-Based Policymaking Act of 2018
- The 21st Century Cures Act
- The Substance Use-Disorder Prevention that Promotes Opioid Recovery and Treatment (SUPPORT) for Patients and Communities Act

For each piece of legislation, the team summarized potential data infrastructure needs that could be addressed, in part, by the portfolio, identifying ten such needs. Three senior staff members then independently reviewed the 43 award descriptions

and coded the high-level data infrastructure needs being addressed by each.

Technical Expert Panel. ASPE convened a TEP of seven experts (see Appendix D) to review the analyses presented in this chapter. The intent of the TEP was to obtain external stakeholder viewpoints from individual attendees as opposed to collective advice or recommendations resulting from consensus building activities. The TEP considered three questions:

1. What are potential refinements or areas of revision to the Strategic Framework, particularly to the five articulated PCOR data infrastructure functionalities?
2. What metrics are needed to understand the impact of the portfolio to building data capacity for PCOR? How can the Strategic Framework inform these metrics?
3. Based on the results of this assessment, are there other areas or gaps ASPE should address or prioritize to guide the work of the portfolio moving forward?

Individual TEP inputs were gathered over the course of three meetings (two virtual and one in-person) and are incorporated throughout this report.

Limitations. There are several limitations to this assessment:

1. The analysis only includes a subset of the awards funded through the OS-PCORTF portfolio.
2. The analysis is based on self-reported quarterly progress report data. Therefore, the analysis relies on the judgment of the award leads in identifying which milestones their award addresses.
3. The FY 2019 awards are in their beginning stages and do not yet have self-reported quarterly progress reports submitted by award leads. Therefore, reviewers used statements of work to assess potential contributions. The authors recognize that as the award work progresses, there might be changes to the original scope; however, this analysis does not account for those potential changes.

Findings

How have the awards addressed the Strategic Framework?

This section provides an overview of each Strategic Framework functionality and its relevance to building data infrastructure for PCOR, as well as descriptions of the specific milestones that indicate progress toward achieving that functionality. Findings from the quantitative analysis of award data are included under each functionality along with an illustrative case example of an award that demonstrates achievement toward the functionality.



FUNCTIONALITY 1: USE OF CLINICAL DATA FOR RESEARCH

Background

The functionality of Use of Clinical Data for Research refers to enabling researchers to utilize and analyze clinical data that are routinely collected for care delivery or stored in clinical registries. These data, while rich in clinical information, are often siloed, making it challenging for researchers to utilize. Promoting the development and use of standards, policies, services, and analytic tools can allow researchers to access, aggregate, and analyze data from multiple sources in innovative ways.

This functionality also encompasses the infrastructure to ensure that clinical data are of high quality and fit-for-use in research. This is accomplished through the development of standards, policies, and services to assess the data's completeness, comprehensiveness, and representativeness to the population being studied.

Progress Milestones

The Strategic Framework identified six milestones to measure progress toward achieving the functionality of using clinical data for research:

1. **Establish services and tools that can be leveraged nationally to support data access, querying, and use, including privacy-preserving analytics and queries.** The government is in a unique position to support the development of open-source tools and services to facilitate researchers' ability to access, use, and query data.
2. **Develop support services and tools that can be leveraged nationally and are not likely to be developed by the private sector; these tools would test the quality of unstructured and structured data to answer PCOR questions.** Once standards and policies are available to support high quality data for research, researchers will need services and tools to test adherence to these quality standards.
3. **Develop standards that support secure, electronic query of structured data across clinical research and delivery systems, including standards for open-source access.** Standards that support queries across EHRs or other clinical information systems can help researchers overcome technical barriers to accessing clinical information.
4. **Develop and test metadata standards that describe data quality.** Metadata standards establish a common understanding for assessing whether data meets quality and fitness for use for specific research purposes.
5. **Create a policy framework for privacy-preserving access and querying of clinical data by researchers conducting patient-centered research, and policies that govern the use of the services that support data access, querying, and use.** Clear policies ensure that data are accessed and used without compromising patient privacy and data security and in accordance with current rules and regulations that govern data use.
6. **Develop a policy framework for ensuring clinical data used for research is of "research grade."** A policy framework lays out the necessary procedures and processes for ensuring that that data are representative, accurate, timely, and appropriate for use in research.

Assessment of Progress

As shown in Exhibit 3, the two most commonly addressed milestones relate to the development of services and tools—with 60 percent of the 43 awards indicating that they support services and tools for data access and 42 percent of awards indicating that they developed services or tools to test the quality of unstructured and structured data for patient-centered research. For example, the FDA’s Standardization and Querying of Data Quality Metrics award developed a web-based platform of data quality metrics to evaluate the fitness of data across data sources. The data quality metrics describe data characteristics using common terms despite how the data are defined locally by a research network. Researchers can use these metrics to find the right data from among diverse research networks (e.g., Sentinel, the National Patient-Centered Clinical Research Network (PCORnet), Medical Device Epidemiology Network (MDEpiNet), etc.) to answer their study questions.

The milestones least commonly addressed relate to developing policy frameworks and/or policies for privacy-preserving access and querying of data and ensuring the quality of clinical data is of “research grade.” However, there is some progress being made. An example of an award that addresses the latter milestone is the cross-agency project “Developing a Strategically Coordinated Registry Network (CRN),” which developed a formal partnership agreement for CRN participation as well as a data governance model

Exhibit 3. Progress toward Using Clinical Data for Research



Note: Percentages do not sum to 100% as awards can address more than one milestone or none of the milestones.

for how data will be shared across the network. Enhancing the interoperability between networks helps to ensure that clinical outcome data collected by individual participating registries can be more easily aggregated in meaningful ways and used for CER studies evaluating outcomes across registry domains.



FUNCTIONALITY 2: STANDARDIZED COLLECTION OF STANDARDIZED CLINICAL DATA

Background

Health data collected as part of a clinical encounter have the potential to support clinical quality improvement and large-scale clinical research studies. However, differences in clinical data definitions and terms across health information technology (health IT) systems and the resulting variability can challenge the meaningful interpretation of study results and use of

the results to improve patient outcomes. In order to support comparability and analysis across data sources, researchers need standard definitions of the data.

Implementing CDEs across data sources supports standardization and improves data quality for linking data across multiple sources and studies (e.g., patient registries and clinical research data with EHR data).¹¹

NIH defines a CDE as “a data element that is common to multiple datasets across different studies.”¹² NIH and FDA have conducted a significant amount of work to develop CDEs for research purposes, such as PCORnet, MDEpiNet, and OMOP. The aim of this functionality is to build upon these ongoing initiatives to increase the use of CDEs by identifying clinical areas in which CDEs are lacking and facilitating access to existing CDEs to encourage uptake. Relatedly, common vocabulary standards (e.g., Logical Observation Identifiers Names and Codes (LOINC[®])) facilitate interoperability across disparate data sources and systems and are necessary to interpret and exchange information with the same clinical meaning.

Progress Milestones

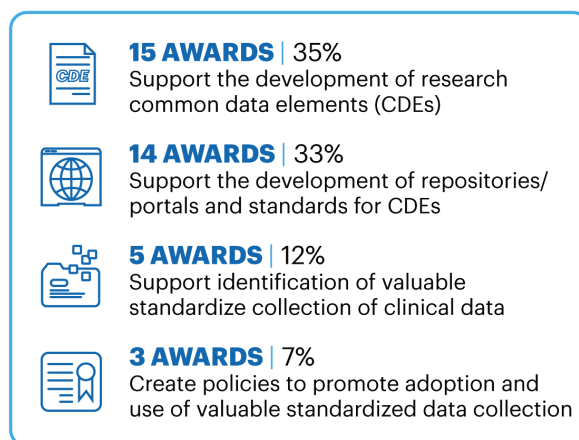
The OS-PCORTF strategic roadmap seeks to increase the development, availability, and use of CDEs through four specific milestones:

1. **Develop research CDEs in specific gap areas and support development of a governance structure for CDE harmonization.** While new CDEs are needed to fill existing gaps, there is also a need to develop and encourage consensus-building around existing CDEs.
2. **Develop repositories of CDEs and standards and services for CDE utilization.** Creating publicly available repositories can support more widespread adoption of CDEs. For example, the NIH CDE Resource Portal includes collections of available CDEs, guidance for selecting CDEs, and notably, the [NIH CDE Repository](#).
3. **Support research and/or crowd-sourced methods to determine which of the standardized collection components and services are most valuable.** Additional research or evaluation work is needed to assess which CDEs are being adopted and used, for what purposes, and where gaps need to be addressed.
4. **Create policies to promote the adoption and use of standard CDE services.** Specific policy efforts, such as NIH’s funding and program guidance¹³ can encourage and accelerate the use of CDEs.

Assessment of Progress

As shown in Exhibit 4, thirty-five percent of awards supported the development of CDEs and development of governance structures for CDE harmonization. An example of this work includes the National Institute on Drug Abuse (NIDA) award to enhance emergency medicine opioid data infrastructure, which is identifying CDEs for opioid use disorder (OUD) for implementation in emergency departments (ED). Increasing CDE use in the ED builds data capacity for studying the opioid epidemic and supports researchers’ ability to answer questions such as how many ED providers provide naloxone or administer buprenorphine for OUD.

Exhibit 4. Progress toward Standardized Data Collection



Note: Percentages do not sum to 100% as awards can address more than one milestone or none of the milestones.

One-third of awards supported the development of CDE repositories and standards and services to increase CDE use. NIH’s Creation of LOINC Equivalence Classes developed and published a hierarchy, or equivalence classes, of clinically relevant LOINC terms to support standardized data aggregation and retrieval. These “equivalencing” classes address interoperability barriers created when clinical and laboratory terms are mapped to LOINC terms with different levels of granularity. For example, the LOINC group file contains 36 parent groups, containing 5,650 groups that organize 24,075 unique LOINC terms. The equivalence groups are available in the current LOINC release 2.64.

Finally, five awards (or 12 percent) support research to identify which CDEs and CDE services are most valuable, and three awards (or 7 percent) are developing policies to promote the use of CDEs.



FUNCTIONALITY 3: LINKING CLINICAL AND OTHER DATA FOR RESEARCH

Background

Electronic clinical data is available through a variety of sources, including EHR data and patient registries (e.g., specific diseases, medical products, interventions and procedures). These data contain a wealth of information on diagnoses, procedures, vital statistics, medications, laboratory test results, and other important areas that can be leveraged for patient-centered research.

The utility of these data sources can be enhanced further by linking clinical data to other datasets, such as claims, PPI data (e.g., survey datasets, social determinants of health screenings), provider files (e.g., organization or physician characteristics), community program data, and environmental data. Specifically, linking data—combining data from two or more data sources by matching unique individual identifiers—can allow researchers to follow patients longitudinally and across health care settings. This also enables researchers to study the impact of various, multi-pronged interventions on patient-centered outcomes.

Progress Milestones

The OS-PCORTF strategic roadmap identified three specific milestones to measure progress toward achieving the functionality of linking clinical data to other data for research:

- **Leveraging existing standards as well as supporting the development and testing of new standards around linking patient data.** Linking data across different sources where patient treatments and outcomes are represented in a standard way can enable more patient-centered analysis.
- **Creating a policy framework to facilitate linkages across patient data sources.** Policy frameworks provide guiding principles and can help to ensure consistency and standardization in the linkage process while also safeguarding patient privacy and data security.

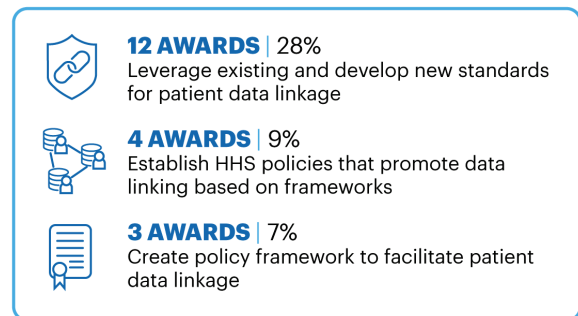
- **Establishing HHS policies to promote data linking based on this policy framework.**

Extending policy frameworks into concrete, actionable policies to promote data linkages across sources and sectors—for example, establishing policies that foster partnerships or collaborations with the private sector—can enable many types of research.

Assessment of Progress

As shown in Exhibit 5, 12 out of the 43 awards (or 28 percent) indicated that their work supports leveraging existing standards and supports the development and testing of data linkage standards. The CDC award to create a surveillance network for maternal, infant, and child health outcomes will result in a data platform to collect linked maternal and infant data among women treated for OUD during pregnancy. The linked data will improve researchers' understanding of maternal, infant, and child outcomes of medically supervised withdrawal during pregnancy, which can ultimately inform and improve clinical decision-making.

Exhibit 5. Progress toward Linking Data



Note: Percentages do not sum to 100% as awards can address more than one milestone or none of the milestones.

Three awards (or 7 percent) are working to create a policy framework to facilitate patient data linkage (see Exhibit 6 for examples of data sets being linked), and four awards (or 9 percent) are working to establish HHS policies around data linkage based on this framework. The CMS and NIH cross-agency project Technologies for Donating Medicare Beneficiary Claims Data leverages Sync for Science and the Blue Button 2.0 API privacy and security policies and standards to facilitate individuals' ability to donate their claims data to research studies. This work will support data donation to the *All of Us* Research Program, increasing the data available for PCOR.

Exhibit 6. Linked Data Files Produced by the OS-PCORTF Portfolio

- 2014 NHCS and NDI linked with the 2014 National Vital Statistics System restricted use mortality files on drug overdose deaths (NVSS-M-DO) file
- 2014 NHCS inpatient and ED claims data linked to the 2014 and 2015 NDI
- 2014 NHCS inpatient and ED claims data linked to 2014-2015 CMS Master Beneficiary Summary File
- 2016 NHCS inpatient and ED claims, and EHR data linked to the 2016 and 2017 NDI



FUNCTIONALITY 4: COLLECTION OF PARTICIPANT-PROVIDED INFORMATION

Background

There is increasing recognition of the importance of PPI in supporting research on patient-centered care as PPI provides an essential perspective into patients' symptoms and experiences that cannot necessarily be captured in the regular course of clinical care.¹⁴ Likewise, PPI can inform multiple aspects of the care process—from prevention and diagnosis to treatment and long-term care.

The types of data that are considered PPI have evolved over the years and now includes patient-generated health data (PGHD) (e.g., data from wearable devices), patient-reported outcomes (PRO) data (e.g., pain, functionality), and patient-reported health data (PRH) (e.g., hospitalization events). These data can also inform researchers' interpretation of clinical trial results as well as support effective drug and device surveillance.

Progress Milestones

The OS-PCORTF strategic roadmap seeks to address barriers to the collection and use of PPI through three milestones. Each of these milestones measure progress toward achieving the collection and use of PPI:

1. **Develop tools and services to facilitate PPI data collection, exchange, and donation for patient-centered research.** Pilot projects that demonstrate the use of FHIR APIs and patient-facing apps to collect and share PPI with clinicians contribute valuable insights to the application of novel PPI data collection tools and services.
2. **Develop standards for PPI data by leveraging existing standard vocabularies.** The capacity of EHRs to capture PPI in a standardized form is currently limited. Efforts to standardize the collection of PPIs by mapping data to existing standards vocabularies increase the ability to integrate and analyze the data.

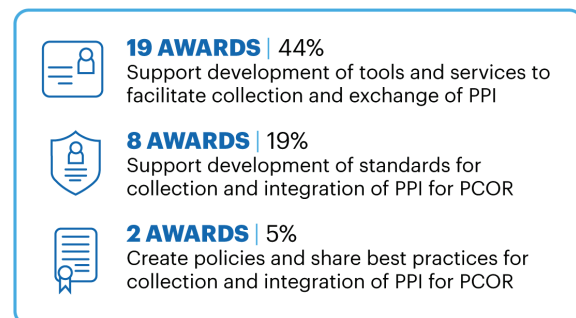
3. **Create policies and best practices for PPI data collection and use.** Guidance is needed to address current limitations of PPI data collection, including: 1) privacy and security concerns, 2) the volume of data and potential burden on providers, and 3) validity of the data and fitness of use for clinical or research purposes.¹⁵

Assessment of Progress

As shown in Exhibit 7, 19 out of 43 awards (44 percent) indicated that they are addressing the first milestone related to developing tools and services and nearly a fifth (or 19 percent) contribute to standards development. The NIH award to strengthen methods to collect, validate, and integrate PRH with EHR data is developing tools and standards for PPI use in patient-centered research. The work includes the development of standards and a query tool to assess the completeness, concordance, and fitness-for use of patient-reported data in EHR data collected for

the ADAPTABLE Trial. They have also developed LOINC patient-reported data elements. Two awards (or 5 percent) are working to create policies or best practice guidance for PPI data collection and use.

Exhibit 7. Progress toward the Collection of PPI



Note: Percentages do not sum to 100% as awards can address more than one milestone or none of the milestones.



FUNCTIONALITY 5: USE OF ENHANCED PUBLICLY FUNDED DATA SYSTEMS FOR RESEARCH

Background

HHS collects and monitors data across its multiple departments to administer, monitor, and evaluate programs and to inform policymaking. However, datasets and the policies that govern them, collectively referred to as a data system, are not always optimized to support their use for evidence-building. This functionality addresses the need to enhance these federal datasets for research by facilitating the retrieval, linkage, aggregation, and use of this data.

The OS-PCORTF strategic roadmap identified two milestones to measure progress toward achieving the functionality of use of enhanced publicly funded data systems for research:

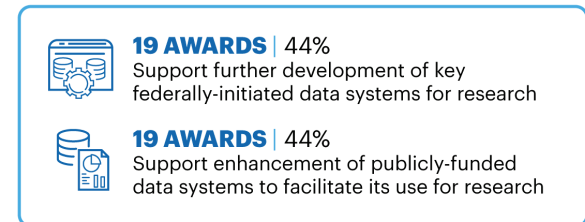
1. **Support the further development of key federally initiated data systems for research.** This includes policies to facilitate the use of administrative data, as well as the need for a practical implementation guide for federal security standards.
2. **Support the enhancement of strategic publicly funded data systems to facilitate their access and use, and ease retrieval of data for research purposes.** Researchers use publicly funded data systems (such as CMS data and the NDI) to assess key patient-centered outcomes, including hospitalizations, ED visits, and mortality. However, these and other federal data systems need to be enhanced so that researchers can more easily access and use the data.

Assessment of Progress

As shown in Exhibit 8, nearly half (44 percent) of the 43 awards indicated that their work is supporting the enhancement of a publicly-funded or federally initiated data system for research. Two more recent awards highlight work to enhance federal data systems to support emerging PCOR needs. An AHRQ and ASPR award aimed at assessing and predicting medical needs in a disaster expands the HCUP database to include new quarterly ED and in-patient data to study medically-related disaster response and recovery interventions. Another AHRQ award will develop a data platform for a valid, reliable, and standardized set of social determinants of health data sources across geographic areas from existing HHS federal datasets that currently lack standardized

metrics or estimates at the small-area level. These new federal data assets will reduce the resources researchers currently spend linking multiple datasets to create data files suitable for analysis.

Exhibit 8. Progress Toward Using Federal Datasets for Research



Note: Percentages do not sum to 100% as awards can address more than one milestone or none of the milestones.

Summary of Progress toward Achieving Strategic Framework Functionalities

While the work conducted by the FY 2016 to FY 2019 portfolio has spanned across all five functionalities, the highest volume of activity has focused on using clinical data for research. This functionality is wide-ranging, addressing many ways of facilitating the use of electronic clinical data for research from supporting data access to ensuring data quality and fitness-for-use. The portfolio is meeting these milestones in a variety of ways, from developing mapping tools that can allow researchers to query data across research networks that are utilizing distinct common data

models (CDMs), to developing a data governance model that may help registries who are searching for more efficient and sustainable methods of data collection.

Roughly half of the awards address the functionalities Standardized Collection of Standardized Clinical Data, Enhanced Publicly-Funded Data Systems for Research, and Collection of PPI. While the functionality related to Linking Clinical and other Data for Research is being addressed by just 12 awards, these awards are producing linked data files to support an expanded set of PCOR inquiries.

How Has the Portfolio Addressed Gap Areas Identified in the 2017 Evaluation?

This gap analysis assesses the extent to which the OS-PCORTF portfolio addresses gap areas identified by the 2017 Evaluation. The section below is organized in three parts: 1) an overview of the gap areas identified by the 2017 Evaluation; 2) findings from the gap analysis; and 3) examples of award products that demonstrate how the portfolio addresses the 2017 Evaluation gap areas.

Background

Overall, the 2017 Evaluation found that the FY 2012 to FY 2015 awards had made progress toward achieving each of the five Strategic Framework functionalities, with the most gains in the Standardized Collection of Standardized Clinical Data functionality. Additionally, the 2017 Evaluation identified five strategic areas in which ASPE could further progress toward building federal data capacity efforts. Specifically, it identified five strategic gap areas:¹⁶

- **Implementing Standards** – Develop best practices to develop, implement, and maintain data standards so that health care and research institutions can reduce the time and costs incurred when implementing and updating standards.
- **Enhancing Data Governance** – Additional effort is needed to address ongoing barriers to increased data capacity. Although this issue remains challenging, it is critical to the efficient use of the research-oriented data infrastructure across individual and organizations' boundaries of control and ownership.
- **Improving Data Quality** – Promote focus on data quality and increase the quantity and accessibility of electronic health data to improve the efficiency and effectiveness of PCOR; also support core functions and improvements in data interoperability.

- **Balancing Access with Enhancing Privacy and Security** – Spur strategies that enhance privacy and security and inform how research and health care entities can better balance data access with security. Strategies include employing innovative technologies that offer researchers access to data, securely and privately, as well as educating the public about the benefits of making available their anonymous health care data.
- **Disseminating Research Findings** – Improve mechanisms for dissemination of OS-PCORTF-sponsored research so that end-users can better gauge federal efforts to build data capacity.

Assessment of Progress

Informed by the 2017 Evaluation, ASPE, in coordination with the OS-PCORTF Leadership Council, has pursued new portfolio work to address the gap areas described above. Since 2017, the portfolio has grown significantly. At the portfolio-level, the results of the gap analysis indicate that progress has been made to address these five strategic gap areas. Table 3 provides a detailed overview of results from the gap analysis.

Nearly all (N=37) of the 43 awards have objectives describing the dissemination of products. Dissemination activities included posting code to open-source platforms like GitHub, publishing implementation guides and data standards, posting pilot study results and lessons learned to agency websites, and submitting study findings to peer-reviewed journals.

Table 3. Descriptive Table of Gap Analysis Results

Strategic Gap Areas	Brief Description of Gap Analysis Criteria	Number of Awards
Disseminating Research Findings	Awards with plans to publish results; awards that support researcher education	37
Implementing Standards	Awards that develop and/or pilot data content and exchange standards and implementation guides	27
Improving Data Quality	Awards that improve data completeness, validity, and reliability; awards that develop and/or pilot approaches to analyzing unstructured data; awards that develop data linkages	22
Balancing Access with Enhancing Privacy and Security	Awards that develop and/or pilot security, access, and data donation standards	6
Enhancing Data Governance	Awards that develop data access policies or frameworks; awards that develop and/or pilot data provenance standards	5

Roughly 60 percent of awards addressed implementing standards and over 50 percent focused on improving data quality. Work in these areas is critical to facilitating increased use of existing federal data assets and the use of both clinical data and new data types (e.g., PPI) for CER and PCOR. For example, NIDA will establish a new research network and patient registry designed to collect outcomes data on medication-assisted treatment for OUD. This award includes feasibility and validity testing of OUD CDEs to enhance the collection of standardized addiction-related data for patient-centered research on opioid treatment.

A lower percentage of awards have addressed the strategic gap area of enhancing data governance (17 percent of awards) and balancing access with enhanced privacy and security (12 percent of awards). However, the products produced by these awards increased the data accessible for research purposes. For example, a series of CMS awards focused on improving Medicare beneficiaries’ ability to access and donate their claims data have matured from pilots to tools and services that are publicly available and in use today (i.e., the CMS Blue Button 2.0 API). The collaboration between CMS and NIH will further enhance the Blue Button service using the Sync for

Science FHIR specification to enable Medicare beneficiaries to donate their claims data to the *All of Us* Research Program.

To further demonstrate the impact of the portfolio, Exhibit 9 provides examples of five different awards and describes how the award products address the gap areas. To highlight products that are completed or nearing completion, these examples were selected from among the awards that met criteria for that gap area and that completed activities in FY 2019. In the table below, examples are grouped by gap area and consist of an overview of the award along with a description of the award products that specifies how the gap is addressed.

Exhibit 9. Awards and Products that Address the Strategic Gap Areas Identified from 2017 Evaluation



Disseminating Research Findings: Across the portfolio the most gains were made in this gap area through efforts to publish and make publicly available tools, services, and other products to end-users. .

Award: Harmonization of Clinical Data Element Definitions for Outcome Measures in Registries (AHRQ)

The goal of the award was to develop a consensus set of clinical data element definitions for outcome measures for use in AHRQ's Registry of Patient Registries. Harmonized clinical outcome definitions are necessary so that outcomes can be meaningfully compared and aggregated between and among registries, clinical research, and quality reporting. Harmonized clinical outcomes enhance the utility of outcome data for analytic purposes and value-based care programs.

Products: AHRQ led the development of five standardized libraries of clinical data element definitions and value sets for five clinical topic areas: [atrial fibrillation](#), lumbar spondylolisthesis (lower back pain), lung cancer, [asthma](#), and [depression](#). AHRQ has continued to work with several professional societies to seek inclusion of CDEs and value sets in NLM's Value Set Authority Center and the NIH CDE Repository. AHRQ also published a [white paper](#) that describes their approach to data element development, harmonization, and lessons learned that can inform development of data element definitions in other clinical topic areas.



Implementing Standards: The portfolio made significant progress in addressing this gap area by engaging in work related to the development of standardized content, exchange, and terminology standards and accompanying implementation guidance.

Award: OneSource: Source Data Capture from EHRs: Using Standardized Clinical Research Data (FDA)

The goal of the OneSource award was to create a standard and reproducible approach to using clinical data from EHRs at the point of care to populate electronic data capture (EDC) systems for use in FDA-regulated clinical trials. OneSource was piloted in ongoing I-SPY 2 breast cancer clinical trials. OneSource addresses inefficiencies in the research process by creating interoperability between healthcare and clinical research systems that were previously siloed. OneSource also enhances the data infrastructure needed for Real World Evidence (RWE) and increasing the clinical trials efficiency by improving the data quality through elimination of duplicate data entry and multiple data transformations, which supports better tracking of clinical study outcomes and enhancing clinical decision-making.

Products: OneSource Clinical Checklists are tools that, when available, will seamlessly integrate EHR data into EDC systems using consensus-based FHIR standards. OneSource is a real-world demonstration of the FDA's guidance Electronic Source Data Capture in Clinical Investigation¹⁷ and Use of Electronic Health Record Data in Clinical Investigations¹⁸ and provides a framework for future implementations in patient-centered clinical research and trials.



Improving Data Quality: Across the portfolio, considerable progress was made to increase the quantity and accessibility of electronic health data by engaging in work to improve data completeness, assess fitness-of-use of clinical data for PCOR, and increase analysis-ready data.

Award: Development of a Natural Language Processing Web Service for Public Health Use (CDC/FDA)

The goal of this award was to develop a publicly available web service to support conversion of unstructured clinical information to structured data (i.e., ICD-10-CM, LOINC, SNOMED CT, and MedDRA). The Natural Language Processing (NLP) web service, or Clinical Language Engineering Workbench (CLEW), was piloted in the FDA's Vaccine Adverse Event Reporting Systems (VAERS) and FDA Adverse Event Report System (FAERS) and CDC's Electronic Mapping, Reporting, and Coding (eMaRC) Plus used by central cancer registries.

Products: [CLEW](#) provides open-source NLP and machine learning tools. CLEW contains NLP architectures and tools to process spontaneous report narratives, extract clinical and temporal information from text, format the data for presentation, and map unstructured cancer pathology data and safety surveillance data into structured data, increasing the availability of analysis-ready data. The [CLEW source code and documentation](#) are available on the CDC's GitHub page.



Balancing Access with Enhancing Privacy and Security: At the portfolio-level, there has been less project work focused on the development and use of innovative technologies that offer researchers access to data, securely and privately.

Award: Improving Beneficiary Access to their Health Information through an Enhanced Blue Button Service (CMS)

The goal of the award was to develop a service to allow Medicare beneficiaries to access their claims data as well as share that data with health applications they trust, including donating that data to researchers. Blue Button 2.0 officially launched in March 2018. Blue Button 2.0 contains four years of Medicare Part A, B, and D data for 53 million Medicare beneficiaries. Blue Button 2.0 facilitates patient access and empowerment which are two major goals of the MyHealthEData initiative and the Cures Act.

Products: The [Blue Button 2.0](#) API uses HL7 FHIR to ensure claims data are in a structured format and can be easily accessed and integrated with other applications, such as research platforms. The OAuth standard is used for beneficiary authorization, giving patients secure access to their data. A collaboration with NIH will further enhance Blue Button to allow Medicare beneficiaries to donate their data to the *All of Us* Research Program. Blue Button 2.0 also initiated an app developer program and application vetting process.



Enhancing Data Governance: At the portfolio-level, there has been less project work focused on addressing policy and data governance barriers to access and use data assets and research-oriented data infrastructure.

Award: Enhancing Data Resources for Studying Patterns and Correlates of Mortality in Patient-Centered Outcomes Research: NDI Workshop and Strategy Paper (CDC)

The goal of the NDI Workshop was to develop a long-term strategic plan for access and use of NDI data. The NDI Strategic Plan charted a course to improve efficiency, timeliness, and quality of the NDI.

Products: The NDI Strategic Plan will serve as an internal resource for the CDC that describes opportunities to minimize barriers to NDI data access. The plan presents findings from assessments of: 1) non-economic barriers to access and use, and 2) the legal/regulatory environment and perceived policy barriers to access and use.

How Has the Portfolio Evolved to Address the Data Infrastructure Needs of Emerging Policy Priorities?

The contributions of the OS-PCORTF portfolio should not be assessed within a vacuum but rather within the current policy landscape, which is shaping the way data are being collected, curated, and used for conducting patient-centered research. The original mandate of the Affordable Care Act (Section 937(f) of the Public Health Service Act), emphasized the development and use of clinical registries and health outcomes research networks and included funding provisions for ASPE, the Patient-Centered Outcomes Research Institute (PCORI), and AHRQ to further develop the nation's PCOR data infrastructure. These priorities have been applied to research needs that have emerged (e.g., in the form of the opioid crisis) and legislative priorities have followed suit to support ongoing development. The section that follows presents an assessment of how the portfolio has evolved to address the data infrastructure needs of key policy priorities that emerged with the enactment of four important pieces of legislation in recent years identified by ASPE as having implications for PCOR data infrastructure.

The implications of each of the four key legislative mandates for patient-centered and effectiveness research are summarized below:

- MACRA has accelerated the movement of the health care system away from volume-based care to value-based care. Under value-based care programs, health care providers are incentivized with payments to provide high-quality, patient-centric care and must gather and analyze data to report on key quality and cost measures, including patient-reported outcome measures.
- The Foundation of Evidence-Based Policymaking Act of 2018 builds upon the momentum and recommendations of the Commission on Evidence-Based Policy Making (CEP) to develop a strategy for increasing data availability and use to support evidence-based policymaking in government programs.^{19,20} The CEP's activities culminated in a final report.²¹ that included several recommendations that have cross-cutting relevance to the OS-PCORTF portfolio's goal—

including enhancing researchers' access to high-value federal datasets to support evidence-building and enhancing security of this access (e.g., through temporary data linkages rather than repositories). The report also emphasized the need for robust privacy protections for patient data.²² Now, in its next instantiation, the Foundations of Evidence-Policymaking Act requires all federal agencies to lay out their plans for "identifying and addressing policy questions relevant to the programs, policies, and regulations of the agency."²³ Most relevant to PCOR data infrastructure, these plans must include: a list of data the agency intends to collect, use, or acquire to facilitate the use of evidence in policymaking; a list of methods and analytic approaches to developing evidence; and a list of challenges to developing evidence in support of policymaking, including any statutory or other restrictions to access relevant data.²⁴

- The Cures Act, enacted in December 2016, is a key piece of legislation that includes provisions across a range of important topics that intersect with the charge of the OS-PCORTF portfolio to build data capacity for patient-centered outcomes research. One overarching theme of the Cures Act is the push toward improving the interoperability of health information along with a very specific focus on accelerating individuals' ability to access and share their health information (which could include for research purposes).²⁵ It specifically allocates funding for *All of Us* Research Program, which puts a spotlight on initiatives that empower patients to engage in participatory research toward the development of individualized care. The Cures Act also designates funding to support research in clinical priority areas including mental health, opioid abuse, and cancer, as well efforts to speed the regulatory review processes of breakthrough medical devices.

- The SUPPORT Act, enacted in 2018, seeks to address the nation’s opioid overuse epidemic through provisions around prevention, treatment, control of the supply of illicit drugs, and research around opioids. The SUPPORT Act also directs government agencies to strengthen surveillance data infrastructure and conduct studies on aspects of the opioid epidemic.²⁶ This includes improving the timeliness and specificity of data collection and reporting (e.g., data around health outcomes, supplementary data around causes, risk and protective factors, comorbidities and disparities, geographic data, identifying gaps in key metrics, improving data linkages, and further utilizing and developing prescription drug monitoring programs) to inform more real-time public health response to the epidemic. Exhibit 10 provides a summary of the types of opioid-related issues being addressed by the portfolio.

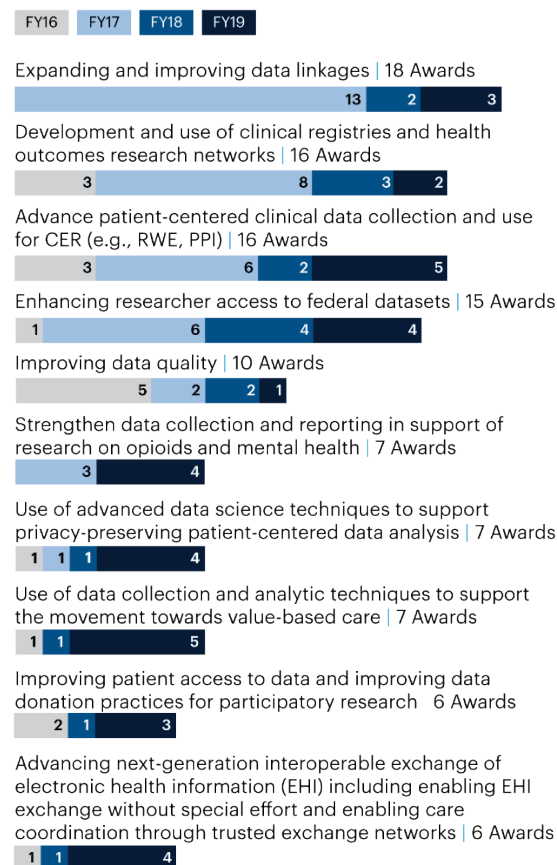
Exhibit 10. Opioid Topics Addressed by the Portfolio Award Products and Description

- Medication-assisted treatment during pregnancy
- Co-occurring disorders
- Modeling of opioid-related health outcomes using synthetic data
- Interoperable electronic care plan for pain patients with opioid use disorder
- Mortality and other health outcomes associated with opioid poisonings
- Opioid use disorder treatment in emergency department settings

Content analysis of these four legislative mandates revealed 10 key data infrastructure themes. Exhibit 11 presents the number of awards whose objectives support each of these 10 themes. The exhibit also includes the breakdown of total awards by the fiscal year of funding.

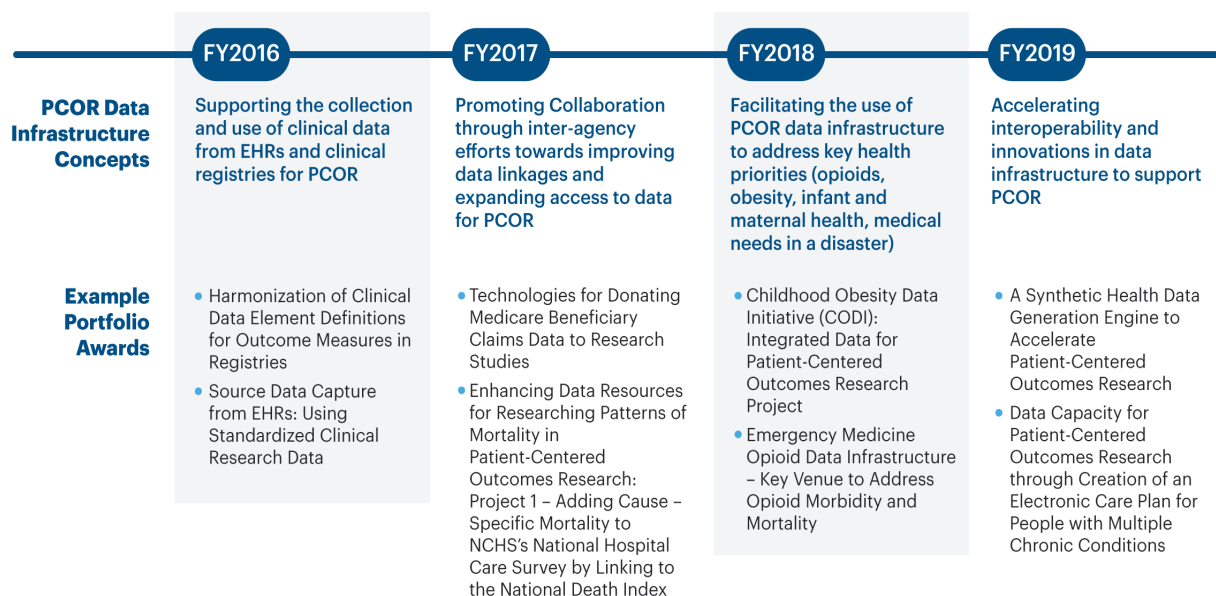
This analysis shows the evolution of the OS-PCORTF portfolio to support the data infrastructure needs of a shifting policy landscape. As seen in Exhibit 11, expanding and improving data linkages was the data

Exhibit 11. Legislative Mandate Data Infrastructure Needs Supported by OS-PCORTF Award (Awards per Fiscal Year)



theme that projects most commonly addressed (N=18 awards). Development and use of clinical registries and health outcomes research networks was also highly represented (N=16), as were awards focused on advancing patient-centered clinical data collection and use for CER (e.g., real-world data (RWD), PPI) (N=16); and efforts to enhance researcher access to federal datasets (N=15). The least commonly addressed data themes were related to advancing next-generation interoperability of electronic health information through APIs (N=6) and improving patient access to data and data donation practices for participatory research (N=6). These lower numbers are attributable to a small number of awards with this focus in FY 2016 through FY 2018 but have been bolstered significantly in FY 2019 to equal or exceed investments in other priority areas. These latter two

Exhibit 12. Evolving Data Infrastructure Concepts of the Portfolio and Representative Awards from FY 2016–2019



priorities support interoperability, patient access, and information-blocking provisions of the Cures Act. By FY 2019, the portfolio was addressing all 10 themes.

Many of the awards in FY 2018 and FY 2019 build on foundational work of awards in earlier years. For example, one of the awards funded in FY 2019, Identifying Co-Occurring Disorders among Opioid Users Using Linked Hospital Care and Mortality Data, is a capstone to a prior award funded in FY 2018. The FY 2018 award provides enhanced algorithms to more accurately identify a hospital patient’s use of opioids in any form and the specific opioid agent taken. The FY 2019 award will build on the prior award’s algorithm methodology to flag evidence of disorders that co-occur with opioid use. This work will allow researchers to conduct retrospective analysis to determine the extent to which opioids were involved in hospital encounters preceding post-discharge deaths.

Based on the synthesis of legislative mandates and supporting data infrastructure needs, the team also identified four overarching data infrastructure concepts for characterizing the work funded during FY 2016 through FY 2019 (Exhibit 12).

1. The FY 2016 awards focused on infrastructure needs related to collecting and using clinical data from EHRs for evidence generation and patient-centered research. They include a cross-agency award that developed an NLP web service that can be accessed by researchers to support conversion of unstructured clinical information to structured and standardized coded data, as well as an AHRQ award that developed libraries of clinical data element definitions that can be used to represent outcomes measures for five clinical topics: atrial fibrillation, depression, asthma, lung cancer, and lumbar spondylolisthesis using EHR-derived data.
2. The objectives of awards funded in FY 2017 promote inter-agency collaboration around linking datasets and making data more accessible for research. Efforts included the CMS and NIH collaboration to enable Medicare beneficiaries to donate their claims data to the *All of Us* Research Program via the Sync for Science and Blue Button 2.0 API.
3. The awards dispersed in FY 2018 focus on using data infrastructure to address important health topics including the development and testing of linkage and de-duplication tools and services to link pediatric clinical data to weight management

program data in support of childhood obesity research.

4. Finally, in FY 2019, the awards focus on supporting the development of innovative data resources and analytic approaches, including work that enhances Synthea, a synthetic software engine that produces synthetic data for research by generating modules for opioid, pediatric, and complex care use cases.

Discussion: Areas for Future Work

Results of the portfolio assessment were shared with a technical expert panel (TEP) of external stakeholders that included industry and research network representatives. These experts convened with ASPE and federal agency partners involved in the OS-PCORTF portfolio to reflect on strategic and operational opportunities to enhance the work of the portfolio going forward. The TEP provided their individual viewpoints on the portfolio assessment in three areas: 1) potential areas of refinement to the Strategic Framework; 2) PCOR data infrastructure needs that should be prioritized in future work of the portfolio; and 3) metrics that can be used to assess the impact of the portfolio going forward.

What are potential refinements or revisions to the Strategic Framework? The Strategic Framework was originally developed in 2014 and milestones associated with each of the five functionalities were articulated in 2015. The majority of TEP members felt that the Strategic Framework stood the test of time and provided useful guide posts for defining patient-centered research data infrastructure needs. However, the TEP identified the following potential refinements, which will allow for the Strategic Framework and associated milestones to more effectively support data infrastructure efforts moving forward:

- Ensure that the Strategic Framework accounts for the broader ecosystem in which PCOR and CER takes place by considering the external factors which can influence the data sources, data infrastructure, and types of patient-centered research inquiries that are carried out. The TEP highlighted factors such as financial and policy drivers in health

care, which can influence the types of clinical data that are captured in EHRs including MACRA quality reporting requirements, as well as the incentives that promote the sharing and use of data for patient-centered research and care delivery by health care organizations.

- Draw out the patient-centered aspects of the framework more explicitly and recognize that social risk factors play an important role in patient health outcomes. Specifically, social determinants of health (SDOH) is currently included as a data source within the Strategic Framework but going forward SDOH should feature more prominently within each of the five functionalities of the framework. For example, the portfolio could identify potential needs around standardized collection of SDOH, linking of SDOH to other clinical data sources, and enhancing researcher access to SDOH data within federal data assets.
- Identify and integrate cross-cutting barriers which can impede progress in patient-centered research. These barriers are likely to relate to the policies and governance components of the Strategic Framework as opposed to the more technologically focused components of the Strategic Framework (standards and services).
- Establish and convene working groups for each of the five functionalities. The working groups could be comprised of experts across the federal government who can advise on relevant federal initiatives and ongoing efforts that could be leveraged by the portfolio, metrics for assessing progress, as well as evolving needs related to that functionality.
- Incorporate data provenance into the Strategic Framework as a key component alongside standards, services, policies, and governance. The TEP noted that data provenance, or the originating source and pathway of data, has implications for data quality (e.g., timeliness, completeness, and accuracy), accessibility and manipulability, and therefore must be considered by all portfolio activities that target data infrastructure.

What are the other areas or gaps that should be addressed or prioritized in future work of the portfolio? In considering the results of the portfolio assessment as well as the ongoing data infrastructure needs of the patient-centered research community, the TEP identified a few targeted areas of focus for the portfolio moving forward:

- Address the non-technical barriers that impede the use and sharing of data for research including data governance, and privacy and security protections. Specifically, the TEP noted that the portfolio can do more to ensure that the governance and policy artifacts developed by more mature projects are shared and utilized by newer project teams. Another potentially useful product of the portfolio could be a master data use agreement that could be shared with and adapted by researchers who are working to secure data for their research projects. The TEP also acknowledged that often local organizational politics and policies influence governance challenges faced by researchers and suggested that it would be beneficial to explore these local dynamics and document the common challenges health organizations face related to sharing and use of data for research. The findings could then inform development of more useful data governance and policy frameworks. Similarly, the portfolio award teams can work with the pilot sites and “on the ground” research teams that are implementing and testing project products to surface the legal, financial, governance, and technical challenges encountered and work with ASPE to identify the entities that can help address these challenges.
- Provide targeted support and guidance to project teams with products that have been tested and are ready for translation and broader uptake. Guidance can take the form of identifying the most appropriate pathways for dissemination and development of a marketing plan to get products in the hands of a broad range of end-users.
- Engage end-users during the planning phases of project awards to support translation and sustainability efforts. Early input of the end-user community can help to identify and mitigate local data governance

challenges and maximize the chances that implementation sites sustain the use of products beyond the award funding period. The stakeholder input can also elucidate how the project products helped advance the research community from point A to B and how it facilitates their movement forward to point C after the project is over.

What metrics are needed to understand the impact of the portfolio on building data capacity for PCOR?

ASPE recognizes the importance of having metrics to assess the real-world impact of the portfolio on the research community. The TEP identified three categories of metrics for tracking progress and impact of the portfolio: 1) portfolio-wide metrics, 2) project-specific metrics, and 3) dissemination and translation metrics.

Example portfolio-wide metrics or activities which can be used to assess how the portfolio is advancing data infrastructure for PCOR include:

- Developing a more prescriptive roadmap that outlines a timeline for achieving a series of progressive goals. TEP participants recognized that this roadmap must be balanced with the need for a nimble portfolio that can also be responsive to evolving policy priorities.
- Identifying, tracking, and quantifying artifacts that are developed by one project award and then leveraged by another. In some cases the artifacts might be the end-products of an award such as a service or linked data set, or it may be an interim product such as a methodological report, lessons learned report or data linkage algorithm.
- Convening an external stakeholder council of end-users that can be consulted as part of funding decisions to ensure that portfolio investments are directed to projects that address the most acute data infrastructure needs of the research community.

Award-specific metrics can be used to assess whether each individual project award achieved its articulated objectives. The TEP noted that these metrics should be established at the award conception and should align with the aims specified in the scope of work. A suggestion was also made that each award team could develop logic models that describe the award aims, outputs, targeted end-users, and the research

infrastructure problem addressed. These logic models could be used to track the progress and trajectory of awards.

The final category of dissemination and translation metrics are critical for assessing how the portfolio is supporting the research community. Traditional dissemination metrics include number of publications,

conferences and other presentations as well as citation counts (i.e., the number of times a publication is cited by other researchers), translation and uptake metrics could include use of website analytics (e.g., measuring website traffic and counting the number of downloads of posted material), and tracking the ways in which other research initiatives leverage award outputs.

Conclusion

The five functionalities of the OS-PCORTF Strategic Framework have been well addressed by the 43 awards initiated between FY 2016 and FY 2019. These awards have also made headway in closing gaps identified by the prior 2017 Evaluation, particularly around the dissemination of products, as well as in implementing standards and improving data quality. While fewer awards have addressed enhancing data governance and balancing access with enhanced privacy and security, the products of these awards have made significant contributions in increasing data access.

These awards have not been pursued in a vacuum; the portfolio and federal policy have evolved in parallel. In particular, four legislative mandates emerged during this period: MACRA, the Foundation of Evidence-Based Policymaking Act of 2018, the Cures Act, and the SUPPORT Act. While the portfolio was more narrowly focused in the early years, it has expanded to cover the full breadth of emerging policy priority areas.

In considering future opportunities, TEP members used the results of this portfolio assessment to make the following recommendations.

- Consider revising the Strategic Framework to incorporate financial and social factors which can influence research data infrastructure as well as the barriers which can impede the advancement of infrastructure.
- Emphasize the role of data provenance as a key component that influences data usability for research.
- Focus future portfolio efforts on elucidating and addressing non-technical barriers to use of data for research including data governance, privacy, and security.
- Engage end-users early and regularly to appropriately target project award efforts, identify clear metrics, and promote sustainability of end products.
- Develop portfolio-wide, award-specific, and dissemination and translation metrics to assess the impact of the portfolio at both the micro and macro-level.

CHAPTER 2

PORTFOLIO SYNERGIES AND OPPORTUNITIES WITH STRATEGIES

Chapter Executive Summary

Patient-centered outcomes research seeks to prioritize the patient in the design, conduct, and reporting of research, with the goal of generating new scientific evidence that informs decision-making and patient care and improves health outcomes. As discussed in Chapter 1, the Strategic Framework has been used to guide ASPE's awards to agencies across HHS, which are working to enhance the ability of researchers to conduct such research. Concurrently, data are increasingly recognized by HHS agencies as an important resource that supports many aspects of each agency's mission and many departmental priorities. As a result, HHS and several agencies have recently published materials that guide data capacity-building. Chapter 2 is intended to engage stakeholders in an initial discussion about potential updates to the Strategic Framework, and the implications of these agency data initiatives for patient-centered data in HHS. It does so by examining the HHS Data Strategy, OS-PCORTF Strategic Framework, and other selected agencies' data initiatives to identify areas that are pertinent to the goals of patient-centered outcomes research, as well as opportunities for continued intradepartmental collaborations on building data capacity for research.

The OS-PCORTF functionalities and its portfolio are aligned with the HHS Data Strategy. Further, the awards have created products that support agencies' data strategies.

- The awards have created tools, guides, frameworks, and other solutions that contribute to the data strategies of multiple agencies. As such, resources from these awards should be disseminated to and used by HHS agencies to accelerate the pace of patient-centered outcomes research. Lessons learned from the awards could be applied by HHS agencies pursuing their own projects and serve as building blocks for new or related work within an agency.

The analysis indicates that there is a recognized need for interoperability and data-sharing among HHS agencies to support patient-centered research and knowledge generation.

- Robust research relies on equally robust data infrastructure in which: health information can be shared among multiple stakeholders (e.g., patients, providers, hospital systems, federal agencies); the health information consists of high-quality data; and the researchers have access to useful tools to gather and analyze this data to generate new knowledge. Finally, the health information being shared and analyzed for research must be protected and is a key priority area for future HHS data infrastructure projects.

A common theme across HHS data frameworks is the importance for agencies to learn from and collaborate with one another in pursuing their research and data priorities.

- Most, if not all, awards involve cross-agency collaboration either through joint agency funding or through the use of TEPs or steering committees. These partnerships demonstrate how multi-agency coordination can positively impact efforts to build data infrastructure and support the researchers conducting PCOR. Such collaborations would be beneficial to continue in the future, and these projects may offer best practices for doing so.

The results of this framework assessment were shared with the TEP. Feedback from the TEP will be used to inform initial discussions about the direction of future work. The TEP was asked to consider the following:

1. What opportunities exist for the OS-PCORTF portfolio to continue to build data capacity for research?
2. Are there ways for the portfolio to further align with other HHS agencies to build data capacity for PCOR?
3. Given the areas of alignment among the frameworks, what are the implications for future patient-centric initiatives, which the portfolio could support?

The TEP offered their thoughts, for the portfolio to seek opportunities to lead, collaborate, and support the development of data infrastructure and other initiatives that support patient-centric research.

- The portfolio has played a consistent role in bringing agencies together in partnerships and knowledge-sharing. There are several other avenues in which the portfolio leadership can advance the field, such as dissemination, convening, and formation of a brain trust to further awareness of and participation in PCOR among diverse research stakeholders.
- Having already built numerous relationships with agencies and researchers via the OS-PCORTF awards, the portfolio leadership is well positioned to continue its coordination and promotion efforts of patient-centric efforts across the department, as well as facilitation of PCOR collaboration with other departments (e.g., DHS, VHA).
- The TEP suggested that in areas that are outside of the portfolio's purview and/or already being addressed by other agencies, the portfolio could contribute to these efforts in a supportive way (e.g., contributing to the development of a federal catalog of data assets and other aggregate resources).

Introduction

The objectives of this chapter are to: 1) discuss ways the OS-PCORTF Strategic Framework aligns with the HHS Data Strategy; 2) draw connections among the Strategic Framework, HHS Data Strategy, and HHS agency-specific strategies; and 3) to identify opportunities to strengthen data capacity for PCOR with continued activities across agencies.

HHS agencies routinely collect, link, and analyze data that can be used to generate new scientific knowledge about Federal programs and the patient populations these programs serve. These data are foundational to research that expands knowledge about the outcomes and effectiveness of health care treatments and interventions. As a consumer, producer, and regulator of key national health data, HHS is uniquely positioned to coordinate its programs to build national data capacity in support of the

mission, statutory authorities, and annual priorities of each HHS agency and the Department as a whole.

This chapter reviews the data strategies of HHS and seven HHS agencies and two proposed rules and identifies commonalities across these strategies. Drawing connections across data strategies reveals the areas in which agencies are aligned around the development of PCOR data infrastructure. This chapter also highlights areas where the OS-PCORTF portfolio has made advancements and where additional opportunities exist for the future, especially for further collaborative efforts.

The OS-PCORTF Portfolio. Awards made under the OS-PCORTF portfolio are intended to expand data capacity and available infrastructure, ensuring that rigorous research studies can be conducted to improve the safety and effectiveness of interventions across patient populations. Within the portfolio, and across the health system, PCOR-targeted efforts can

take many forms, such as research that accelerates drug development and testing for conditions like cancer or research that enhances understanding of public health issues like OUD. Patient-centered research data infrastructure activities under the OS-PCORTF can also involve the creation of tools that allow data to be gathered directly from patients, shared among research networks and registries, and linked with other data sources to produce larger, more comprehensive patient records. This movement of health information improves clinicians' ability to answer questions that patients care about, such as the risk of hospital readmission or death after a heart attack.²⁷

The OS-PCORTF is also intended to provide for the coordination of relevant federal health programs to build data capacity for clinical comparative 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.

To guide the OS-PCORTF awards, ASPE focuses on making progress toward the five Strategic Framework functionalities, and filling infrastructure gaps identified by a 2017 Evaluation, as described in Chapter 1. ASPE targeted these five functionalities because they represent the tools necessary for health information capture, sharing, and analysis, all of which underpin patient-centered outcomes research.

The HHS Data Strategy. The HHS Data Strategy, released in 2018, was part of an ongoing effort “to expand the capacity of HHS’ data resources; promote synergy across data systems; ensure the efficiency, quality, utility, and timeliness of data collection systems; and address high-priority gaps in data.”²⁸ HHS is focused on optimizing the use of its data to support knowledge generation and evidence-based decision-making leading to better health outcomes for patients.

HHS has identified a six-priority data strategy that includes strategies for achieving its stated goals. Individual agencies within HHS also have their own data strategies to target their efforts in ways that support their unique missions, perspectives, goals,

and populations. The HHS Data Strategy is distinct from the Federal Data Strategy, which articulates a ten-year vision of federal data priorities.²⁹ In this chapter, we focus on the HHS strategy and those of its agencies, given the lens of patient-centric health research.

Patient-Centered Goals and Agency Activities

In addition to the agency-level data strategies considered in this chapter, many agencies have patient-centric goals and activities designed to enhance data infrastructure for patient-centered outcomes research. These are not always explicitly mentioned in their data strategies but are often addressed by them. Such activities include:

- The NIH Health Care Systems Research Collaboratory was initiated in 2012 to create new infrastructure that supports and accelerates pragmatic clinical trials. The Collaboratory helps researchers and health care delivery organizations by: 1) offering supportive infrastructure for multi-site research; 2) encouraging connections and collaborations; and 3) disseminating best practice tools and resources to guide researchers in the design and conduct of clinical trials. Clinical trials are a key mechanism for assessing treatment options for patients and generating evidence-based knowledge to guide patient and clinician decision-making.³⁰
- The FDA was mandated by the Cures Act to undertake new patient-focused drug development processes, including soliciting patient engagement in drug development, and incorporating real-world evidence (RWE) into the testing and approvals process. The FDA has since released a plan describing its implementation of new methodologies, measures, and other strategies that will ensure the patient perspective plays a larger role in the drug development lifecycle.³¹
- In 2017, the Centers for Medicare & Medicaid Services Innovation Center (CMMI) announced its intent to increase its orientation toward patient-centeredness and value-based models of care.³² This included a request for information (RFI) to gather stakeholder feedback on interoperable solutions to sharing health care data. The RFI was followed by a 2018 announcement of new

payment rules focused on whole patient care and improved patient outcomes, rather than fees paid for health services rendered.³³

These patient-centric approaches demonstrate the types of activities underway at individual agencies. In undertaking these activities, the agencies share the overarching goal of generating meaningful health insights for patients and clinicians to apply to health decision-making. The frameworks and data strategies analyzed below indicate there is further potential for cross-agency collaboration in the development of new tools, processes, and infrastructure to enable researchers to collect, link, and analyze data.

Methods

To identify common areas of interest across HHS agencies and potential opportunities for building the data infrastructure to support patient-centered research, the team reviewed the overarching HHS Data Strategy (Exhibit 13), the OS-PCORTF Strategic Framework functionalities (Exhibit 14), six agency-level data strategies (Exhibit 15), two Proposed Rules and an HHS report assessing the current state of data sharing within the department (Exhibit 16). The full HHS Data Strategy framework and agency-specific priorities, opportunities, and strategies discussed in this chapter are presented in Appendix E and F, respectively.

These data strategies, reports, and rules come from agencies that are frequent OS-PCORTF collaborators and who routinely collect, link, and analyze data to generate new scientific knowledge on health outcomes.

The six HHS Data Strategy priorities (Exhibit 13) can serve as a framework for identifying common themes across the agency strategies and proposed rules to identify areas of common interest. Agencies' high-level strategies and themes, as well as their components, can be used (via content analysis) to determine areas of overlap with the HHS Data Strategy. Next, the OS-PCORTF functionalities (Exhibit 14) help to contextualize the PCOR-related activities of the other strategies. Through this process, areas of overlap and unique attributes under each strategy become evident.

The HHS Data Strategy consists of six priority areas and identifies specific opportunities for addressing each priority (as shown in Exhibit 13). For example, under Priority 1, Improving Access to HHS Data, there are two specific opportunities for increasing data access both within HHS agencies and externally.

Exhibit 14 shows the five functionalities of the OS-PCORTF Strategic Framework that help categorize the types of contributions awards make to data infrastructure. These functionalities are components of a larger Strategic Framework for building data capacity of patient-centered outcomes research (see Appendix F).

Exhibit 13. HHS Data Strategy

Priority 1 – Improving Access to HHS Data

- Opportunity 1: Increase the accessibility of HHS data to internal and external users while ensuring that the information is used responsibly
- Opportunity 2: Increase awareness within the Department about available HHS data resources and research

Priority 2 – Enhancing Administrative Data for Research

- Opportunity 1: Expand the use of administrative data in the Department
- Opportunity 2: Improve the quality of administrative data for research

Priority 3 – Increasing Data Linkages across Diverse Data Assets

- Opportunity 1: Apply existing departmental knowledge and lessons learned from data linkages
- Opportunity 2: Improve the capacity to link HHS data internally and with other data sources

Priority 4 – Modernizing Privacy Protections

- Opportunity 1: Without eroding privacy protections, increase data-sharing through better communication and coordination with experts
- Opportunity 2: Assist in standardization of departmental privacy policy practices

Priority 5 – Increasing Data Policy Coordination and Information Sharing Across the Departments

- Opportunity 1: Increase coordination in the Department regarding data collection, system and software investments, and data management and governance
- Opportunity 2: Inform policymakers and researchers about the value and uses of HHS data

Priority 6 – Building a 21st Century Data-Oriented Workforce

- Opportunity 1: Enhance the data science capability of the current HHS data workforce
 - Opportunity 2: Reinforce capacity to explore the application of data science and alternative data to HHS research and program evaluation
 - Opportunity 3: Invest in the future of data science
-

Exhibit 14. The OS-PCORTF Strategic Framework Functionalities

1. Use of Clinical Data for Research
2. Standardized Collection of Standardized Clinical Data
3. Linking Clinical and Other Data for Research
4. Collection of Participant-Provided Information
5. Use of Enhanced Publicly Funded Data Systems for Research

Six agency-level data strategies, two proposed rules, and a departmental report were reviewed (Exhibits 15, 16) and categorized in terms of their overlap with the six HHS Data Strategy priorities and the five OS-PCORTF functionalities. The exhibits include the names of the relevant data strategies, as well as a brief summary of each agency's mission to contextualize their perspective.

After determining the areas of overlap and divergence, the OS-PCORTF portfolio was reviewed to identify example awards that demonstrated implementation and development of particular priorities and/or opportunities for broader use and/or opportunities for cross-agency collaboration.

To determine overlap among agency goals, the high-level strategies and sub-themes were extracted from each strategy document. A content analysis of each agency strategy and proposed rule was then done to determine the extent to which the specifics of the objectives overlapped with one another and with the HHS Data Strategy.

The findings and synthesis were then discussed with a technical expert panel (TEP) who provided overall feedback on the chapter and were asked questions relating to next steps for the Department in continuing to build data capacity for PCOR. Their feedback informed the sections that follow, and their suggestions for future activities are discussed in the conclusions.

Exhibit 15. HHS Agency Data Strategies

National Institutes of Health (NIH)

- *Mission:* lead research and discoveries to understand, prevent, diagnose, and treat patient diseases and disorders and further science and medicine³⁴
- *Data Strategy:* Strategic Plan for Data Science³⁵

National Library of Medicine (NLM)

- *Mission:* lead, conduct, and support research in biomedical information science, informatics, and data science that ultimately improves patient health³⁶
- *Data Strategy:* A Platform for Biomedical Discovery and Data-Powered Health: National Library of Medicine Strategic Plan 2017–2027³⁷

Food and Drug Administration (FDA)

- *Mission:* ensuring the safety, efficacy, and security of human and veterinary drugs, biological products, medical devices, food supply, cosmetics, and products that emit radiation to protect patient health³⁸
- *Data Strategies:*
 - Digital Health Innovation Action Plan³⁹
 - Framework for FDA’s Real-World Evidence Program⁴⁰
 - Sentinel System Five-Year Strategy 2019–2023⁴¹

Centers for Disease Control and Prevention (CDC)

- *Mission:* protect communities from health, safety and security threats, both foreign and in the U.S.⁴²
- *Data Strategy:* Surveillance Strategy⁴³

Exhibit 16. Reports and Proposed Rules

Centers for Medicare and Medicaid Services (CMS)

- *Mission:* empowering patients, focusing on results, unleashing innovation for Medicare and Medicaid populations⁴⁴
- *Proposed Rule:* Interoperability and Patient Access Proposed Rule⁴⁵

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

- *Mission:* coordination of nationwide efforts to implement and use the most advanced health IT and the electronic exchange of health information to support both research and clinician and patient decision-making⁴⁶
- *Proposed Rule:* 21st Century Cures Act: Interoperability, Information Blocking, and the ONC Health IT Certification Program Proposed Rule⁴⁷

Office of the Chief Technology Officer (CTO)

- *Mission:* testing and validating solutions to solve challenging problems in the delivery of health and human services and promoting innovation across the department⁴⁸
- *Report:* The State of Data Sharing at the U.S. Department of Health and Human Services⁴⁹

Findings

The data strategies analyzed in this chapter are predicated on the idea that an immense amount of electronic patient health data exist and must be made more available and useful to patients, clinicians, and researchers alike. Each strategy outlines the ways in which a given agency will use health data to meet their missions, research priorities, and goals for patient-centeredness. This section explores the alignment of the OS-PCORTF functionalities with HHS and agency-level data priorities and identifies opportunities to build on existing work and

infrastructure to continue advancing patient-centered research.



Functionality 1: Use of Clinical Data for Research

Robust clinical research relies on data from multiple sources like EHRs, claims, and registries. This kind of research also requires a data infrastructure to ensure that data is high quality and fit for use in research. As agency partners work to address this functionality within their OS-PCORTF awards, they are also advancing similar department priorities: for example,

Priorities 1 and 2, which relate to “Improving Access to HHS Data” and “Enhancing Administrative Data for Research.” The HHS Data Strategy specifically highlights opportunities to improve the quality of administrative data for research through the development of a framework for administrative data collection. These priorities are echoed across the agency-specific data strategies and goals, such as the NIH Strategic Plan for Data Science, which includes a priority of using clinical data for research by prioritizing promotion of a FAIR data ecosystem in which data can be findable, accessible, interoperable, and reusable.



Functionality 2: Standardized Collection of Standardized Clinical Data

Research is often challenged by issues of variability in clinical data definitions used across the health IT system. Lack of standards or failure to use standards makes data more difficult to correctly interpret, which can impact the interpretation of study results and the use of results to improve patient outcomes. As such, the need for data standards is a cross-cutting issue that underpins the majority of agency-specific data strategies. Data standards and linkages are necessary components for data-sharing, which in turn is necessary for analysis of clinical data (functionality 1).

The functionality of standardizing data collection of clinical data aligns with HHS Priority 2 “Enhancing Administrative Data” via data collection standards and Priority 3, creating standards to support interoperable data-sharing and “Data Linkages.” NIH is pursuing this priority in their own objective to “Leverage Ongoing Initiatives to Better Integrate Clinical and Observational Data into Biomedical Data Science.” To do so, they are defining CDEs and encouraging their use as a tool to improve data accuracy, reliability, and interoperability among different datasets that may be useful for research.



Functionality 3: Linking Clinical and Other Data for Research

When using data from diverse sources, researchers need tools to access and combine this health data across time and the care continuum. Linking data

increases the scope of a given dataset and allows researchers to capture the range of variables needed to generate clinically impactful research insights. Recognition of the need to link data is likewise reflected in Priority 3 of the HHS Data Strategy: “Increasing Data Linkages across Diverse Data Assets.”

HHS has identified three opportunities to support this departmental priority: 1) developing a repository for methods and best practices around linkage strategies, barriers, and opportunities (pooling HHS departmental knowledge); 2) promoting data linkages between HHS agencies (e.g., exploring the use of storage platforms such as data lakes to facilitate linkages); and 3) promoting data linkage to nonfederal data (e.g., developing waivers, promoting interoperability standards, and technical assistance to create linkages between HHS data, state and local data, and data from the private sector). All of these are opportunities that the OS-PCORTF portfolio has pursued.⁵⁰

The FDA is pursuing linkage solutions as a priority under their framework for RWE. FDA uses RWD from a number of sources, including EHRs, patient registries, claims and billings data, as well as data from mobile devices, wearables, and other biosensors to generate RWE in support of regulatory decisions, health care coverage decisions, and development of innovative treatments.⁵¹ Linking data from across these sources improves FDA’s ability to conduct meaningful research that enhances patient safety and effectiveness of regulated products.



Functionality 4: Collection of Participant-Provided Information (PPI)

There is growing interest across the health system in the capture of PPI. It is seen as an opportunity both to involve patients in research and to enrich the health data available for analysis. This functionality is related to both HHS Priority 1 to “Improve Access to HHS Data” and Priority 3 to “Increase Data Linkages across Diverse Data Sets.”

The FDA also plans to use PPI or what it calls RWD—defined as “data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources including patient-generated data.”⁵² The FDA’s framework for its Real-World

Evidence Program, its Sentinel System Five-Year Strategy 2019–2023, and National Evaluation System for health Technology (NEST) describe a need to assess the fitness-of-use of RWD for RWE generation. The FDA's plans intend to address gaps in RWD sources, such as data from mobile devices, wearables, and PRO tools. For example, the goal of the FDA's Pre-Cert Pilot Program is to streamline digital product regulatory oversight, speeding the time to market of novel PPI data collection tools.

NIH's Strategic Plan for Data Science sees similar value in PPI. NIH plans to "Leverage Ongoing Initiatives to Better Integrate Clinical and Observational Data into Biomedical Data Science," including collecting observation data from patients. The *All of Us* Research Program is one example in which PPI and sensor data are being collected to enrich the research datasets and analyses.



Functionality 5: Use of Enhanced Publicly Funded Data Systems for Research

Given the abundance of rich federal datasets, this functionality focuses on ways to further enhance, connect, and ensure these datasets are being used for health research. This functionality supports the first three priorities of the HHS Data Strategy: 1) improving access to HHS data; 2) enhancing administrative data for research; and 3) increasing data linkages across diverse data assets. Opportunities in these areas involve increasing awareness of data, streamlining processes for accessing data, improving the quality of administrative data, and improving the capacity to link HHS data internally and with other data sources.

These priorities tie closely to the aspects of this functionality related to facilitating access, retrieval, and linkage of federal datasets. The CDC Surveillance Strategy illustrates this priority at the agency-specific level, targeting the enhancement of its public health surveillance capabilities. This includes increasing the timeliness and usability of data for research through the development of new tools and technologies, as well as consolidating systems, eliminating unnecessary redundancies in reporting, and reducing reporting burden associated with surveillance data.

Discussion

Given the areas of mutual interest among agencies, there are opportunities for collaboration in their approaches to research and PCOR. The section examines shared agency interests in terms of each HHS priority, and provides examples of OS-PCORTF awards that could be used and/or expanded upon by HHS agencies, as they collectively pursue their research and data infrastructure priorities.

Areas of Alignment and Opportunity, by HHS Priority

There is a high degree of alignment between the OS-PCORTF Strategic Framework functionalities and the HHS Data Strategy, as well as between the agency-specific data strategies. The significance of these areas of alignment are twofold: 1) the projects conducted by OS-PCORTF awardees have contributed numerous technical solutions to shared data infrastructure that benefit multiple agencies and support multiple data strategies; and 2) there are opportunities to continue to support inter-agency cooperation as PCOR work proceeds, e.g., increasing access to HHS data and investing in data linkages. Finally, the proposed rules and CTO Report point to upcoming areas (e.g., privacy protections) that may also lend themselves to inter-agency collaboration.

The section below highlights specific opportunities for collaboration under each priority of the HHS Data Strategy, beginning with the HHS priorities where there is the greatest alignment across agency-specific frameworks.

HHS Priority 1: Improving Data Access

Many agencies are working to improve data access. There is a vast amount of electronic health data that exists in both distributed systems (e.g., health systems) and consolidated systems (e.g., CMS datasets). Analysis of these types of "big data" are essential for the scientific research community; however, they are not always readily available for research and analysis. Distributed research networks are one solution to the problem of aggregating large datasets of health information. These research networks support data querying and other functions across institutions. Data donation is another solution to help transfer health information into the hands of researchers, while giving patients autonomy in managing this data transfer.

However, there are several technical challenges associated with distributed research networks, which two OS-PCORTF awards worked to address. The FDA's Utilizing Data from Various Data Partners in a Distributed Manner award developed a mechanism to allow researchers within a distributed data network to access data from multiple organizations and analyze it within a single dataset. FDA's Cross Network Directory Service built an interoperable service to make it easier for researchers to run queries across multiple data research networks, and share analytic capabilities and knowledge across those networks. Both solutions enhance the shareability and searchability of data, while allowing research networks to maintain control of patient-level data. This increases the availability of data to answer research questions while also protecting patient privacy.

Another OS-PCORTF solution to increase access to data comes from the CMS Blue Button 2.0 API award. The goal of this award was to provide a safe and secure mechanism for Medicare beneficiaries to donate their claims data to a source of their choosing, including to research studies. The award combines recent advances on two fronts: 1) the development of the CMS Blue Button API—an easy and secure way for beneficiaries to share data, and 2) use of FHIR. A subsequent OS-PCORTF award to CMS and NIH (the Leveraging the Sync for Science™ initiative) extends this work by providing participants with a mechanism for also donating EHR data. These awards advance the priorities of multiple agencies to improve researcher access to patient health data. They also align with the goals of the ONC proposed rule to offer an easy solution for patients to access their own data through APIs.

Opportunity: The FDA awards enhance research and analytic power through distributed networks, while the Blue Button award developed and successfully demonstrated technical solutions to facilitate patient participation in research. These awards demonstrate the range of different solutions to increase researcher access to high-value datasets. Importantly, both utilize open-source technology, meaning that other researchers and end-users can leverage the tools developed by these awards for their own purposes.

HHS Priority 2: Enhancing Administrative Data

Many agency-specific data strategies seek to enhance administrative data. Clinical data sources can be enhanced through linkage to administrative sources, providing additional data that broadens their scope (e.g., laboratory tests, specifics of health insurance coverage, or the timing of events), as well as through improvements to the quality and timeliness of the data themselves. The OS-PCORTF NDI-related awards—mentioned in the discussion under Priority 3—sought to harmonize, connect, and enrich the federal mortality data for PCOR. Specifically, the four complementary awards to CDC, FDA, and CMS involve linking claims, EHR data, survey data, and other sources to transform the NDI into a nationally representative resource available for clinical and claims-based research. To fill gaps, CMS will link the NDI data on cause and manner of death with claims data to produce updated research files with Medicare descendants across a broader range of years than previously available. Linking health care claims data that include patient diagnoses with data on the cause and manner of death enables many types of PCOR research, including descriptive epidemiology, predictive modeling to identify high-value intervention targets, and CER.

Opportunity: The NDI awards may serve as a model for other agencies in terms of the lessons learned around solutions to technical hurdles as well as a demonstration of how shared priorities may be best addressed through coordinated, multi-agency efforts.

HHS Priority 3: Promoting Data Linkages

Across agencies there is widespread interest in promoting data linkages, and there are multiple active OS-PCORTF awards and cross-agency collaborations underway that are tackling this issue. Information on patient health care services is captured and stored across health care entities, and therefore tools are needed to connect these disparate records into datasets that can be used for PCOR. Data linkages can enrich a data source by filling gaps and providing robust cross-sectional or longitudinal patient profiles to enhance research and improve access to clinical information.

For example, mortality is an important outcome for many public health inquiries, clinical trials, and assessments of care quality. Mortality data supports research into the causes and risk factors of disease

and the effectiveness of a wide range of interventions and drug therapies. However, there are a number of technical, legal, statutory, and operational issues, along with financial barriers, that limit use of mortality data for PCOR. Two OS-PCORTF CDC awards focused on enhancing the NDI by: 1) linking NDI data to other sources of mortality data (e.g., vital records and national surveys); and 2) expanding the information available in the NDI on cause and manner of death, which are key indicators for research.

The FDA's Sentinel Initiative has plans to vastly expand mortality data linkages among state and federal databases to support PCOR. The planned data sets that will be linked to Sentinel include: the NDI, SEER (Surveillance, Epidemiology, and End Results) cancer registry, rare-disease registries, EHRs, and research networks like PCORnet. This work will increase the availability of vital health information for outcomes research, including for specific diseases like cancer.

Opportunity: As illustrated by the examples above, data linkage naturally draws together the HHS agencies in order to leverage and expand federal data assets for research. These examples underscore the importance of knowledge sharing and cross-agency use of existing tools to support data linkages. Agencies like NLM, whose strategic data plan focuses on building tools and partnerships, could be an additional strategic partner in data linkage activities.

HHS Priority 4: Privacy Protections

Privacy protections are part of the OS-PCORTF Strategic Framework functionalities as well as the recent policy/guidance documents. They are also included as a subcomponent of several agency-specific data strategies. For example, Goal 5 of the NIH Data Strategy focuses on "Enacting Appropriate Policies to Promote Stewardship and Sustainability," with Objective 5-2 defining stewardship as planned data management and security policies to protect patient privacy. Although privacy protection is not a major category in most agency data strategies, this does not appear to be a result of inattention to privacy. Rather, as agency data strategies outline overarching goals like data linkages and enhanced sharing of administrative data, privacy, and data protections are an important underpinning of those goals.

Agency-specific data strategies recognize a need for technology-based protections that maintain privacy and facilitate data-sharing among health organizations, patients, and federal agencies for the purposes of research. Moreover, relevant policy frameworks need to be updated as they may not sufficiently balance the dual need for privacy and data fluidity. Two OS-PCORTF awards led by ONC and CDC recently created frameworks that aim to lessen confusion over the laws, policies, and technical infrastructure needed to support PCOR. These frameworks for Privacy and Security (ONC) and Legal and Ethical Issues (ONC, CDC) offer practical guidance to researchers on how to safely use and protect health data, including a sampling of relevant federal and state laws and issues that should be considered. Topics discussed in the frameworks include consent procedures for special populations and proper data de-identification practices. While applicable across the research landscape, these frameworks may also be useful for specific agencies who are considering revising their privacy policies and/or pursuing technical solutions as part of their efforts to link and share data. The CMS and ONC proposed rules and the CTO Report make clear that privacy considerations must be at the forefront of activities to enhance interoperability and to better leverage federal data assets for PCOR.

Opportunity: The three policy frameworks created under the OS-PCORTF awards could serve as an area for multi-agency collaboration through sharing of best practices and practical guidance around privacy protections and the use of data for research.

HHS Priority 5: Interdepartmental Data Policy Coordination and Information-Sharing

Interdepartmental data-sharing is identified as a priority in half (three of six) of the agency-specific data strategies. However, there are many areas of overlap between this priority and Priorities 1 and 3 related to increasing access to HHS data and creating data linkages. Moreover, this priority encompasses the need for policy frameworks and data governance strategies that require coordination among agencies, which intersects with Priority 4 (privacy protections). One issue garnering cross-agency focus that is also likely to benefit from cross-agency coordination is how to incorporate the patient perspective more readily into health care and health research. PGHD

and patient-reported outcomes (PROs) can capture the patient's unique perspective and experiences outside the clinical setting, complementing clinician assessments to assist with diagnosis and treatment, as well as with PCOR. However, there is a need to develop and adopt the infrastructure, standards, and governance for PRO and PGHD for use in research.

One portfolio award, led by ONC, developed a policy framework for the use of PGHD in research and care delivery and pilot tested a related tool for data collection. The policy framework addresses what is needed to support PGHD capture and use, including collection tools, data donation policies, policies to address regulatory gaps, and interoperability of data across health information systems and devices. In implementing the tool for data collection, one pilot site used an app-driven approach to capture PGHD that was integrated into physician workflows to achieve better care coordination and population management for diabetes patients. The other pilot site tested a technical platform for capturing PGHD to support care for orthopedic surgery, behavioral health, bariatric surgery, and stroke. The PGHD award demonstrates both the necessary framework and technology for the capture, use, and sharing of PGHD in clinical care delivery and research models.

Opportunity: These types of targeted frameworks and tools highlight challenges and solutions to emergent issues in PCOR. However, there is a need for additional conversations and formalized approaches for leveraging new sources of health data, like PGHD and PROs, and coordination across departments to determine the optimal ways to collect and use these data. The agency-specific data strategies prioritize enhancing what is available for health research while allowing research networks, federal agencies, and other participants to control access to their data. Coordination and governance structures can provide a balance between analytic requirements, patient privacy, and confidentiality and proprietary considerations inherent in PCOR-related data-sharing.

HHS Priority 6: Workforce Development

Workforce development is a priority across four out of six agencies, although it is not a focus of the OS-PCORTF Strategic Framework functionalities or the ONC and CMS proposed rules. The agency-specific data strategies articulate a need for a highly trained

workforce that is capable of addressing agency data and research priorities and the technical complexities of bringing them to fruition. This manifests differently by agency; for example, FDA cites the need for software development related to medical devices and digital health, CDC has a specific interest in technology experts who can focus on systems and surveillance-related challenges, and NIH identifies a need for data scientists and a highly trained workforce for biomedical research.

Opportunity: Given the high demand for expertise in data, data science, and health IT, it is likely that agencies may be competing for talent within the same pool until workforce capacity can be substantially increased. In the interim, collaboration may be advantageous from a technical and practical perspective so that existing expertise can be shared across and/or can benefit multiple agencies. To support the training of an expanded health IT workforce, the use of implementation guides developed under several OS-PCORTF awards may serve as resources.

Opportunities Moving Forward

Looking across the frameworks, the awards made under the OS-PCORTF portfolio are supportive of HHS priorities, and the project work undertaken by multiple agencies has contributed to research data infrastructure. Portfolio awards include multi-agency collaborations and have created tools, frameworks, and other solutions that are relevant across HHS agencies. As such, these resources should be broadly leveraged to accelerate the pace and scope of PCOR infrastructure development—they can serve as lessons learned for HHS agencies pursuing their own projects, as building blocks for new or related work within an agency, and as models for inter-agency collaborations to address PCOR infrastructure priorities.

Future opportunities must recognize the benefits of data interoperability across HHS agencies. Furthermore, there is recognition across the agencies that the data necessary for robust PCOR must be shared from multiple data sources inside and outside the federal government; that high-quality data are standardized and can be linked together to enhance its research potential; and that the systems that support data and analysis require continued investment in both technology and workforce. Finally,

future opportunities must include strong privacy, and the proposed rules and CTO Report make it clear that this will be part of the next phase of work.

These cross-framework themes reflect the need for HHS agencies to learn from and collaborate with one another in pursuing their research priorities. Because multiple agency data strategies prioritize enhanced data access and sharing, there is mutual benefit to collaborating on the tools and data infrastructure to instantiate the shared vision of robust data to fuel treatment and innovation. The OS-PCORTF portfolio has contributed solutions to the data infrastructure needs of multiple agencies. Moreover, it offers multiple models of success that demonstrate how multi-agency activities can positively impact data infrastructure and support the researchers conducting PCOR.

TEP Question 1: What opportunities exist for the OS-PCORTF portfolio to continue to build data capacity for research?

Lead. The portfolio has played a consistent role in bringing agencies together in partnerships and knowledge-sharing. In addition to leading by example with the OS-PCORTF portfolio, there are several other avenues in which portfolio's leadership can advance the field.

Dissemination. Dissemination will continue to be an important function, especially given that many awards are maturing and their tools and products will become ready for broader dissemination. TEP members discussed the need to raise awareness of the OS-PCORTF among agencies who conduct research and those who contribute data but do not pursue their own research agendas. Both types of agencies would benefit from discussion of the ways their agency can or already does contribute data, with the idea that they may be able to leverage existing program data resources to support PCOR. In addition, making agencies aware of the potential value of their data contributions can provide the impetus for data harmonization.

Building a visible web presence for the OS-PCORTF portfolio is one important strategy for disseminating products and enabling their use for patient-centered research. Raising awareness via the conference circuit is another tried-and-true strategy for engaging with

researchers on the utility of the OS-PCORTF awards. The PCOR Resource Center continues to be an avenue for publishing papers and disseminating information and could be further leveraged to highlight project capabilities.

Further work is needed to conduct outreach and engage experts on the topic of sustainability. The agencies who received OS-PCORTF awards were encouraged to build interoperable tools that could be widely deployed beyond the life of each award. A strategic discussion is warranted regarding how to disseminate award products more widely to increase the likelihood that they are adopted and used across agencies and by outside researchers.

Formation of a Brain Trust/PCOR Innovation Core. The TEP discussed ways for portfolio to help create an OS-PCORTF brain trust to serve as a bulwark for PCOR, given the wealth of knowledge that is being cultivated by OS-PCORTF awardees and the potential for turnover within the federal government. There is a need to ensure that the knowledge and lessons of the portfolio are retained and expanded upon among the agencies.

A training fellowship or research opportunity could recruit fellows who could rotate through agencies and projects. This would help extend the reach across the OS-PCORTF strategic framework pillars and across the agencies. There are also opportunities to support researchers new to PCOR through small grants and to recruit academics on sabbatical. Fellows and students often bring research teams together and facilitate collaboration and learning.

There is the potential to build on existing fellowship programs in the department (e.g., CDC Public Health Informatics Fellowship, NLM Fellows, FDA Fellows, ORISE Fellows, Presidential Management Fellows), for example, by recruiting fellows in these programs to work on patient-centered outcomes research data initiatives.

To enhance learning opportunities, OS-PCORTF awardees could be called upon to educate these fellows and/or new award leads by holding seminars on their projects and related data concepts (e.g., FHIR 101). Another important aspect of training would include sustainability planning, as this is not something researchers or federal staff usually do but

is necessary for project planning. ASPE could develop an orientation packet for new OS-PCORTF leads with input from the veteran leads.

TEP Question 2: Are there ways for the portfolio to collaborate with other HHS agencies to build data capacity for PCOR?

Coordinate. In keeping with its statutory mandate, ASPE is tasked with coordinating PCOR efforts. The previous sections identify numerous opportunities for continued research and infrastructure development, including: pursuing additional data linkages; continuing to develop solutions that improve data access for research while engaging patients (e.g., APIs that facilitate data donation from patients); and overcoming technical hurdles to data-sharing across agencies. These types of solutions speak to the broader need to share information and access across diverse sources, including agencies, and therefore require multi-agency cooperation.

Formation of an External Stakeholder Council. TEP members suggested that ASPE consider an active role in supporting department-level thinking on funding new work. For example, the TEP discussed convening an external stakeholder council that could meet throughout the lifecycle of designing and implementing new projects, the goal of which would be gathering input on what projects are most important to the research and PCOR communities.

The findings of the Council could then assist ASPE in assessing future applications for OS-PCORTF funds. For example, ASPE could consider supporting awards that do an end-to-end scoping and include enhanced messaging around PCOR. The TEP suggested ASPE could require or prioritize partnerships among researchers working in similar areas. In addition, to ensure funded awards meet certain expectations (e.g., for outputs, contributing to a longitudinal network across the government/researchers), there may be a need for a longer development time at the front end. This time would be used to make sure funded awards have a work plan, metrics, and outcomes before work commences.

TEP Question 3: Given the areas of alignment among the frameworks, what are the implications for future patient-centric initiatives?

Support. Opportunities are identified in the chapter that are outside ASPE's statutory purview but that will benefit patient-centered research as a whole. For example, multiple agency data strategies recognize the need for increasing the research and high-tech workforce. While the portfolio is not involved in workforce development, maintaining awareness of these activities and potentially offering advice to those agencies (e.g., NIH) making such investments will improve the field of patient-centered research.

A Federal Catalog of Data Assets. The TEP discussed the benefit of building broader capacity across HHS via a catalog of data assets, and discussed the ways in which ASPE and the OS-PCORTF could contribute. The TEP identified multiple efforts underway to this end, largely through the HHS Data Council as a result of the Evidence-Based Practice Act of 2018. For example, there are federal data catalogs in which data and meta-data must be public, all of which have supportive infrastructure and governing bodies. There are subcommittees assigned to each HHS data priority and tasked with pursuing pragmatic solutions to technical and workforce challenges; NIH and CDC also have active task forces on workforce. As well as one focused on data use agreements so people can acquire the data they need.

Another subcommittee on data has a contractor designing a prototype for a federal data catalog. Therefore, they would be a good contact for ASPE to connect with to ensure that data resources from OS-PCORTF are reflected.

Increasing Transparency and Contributing Data. These efforts emphasize the myriad activities, and the need to support existing efforts, rather than reinventing the wheel. For example, NIH has multiple clinical data element repositories, and it would be wasted effort for every agency to do this work. The TEP suggested that the OS-PCORTF could help connect agencies to help them get more mileage out of their activities, and to avoid pursuing duplicative work.

The OS-PCORTF could help by cataloging the specific data assets from its portfolio, including the data sources that were used and by whom, which would

later feed into the federal data asset catalog that is underway. In the course of its own work (e.g., projects, dissemination and outreach), the OS-PCORTF could help steer people to the data owners, disseminate information about the gold standards, and help instill best practices for data use and sharing among the broader research community. Relatedly, ASPE could help identify which HHS data sets that are fit for use in supporting patient-centered outcomes research

Based on the findings in this chapter, it is clear that ongoing discussions with internal and external stakeholders will be useful to forge a path forward. There is both agency-level and department-level interest in fostering conversations about the direction the HHS department should take in building data capacity for patient-centered outcomes research. Furthermore, there may be opportunities to fine-tune the portfolio's Strategic Framework and make strategic decisions with regard to future OS-PCORTF awards and the agency's role in supporting patient-centric research across the department.

Conclusions

This chapter takes a broad look at the data strategies and frameworks guiding HHS and its agencies and how these intersect with the functionalities of the OS-PCORTF Strategic Framework. Across all HHS agencies there is strong interest in sharing and utilizing health data for patient-centered research that improves patient health and outcomes. Likewise, there is a shared understanding around continued efforts to enhance data quality, the need for robust data infrastructures to improve data timeliness, availability, and interoperability and resolve technical and policy challenges.

The HHS Data Strategy, as a Department-level statement of priorities, serves as the main point of comparison for the agency-specific frameworks and the OS-PCORTF Strategic Framework functionalities. Unsurprisingly, there is a significant amount of overlap across these strategies. Certain priorities, such as increasing access to HHS data and improving data linkages, have almost universal representation across agencies and frameworks. Other priorities, such as workforce development, are being pursued by certain agencies and oriented toward agency-specific needs (e.g., medical devices, public health surveillance). Privacy protections are integrated throughout the priorities articulated in the agency frameworks and are an explicit priority in the recent proposed rules issued by CMS and ONC, as well as in the CTO Report assessing HHS' data-sharing tools and processes.

Looking into the future, there are numerous opportunities for the portfolio leadership to lead, coordinate, and support HHS activities. Broadening the agency's dissemination of the OS-PCORTF products, as well as broadening awareness of PCOR itself among potential agency partners and other stakeholders, was a key theme during the TEP discussion. Facilitating departmental-level thinking about PCOR strategy, including among external stakeholders, and allowing that to inform the agency's work was another key suggestion for leveraging ASPE's existing relationships. Finally, there are areas where other agencies are natural leads and ASPE could play an important supporting role, such as in contributing data and knowledge to help populate the multiple federal repositories being built to support stakeholders in identifying, accessing, and using data for patient-centered research.

CHAPTER 3

PORTFOLIO CASE STUDIES

Chapter Executive Summary

The following case studies describe the activities and products for 12 OS-PCORTF awards. During a series of thematic webinar presentations, award leads and key collaborators reflected on their projects' contributions to enhancing the data infrastructure for patient-centered research and anticipated impact on the current and emerging data science field. The cases studies were organized according to four themes: 1) data interoperability and novel approaches to enhancing researchers' access to PPI; 2) data standardization; 3) enhancements to the NDI; and 4) data linkages. Case studies directly incorporate content and key learnings expressed during the webinars by award team representatives. The first three case studies describe work related to advancing data interoperability and novel approaches to enhancing researchers' access to patient-reported and patient-donated data.

- A team led by the NIH and collaborators from Duke University are developing, testing, and evaluating methods to validate and standardize patient-reported information with data obtained from EHRs in the context of the ADAPTABLE clinical trial.
- Award teams from AHRQ and ONC are developing and testing technical specifications, applications, and technical infrastructure to collect and seamlessly integrate PRO data into EHRs.
- CMS and NIH partners affiliated with the Sync for Science Program (S4S) have collaborated to leverage CMS' Blue Button 2.0 API and FHIR specifications to enable Medicare beneficiaries to donate their EHR and Medicare claims data for scientific research studies.

The next three case studies describe work related to data standardization.

- The CDC and FDA co-led an effort to standardize unstructured clinical text data through an NLP web service. The goal of their work is to improve the accuracy, timeliness, and completeness of the data and to provide simple, user-friendly tools for those with varying levels of NLP expertise.
- An FDA team has partnered with the Harvard Medical School to develop and test metadata standards and technical specifications to help researchers assess fitness-of-purpose of data across distributed research networks
- A multi-agency partnership of teams at the FDA, NIH National Center for Advancing Translational Sciences (NCATS), NLM, National Cancer Institute (NCI), and ONC collaborated to standardize queries and results processed through four existing CDMs and allowing export to open, consensus-based standards (HL7 FHIR and CDISC Study Data Tabulation Model in support of patient-centered outcomes research and submission to FDA.

The next three case studies describe work related to the NDI.

- A team led by the CDC has identified barriers preventing researchers from accessing the NDI. Logistical modifications to the NDI, and longer-term solutions, are already being planned and implemented based on the findings of the team's analysis activities.
- Another CDC team is working to develop the data infrastructure necessary to conduct linkages between mortality data and claims data or EHR data.
- The FDA is leading work to develop standard, repeatable, and efficient technical solutions for linking the NDI's death and cause of death data to large commercially and publicly insured populations. One award objective involves demonstrating the feasibility of linkage by examining a use case for evaluating the associations between select medications and death or cause of death.

The final three case studies describe work related to data linkages.

- A team at ONC is addressing the challenges around matching patient data across research, claims, and clinical datasets by improving matching algorithms to support PCOR.
- Teams at FDA, NLM, and ONC are addressing the need for a comprehensive approach to women's health by creating a new strategically CRN capable of addressing clinical questions on technologies uniquely affecting women.
- Teams at the CDC and ONC are collaborating to create resources to help researchers navigate the legal requirements and ethical considerations related to data use for research.

These awards also contribute to progress across multiple functionalities. Appendix G depicts the functionalities each award addresses.

The case studies underscore the importance of focusing product development toward addressing end-user need and identifying strategies to support the use of those products by the intended end-user community. In their discussion about the direction of future work, the TEP offered input on strategies to ensure that the work of portfolio has enduring value and maximizes the opportunities and mechanisms to support engagement between awardees and end-users. Progress in these areas is a key component of promoting the translation of portfolio products, outputs, and lessons learned to real-world implementations with the goal of achieving long-term sustainability. The TEP offered input on the following:

What Activities Can Be Pursued to Ensure That The Work Has Enduring Value?

- The TEP discussed ways ASPE could build in a framework to help guide awardees in planning for product sustainability. This framework could be integrated into the application process whereby the applicant articulates their plan for sustainability with a particular emphasis on the value of the work to patients, the end-user, and to the organization that collects or owns the data.
- Recognizing the importance of translation, the TEP provided ideas for increasing the spread and uptake of portfolio products. One suggestion for improving translation is to identify a Community of Practice (CoP) a priori for the award and the resulting assets. A CoP can help steer decision-making around the product dissemination, and identify challenges and barriers to address future product development. A CoP can also facilitate engagement and new partnerships among stakeholders who have not worked together previously, but who share a common data or infrastructure need.

- Because many of the awards in the portfolio are highly technical in nature, product uptake may improve if the general researcher community had a better understanding the products and how they can be applied in other contexts. OS-PCORTF awardees could be encouraged to develop resources such as a demonstration module that could quickly help potential users determine whether the product is a good fit for their needs.
- Identify interim products, such a methodologies and make them publicly available; for example, linkage algorithms should be made available, not just the resulting linked data set.

What Non-Traditional Opportunities Or Platforms for Products, Outputs, Or Learnings Could Be Leveraged to Support Engagement between Researchers And End-Users?

- Professional associations can serve as the vehicle for engaging providers during the early stages of a project to assess interest and secure buy-in for the resulting products.
- The OS-PCORTF portfolio should look for opportunities to promote marketing or business development during pre- and post-award. First, as part of idea generation and concept planning, there should be some intelligence gathering to gauge the need and interest in the work prior to pursuing an award. Second, to raise awareness among end-users, products should be promoted and marketed upon completion. Building this information gathering into idea generation and dissemination may improve translation of products that are designed to fit a stated need.
- Code-a-thons, datathons, and challenge competitions are a widely-used industry tool for quickly testing and developing innovative solutions to technical challenges. ASPE should continue to encourage the use of these low-cost collaborative learning activities to test and disseminate OS-PCORTF products that allows for some real-world testing and offers opportunities for refinement.
- In coordination with its agency partners, ASPE could consider creating a dissemination platform or website that describes all of the products developed through the OS-PCORTF and content areas they address to help people identify products that are fit-for-purpose.

Data Interoperability and Novel Approaches to Enhancing Researchers' Access to Patient-Donated Data for Real-World Evidence

Researchers' access to patient data must be balanced with patient privacy and security protections. Findings from the 2017 Evaluation pointed to the need to spur innovative technical solutions that support researchers' access to data in privacy preserving ways, as well as a need for increased patient education about the value of their data for both clinical decision-making and research. To address this gap, ASPE has invested in awards that utilize novel approaches to PPI data collection and data donation. The following three awards illustrate several such approaches. One award is developing, piloting, and testing tools and metadata standards to validate and assess the fitness-of-use of PRO data in large pragmatic trials. Another award is developing and testing apps and implementation guides to collect and integrate PRO data into EHRs using APIs. The third award is also using APIs, but to facilitate Medicare beneficiaries' ability to donate their data for research.

Case Study: Use of The ADAPTABLE Trial to Strengthen Methods to Collect, Validate, and Integrate Patient-Reported Information with Electronic Health Record Data



Agency & Partners: NIH, Mid-South Clinical Data Research Network (Vanderbilt University Medical Center), Harvard Pilgrim Healthcare, Duke University, Indiana University-Regenstrief

Webinar Presenters (Title, Affiliation): Wendy Weber (Branch Chief for the Clinical Research in Complementary and Integrative Health Branch in the Division of Extramural Research at the National Center for Complementary and Integrative Health, NIH); Emily O'Brien (Assistant Professor in the Department of Population Health Sciences and Assistant Professor in the Department of Neurology, Duke University School of Medicine)

In the words of one of the lead scientists on ADAPTABLE, Dr. Emily O'Brien, "There is increasing interest ... in using patient-reported data to identify clinical events of interest that occur outside of the health care system in which the patient was enrolled or received care." As people seek to use these data to support patient-centered outcomes research and care delivery activities, there is a growing need to facilitate electronic knowledge-sharing in the form of standards and best practices for PRH data capture and integration.

While conventional PROs capture information based on individuals' experiences or symptoms such as fatigue, sexual function, or functional status, PRH data consists of patient- and/or caregiver-reported health information that is often

ADAPTABLE (Aspiring Dosing: A Patient-centric Trial Assessing Benefits and Long-Term Effectiveness), the first pragmatic clinical trial conducted in PCORnet, will assess the effectiveness of two different daily doses of aspirin. The three-year comparative effectiveness trial has enrolled approximately 15,000 patients across six large health care systems.

captured in EHRs such as medications, hospitalizations, and comorbidities. PRH data can enhance and support clinical decision-making by providing a timelier, complete, and accurate picture of the influences contributing to an individual's health outcomes or lived experience. Patient reporting can also be instrumental in supplementing electronic records with important clinical and demographic information that is of use to investigators. Researchers conducting pragmatic clinical trials (PCT) may only have access to electronic data made available to them at their single trial site, when the patient may receive care from multiple other sites or specialists outside of the enrolling health system. Especially in these cases, integration of PHR measures can help to bolster the completeness and accuracy of EHR data.

Based on the ADAPTABLE trial (see text box), O'Brien and collaborators from the NIH designed and conducted a research project to: 1) evaluate the face validity of patient-reported data through systematic comparison to EHR data; and to 2) develop, pilot, and evaluate methods to validate and integrate patient-reported information with data obtained from the EHR. Distinctively, this trial relied on the capture of PRH data both at baseline and during follow-up throughout conduct of the trial.

As an initial step, the team conducted a literature review on the topics of data and metadata standards for patient-reported data being recorded in the EHR. This included literature that compared EHR and PRH data and studies that assessed existing data and metadata standards of PRH measure completeness, consistency, and fitness-for-use in EHR-based research. The team also documented any recommendations for future investigators who might seek to reconcile PRH and EHR data for use in their own research. Ultimately, this review effort led the team to conclude that this is a vastly underexplored aspect of research. Responding to the dearth of guiding literature, the team proposed an evaluation for the trial guided by their intention of building knowledge around data standards, and generating guidance that could inform future PRH data use for research.

The team selected a core set of clinical and demographic variables for comparison between EHR-derived and patient-reported data. Ultimately, they found high levels of agreement for data elements related to clinical events of interest (i.e., hospitalizations, chest pain, or a myocardial infarction). The team also observed strong agreement for data elements related to race and smoking status, but not among ethnicity data. O'Brien suspects this relates to the lack of consistent reporting for ethnicity variables in the EHR. The team also observed that patient-reported data tended to be more current than the EHR data, which they attributed to EHR data latency issues at individual trial sites. Taken together, the findings demonstrate the value of using these data to supplement clinical data in EHRs.

The team is disseminating these findings among a broad audience via multiple mechanisms. As of July 2018, multiple award outputs including the literature review findings and a report on PRH data/metadata standards in EHR-based trials were posted in the NIH Health Care Systems Research Collaboratory's Living Textbook. This unique resource offers a collection of knowledge regarding "special considerations, standard approaches, and best practices in the design, conduct, and reporting of pragmatic clinical trials."⁵³

The team developed a peer-reviewed manuscript that draws a direct comparison between EHR and PRH data (helping to demonstrate the validity of the latter) and offers recommendations for ways to integrate data from these two sources. The team has also produced two white papers on optimal approaches for data integration; these were based on roundtable sessions with leaders in the PRO, informatics, clinical trials, and federal agency communities. During meetings, the group explored topics related to: 1) best practices for capturing PRH data in PCTs; and 2) analytic approaches for integrating these data, once captured. Roundtable discussions led to the production of two papers that were submitted to the *Journal of the American Medical Informatics Association (JAMIA)* for possible publication.

The team has also produced a patient data assessment tool using PopMedNet that allows investigators to compare PRH and EHR information using a menu-driven query tool. According to O'Brien, feedback from the test site at Vanderbilt has offered positive indications that this patient-reported data assessment tool will support the conduct of PCTs; however, work is needed to encourage investigators' awareness and uptake of the tool. The team concluded that integrating EHR data into CDM tables can present challenges, and that PRH data appear to have limited sensitivity and specificity for some endpoints (i.e., ethnicity, race, and smoking status), relative to the "gold standard" set by EHR data.

Finally, the team applied through LOINC to create 50 new data elements in a standardized format so that data collected through the ADAPTABLE trial could be available for PCT researchers. LOINC accepted this proposal and an ADAPTABLE PRO data element panel has since been added to LOINC (June 2018 release) as well as the NIH CDE Repository (October 2018).

In accordance with the original aims, the team successfully evaluated and described the completeness, consistency, and fitness of PRH data for use in EHR-based research. Their work sheds light on existing gaps in the literature related to the documentation of data quality standards and has disseminated data elements and best practices for integrating PRH and EHR data. Their efforts also demonstrated that PRH data enable capture of relevant information about clinical events that occur beyond the point of care. As such, findings support the claim that PRH data can effectively supplement EHR data in ways that may enhance future clinical trial activities. This helps to demonstrate the value of patient-reported data to support robust and rigorous clinical trial activities. Beyond supporting efforts to "reform, strengthen, and modernize the nation's health care system" and to "foster sound, sustained advances in the sciences" in alignment with the stated HHS Strategic Goals.⁵⁴

Table 4 below describes the publicly available outputs of this award.

Table 4. Use of the ADAPTABLE Trial to Strengthen Methods to Collect, Validate, and Integrate Patient-reported Information with Electronic Health Record Data Award Products and Description

Publicly Available Product	Brief Description	Intended Audience	Link
Report on data and metadata standards for PRH data	Literature review and accompanying report on data and metadata standards for PRH data in EHR-based trials, including description of the standards and a dictionary describing each metadata element	Informatics research community, researchers conducting pragmatic clinical trials	http://rethinkingclinicaltrials.org/news/july-23-2018-new-report-summarizes-patient-reported-health-data-and-metadata-standards-from-the-adaptable-trial
Patient-reported data assessment tool	Patient-reported data assessment tool developed on the PopMedNet platform to enable efficient evaluation of concordance of patient-reported and clinical data elements	Researchers conducting pragmatic clinical trials, particularly those using national research networks	<i>Pending</i> (Folder location once available: https://github.com/PopMedNet-Team/ADAPTABLE)
Technical documentation for the patient-reported data assessment tool	Technical and user documentation for the patient-reported data assessment tool	Researchers using the patient-reported data assessment tool to evaluate concordance	<i>Pending</i> (Folder location once available: https://github.com/PopMedNet-Team/ADAPTABLE)
Addition of 50 ADAPTABLE patient-reported elements to LOINC	Addition of 50 patient-reported data elements from ADAPTABLE to LOINC to extend the application of research results to other population-based studies	PRH data users, researchers conducting pragmatic clinical trials, researchers using LOINC elements	https://r.details.loinc.org/LOINC/89070-7.html https://github.com/ADAPTABLETRIAL/REFERENCE/wiki
Data element update to NIH CDE Repository	Inclusion of LOINC ADAPTABLE Patient-Reported Data Elements in NIH CDE Repository	Informatics research community, PRH data users	https://cde.nlm.nih.gov/cde/search?selectedOrg=External%20Forms&classification=ADAPTABLE
Manuscript on patterns of concordance between PRH data and EHR data	Report of results from the evaluation of concordance between patient-reported and EHR data elements	Researchers conducting PCTs and other communities interested in the use of PRH data for research	<i>Pending</i>
NIH Collaboratory Roundtable White Paper #1	Companion perspective (with White Paper #2) titled "Capture of Patient-Reported Health Data in Pragmatic Clinical Trials: Report From an NIH Collaboratory Roundtable"	General medicine and/or informatics research communities	<i>Pending</i> (resubmitted to JAMIA)
NIH Collaboratory Roundtable White Paper #2	Companion perspective (with White Paper #1) titled "Analysis of Patient-Reported Health Data in Pragmatic Clinical Trials: Report From an NIH Collaboratory Roundtable"	General medicine and/or informatics research communities	<i>Pending</i> (resubmitted to JAMIA)

Case Study: Advancing the Collection and Use of Patient-Reported Outcomes through Health IT



Agency & Partners: AHRQ and ONC

Webinar Presenters (Title, Affiliation): Janey Hsiao, (Health Scientist Administrator at the Division of Health IT within the Center for Evidence and Practice Improvement, AHRQ) and Stephanie Garcia (Award Lead, Patient-Centered Outcomes Research Portfolio Manager, ONC)

Patient perspectives are central to health care decisions about prevention, diagnosis, treatment, and long-term care. PROs offer a complementary perspective to clinician assessments and may provide greater insight into health status, symptom burden, adherence, and quality of life. The incorporation of patient perspectives into health care research activities facilitates evidence generation regarding treatments and outcomes for diverse populations, including those often underrepresented in traditional research. Further, the consideration and integration of patient viewpoints allow investigators to better answer questions that are important to patients and their families. Resulting research outputs are then more directly relevant and useful for informing clinical care and improving health outcomes.

To effectively integrate patient perspectives into health care research and care delivery, standards and infrastructure need to be in place to support PRO data capture. To address the issue, AHRQ and ONC partnered on a project to enable standardized PRO data capture. Regarding standardization, ONC developed the PRO FHIR Implementation Guide that articulates how the FHIR specification can support the exchange of PRO data. Specifically, it includes a model illustrating those capabilities, technical specifications, implementation guides, and use cases that can inform the electronic capture and exchange of PRO data using the FHIR standard. The PRO FHIR Implementation Guide is publicly available.⁵⁵ Due to the expansive range of PRO domains, AHRQ and ONC selected physical function as the use case of the project. However, the PRO FHIR Implementation Guide can be applied to other domains of PROs.

AHRQ subsequently developed and pilot-tested user-friendly PRO data collection apps to support standardized PRO data collection and integration into EHRs in clinical settings. This work involved both modifying an existing app and developing a new app utilizing the PRO FHIR Implementation Guide.

AHRQ modified an existing OBERD⁵⁶ app to administer the PROMIS[®] physical function measures via computer adapted testing on a tablet. Using a FHIR server, the PROs are collected and integrated into EHRs in real time. The EHR vendor partners for the pilot test included Cerner and NextGen. Feedback from users of the app has been positive. One clinician enthusiastically endorsed the OBERD app, noting that the survey was extremely useful to the clinician-patient encounter and prompted important discussions that might not have otherwise occurred.

The new app for collecting PRO data was developed through a challenge competition supported by AHRQ, called the AHRQ Step Up App Challenge. Competitors were tasked with developing a user-friendly app capable of collecting standardized PRO data in various ambulatory care settings, including primary and specialty care. In phase one, the competitors submitted detailed proposals, and in phase two they developed the app using the FHIR implementation guide. The phase two grand prize was awarded to PRISM (PROMIS Reporting and Insight System from Minnesota). PRISM is intended to enhance the quality of clinical discussion between health care providers and patients with opportunities for engagement both inside and outside the clinical setting.⁵⁷ Beyond its data collection functionality, the PRISM app offers key features such as score trending, personalized patient recommendations, and educational materials. It also offers peer-group comparisons, which allow patients to compare their reported outcomes with those reported by other patients of similar age, sex, diagnoses, etc. The PRISM platform supports any PRO instrument and leverages FHIR for real-time data integration with EHRs. The PRISM team is working with AHRQ's contractor, MedStar, to pilot-test the app through private/public partnerships in the final, and currently active, phase of the competition.

The team identified a number of key lessons learned while completing this work. First, while EHR vendors are moving in the direction of adopting the FHIR standards, using a FHIR server is an efficient and secure way to integrate PRO data. Understanding the needs of the industry, as well as the needs and perspectives specific to individual sites that

plan to collect standardized PRO data, is essential for securing organizational commitment. Leveraging existing resources that organizations have invested in and using data collection instruments that align with organizational priorities will help mitigate compatibility issues.

Gaps and priorities discovered through the project work show where future work can continue. For example, there is still work to be done to enhance the utility of PRO data for providers and preserving patient privacy. In addition, providers reported feeling that PRO data are valuable but not always clinically useful. Providers also expressed that they do not always know how to interpret PRO data, when to take action, and where their liability begins and/or ends. Providers would also like to monitor how patients are doing over time and would like more customized data collection instruments. Lastly, funding is an essential part of PRO data collection and integration, and limited funding can hinder the PRO collection and integration capabilities of implementing sites.

The collection and use of PROs is valuable because these data elements are highly personal and relevant for each individual patient. In order for PRO collection and use to become common practice, patients and providers need to be engaged in the design of customized data collection methods. The pilot tests provide early evidence of the successful use of apps based on (the publicly available and pilot-tested) PRO FHIR Implementation Guide, as well as the impact of the team's work. The FHIR PRO Implementation Guide should continue to be tested, and as it matures, this will increase the likelihood that developers will use it and extend its application to capture different PRO domains. For example, the pilot sites have expressed interest in expanding the use of PRO data in pain management and oncology practices. Ultimately, the project products can help to support researchers who aim to explore innovative research questions, investigate the comparative effectiveness of various treatment options, and inform practices to better meet the needs of specific patient populations.

Table 5 below describes the publicly available outputs of this award.

Table 5. Advancing the Collection and Use of Patient-Reported Outcomes through Health IT Award Products and Description

Publicly Available Product	Brief Description	Intended Audience	Link
PRO Fast Healthcare Interoperability Resources (FHIR) [®] Implementation Guide (and pilot test)	Series of guidelines that can inform electronic capture and exchange of PRO data, in accordance with the FHIR standard	Health information technology developers	https://build.fhir.org/ig/HL7/patient-reported-outcomes/
Challenge competition winning app - PRISM: PROMIS Reporting and Insight System from Minnesota (and pilot test)	Tool that offers a range of functionalities (data collection, score trending, personalized recommendations, educational materials, etc.) – and leverages FHIR for real-time data integration with EHRs	Clinicians and patients	https://apps.apple.com/us/app/prism-for-minnesota/id1454714605 https://protect2.fireeye.com/url?k=ef4a00bc-b31e2997-ef4a3183-0cc47a6d17ccb35cdf3a2bd4855&u=https://play.google.com/store/apps/details?id=com.perkmotivation.PrismForMedstar&hl=en_US
AHRQ Step Up App Challenge Report	Describes the end-to-end challenge design and operations process as well as key success metrics and outcomes	People who want to host a challenge competition	<i>Pending</i>
AHRQ Step Up App Challenge Summary	Overview of the challenge competition	People who want to know about the challenge competition at a high level	<i>Pending</i>
AHRQ PRO Applications Pilot Test Report	Report on the pilot test approach and lessons learned	People who want to implement the PRO apps	https://digital.ahrq.gov/sites/default/files/docs/citation/ahrq-step-up-app-challenge-summary-2019.pdf
PRISM App Open Source Code	The code that was used to develop the PRISM app	App developers	<i>Pending</i>

Case Study: Technologies for Donating Medicare Beneficiary Claims Data to Research Studies



Agency & Partners: NIH and CMS

Webinar Presenters (Title, Affiliation): David Kreda (Consultant to the Sync for Science (S4S) Programs, Harvard Medical School, Department of Biomedical Informatics), Lori Pettebone-Koraganie (Blue Button 2.0 Program Manager, CMS, Office of Enterprise Data & Analytics)

As part of its Strategic Plan for Data Science, the NIH articulated its aims to help modernize the data ecosystem by supporting “storage and sharing of individual datasets” and “better integrat[ing] clinical and observational data” for use in research activities.⁵⁸ Further, NIH is pursuing efforts to enhance data management, analytics, and tools.⁵⁹ One such effort has been initiated by a team involving researchers from the NIH, as well as those from the ONC and Harvard Medical School’s Department of Biomedical Informatics. Since 2016, this group has aimed to coordinate adoption of data standards and a HIPAA-compliant consumer-facing workflow to enable consumers to transfer health information with a few clicks—a design meant to encourage and facilitate individuals’ donation of their health care data for research and use consumer app. The Harvard team worked collaboratively with an initial group of volunteer EHR vendors, including Allscripts, Cerner, eClinicalWorks, and Epic to implement the proposed design.⁶⁰

Exhibit 17 shows the S4S concept for patient-mediated health record data-sharing. The workflow begins with an end-user who is working with a consumer-facing app that uses the S4S API. When the app asks the user to share clinical data, she selects a provider from a list, and the app redirects the user to that provider’s patient portal, where her health information is stored. The HIPAA-compliant workflow requires that a vendor’s system indicate data elements that would be shared (or it may also offer a way to select a subset of the indicated data) for a period of time, such as one year into the future. When the user completes the approval process, she gets redirected back to the app and her data will be transferred to the selected app, if she provided approval. Because she may have more than one health care provider, she can use the app to select additional providers to retrieve data from each, allowing a more complete longitudinal medical record to be compiled.

Exhibit 17. S4S Patient-Mediated Data-Sharing



S4S specifies implementing a simple user-facing workflow coupled to a standards-based API mechanism. An EHR vendor enhances its patient portal with approval screens and the API so patients are able to share their clinical data electronically with third-party apps. The API must also be implemented by any app wanting to receive a patient’s data.

When researchers implement the API in a research app, they are able to request that participants share baseline clinical data without needing clinical staff time to help participants or, no less important, requiring providers to operate a clinical data warehouse.⁶¹ According to David Kreda, “The idea was to construct a fully digital process ... in most respects like e-commerce behavior, in which you can go to a portal and share your data.”

As Kreda notes, the S4S design shows how to operationalize “patients exercising their HIPAA rights to share their health data as they choose” efficiently and elegantly and provides EHR vendors with a way to satisfy the ONC 2015 Edition EHR technology certification requirement to publish an API.⁶² While EHR vendors were free to implement their own API to obtain certification, S4S proposed using the HL7 FHIR specifications and use of OAuth, a web standard authorization protocol used to access data. In promoting S4S, Kreda noted, “NIH opted for an all-in-one solution that could address research needs, consumer engagement, and clinical uses so as to be seen by vendors as a good bet for the long term. Now that new regulations addressing the Cures Act specify FHIR and OAuth for the next generation API, this is proving true.”

The initial S4S data scope is the Common Clinical Data Set, which covers structured data for patient medications, allergies, demographics, immunizations, problems, procedures, vital signs, smoking status, laboratory data, and structured documents, among other data.⁶³ The data are set to expand under the latest ONC proposal to satisfy the Cures Act. The expanded dataset, known as the U.S. Core Data for Interoperability (USCDI), will initially add free-text clinical notes and provenance but is expected to expand in subsequent certification cycles.

The *All of Us* Research Program, a historic effort to collect data from an unprecedented number of U.S. residents, has piloted S4S with 10 health care providers, and approximately 100 participants have shared their data.⁶⁴ Data quality assessment is underway. This includes assessing what data was transferred and how it compares in scope and quality to academic centers that sent clinical data in Observational Medical Outcomes Partnership (OMOP) file transfers. Although the pilot validated the technology and workflow, securing the willingness of participants to share data was less successful, perhaps because the pilot design, as well as the value proposition for patients, and larger *All of Us* workflow issues, impinged upon recruitment. On the consumer side, S4S is being used for clinical data acquisition by Apple’s [Health Record](#) app. Approximately 500 health care organizations agreed to honor patients who wish to send data to their iPhones.⁶⁵ As noted by Dr. Shafiq Rab (Chief Information Officer, Rush University Medical Center) on the Apple Health Record website, “The ability to obtain your health records from anywhere via your iPhone is nothing less than magic. Apple has put the power in the hands of patients, who are the most important stakeholders.”

As shown in Exhibit 18, this project work has resulted in the creation of numerous deliverables.

Exhibit 18. Award Deliverables

CMS Blue Button 2.0 + Sync for Science

SYNC FOR SCIENCE

syncfor.science

S4S Test Suite ✓
 ✓ **Test Suite**
 FHIR STU3 EOB and Coverage resources
 9 CMS profiles
 Automated sign-in to CMS Sandbox
 Initiated migration to MITRE’s Inferno Testing Tool
 Supported BB Team adoption of S4S APIs and payloads

Infrastructure ✓
 Created STU3 compatible infrastructure
 Deployed Test Suite and demos onto new Infrastructure
 Extend Research App API framework for researchers to use S4S

Data Visualization ✓
 App supports organize, filter, and search by data values
 Timeline and time series graphing, comparison, benefit views

Financial Data ✓
 ✓ **Dissemination**
 Online documentation of Financial data
 Online notebook-based tutorial for developers
 Test Suite deployment of clinical and financial data
 Presented Financial Data at HL7 Dev Days meetings
 Extend Research App API for researchers
 Dynamic client registration specification (Grieve/Mandel/Scrimshire)
 Argonaut specification for asymmetric client secrets

All of Us Research Program - CMS BB 2.0 - S4S Next Steps 09-Apr-2019

Blue Button and S4S. The CMS and NIH effort aims to incorporate the Blue Button 2.0 API effort into the broad S4S model, as both use the same FHIR and OAuth standards. Medicare beneficiaries could, for example, share their claims data from the MyMedicare.gov portal with apps like *All of Us* with the same S4S experience.⁶⁶ Kreda explains the value: “In some cases, it’s precisely the claims data that shows the continuity of care for an individual—even when clinical data are missing. So, that type of data can actually be used to stitch together what else is missing in the clinical data record.”

Infrastructure/Compliance Testing. To support EHR vendors as they implemented S4S, the Harvard S4S team built the S4S Test Suite to perform automated testing of vendor systems. According to Kreda, “It attaches to the API endpoints of a provider patient portal and can be used to ensure that all the data payloads and the OAuth responses are compliant with requisite standards.” As part of the CMS and NIH effort, the S4S Test Suite was expanded to cover CMS benefits and claims data types. This work will be available on an ongoing basis, by migrating to an ONC-funded test tool called Inferno, which will be used by developers to verify their API meets certification requirements under the 21st Century Cures Act: Interoperability, Information Blocking, and the ONC Health IT Certification Program Proposed Rule.^{67,68}

Shared Clinical and Financial Data Visualization. The Harvard S4S team has developed the Discovery app, an open source, data visualization web app for users to see data they contributed from both clinical and beneficiary portals in consolidated forms with the ability to zoom in on data elements using filtering by type, source, and time. Kreda adds that the focus is on “the ability to supply a narrative that an individual could make sense of their data” minus any health care interpretation. The hope is that presenting shared data in organized ways provides one aspect of value for patients who donate their data to research.” The app will also provide an extensible foundation for adding features, such as allowing patients to electronically annotate their data and send to researchers or their health care providers.

Table 6 below describes the publicly available outputs of this award.

Table 6. Technologies for Donating Medicare Beneficiary Claims Data to Research Studies Award Products and Description¹

Publicly Available Product	Brief Description	Intended Audience	Link
S4S API	FHIR API details to support the sharing of data via portals. Currently, they support the following use cases: clinical data, financial data, and imaging data	Consumers, Researchers, Clinicians, Developers	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	FHIR STU3 Explanation of Benefits and Coverage resources 9 CMS profiles Automated sign-in to CMS Sandbox Initiated migration to MITRE’s Inferno Testing Tool	App Developers	Inferno Testing Tool for testing server compatibility with S4S: https://github.com/sync-for-science/test-suite/wiki/Inferno-Migration-Guide Framework for testing S4S API implementations:

¹ Note that the content from this table has not been reviewed by the project lead.

Publicly Available Product	Brief Description	Intended Audience	Link
	Supports Blue Button Team adoption of S4S APIs and payloads		https://github.com/sync-for-science/test-suite
Research App API	Test Suite and demos onto new Infrastructure	Researchers	List of endpoints available to any clients working with the Research App API:
	Research App API framework for researchers to use S4S		https://github.com/sync-for-science/research-app-api
Blue Button 2.0	API software which allows beneficiaries to share their Medicare data with researchers. Includes Medicare parts A, B, and D and 4 years of claims history	Medicare Beneficiaries	https://bluebutton.cms.gov/
Blue Button API	Instructions for understanding and using the CMS Blue Button 2.0 API to connect patient health data to health apps and services	Developers	https://bluebutton.cms.gov/developers/
BlueButton 2.0 updates	Published updates to the open source code on GitHub	Developers	https://github.com/CMSgov
BlueButton 2.0 Sandbox	Front-end platform in Amazon Web Services	Developers	https://sandbox.bluebutton.cms.gov
MyMedicare.gov BlueButton Data Pipeline	Provides an Extract-Transform-Load data pipeline, which moves data from CMS' Chronic Conditions data warehouse into the CMS Blue Button FHIR Server.	Developers	https://github.com/CMSgov/bluebutton-data-pipeline

Data Standardization

Standardized data are foundational to effective and efficient data exchange, linking, analysis, and aggregation of clinical and other data for research and clinical decision-making. The implementation of standards was identified as an area of future work by the 2017 Evaluation that ASPE has addressed through awards intended to develop, implement, and maintain standards that improve the uniformity and consistency of data for research and clinical care. To demonstrate progress in data standardization, the following three case studies exemplify these data standardization efforts through the: 1) development of NLP processing techniques to standardize unstructured data for medical product safety surveillance and to standardize data capture in state-based cancer registries for research; 2) the creation of data quality metrics to assess fitness-of-use of data across distributed research networks to answer patient-centered research questions; and finally 3) the creation of tools that can harmonize data queries and results delivery across research networks with different CDMs, thus increasing the data available for PCOR.

Case Study: Development of a Natural Language Processing (NLP) Web Service for Public Health Use



Agency and Partners: CDC and FDA

Webinar Presenters (Title, Affiliation): Sandy Jones (Public Health Advisor with the Cancer Surveillance Branch, Division of Cancer Prevention and Control, CDC); Mark Walderhaug (Microbiologist, Center for Food Safety and Applied Nutrition, FDA)

Narrative text data is commonly found in clinical reports, pathology reports, and EHRs but is difficult for researchers to extract and analyze. Converting narrative text into standard, coded data for surveillance and research purposes requires expensive and time-consuming manual intervention for individual researchers. Through a collaborative project between the CDC and the FDA, researchers developed an NLP Web Service that will provide researchers with access to NLP tools to convert unstructured clinical data into standardized coded data. An NLP tool helps fill the gaps in human-to-computer communication by allowing computers and other technologies to interpret and understand human language. NLP tools assist computers in processing, analyzing, and generating larger amounts of data than humanly possible.⁶⁹ Sandy Jones from the CDC reports, "There is still a significant amount of clinical records that have unstructured text. We find that the use of natural language processing techniques...can increase the completeness, timeliness, and accuracy of the data." Additionally, Jones's team is developing techniques for the NLP tool end-users that can be utilized across multiple clinical and public health domains.

The NLP Workbench, Clinical Language Engineering Workbench (CLEW), is designed for three types of users: novices, intermediates, and advanced users. Novice users, who know very little about NLP but could benefit from NLP tools,

CLEW Requirements:

- Build on existing efforts
- Modular, freely available, open source
- Able to satisfy multiple use cases and applications for various clinical subdomains
- Encode data to multiple terminologies
- Provide interoperability between tools
- Service-oriented architecture using RESTful services

will have access to educational materials and a simple, intuitive platform with the ability to upload data and receive results.

Intermediate users, such as IT developers or organizations that are IT competent but lack specific NLP expertise, will be able to locate NLP Web Services from an end-user application. Advanced users with NLP expertise will be able to test different solutions and share their work with other NLP experts through the CLEW.

The project began with an environmental scan and literature review to identify published and existing open-source NLP algorithms, NLP methods, NLP tools, and other materials currently utilized by federal agencies, public health agencies, academic centers, commercial vendors, and PCORnet. Through this process, the team discovered several widely used NLP systems such as cTAKES, GATE, MedEx, MetaMap, and Stanford CoreNLP, as well as significant development activity around GATE and UIMA frameworks. Popular components

featured in these systems include time extraction, named entity recognition, abbreviation normalization, and co-reference resolution. The project team published a systematic review of their findings in the *Journal of Biomedical Informatics* to contribute to the literature.⁷⁰

After the environmental scan, the project team designed and developed the NLP Workbench Web Service, CLEW. The team built CLEW to the specifications outlined in the text box, based on best practices discovered through the environmental scan. The resulting cloud-based platform provides open-source NLP and machine learning tools. CLEW also provides clinical NLP services, as well as the opportunity to develop, refine, and experiment with NLP models. CLEW users can share new tools with the wider clinical NLP community, assemble NLP tools into a processing workflow, and generate training files for feeding machine-learning algorithms to develop language models.

The LAPPS Grid Platform was identified as the best fit for the CLEW through the environmental scan. A LAPPS Grid Platform is an open, interoperable web service platform for NLP research and development.⁷¹ The LAPPS Grid

Platform provides a user-friendly interface for NLP engineers to create pipelines from an expansive, preexisting catalogue of NLP tools, such as GATE, Stanford NLP, OpenNLP, and others. Another useful component of the LAPPS Grid Platform is the LAPPS Interchange Format (LIF). LIF is used to transform data between multiple frameworks and tools. For example, if GATE is used for sentence recognition, OpenNLP may be used for abbreviation normalization. "This was very advantageous and allows you to pick the best from different frameworks," says Jones.

Once CLEW was developed, it was tested via two pilot projects, one led by the CDC, and the other led by the FDA. The CDC chose to focus on cancer pathology for the pilot project domain because of the high percentage of cancer cases that require a pathologic confirmed diagnosis and report. In addition, pathology reports frequently use inconsistent terminologies, and laboratories are not required to store or collect data in a standardized method.⁷² Information from five key data elements (histology, primary site, behavior, laterality, and grade) were extracted from datasets provided by four national laboratories. "We also created a service to code the data for these five elements and mapped them to the national ICD oncology standard coding system," Jones reports. 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. Reports are submitted to spontaneous reporting systems" such as the FDA's VAERS and FAERS.⁷³ Together, VAERS and FAERS contain an abundance of structured and unstructured data. In 2016, VAERS received approximately 54,000 reports, and FAERS received 1.7 million reports detailing clinical information, temporal information, and temporal relationships. The project team found that their 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 dataset for training NLP models and uploaded the solution to GitHub for broader use.^{74 75}

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.

NLP tools have the potential to enhance the usefulness and value of patient data while advancing surveillance and research in myriad ways. "The ultimate patient benefit is having narrative data accurately extracted into structured information for the development of better public health models," Jones concludes. Additionally, researchers will have access to tools and datasets that enhance analytic capabilities, which in turn will accelerate clinical innovation. Federal agencies, public health agencies, and the research community have access to extensive and useful NLP web services, regardless of NLP expertise, including examples of inputs, outputs, schemas, formats, and computing environments.

Table 7 below describes the publicly available outputs of this project.

Table 7. Development of a NLP Web Service for Public Health Use Award Products and Description

Publicly Available Products	Brief Description	Intended Audience	Link
Clinical Language Engineering Workbench (CLEW)	A cloud-based, open source, NLP Workbench Web Service that hosts NLP and machine learning tools, clinical NLP services, and the opportunity for tool development	NLP Tool Users	https://www.cdc.gov/cancer/npcr/informatics/nlp-workbench/index.htm
CLEW Prototype Source Code and Documentation	Software code and installation instructions	NLP Users	CDC GitHub: https://github.com/CDCgov/NLPWorkbench FDA GitHub: https://github.com/FDA/
Environmental Scan and Literature Review	Compilation of all identified open-source NLP and machine learning tools, frameworks, and systems	Publication	Kreimeyer K, Foster M, Pandey A, Arya N, Halford G, Jones SF, Forshee R, Walderhaug M, Botsis T. (2017). Natural language processing systems for capturing and standardizing unstructured clinical information: a systematic review. <i>Journal of Biomedical Informatics</i> .
Annotated Dataset Created for use in Training NLP Models	Publication on how to create an annotated dataset for training NLP models	Publication	Foster M, Pandey A, Kreimeyer K, Botsis T. (2018). <i>Vaccine</i> . (29): 4325-4330.
CLEW Technical Report	Description of technical architecture design of CLEW and pilot use cases	NLP Users	<i>Pending: To be posted on ASPE website</i>
CLEW Lessons Learned Document	Compilation of all lessons learned from the project	NLP Users	<i>Pending: to be posted on ASPE website</i>
CLEW Users Guidance Document	Guidance document for NLP users to install and use CLEW	NLP Users	<i>Pending: to be posted on ASPE website</i>
CLEW Final Report	Final Report to ASPE about overall project accomplishments, challenges, and future activities		<i>Pending to be posted on ASPE website</i>

Case Study: Standardization and Querying of Data Quality Metrics and Characteristics for Electronic Health Data



Agency and Partners: FDA

Webinar Presenters (Title, Affiliation): Jeff Brown (Associate Professor in the Department of Population Medicine, Harvard Medical School and the Harvard Pilgrim Health Care Institute); Michael Nguyen (FDA Sentinel Program Lead, Deputy Director of the Regulatory Science Staff in the Office of Surveillance and Epidemiology at the Center for Drug Evaluation and Research, (CDER)).

The last several decades have seen an increase in public support for health data networks. However, there does not yet exist a central place to access or learn about growing data systems. Additionally, there are no overarching operational standards or metrics for describing the data. Individual networks offer varying data quality processes and definitions, which can limit researchers' ability to integrate data or to assess dataset fitness-for-purpose.

Jeff Brown (Harvard Pilgrim Health Care Institute) and Michael Nguyen (FDA) are leading a team of researchers at the FDA in efforts to address this issue. Brown describes this award as a pilot effort to operationalize some of the strong theoretical work guiding the field, which includes a seminal Data Quality Harmonization Framework developed by Dr. Michael Kahn (Professor of Epidemiology in the Department of Pediatrics, at the University of Colorado Denver) and colleagues.⁷⁶

Brown explains that the Data Quality Metrics (DQM) award serves as the "first step" to move from theory into practice by addressing gaps in data quality reporting, and by linking the framework to an "agnostic data quality data model." The award team has realized its work through design of a web platform that incorporates standards laid out in the data quality framework. "We're trying to come up with a way of implementing, through some infrastructure, so we leave behind something that the community can then use," says Brown.

Award work has been guided by a broad aim to create an infrastructure, which enables curation and exploration of standardized DQM across networks. The team loosely refers to this as a "database fingerprinting framework" that captures the unique characteristics of different databases. Using the team's newly developed tools, researchers can review unique database "fingerprints" and use related information to inform their decision-making regarding which data to use for a research activity.

To realize this aim, the team is creating a prototype DQM website, which provides data networks with a place to contribute and share their data quality checks, metrics, and metadata. Essentially, the DQM offers a platform for "describ[ing] data characteristics using common terms, despite how the data are defined locally" for disparate data sources. Brown notes that this approach does not require disruptions or amendments to existing network-specific processes. Using this website, end-users will be able to determine whether "the data fields in different networks have the same meaning, and if they can be used harmoniously." The platform utilizes a unique approach, whereby information is captured in a standard format from any source. It can then be pulled into a data visualization tool that helps to evaluate fit-for-purpose of different data sources. This generic DQM model can be populated with data points from different inputs. That way, even if the data do not necessarily adhere to a CDM, people can compare data from disparate sources.

Notably, as part of the DQM website, the team is developing and prototyping a new piece of architecture that allows functionality for authoring new metrics. On the site researchers can describe what they did to author the metric and then others can build or comment on it. It also allows investigators to explore available data and to identify the right data sources for specific studies, despite the unique governance policies to which each individual data network adheres. Data sources ideal for use in one context may not be fit-for-use in another. It can be difficult for researchers to tell the difference or to select a data source, unless they have "deep expertise in each of the networks and each of those sources. We need for not everyone to be a deep expert in all of those sources to be able to use the data," notes Brown.

For website development, the team is soliciting stakeholder feedback in order to ensure that the final product will reflect and respond to the needs, values, and interests of key end-users. Brown shares that, “We want to create the beta version of the platform with use cases, [to] show ... to the community to get feedback and then leave behind open-source tools that will enable the continuation of this work.” While piloting the tool, the team specifically solicited end-user feedback on the drivers that would result in (or barriers that would detract from) researchers participation in data contribution or website engagement activities. Ultimately, these points of feedback were considered and incorporated into the website design process.

Brown was quick to acknowledge that a number of key players comprise this community, and that they are doing significant and valuable work. “What we’re trying to do,” he said, “is find a way to bring that community together.” He believes the tool will support efforts to do this by facilitating the sharing of data quality-related learnings across stakeholder groups (including those involved with various data networks and those involved in research activities who seek to make use of the data).

While their work has already led to the achievement of their project specific aims, Brown and others have not lost sight of the broader potential for impact and real-world application. Referencing the substantial investments already made by HHS and others to support these networks and enhance their data sources, Brown comments on the potential for this work to enable the production of RWE. “If we’re willing to push forward and use real-world data to generate real-world evidence for decision-making, we have to do a better job of understanding the data.” His team’s effort reflects one approach to fostering general understanding of the underlying data, and of when to use them for valuable research efforts. Input from stakeholders demonstrate community recognition of this work’s value, and the results of expanding the platform may directly help policymakers to understand the data sources forming the basis of RWE.

Moving forward, the team has recommended creating a national coordinating center to support this work; doing so might involve creating a task force to oversee and curate the growing body of data quality metrics. Additionally, they hope that people will be increasingly incentivized (perhaps in the form of proliferating funding opportunities) to share their data, articulate their data characteristics, and create quality metrics that can be used to improve and enhance data use for high-impact patient-centered research.

Table 8 describes the publicly available outputs of this award.

Table 8. Standardization and Querying of Data Quality Metrics and Characteristics for Electronic Health Data Award Products and Description

Publicly Available Products	Brief Description	Intended Audience	Link
Data Quality Metrics Authoring and Querying Platform	Cloud-based, open-source tools (web application, flexible data model, visualization templates) that enable data quality metric authoring, capture of data quality metric output (measures), and support the evaluation and visualization of supplied measures	<ol style="list-style-type: none"> 1. Data quality stakeholders interested in development and implementation of data quality metrics generally and within multi-site distributed networks specifically. 2. Researchers and funders interested in evaluating fitness for use of various data sources and comparing data sources within to assess fitness for use. 	<i>Pending:</i> will be published on https://github.com/PopMedNet-Team/DataQualityMetrics by December 31, 2019

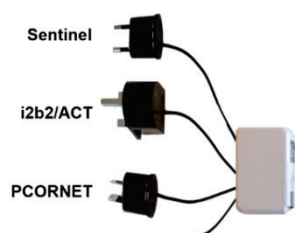
Case Study: Harmonization of Various Common Data Models and Open Standards for Evidence Generation



Agency & Partners: FDA, NCI, NCATS, NLM, ONC

Webinar Presenters (Title, Affiliation): Mitra Rocca (Associate Director of Medical Informatics, FDA); Scott Gordon (Senior Health Informatics Officer, FDA); and Kenneth Gersing (Director of Informatics in the Division of Clinical Innovation, NCATS)

CDMs are standardized, agreed-upon structures that are used to transform data obtained in observational databases into a comparable format with shared vocabulary. This allows entities with disparate data capturing systems and languages to harmonize their data, making it possible to perform systematic analyses. Various networks currently utilize their own CDMs. This multi-agency collaboration seeks to advance patient-centered outcomes research by developing a method to harmonize the CDMs of four such networks. CDM harmonization will allow researchers to ask questions using larger amounts of RWD, such as EHRs data or administrative claims data, than is currently possible by leveraging open standards and controlled terminologies.



The project team aimed to develop a common data architecture to serve as the intermediary between four frequently used CDMs illustrated in the figure: 1) the FDA Sentinel; 2) The Accrual to Clinical Trials (ACT) Network, which is housed alongside Informatics for Integrating Biology at the Bedside (i2b2) and is a federated network of sites from the National Clinical and Translational Science Award (CTSA) Consortium that has been created to significantly increase participant accrual to multi-site clinical trials; 3) PCORnet; and 4) OMOP). As part of developing this architecture, the project team collaborated with ONC to expand the Data Access Framework research FHIR standard by adding new extensions to the existing resource. An additional collaboration with NCI focused on registering data elements from these four CDMs

within the NCI Cancer Data Standards Registry and Repository (caDSR).

“The proposed architecture will allow a researcher to ask a question once and receive results from many different sources using different CDMs. Results will be available in multiple formats: for example, HL7 FHIR, and CDISC standards. CDISC SDTM is currently used by FDA for clinical trials study data submissions and is developed by the Clinical Data Interchange Standards Consortium.” says Dr. Mitra Rocca from the FDA. The adapter that will be built during the next phase of this project will allow the CDM data to be converted in accordance with HL7 FHIR standards, and a single query can be issued to all CDMs simultaneously (Exhibit 19).⁷⁷ The data format will appear as needed and be given to the FDA, the CDC, or other researcher submitting the query. “When you are in Microsoft Excel, you can say, ‘I want to save my file as a text file, as an Excel file, as Google Sheets, etc.’ The export is a variable of what you will select,” says Ken Gersing from NCATS. The main goal of this project, Rocca explains, is therefore “to develop a general framework, tools, processes, governance, and standards” necessary to enable this transformation and to facilitate the “curation, maintenance, and sustainability” of that architecture.

The project team created several tools, modules, and mechanisms. After developing the technical infrastructure, including the selection of the Extract Transform and Load tool, the project team composed a summative report detailing the selection process for the intermediary model. The team also created a “Query Builder,” which serves as the front-end interface that offers researchers a simple way to construct and issue their questions. The “Query Transformation” module transforms the original query into a version of the question that is compatible with each CDM. The Common Data Model Harmonization (CDMH) 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 CDCISC STDM format. “There is future possibility for HL7 FHIR exports, and much of the foundational work is underway which can support future implementation,” says Gordon.

The export tools rely on a transformational mapping process, which define the rules for transforming and exporting data through the CDMs. The project team created the CDMs-to-Biomedical Research Integrated Domain Group (BRIDG) mappings that align the data models and show the alignment of data elements and the existing gaps. The CDMH-BRIDG mapping process can be used to transform data to various output versions. CDMH/BRIDG-to-SDTM mapping provided rules for exporting results to the CDISC SDTM format, which is required for submissions to the FDA. CDMH/BRIDG-to-FHIR mapping illustrated the alignment of elements and gaps and allowed the results to be

implemented in FHIR. The project team initiated the development of the FHIR extensions and profiles needed to represent the harmonized CDM, including building out the extensions that filled previously identified gaps. Finally, RxNorm to national drug codes mapping enabled more complete content for FDA submission purposes. These mappings are nearing completion with the NCI caDSR. The BRIDG conceptual model has recently been updated with data elements identified in the gaps and is now fully aligned with CDMs.

These project products offer opportunities to various agencies within HHS by offering “benefits to various federal agencies and PCOR researchers,” says Ken Gersing. On an HHS-wide level, CDMH can offer tools to enhance interoperability, using nationally mandated standards, and could be utilized to help create a shared infrastructure across all HHS agencies. The FDA can benefit from enhanced regulatory decision-making through increased access to RWD, and the NIH can have the ability to facilitate research across multiple networks. The repository of CDM cross maps is beneficial to NCI, and it has also been made available for public use.

In the future, CDC could use these tools to support enhanced public health surveillance. More broadly, the research community will benefit from access to a multiple use case query tool and from tools that can support cohort discovery, natural history, protocol design, and amendment.

According to Gersing, the team “is harmonizing the CDMs, so that individual investigators can ask a single question that goes out to four CDM networks. We did succeed, but the lessons learned were that there is a lot of variation of each individual CDM, based on the creation of the data itself and technological variations.” For example, if CDMs use different databases (such as Oracle or MS SQL Server), the query will need to be written in multiple ways and tailored to each database. Ultimately, the team would like to decrease complexity and eliminate the need to build separate queries for each data model and proprietary database.

Gersing notes that HL7 FHIR is a widely appealing data exchange standard. Since it has proven particularly useful for academics, modifications have been made to increasingly respond to their research needs and interests. Explaining why the model is so distinct and appealing, Ken notes, “For example, in an EHR system, most data models are limited to examining a small set of tables. However, the HL7 FHIR data model can pull all of the tables within an EHR. Instead of drinking through a straw, it’s like drinking through a firehose—good or bad.” That said, he recognizes that this decrease in limitations associated with access to specific data elements may contribute to a parallel loss of data integrity and granularity.

Multiple agencies continue to move the model forward, and the project team hopes the work will be fully functional within 18 months, and eventually maintained in a primary home.

Table 9 describes the publicly available outputs of this award.

Exhibit 19. Overview of CDM Harmonization

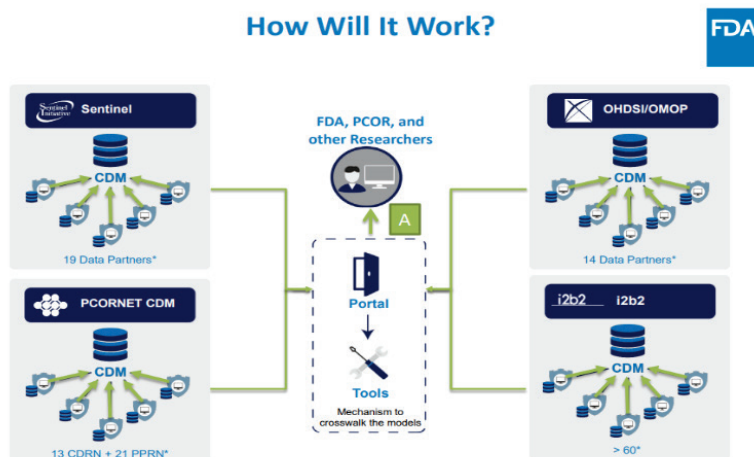


Table 9. Harmonization of Various Common Data Models and Open Standards for Evidence Generation Award Products and Description

Publicly Available Products	Brief Description	Intended Audience	Link
NIH CDE Repository- CDMH Common Data Elements	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	Researchers	https://cde.nlm.nih.gov/cde/search?selectedOrg=NCI&classification=PCORTF%20CDMHhttps://cde.nlm.nih.gov
CDMH FHIR Implementation Guide	Supports the mapping and translating observational data extracted for PCOR purposes into FHIR format	Researchers, Developers	https://build.fhir.org/ig/HL7/cdmh/
SDTM export tool	Exports record level results in the CDCISC STDM format	Researcher/Institution	https://www.cdisc.org/standards/foundational/sdtm
CDMs-to-Biomedical Research Integrated Domain Group (BRIDG) mapping	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	Researcher/Institution	https://bridgmodel.nci.nih.gov/
CDMH/BRIDG-to-SDTM mapping	Provides the rules to export results in support submissions to FDA	Researcher/FDA	https://github.com/cdmhproject/cdmh
CDMH/BRIDG-to-FHIR mapping	Proves the alignment and gaps of the CDMs to existing HL7 FHIR resources	Researcher/Institution	<p>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</p> <p>OMOP: https://cdebrowser.nci.nih.gov/cdebrowserClient/cdeBrowser.html#/search?programArea=0&contextId=6CB969CC-DD4B-1016-E053-F662850A40C7&classificationSchemeId=339F8634-199C-3A8A-E050-BB89AD431025</p> <p>PCORnet: https://cdebrowser.nci.nih.gov/cdebrowserClient/cdeBrowser.html#/search?programArea=0&contextId=6CB969CC-DD4B-1016-E053-F662850A40C7&classificationSchemeId=66589E50-F300-4B2E-E053-F662850A5342</p>

Publicly Available Products	Brief Description	Intended Audience	Link
			meld=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
BRIDG Updated with CDMH Data Elements	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	Researcher	https://bridgmodel.nci.nih.gov/download-model/bridg-releases
Registration of cross-mappings within the NCI Cancer Data Standards Registry and Repository (caDSR) and Enterprise Vocabulary Services (EVS)	<i>Pending</i>	<i>Pending</i>	https://cdebrowser.nci.nih.gov

Enhancing National Death Index Data

Mortality is a key outcome of interest for research on the effectiveness of health care interventions. The NDI is the only centralized data source containing information on both fact and cause of death for all deaths occurring within the U.S. The portfolio has made several investments to enhance the NDI to accelerate the availability and utility of NDI data for linkage to other research datasets. The three awards described in the next set of case studies demonstrate the broad range of NDI enhancements that support the study of emerging health topics and HHS policy priorities. To improve the timeliness and quality of data reported to the NDI, one award developed a strategic roadmap to address barriers to data collection and reporting of data from state data infrastructures to the NDI. Another award created the first-ever data linkage of EHR data from a nationally representative survey to the NDI and separately linked Medicare claims data to the NDI. A third project developed tools to link NDI data with data from claims-based research networks such as Sentinel to better study the association between medical products and mortality.

Case Study: Enhancing Data Resources for Studying Patterns and Correlates of Mortality in Patient-Centered Outcomes Research: Project 4 — NDI Workshop and Strategy Paper



Agency and Partners: CDC

Webinar Presenters (Title, Affiliation): Paul Sutton (Deputy Director of the NCHS Division of Vital Statistics, CDC)

The NDI is a federal centralized database of death record information provided by individual states’ vital statistics offices. The NCHS established the NDI as a resource to aid epidemiologists and other health and medical researchers with their research queries.⁷⁸ The NDI serves as a companion to the publicly available, de-identified NVSS mortality data sourced directly from states’ death certificates.⁷⁹ The NDI, unlike the NVSS, includes direct identifier information needed to support data linkages. It is also the only central data source offering information on the fact and cause of death for all deaths occurring in the United States.

The NDI is heavily utilized by public health researchers and is widely recognized as a valuable and rich resource for mortality data. However, a variety of barriers limit the use of the NDI. The award team, sought to identify the non-economic barriers (i.e., technical, operational, policy, etc.) facing researchers who seek NDI access. Although this award focused on non-economic barriers, it is important to recognize that cost is a significant barrier to many NDI end-users. For reference, NDI staff are funded by the revenue, and states are compensated for their mortality data.

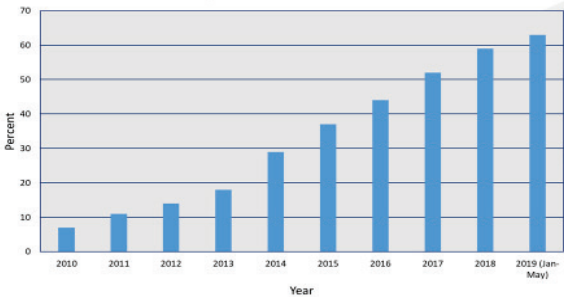
Initially, the team conducted an assessment for the purposes of identifying non-economic barriers potentially restricting NDI access. The team looked closely at factors that could influence access, particularly state-level statutes, regulations, and policies. Based on findings from the assessment, they generated an internal, summative report. This report informed a discussion to identify barriers.

Through the assessment, the team identified several barriers to NDI access. First, access to the NDI is granted only through an antiquated, 25-page paper-based application that is inconvenient and time-consuming. Supplemental documentation, such as the fee worksheet used to estimate the cost of the NDI data requested, is particularly burdensome. Additionally, the long application approval process can discourage applications from teams with shorter duration studies. Lastly, IRB approval for use of NDI data may not be necessary in select cases, and the process is currently under review.

Concerns of NDI data quality, timeliness, and availability have been cited as other barriers to access. End-users also observed issues related to data reuse, security, and linkage algorithms related to technical processes for very large files (over 500,000 records) provided by researchers seeking to utilize the NDI. Currently, a restriction limiting the use of NDI data for activities classified as medical research can severely limit the scope of its use. Individual states own all of the mortality data stored in the NDI. As a result, the various statutes, regulations, policies, and rules of each jurisdiction heavily influence NDI data submission and utilization. “To maintain a national dataset, we use the lowest common denominator, or the most restrictive rules that exist,” says Sutton. Additionally, the team discovered that inappropriate application and misinterpretation of HIPAA rules to vital statistics poses challenges for effective NDI data use.

While the timeliness of reporting to the NVSS has significantly, and continuously, improved since 2010 (Exhibit 20), the NDI is only updated twice a year—once in January and once in the fall— after the data have been finalized. Updating the data more frequently could

Exhibit 20. Percent of Mortality Records Transmitted to NCHS within 10 Days of the Date of Death



vastly improve data timeliness. NDI end-users also expressed a concern about the complexity of the outputs received from NDI.

The goal of the project was to identify barriers to NDI access. The impact of this work is already evident, as changes to address identified barriers are being implemented. For example, the NDI application is moving from a paper-based system to an electronic system. In addition, a pilot process to reduce the cost of NDI access for NIH grantees is being planned and should be operational in the near future. A transition to a quarterly, rather than biannual, reporting system will ensure that mortality data are timelier and more frequently available to researchers. Lastly, the NDI system IT infrastructure is increasingly stable, and more compatible with common servers. The award team hopes that the cost and contract structures are further examined so that appropriate alternative models can be developed and operationalized to reduce non-economic barriers. “Strengthening the NDI will strengthen the entire NVSS, as they are all linked together,” concluded Sutton.

Table 10 below describes the publicly available outputs of this award.

Table 10. Enhancing Data Resources for Studying Patterns and Correlates of Mortality in Patient-Centered Outcomes Research: Project 4 — NDI Workshop and Strategy Paper Award Products and Description

Publicly Available Products	Brief Description	Intended Audience	Link
National Vital Statistics System (NVSS) Database	Database which holds de-identified mortality data sourced directly from states’ death certificates.	Epidemiologists, Medical or Public Health Researchers	https://www.cdc.gov/nchs/nvss/index.htm
NDI Barriers report	Report which disseminated findings from barrier improvement workshop.	Researchers	<i>Pending</i>

Case Study: Enhancing Data Resources for Studying Patterns and Correlates of Mortality in Patient-Centered Outcomes Research: Project 1—Adding Cause-Specific Mortality to NCHS’ National Hospital Care Survey by Linking to the National Death Index



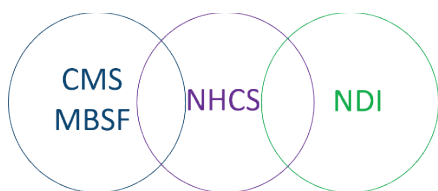
Agency and Partners: CDC

Webinar Presenters (Title, Affiliation): Carol DeFrances (Deputy Director of the Division of Health Care Statistics, CDC NCHS)

DeFrances states, “Mortality, and its reciprocal, survival, is the only health outcome relevant to all patients, settings, and disorders.” However, data platforms such as EHRs, claims databases, and other major population-based data platforms do not consistently include information on the cause, manner, or fact of death. A key objective of the OS-PCORTF is to create and support methodologies and infrastructures necessary to leverage relevant health databases to improve the quality of patient-centered outcomes research. It is therefore critical to develop the capacity to link these datasets in order to help fill in gaps and improve the quality of data. This ongoing OS-PCORTF award aims to develop the data infrastructure necessary to conduct these linkages.

This project is a collaboration between the Division of Health Care Statistics and the Data Linkage Program of the Division of Analysis and Epidemiology at NCHS. Specifically, the project seeks to link data collected in NCHS’ NHCS,

Exhibit 21. Datasets to Be Linked by the Project



including claims data and EHR data, with data from the NDI and the CMS Master Beneficiary Summary File (MBSF) (Exhibit 21). The NDI is a centralized database of death record information gathered from states' vital statistics offices.

The MBSF includes information on Medicare beneficiary characteristics, program enrollment type, and summarized Medicare cost and utilization. These linkages will improve the ability of researchers to study mortality and health care utilization following hospital care.

For each linkage, a number of demographic identifiers are used to ensure that records are properly matched for an individual (see Exhibit 22). Patient records are considered eligible for linkage if they meet the minimum threshold for completeness. Patient records are only included in the CMS and NDI linkage for “patients with at least one inpatient or emergency department encounter record reported by participating hospitals. Patients for whom only outpatient department encounters were reported are excluded from the NDI linkage” explains DeFrances. The methodology to conduct the linkages was created through a collaboration between the Data Linkage Program and NORC at the University of Chicago. It includes both deterministic and probabilistic approaches and is based on the foundational Fellegi-Sunter linkage methodology. The probabilistic approach estimates the likelihood that each pair is a match based on its computed pair score (the weighted sum of agreement/non-agreement status of each component variable comparison). The award team published linkage methodology and analytic guidelines for both the NHCS-NDI data linkage process and the NHCS-CMS data linkage process.^{80,81,82}

Thus far, results from data linkages have proven the success of these methodologies. Data from 95 hospitals were included in the 2014 NHCS linkage to 2014/2015 CMS MBSF data. After excluding ineligible patient records, 97.9 percent were matched to the CMS data for the sample of patients ages 65 and older. One-hundred and fifty-eight hospitals participated in the 2016 NHCS linkage to 2016/2017 NDI data. The number of eligible patient records found deceased in the NDI understandably varied by age: 20.4 percent of patients ages 65 and older, 6.2 percent of patients ages 45–64, 1.1 percent were matched for patients ages 18–44, and 0.3 percent were matched for those under the age of 18. Results are also available for the 2014 NHCS data linked to the 2014/2015 NDI, and are available through the NCHS Research Data Center.⁸³ Several additional projects have highlighted the analytical capabilities and uses of the linked data file, including the NHCS Alzheimer demonstration National Health Statistics Report (NHSR) and the NHCS Pneumonia demonstration NHSR.⁸⁴

There are several implications for these linkages. “The project provided longitudinal follow-up post-hospital care by linking two years of NDI data,” explains DeFrances. This allowed the project team to look at “hospitalizations at the end of the calendar year that resulted in deaths early in the following year, as well as estimates of post-discharge mortality at 30, 60, and 90 days.” The project also allowed researchers to identify comorbidities that may be associated with health- and hospital service-related outcomes. Additionally, because NCHS has a Data Linkage program with “substantial statistical and methodological expertise ... this project created the infrastructure to sustain data linkage ... for future years of [NHCS] data collection.”

Exhibit 22. Linkage Identifiers

NHCS-NDI Linkage Identifiers

- Social security number
- First name, last name, middle initial
- Date of birth (month, day year)
- Date of death (according to the discharge information on the hospital record)
- State of residence
- Sex

NHCS-CMS Linkage Identifiers

- Social security number
- First name, last name, middle initial
- Date of birth (month, day year)
- State of residence
- Sex
- Medicare Health Insurance Claim Number
- ZIP code of residence

The team learned several key lessons throughout this process. First, they noted that because linkage is an inference, different methods produce different results. As a result, careful documentation of methods is important to ensure transparency and replicability. Second, high-quality identification used to perform the linkage is critical. Finally, the project team explained that the success of the linkages between NHCS data and NDI and CMS data indicates that there is potential to explore other linkages between NHCS and other external data. As a result, they hope to increase participation in the NHCS, which would provide them more nationally representative data on hospital care utilization and mortality. Additionally, there are further gaps in the data used in the project that the NCHS aims to fill. “This project relied on structure-coded data for studying diagnoses,” explains DeFrances. “However, patient conditions and potential causes of death are not always captured by coded data, such as specific opioids involved in visits to the emergency department or drug overdose deaths.” The NHCS hopes to continue collection of EHR data, which would allow them to access contextual information that is often missed in UB-04 claims condition coding alone.

Several other OS-PCORTF awards have benefited from the foundations built by this work. One award that began in 2018 aims to address the opioid crisis by using structured and unstructured data to examine post-acute mortality and cause of death through NHCS-NDI data linkage with the addition of the DIM file, which includes drug-specific mentions from the death certificates.⁸⁵ Another award seeks to build on this work in order to identify opioid deaths, the drugs involved, and co-occurring mental health disorders. Other new awards are underway to link NHCS data to CMS claims and assessment files, including Part D to examine prescription drug adherence post-discharge, and to the U.S. Department of Housing and Urban Development data, to examine the role of federal social support programs on health outcomes and treatment efficacy for persons with stable housing, with the ability to focus on specific subpopulations, including persons with substance use disorders.

Table 11 below describes the publicly available outputs of this award.

Table 11. Enhancing Data Resources for Studying Patterns and Correlates of Mortality in Patient-Centered Outcomes Research: Project 1—Adding Cause-Specific Mortality to NCHS’ National Hospital Care Survey by Linking to the National Death Index Award Products and Description

Publicly Available Products	Brief Description	Intended Audience	Link
2014 NHCS data linked to the 2014/2015 NDI	Available in the NCHS Research Data Center. Includes information on 2014 NHCS inpatient and emergency department visits and 2014–2015 mortality information. Allows researchers to track mortality up to a year after the hospital visit.	Researchers	https://www.cdc.gov/nchs/data-linkage/nhcs-ndi.htm
Report on 2014 NHCS data linked to the 2014/2015 NDI	The report includes a description of the methods used for linkage and analytic guidance to assist researchers using the 2014 NHCS data linked to the 2014/2015 NDI.	Researchers	https://www.cdc.gov/nchs/data/datalinkage/NHCS14_NDI14_15_Methodology_Analytic_Consider.pdf
2014 NHCS data linked to the 2014/2015 CMS MBSF	Available in the NCHS Research Data Center. Provides the opportunity to conduct a vast array of utilization studies on the Medicare-eligible population.	Researchers	https://www.cdc.gov/nchs/data-linkage/CMS-Medicare-Restricted.htm
Report on 2014 NHCS data linked to the 2014/2015 CMS Medicare MBSF	The report includes a description of the methods used for linkage and analytic guidance to assist researchers using the 2014 NHCS data linked to the 2014/2015 CMS Medicare MBSF.	Researchers	https://www.cdc.gov/nchs/data/datalinkage/NHCS-CMS-Medicare-Linkage-Methods-and-Analytic-Considerations.pdf

Publicly Available Products	Brief Description	Intended Audience	Link
2016 NHCS data linked to the 2016/2017 NDI	Available in the NCHS Research Data Center. Includes information on 2016 NHCS Inpatient and emergency department visits and mortality information from 2016 and 2017. Allows researchers to track mortality up to a year after the hospital visit.	Researchers	https://www.cdc.gov/nchs/data-linkage/nhcs-ndi.htm
Report on 2016 NHCS data linked to the 2016/2017 NDI	The report includes a description of the methods used for linkage and analytic guidance to assist researchers using the 2016 NHCS data linked to the 2016/2017 NDI.	Researchers	https://www.cdc.gov/nchs/data/datalinkage/NHCS16_NDI16_17_Methodology_Analytic_Consider.pdf

Case Study: Enhancing Data Resources for Studying Patterns and Correlates of Mortality in Patient-Centered Outcomes Research: Project 2—Pilot Linkage of NDI+ to Commercially and Publicly Insured Populations



Agency and Partners: FDA

Webinar Presenters (Title, Affiliation): Robert Ball (Deputy Director, Office of Surveillance and Epidemiology—CDER, FDA); Wei Hua (Associate Director, Division of Epidemiology II, Office of Surveillance and Epidemiology, FDA)

The NDI, established by the NCHS, compiles information from state vital statistics offices into a centralized database. Included death records capture important cause of death information, and they are often recognized as the “best source for obtaining death and cause of death information for large population-based epidemiologic studies,” according to Ball.

While EHR data are commonly used for patient-centered outcomes research, EHR and claims databases do not often capture complete data regarding the fact, cause, and manner of death. This is especially true for deaths that occur outside of a hospital setting. This award, led by Ball and other colleagues at the FDA, is intended to enhance the FDA’s ability to evaluate associations between mortality and medical products. To do this, the team aims to “develop reusable, generalizable, and efficient technical solutions for linking EHR data and claims data for large, commercially and publicly insured populations to NDI data.” The team plans to assess feasibility of its approach, based on a specific use case: evaluating the ability to anticipate risk of sudden cardiac death (a mortality outcome) based on prior use of an antiarrhythmic medication (a potential cause of death). If successful, the creation of a mechanism for linking distributed health plan databases to the NDI may enable the FDA to better answer mortality-related questions of interest and importance. Specifically, this linkage mechanism could support the FDA’s distributed research network, Sentinel, which “has transformed the way researchers monitor the safety of FDA-regulated medical products, including drugs, vaccines, biologics, and medical devices has.”⁸⁶ Ball notes that this electronic system, which the FDA uses to pull data from multiple sources, “is made up primarily of commercially insured populations. This is not specifically a Sentinel project, but the success of this project would give FDA the ability to use NDI linkage in the context of the FDA Sentinel system.”

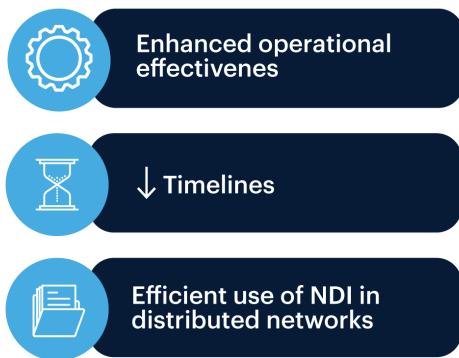
“What we’re trying to do is develop reusable, generalizable methods that multiple health systems could simultaneously use,” Ball explains. He does so, acknowledging that health systems encounter a number of other challenges in the processes of attaining and linking the needed data. For example, in practice, it can be difficult or cumbersome to secure the IRB and NDI permissions necessary to create linkages, particularly if multiple permissions

need to be obtained. It is also important to consider which health plan study subjects' data should be sent to the NDI for linkage and how to obtain NDI data without sharing personal health information with other sites or coordinating centers. Finally, it can be difficult to analyze these data and to provide systems-level, summary results.

The team has already completed work on its first deliverable, which involved finalizing the administrative workflow required to guide this multi-site research effort. Three planned deliverables will include creating distributed linkage processes for data exchange between health plans and NDI+, which includes cause of death codes (in addition to the death fact, date, state, and certificate number information captured in the NDI).⁸⁷ Planned work also involves selecting and retaining "best match" using standardized criteria and delivering a final report. This document will articulate methods, recommendations and lessons learned from this effort in hopes that future studies, using large distributed data networks and requiring NDI linkage, will be able to reference the process developed by the project team.

From their early efforts, Ball and colleagues have already learned that using a central IRB for a multi-site system can help to mitigate administrative burdens. Additionally, while it took the team eight months to complete the NDI

Exhibit 23. Short-term Implications of Project Work



approval process, their creation of templates and recommendations (intended to inform and support future study efforts) should help to shorten this timeline for other research and project teams. Now that this approval has been secured, activities linking multiple study sites to the NDI are officially underway.

The team identified several short-term implications of project work (Exhibit 23). Discussing the longer-term outcomes and impacts, Ball explains that "if the process is successful, and we're able to examine the risk of death associated with medical products through linkage with NDI and commercially insured populations, it will help realize an important part of the FDA's mission to protect public health." The linkages can help to evaluate potential safety issues, and the risk of death associated with certain medications, which may not be

identified in premarket clinical trials. Additionally, he notes that "under the 21st Century Cures Act, FDA has been tasked by Congress to develop an approach that would use real-world evidence to evaluate the efficacy of medical products in certain circumstances. Being able to evaluate the impact of medications on mortality is an important part of that effort."

The hope is that this work will help to enhance operational effectiveness and to minimize timelines for future data-linking efforts. Additionally, if successful, the work will facilitate efficient use of NDI data in distributed networks, including FDA's Sentinel.

Table 12 below describes the publicly available outputs of this award.

Table 12. Enhancing Data Resources for Studying Patterns and Correlates of Mortality in Patient-Centered Outcomes Research: Project 2—Pilot Linkage of NDI+ to Commercially and Publicly Insured Populations Award Products and Description

Publicly Available Products	Brief Description	Intended Audience	Link
Methods and processes to link distributed health plan databases to NDI data	Reusable, generalizable, and efficient administrative and technical solutions for linking large, commercially and publicly insured populations to NDI data	Researchers of studies that use large distributed data network and require linkage to the NDI data	Status: In progress

Supporting Data Asset Linkages

Linking data across data sources can provide a fuller picture of patients' care experiences, thereby enabling more robust analyses. Linking patient data from disparate sources requires technical solutions to reliably match patient records in a secure manner. Across the portfolio, there are some awards focused on actually creating linked datasets such as the NDI data files. However, the portfolio also includes awards that address some of the challenges to data linkages. The following three case studies provide examples of how the portfolio is addressing the challenges to linking data in order to follow patients across the care continuum to capture relevant health outcomes. One award is improving patient matching algorithms and security standards to accurately link and aggregate patient records.

Another award is developing the infrastructure to link registry, EHR, claims, and PRO data to support research on the effectiveness of medical devices. Lastly, a cross-agency award produced tools and resources to help researchers navigate the legal and ethical requirements that govern data use for patient-centered research.

Case Study: Security and Privacy Standards for Patient Matching, Linking and Aggregation (PMAL)



Agency and Partners: ONC

Webinar Presenters (Title, Affiliation): Stephanie Garcia (Patient-Centered Outcomes Research Portfolio Manager, ONC)

Patient data are stored and analyzed across multiple sources such as claims, EHRs, and clinical registries. Tracking and matching patients across these sources, however, is challenging for both health practitioners and researchers, and when done inaccurately can result in incorrect conclusions from PCOR and may threaten patient safety. ONC sought to address challenges regarding matching patient data across research, claims, and clinical datasets for PCOR. Specifically, the project targeted four areas: 1) improving matching algorithms; 2) improving data quality; 3) expanding data sharing; and 4) data standardization. ONC hosted competitions to solicit ideas from the public and engage them to test solutions that address key problems in each of the four target areas.

A standardized or single industry-approved approach for patient data-matching does not currently exist. Notably, this project did not seek to create a standardized process; instead, the research team aimed to create user-friendly tools, engage the community, and increase awareness to support the industry in its efforts to identify the best approach for patient data matching, aggregation, and linking.

Target Area 1: Improving Matching Algorithms

Matching algorithms enable researchers to link and aggregate patient records across datasets to support their research. In order to improve the success rate of patient matching, the team collaborated with the Kaiser Permanente Center for Health Research to develop a gold standard dataset and approach for the evaluation of patient matching algorithms. The project team tested the performance of widely adopted patient-matching algorithms by comparing patient record pairs, identified as matches by the matching algorithms, to the known duplicates in the gold standard dataset, and then computed evaluation metrics. The team leveraged the Identity Matching Adjudicator Collector (IMAC) tool to manage the adjudication process when determining if records were true matches. The IMAC tool is

public and currently used by researchers across academic, research, and commercial settings. In addition, the team developed a FHIR-based, Patient Matching Test Harness tool which allows researchers to: 1) inspect patient match results; 2) quickly create test data for sharing; 3) incorporate results from clinical and claims feeds and PCORnet; and 4) model new patient attributes.

ONC held a Patient Match challenge competition, open to members of the public, to “bring about greater transparency and data on the performance of existing patient matching algorithms, spur the adoption of performance metrics for patient data matching algorithm developers, and positively impact other aspects of patient matching, such as deduplication and linking to clinical data.”⁸⁸ Challenge participants developed their own algorithms and applied them to a synthetic dataset (provided by ONC) that included demographic data elements. The top performing teams were interviewed and asked to discuss the data elements they deemed most useful for performance improvement activities. The competition engaged over 140 participating teams. As a result, ONC extended the availability of the dataset and scoring platform to give others the opportunity to test their algorithms.

Target Area 2: Improving Data Quality

The second target area of the project was to improve data quality at the health care organization level, specifically, managing patient demographics and reducing duplicate records. ONC oversaw a pilot implementation to test the Patient Demographic Data Quality Framework (PDDQF). This framework is based on the CMMI Institute data maturity model and provides guidance on standardizing policies, procedures, and practices across a broad spectrum—from a local community level to an entire health care system.⁸⁹ The team published the PDDQF as an interactive guide, giving users an overview of the framework and walking them through implementation tools and considerations.⁹⁰

Target Area 3: Expanding Data-Sharing

The third target area of the project focused on expanding data sharing to facilitate patient engagement. As part of this effort, ONC collaborated with the OpenID Foundation to form the Health Relationship Trust (HEART) working group, which seeks to create security layers in order to allow for greater patient control over data sharing.⁹¹ The working group developed several specifications based on existing, broadly accepted technologies such as OAuth, User-Managed Access (UMA), FHIR, and OpenID Connect. Earlier specifications did not provide robust guidance for securely handling patient data; therefore, as part of its work, the HEART working group provided details on how these servers should be implemented in a health care environment. Following this, the project team supported the Move Health Data Forward Challenge, which invited participants to create an API solution using the HEART specifications to allow individuals to securely authorize movement of their health data to destinations of their choice, thereby enabling data sharing for research.

Code-a-Thon Tracks:

1. Identity Management and APIs
2. Metadata Tagging and Policy Expression
3. Data Aggregation and Linkage

ONC also developed a Patient Data Toolkit with applications that use FHIR for multiple use cases to create longitudinal patient records. The tools were then tested during the “Proving the Potential: A Health Data and Standards Code-a-Thon.” Examples of these tools include a FHIR-based API to make data accessible to tools built for older standards, and a tool that can merge two FHIR-based patient records. Several of these tools are currently available on the HealthIT.gov website.⁹²

To explore novel approaches to data-sharing, ONC held a “Use of Blockchain in Health IT and Health-Related Research” challenge,⁹³ and a “code-a-thon” that were used to gather feedback from the public about the application of Blockchain in health care.⁹⁴ The challenge yielded approximately 77 white papers exploring the topic. During the “Code-a-Thon” contestants used open-source distributed ledger technology to propose solutions for one of three health IT tracks (see textbox).⁹⁵

Target Area 4: Data Standardization

The fourth target area focuses on data standardization. ONC developed a Patient Record Scorecard Application that calculates a completeness score for an FHIR-based patient record. The score is based on the extent to which data fields, defined by FHIR, are complete in the EHR, and includes negative assertions such as “no allergies.” The ONC project team then supported the “Oh, the Places Data Goes” challenge to identify real-world provenance problems, such as tracing a data element’s origin, and establish viable solutions that demonstrate high technological merit, availability, feasibility, implementation, and impact.⁹⁶

ONC collaborated with the Federal Health Architecture to sponsor a workshop with public and private sector stakeholders to review challenges, share successes, and generate new ideas for provider directory standards and solutions.⁹⁷ The workshop resulted in follow-on activities including development of a National Plan and Provider Enumeration System (NPPES) open API; an OAuth Client and Server Library to manage and track the authentication and authorization for users of the Open API for NPPES; and a Provider Data Toolkit for creating and validating health care provider data for integration with health information exchange organizations and research networks.

Lessons Learned and Next Steps

This multi-faceted project yielded many products that facilitate matching of patient data across numerous datasets to support PCOR. The majority of the ONC tools and resources developed and tested through this project are publicly available through GitHub, the HealthIT.gov website, and other avenues, depending on their proprietary status.⁹⁸

The ONC team learned several lessons throughout the course of the project. First, they observed a reluctance from health care providers to discuss challenges regarding matching patient data. They also noted barriers around clinician engagement, during pilot testing, notably resulting from a lack of necessary resources. Second, Implementing new technologies, such as blockchain, in health care contexts can prove challenging because there is limited existing guidance to follow. Third, most patient matching systems require proprietary integration platforms. Standardization across patient matching systems will be necessary to facilitate patient-matching across the health care ecosystem; however, their proprietary nature makes this difficult. Lastly, there is a need for synthetic patient data that reflects real-world situations that can be used to test patient-matching solutions.

The team’s work indicates the need for ongoing work and improvement across each of the four target areas of the project. However, the products developed by the project (which include data security considerations) are already enabling researchers to link patient data across sources, increasing the volume of data available for PCOR.

Table 13 describes the publicly available outputs of this award.

Table 13. Security and Privacy Standards for Patient Matching, Linking and Aggregation (PMAL) Award Products and Description

Publicly Available Products	Brief Description	Intended Audience	Link
Patient Matching Algorithm Challenge (Artifacts)	This patient match challenge competition, open to members of the public, invited participants to develop and apply algorithms to a synthetic dataset provided by ONC.	Public	https://www.patientmatchingchallenge.com/
Patient Matching Test Harness	The Test Harness tool allows researchers to: 1) inspect patient match results; 2) quickly create test data for sharing; 3) incorporate results from clinical and claims feeds and PCORnet; and 4) model new patient attributes.	Researcher end-users	https://github.com/mitre/ptmatc
Identity Matching Adjudicator Collector (IMAC) Tool			

Publicly Available Products	Brief Description	Intended Audience	Link
Patient Demographic Data Quality Pilot (Whitepaper)	The team published, as an interactive guide, an overview of the PDDQF, which is based on the CMMI Institute data maturity model.	Researcher end-users / Public	https://www.healthit.gov/playbook/pddq-framework/introduction/
HEART Working Group Specifications & Move Health Data Forward Challenge (Artifacts)	In collaboration with OpenID Foundation, the team formed the HEART working group, which created specifications and guidelines regarding the secure handling of patient data.	Researcher end-users / Public	https://openid.net/wg/heart/
Tools for the Creation of a Longitudinal Patient Record: <ul style="list-style-type: none"> • FHIR-based Patient Data Toolkit • Record Merge Tool • Claims Data Importer Proving the Potential: A Health Data and Standards Code-a-Thon (Artifacts)	<p>The Patient Data Toolkit includes applications using FHIR for multiple use cases to create a Longitudinal Patient Record.</p> <p>The “Code-a-Thon” tested the tools developed to create longitudinal patient records. Several of these tools are currently available on the HealthIT.gov website.</p>	Researcher or Clinician end-users / Public	https://www.healthit.gov/techlab/testing_and_utilities.html
Use of Blockchain in Health IT and Health-Related Research Challenge (Artifacts) & Blockchain in Healthcare Code-A-Thon (Artifacts)	These activities were held to explore novel approaches to data-sharing. The challenge yielded approximately 77 white papers. Code-a-Thon contestants used open-source distributed ledger technology to propose solutions for set health IT tracks.	Researcher end-users / Public	https://www.cccinnovationcenter.com/challenges/blockchain-in-healthcare-code-a-thon/
Patient Record Scorecard Application	This application calculates a completeness score for an FHIR-based patient record. The score is based on the extent to which data fields, defined by FHIR, are complete in the EHR.	Researcher or Clinician end-users / Public	https://github.com/mitre/scorecard_app
Oh the Places Data Goes Health Data Provenance Challenge (Artifacts)	This challenge was hosted to identify real-world provenance problems, and establish viable solutions.	Researcher or Clinician end-users / Public	https://www.healthit.gov/topic/grants-contracts/announcing-blockchain-challenge
Provider Data Toolkit: <ul style="list-style-type: none"> • National Plan and Provider Enumeration System (NPES) Open API • OAuth Client for NPES Open API • Provider data validation tools Provider Directory Workshop (Artifacts)	The workshop included public- and private-sector stakeholders, and resulted in follow-on activities, including the development of these tools for creating and validating health care provider data.	Research end-users	https://www.healthit.gov/topic/scientific-initiatives/pcor/patient-matching-aggregating-and-linking-pmal

Case Study: Developing a Strategically Coordinated Registry Network (CRN) for Women’s Health Technologies



Agency and Partners: FDA, ONC, NLM

Webinar Presenters (Title, Affiliation): Danica Marinac-Dabic, (Associate Director for the Office of Clinical Evidence and Analysis for the Center for Devices and Radiological Health, FDA); Robin Taylor (Technical Information Specialist, NLM/NIH); Stephanie Garcia (Patient-Centered Outcomes Research Portfolio Manager, ONC)

A CRN consists of multiple data sources, such as EHR data, claims data, and patient-generated data, which can be linked to the registry for health research purposes. The infrastructure of a CRN can support a range of research, from post-market surveillance of therapies and evidence generation to the development of innovative therapies and broadening of indications of existing medical products. However, current registries are often costly to maintain and their singular focus on a disease, condition, or therapy means they are incapable of addressing questions involving multiple therapies or health devices. The purpose of the Women’s Health Technologies CRN (WHT-CRN), a collaboration between the FDA, NLM, and ONC, is to create an efficient platform to generate clinical evidence to support quality improvement and clinical research in multiple device areas unique to women’s health. This CRN addresses the need for an infrastructure to study health technologies in women on a national level and builds on recommendations by the National Medical Device Registry Task Force to develop and use CRNs to study devices.

The WHT-CRN consists of four CRN data domains (registries, claims, EHR data, and PROs) with detailed metadata and descriptive statistics for each CRN describing population demographics, disease presentation, device exposure, follow-up duration, and relevant clinical outcomes captured. Participation in the WHT-CRN is governed by the CRN Partnership Framework, a formal commitment of participation among registry stakeholders, a formal data governance document, and a sustainability plan for continuation of the CRN after the initial funding period.

To promote the development of the Women’s Technology CRN, the project teams developed sets of CDEs for each of four registries focused on specific health conditions (Pelvic Organ Prolapse, Uterine Fibroids, Stress Urinary Incontinence, and Sterilization/Long-Acting Reversible Contraception); harmonized and standardized minimal core datasets for each of these health conditions; developed an initial harmonized set of CDEs across three of the four registries; and developed an HL7 WHT-CRN FHIR Implementation Guide, which was being pilot-tested.⁹⁹

Development of CDEs for Individual Registries

To develop consensus around clinical concepts for each of the four health conditions, the FDA leveraged the existing Medical Device Epidemiology Network (MDEpiNet) Public Private Partnership and its Coordinating Center at Weill Cornell Medical College (www.mdepinet.org). The multi-stakeholder teams convened four clinical working groups based on the clinical areas with representation from regulatory bodies, professional societies, patients, industry, federal partners and academia. The working groups developed four lists of clinical concepts leading to CDEs, approximately 120 in total, including the Unique Device Identification (critical for device-specific research) and finalized the list through a formal multi-round Delphi process. Ultimately, four datasets, including data dictionaries, were developed: one for each clinical area. The informatics working group, led by NLM, translated the clinical concepts into registry domain data elements that included (definitions, permissible values, and context) and created the claims library.

Harmonizing and Standardizing Core Datasets Across Registries

The informatics and clinical working group worked together to harmonize CDEs across all four clinical conditions and CDEs were harmonized across registries. They identified “common concepts” that appeared in two or more of the clinical datasets. They refined that list of common concepts until they arrived at an initial set of data elements harmonized across all three registries.

CDEs in the initial harmonized dataset were tied to terminology with custom value sets in the Value Set Authority Center (VSAC). The CDEs were also represented as a single composite form within the CDE Repository. This core minimum set of data elements was then included by ONC in the HL7 FHIR implementation guide.

Acessa Health™, a medical device company that developed a method to treat uterine fibroids laparoscopically, is already leveraging the core minimum dataset for collecting data on radiofrequency laparoscopic fibroid treatment, demonstrating that this project has the potential to sustain after the current funding period. At least two other manufacturers of devices indicated for the treatment of SUI are in the process of leveraging the AQUIRE- SUI family of registries for evidence generation. Finally, the academic partners have begun exploring the use of WHT-CRN for the next round of proposals to NIH and PCORI which also demonstrates the potential for sustainability beyond the OS-PCORTF funding.

A Data and Tools Report that describes the process of creating the harmonized set of data elements and the creation of the implementation guide, 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, are in progress. These two activities are not public documents but a summary will be included in the final report.

Implementation Guide Development

The ONC developed the HL7 FHIR Implementation Guide: Women's Health Technology Coordinated Registry Network (CRN) Version 2.0 STU 1 Ballot 2¹⁰⁰ to serve as 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 HL7 WHT-CRN Implementation Guide is currently being finalized after a robust comment period.

Importantly, the CRN project leveraged the work of several prior OS-PCORTF awards. It leveraged the U.S. Core Implementation Guide which is based off the ONC 2015 Common Clinical Data Set requirements. The U.S. Core profiles have broad adoption in the real world which is a promising indication for interoperability. The project also leverages two other implementation guides—the PRO FHIR and the Structured Data Capture implementation guides—to support PRO measure questionnaire and questionnaire response. Finally, the project used the Data Access Framework for researchers to support data access from multiple data sources. The project team was also able to leverage the accomplishments of the Medical Devices and Technology Network across a variety of clinical areas.

The project pilot-tested two new modules, Stress Urinary Incontinence and Pelvic Organ Prolapse, for inclusion 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.¹⁰¹ A patient-facing SMART on FHIR mobile application was designed by ONC to collect PROs and was pilot tested under the Stress Urinary Incontinence Module of the AQUIRE registry.¹⁰² The second module pilot tested the core minimum data elements for Pelvic Organ Prolapse and support the implementation and refinement of specifications in the WHT-CRN Implementation Guide in a test environment.

Lessons Learned and Next Steps

The multi-agency project team and its public-private sector partners learned several valuable lessons throughout the course of the project that could be useful for future work. First, a substantial amount of time should be dedicated to the principles of registry development. This project sought to change the culture by working as a CRN, rather than an independent registry developer, and to serve as a model for future CRNs. Cultural change comes through collaboration across disciplines, as demonstrated in the multi-stakeholder working groups. Second, CRN system design depends on what data elements will be captured, how data elements will be captured, and how the data will be used and interpreted. The project team encourages researchers to employ a use case-driven approach to gather valuable insight into how the system will be utilized and what the system will do, while illustrating a practical or

applied vision for the solution. Engaging directly with the end-users as requirements emerge is particularly valuable as it allows developers to hear the challenges, requirements, restraints, and opportunities in the users' own words.

Building a national infrastructure for studying women's health technology provides a formal, integrated method of answering queries. The WHT-CRN now has access to over 550,000 records across the four CRN domains (i.e., registry claims, EHR, and patient-generated data) to support the FDA's Center Devices and Radiological Health and help clinicians make better-informed decisions, more quickly and less expensively. The WHT-CRN has the potential to impact the way that health care is delivered, health technologies are evaluated, how evidence is generated and synthesized and how the lessons learned throughout the lifecycle of technology help promote innovation.

Table 14 below describes the publicly available outputs of this award.

Table 14. Developing a Strategically Coordinated Registry Network (CRN) for Women's Health Technologies Award Products and Description

Publicly Available Products	Brief Description	Intended Audience	Link
Stress Urinary Incontinence Surgery Module	The Stress Urinary Incontinence (SUI) Surgery Module is a new component of the AQUIRE Quality Improvement registry developed in conjunction with Women's Health Technologies Coordinated Registry Network , a project of Medical Device Epidemiology Network (MDEpiNet). The Module is aimed at collecting quality data on all types of SUI surgeries including slings, Burch procedure, and periurethral bulking agents.	Clinicians	https://www.augs.org/clinical-practice/acquire-sui-surgery-module/
WHT-CRN Common Core Dataset	The four clinical core datasets were harmonized to create the WHT-CRN common core dataset. CDEs in the harmonized dataset were tied to terminology with set values in the VSAC and represented as a single composite form within the CDE Repository.	Researchers	https://cde.nlm.nih.gov/
Women's Health Technology Medical Device Epidemiology Network Website	In collaboration with the MDEpiNet, 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.	Researchers	http://mdepinet.org/womens-health-crn/
HL7 WHT-CRN FHIR Implementation Guide	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.	Developers	http://hl7.org/fhir/us/womens-health-registries/2019May/



Agency and Partners: ONC, CDC

Webinar Presenters (Title, Affiliation): Stephanie Garcia (Patient-Centered Outcomes Research Portfolio Manager, ONC) & Alexa Limeres (Confidentiality Administrator, Public Health Analyst, CDC)

The use of patient-level data is central to effective PCOR. To effectively inform policy and practice, this research must be based on data that truly reflects the characteristics and considerations of diverse patient populations. Research involving patient-level data requires additional consideration of the sensitive patient privacy and security issues specific to use of these data. While it is essential for these data to be available to researchers, this access must be provided in ways that ensure patient protection without compromising data granularity or integrity.

Notably, data use and access issues are influenced by broader policy and regulatory structures. Data sources are subject to different and nuanced laws and policies; this can make it difficult for researchers to navigate the complex requirements guiding the use of patient-level data. This causes confusion for researchers, patients, and providers and can dissuade researchers from using these data to conduct their work.

A team of representatives from the CDC and ONC collaborated to create resources to help researchers navigate the legal requirements and ethical considerations related to data use for PCOR. In conducting this work, the ONC team initiated two separate work streams. The first of these focused on providing stakeholders with relevant guidance regarding the responsible use and protection of electronic data. The team created the Legal and Ethical Architecture for PCOR Data, which addresses the legislative and policy features that readers (including researchers) should take into account as they seek data access for their research. This resource, currently available at [HealthIT.gov](https://www.healthit.gov), includes five key chapters.¹⁰³ The chapters provide: 1) an overview of the Architecture; 2) a discussion regarding the legal and ethical significance of the data, including an articulation of PCOR data features and related security or consent issues; 3) content linking different legal and ethical requirements; 4) tools and frameworks facilitating navigation of these legal and ethical requirements; and 5) tools that support stakeholders in mapping research data flows to these legal requirements.

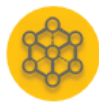
The Architecture was designed for use by a broad audience of stakeholders, including researchers, clinicians, health system representatives, IRBs, contracting officers, compliance officers, and their offices of counsel. The Architecture can also be used to help inform the work of a broader audience that might include students, lawmakers, research funders, policy analysts, patient advocates, and federal or state regulatory bodies. The tool provides decision-support diagrams illustrating key characteristics of data used for research including source, access, content, subject, security, legal status, identifiability, use/purpose, consent/authorization, and the related considerations for each of these characteristics. It also maps the complex and intersecting legal implications of data use for PCOR, and it includes a summary of relevant statutes and regulations, which were current as of the publication date (October 15, 2018). For many diverse groups, the Architecture offers a guide that supports analyses of legal requirements and introduces ethical considerations of note.

ONC led a second work stream, referred to as the Patient Choice Technical Project, which involved developing and testing the technology standards needed to capture electronic consent and enable interoperable exchange of consent across different systems. Issues affecting broad adoption of electronic consent management included: 1) reliance on paper consent directives that do not provide the necessary structured data for end-to-end automated electronic consent management; 2) health IT systems that process consent differently and lack adequate interoperability; and 3) different patient consent models across health information organizations. As previously mentioned, this is further complicated by the many and sometimes conflicting federal and state confidentiality/privacy and consent laws.¹⁰⁴

The project detailed use cases for basic choice consent for treatment, payment, and operations, as well as research. Also included were use cases for capturing granular consent via advanced electronic methods for granular level data segmentation. The project also conducted demonstrations to pilot test the standards for capturing and sharing individual patient consent preferences in the above referenced use cases.

As detailed in Exhibit 24, this project has notable and direct implications for the field. Further, it builds on existing efforts given its alignment with other efforts that encourage data sharing for research, including the shared Nationwide Interoperability Roadmap,¹⁰⁵ and the HHS Open Data Initiative.¹⁰⁶

Exhibit 24. Implications and Impact



Technology-Neutral Common Structure Legal and Ethical Architecture for Researchers

The project provides education, concepts, decision tools, data use scenarios, and a framework to enable researchers to obtain data from health IT systems, clinical trials, and surveys while protecting patient privacy.



Data Flow Maps and Use Cases of Relevance to PCOR

The project explored use cases and analyzed the data flows to identify the relevant statutes, regulations, and policy considerations for PCOR.



Standards and Technical Specifications for Electronic Patient Consent

The project addressed a patchwork of consent requirements for each healthcare classification and a lack of national patient consent standards by investigating standards, use cases, and best practices for basic and granular consent to enable research.



Pilots for Individual Consent

The project conducted a pilot across multiple stakeholders to obtain feedback for basic and granular consent.

Parallel to the ONC work, a collaborating team at CDC developed the Legal and Ethical Framework to Use CDC Data for PCOR. This resource offers a curated package of tools that can be used to inform decisions about using CDC data to support research activities. Notably, the Framework was created by a workgroup that involved several outside contractors with expertise on the legal and ethical implications of data use for PCOR.¹⁰⁷ The team undertook a gap analysis to identify key factors that could be addressed to expand CDC's capacity to support PCOR data infrastructure. This included an examination of ethical considerations and laws governing the release and use of CDC public health data for specific research applications. The team identified three gaps: 1) secondary uses of public health data; 2) sharing of de-identified data; and 3) multiple IRB reviews of public health research.

Notably, this work responds to a prominent concern the team learned from researchers: that researchers and other stakeholders (despite having access to their own counsel) may have trouble applying existing guidance, to inform their specific real-world scenarios; this is particularly true in instances when multiple laws apply and shape data sharing or use issues. For example, teams at the CDC often have to account for multiple legal and ethical parameters because the various datasets in use may be subject to different laws and regulations. Given this layering of multiple legal authorities, the team recognized a need to create a Framework to guide researchers through complicated situations.

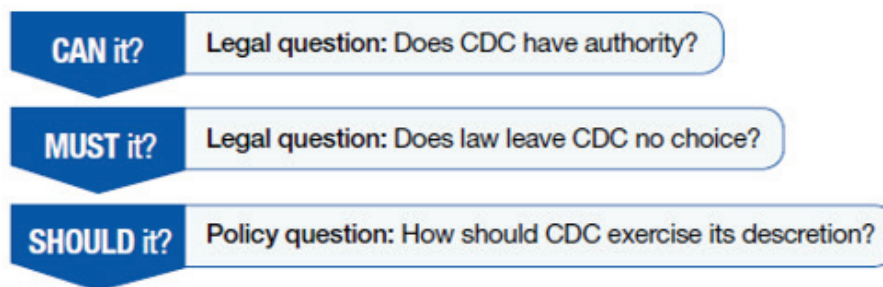
Upon articulating these gaps to other stakeholders, including CDC representatives, the team received feedback that these barriers are particularly prominent for researchers using multiple datasets and or de-identified datasets. Data linkage and de-identification are crucial components of patient-centered research value, which suggests the potential impact of these challenges and the exigency of efforts to address them.

The team took this feedback into account in designing and creating its final work products. These resources contextualize the Framework itself, and they provide background information regarding the use of CDC data for

PCOR, as well as relevant laws and ethical implications. They also present several data use scenarios, which can be referenced by audiences considering potential research projects and collaborations.

One such product is a newly created decision-tree that operationalizes elements of the Framework, and organizes these into an instructive diagram that guides users in applying the Framework. The team also presented additional guidance in the form of a separate legal framework. One tool presented in this resource (Exhibit 25) helps demonstrate different approaches to decision points related to data sharing activities. More specifically, this tool presents readers with a series of questions to consider, and it provides instructive guidance for determining legal authority when navigating a complex regulatory environment.

Exhibit 25. Three Basic Questions When Deciding to Share Data



The team’s extensive work led them to articulate a number of key lessons learned throughout their work. While they acknowledge the need to embrace innovative approaches and technologies, they also commented that these may emerge in unanticipated ways and may not adhere to existing laws and policies. Similarly, project timelines and standards development timelines do not always align. Generally, it is important for researchers and other stakeholders to account for these multiple layers of external forces, which requires ongoing monitoring of relevant laws and regulations.

Given the complexity of the policy and ethics issues involved, the team thoughtfully identified gaps that can be accounted for as people prioritize future efforts to enable data use for research. The team noted that minimum requirements for electronic consent models have not been established, and further testing is necessary to develop these models and related technologies. Additionally, there are gaps in the consistency with which different stakeholders implement existing standards and policies. Finally, the team raised a nuanced point about increased risks facing small, culturally unique populations. Given the likelihood that their data might be easily re-identified, researchers need to take additional steps to ensure the privacy and security of these populations. This added level of effort may ultimately dissuade researchers from working with these groups and their data, rendering this dynamic worthy of consideration.

Responding to some of the key lessons learned, the team built multiple tools into the Framework with the intention of systemizing the process. Their work helped to prove that it is possible to implement technology-enabled parsing of electronic health data, and it provided guidance for those seeking to use sensitive, patient-level data to support research activities. The team articulated the importance of considering multiple laws that apply to a given dataset, and of developing standardized definitions for the field (which will require future effort). Other potential future efforts might include building on the work of these two teams to expand the tools and resources available to researchers, and to develop CDC-wide terms, policies, and definitions to ensure consistency in the work. Ultimately, this work can help to educate researchers about the policy and regulatory environment dictating their use of data to support PCOR. Further, it can enable them to navigate this environment in an informed manner so that they are better enabled to conduct patient-centered and public health research.

Table 15 below describes the publicly available outputs of this award.

Table 15. PCOR: Privacy and Security Blueprint, Legal Analysis, and Ethics Framework for Data Use, & Use of Technology for Privacy Award Products and Description

Publicly Available Products	Brief Description	Intended Audience	Link
Legal and Ethical Architecture for PCOR Data (“Architecture”)	Collection of tools and resources that help researchers and others navigate an overview of the legal requirements related to data use, sharing, and disclosure for PCOR.	Public, Researchers, Lawmakers, Clinicians, Students	https://www.healthit.gov/topic/scientific-initiatives/pcor/legal-and-ethical-architecture-patient-centered-outcomes-research-pcor-data-architecture
Patient Choice Technical Project	Resource which details use cases for basic and granular consent, and for capturing patient consent via advanced electronic methods.	Researchers	https://www.healthit.gov/topic/scientific-initiatives/pcor/privacy-and-security-framework-pcor-psp
Legal and Ethical Framework to Use CDC Data for PCOR	Resource which offers a curated package of tools that can be used to inform decisions about using CDC data to support research activities.	Public, Researchers, Lawmakers, Clinicians, Students	https://aspe.hhs.gov/system/files/pdf/259016/PCOR_Legal_508_2.pdf
Decision Tree	Operationalizes elements of the Framework and organizes these into an instructive diagram that guides users in applying the Framework.	Public, Researchers, Students	https://aspe.hhs.gov/system/files/pdf/259016/PCOR_Legal_508_2.pdf (page 38)
Separate Legal Framework Three Basic Questions When Deciding to Share Data	Demonstrates different approaches to decision points related to data-sharing activities. Presents readers with a series of questions to consider and provides instructive guidance for determining legal authority when navigating a complex regulatory environment.	Public, Researchers, Students	https://aspe.hhs.gov/system/files/pdf/259016/PCOR_Legal_508_2.pdf (pages 35-42)

Discussion

The case studies illustrate OS-PCORTF contributions related to data interoperability, data standardization, enhancing federal datasets, and supporting data linkages. The case studies also underscore the importance of focusing product development toward addressing end-user need and identifying strategies to support the use of those products by the intended end-user community. While these case studies have identified the many ways in which award outputs expand PCOR data infrastructure, a key indicator of their impact on real-world research is the degree to which the products are ready to be disseminated, accessed, and used. In assessing the degree of uptake, the TEP advised conducting an assessment of products that are ready for translation. The TEP

discussed dissemination strategies to ensure the work of the portfolio has enduring value and to maximize opportunities and mechanisms to support engagement between awardees and end-users through novel and non-traditional approaches.

What Activities Can Be Pursued to Ensure That The Work Has Enduring Value?

The TEP suggested several approaches to ensuring the OS-PCORTF products have “enduring value” and a lifecycle that extends beyond the funding period. The TEP discussed ways ASPE could build in a framework to help guide awardees in planning for product sustainability. This framework could be integrated into the application process whereby the applicant articulates their plan for sustainability with a particular emphasis on the value of the work to patients, the

end-user, and to the organization that collects or owns the data. ASPE could provide technical assistance in the form of education or training to applicants who may be unfamiliar with developing sustainability plans.

- *Capstone Awards.* Capstone projects offer an opportunity to strengthen the output for real-world implementation. For example, ASPE could give consideration to Capstone awards aimed at increasing the generalizability of products, such as applying record linkage methodologies to other datasets or encouraging awardees to pursue partnership with new implementation sites in an effort to reduce barriers to highly localized implementations that limit generalizability. Efforts like these can help ensure that the products have utility beyond the immediate team that created the product and that identified gaps are resolved.
- *Promoting Translation.* Recognizing the importance of translation, the TEP provided ideas for increasing the spread and uptake of portfolio products. Some suggested establishing a Community of Practice (CoP) a priori for the award and the resulting products or tools. A CoP can serve several roles in helping the work to achieve sustainability. A CoP, or related community engagement component (e.g., an external stakeholder workgroup) can steer decision-making around product dissemination and future product development. A CoP can also facilitate engagement and new partnerships among stakeholders who have not worked together previously, but who share a common data or infrastructure need. Not all of the portfolio products will lend themselves to translation, so careful assessment will be needed to identify those products with the most potential for successful uptake. A CoP can play a role in determining which products are selected for more targeted translation efforts based on criteria such as the potential impact, generalizability, feasibility, and end-user need.

Because many of the awards in the portfolio are highly technical in nature, product uptake may improve if the general researcher community had a better understanding of the products and how they can be applied in other contexts. Award teams could be encouraged to develop resources such as a demonstration module that could quickly help potential users determine whether the product is a good fit for their needs. Another avenue to support enduring value would be to make interim products, such as methodologies, publicly available; for example, linkage algorithms should be made available, not just the resulting linked data set.

What Non-Traditional Opportunities Or Platforms for Products, Outputs, Or Learnings Could Be Leveraged to Support Engagement Between Researchers And End-Users?

As noted above, a key component of successful translation projects is engagement with the end-user community. Building on existing dissemination efforts to increase OS-PCORTF portfolio product and tool uptake more broadly, the TEP discussed novel and non-traditional approaches to engage with researchers and end-users. Given the diverse set of products arising from the project awards, a “one size fits all” approach to dissemination will not be as effective as an approach that tailors the content, timing, mode, and format of dissemination to end-users. As noted in the portfolio assessment, most projects are disseminating products by posting software code to open-source platforms such as GitHub, publishing implementation guides through standards development organizations, and pursuing publications in peer-reviewed journals or as white papers on agency websites. Many award leads have also raised awareness of work products by presenting at academic conferences. Despite these ongoing award-specific dissemination efforts, ASPE may be uniquely positioned to enhance more traditional dissemination activities, as well as engage with agency partners, external researchers, and private industry in collective dissemination efforts to raise the visibility of the OS-PCORTF portfolio as a whole.

- *Conferences.* In its role as coordinator, ASPE could take the OS-PCORTF “on the road” by supporting efforts to feature the work of the portfolio collectively at various national conferences. For example, ASPE could coordinate an OS-PCORTF booth in addition to the individual agency-specific posters and presentations.
 - *Professional Associations.* While researchers, and increasingly patients, are frequently thought of as the primary end-user for many portfolio products, the provider/clinician community still represents the main source of data collected for research. Professional associations can serve as the vehicle for engaging providers during the early stages of a project to assess interest and secure buy-in for the resulting products.
 - *Entrepreneurs-in-Residence.* Agency leads provided positive feedback on the HHS Entrepreneurs-in-Residence Program, citing their contributions as integral to the success of their award. Using this as a model, ASPE could look for opportunities to leverage the expertise of an Entrepreneur-in-Residence or establish a similar position in marketing or business development. The role would focus on conducting market research to gauge the need and interest of a particular project prior to award. Likewise, this individual could support awardees with raising awareness among end-users upon completion of the work. Building this type of information gathering into the early phases of idea generation may improve translation of products that are designed to fit a stated need.
 - *Extramural Researchers.* The functionality using federal databases for research entails enabling greater access to publicly funded datasets to federal and external researchers alike. ASPE could explore alternative mechanisms for funding (e.g., microgrants, vouchers) external researchers to come and use federal data (e.g., NDI files available through the NCHS Research Data Center) for their own research inquiries. Agency leads could use this as an avenue to gather feedback from researchers about their experiences accessing and using the data and to identify any salient challenges.
- *OS-PCORTF Product Portal.* Requirements under the Foundation of Evidence-Based Policymaking Act of 2018 to identify the data agencies intend to collect, use, or acquire to facilitate the use of evidence in policymaking, can serve as a prototype that ASPE could replicate to more easily facilitate agency and researchers access to federal data assets for patient-centered research. In coordination with its agency partners, ASPE could consider creating a dissemination platform or website that describes all of the products developed through the OS-PCORTF and the functionalities and content areas they address to help people identify products that are fit-for-purpose.
- *Developer and Vendor Engagement.* Code-a-thons, datathons, and challenge competitions are a widely used industry tool for quickly testing and developing innovative solutions to technical challenges. Several awards have conducted code-a-thons and hosted challenge competitions as part of their projects. ASPE should continue to encourage the use of code-a-thons by the agencies as a means to bridge the gap between the researcher and technical community. One potential approach for leveraging and sustaining the outputs of challenge competitions involves creating a website for posting the products of the challenge participants, not just the winners. This website could be marketed or shared with venture capitalists and developers who may be interested in taking-on the work. Relatedly, there are opportunities to leverage existing vendor marketplaces as a means to increase the potential uptake of award products. For example, Logica (formerly the healthcare Services Platform Consortium) is a marketplace and EHR vendor app galleries/marketplaces, which is another way for awardees to get their products into the hands of developers who can market and support provider implementation.

Conclusion

The 12 case studies illustrate OS-PCORTF contributions related to data interoperability, data standardization, enhancing federal datasets, and supporting data linkages as well as making strengthening data capacity across all five functionalities of the OS-PCORTF Strategic Framework.

The case studies also underscore the importance of focusing product development toward addressing end-user needs and identifying strategies to support the use of those products by the intended end-user community. In considering future work, the TEP offered strategic opportunities for accelerating the translation of portfolio products and promoting more targeted approaches to engaging with patient-centered research stakeholders.

As ASPE plans for the future, there are multiple opportunities to bolster the sustainability of the portfolio. Sustainability planning can be integrated into the award planning process to encourage more proactive identification of end-users who will utilize the products upon project completion. This may include establishing a CoP for the award to promote translation of award outputs; conducting market research to identify end-user need; encouraging real-world implementation with multiple partners to increase generalizability; and engaging with the provider, developer, vendor, and investor communities to raise awareness of award products.

APPENDICES

Appendix A. Strategic Framework Milestones by Functionality

Table A-1. Strategic Framework Milestones by Functionality

Functionality	Milestones
Use of Clinical Data for Research	
	<ul style="list-style-type: none"> Establish services and tools to support data access, querying, and use, including privacy-preserving analytics and queries. These services and tools would be leveraged nationally and are not likely to be developed by the private sector.
	<ul style="list-style-type: none"> Develop support services and tools that can be leveraged nationally and are not likely to be developed by the private sector; these tools would test the quality of unstructured and structured data to answer PCOR questions.
	<ul style="list-style-type: none"> Develop standards that support secure, electronic query of structured data across clinical research and delivery systems, including standards for open-source access.
	<ul style="list-style-type: none"> Develop and test metadata standards that describe data quality.
	<ul style="list-style-type: none"> Create a policy framework for privacy-preserving access and querying of clinical data by researchers conducting PCOR, and policies that govern the use of the services that support data access, querying, and use.
	<ul style="list-style-type: none"> Develop a policy framework for ensuring clinical data used for research is of "research grade."
Standardized Collection of Standardized Clinical Data	
	<ul style="list-style-type: none"> Support the development of a set of research CDEs in specific gap areas and support development of a governance structure for CDE harmonization
	<ul style="list-style-type: none"> Support the development of repositories/portals for CDEs, standards for utilizing CDEs for research, and services to allow researchers to easily utilize standardized components.
	<ul style="list-style-type: none"> Support research and/or crowd-sourced methods to determine which of the standardized collection components and services are most valuable.
	<ul style="list-style-type: none"> Create policies to promote the adoption and use of valuable standardized collection components and services.
Linking Clinical and Other Data for Research	
	<ul style="list-style-type: none"> Leverage existing standards and support the development and balloting of needed standards for patient data linkage.
	<ul style="list-style-type: none"> Establish HHS policies that promote appropriate data-linking based on the framework noted in the milestone above.
	<ul style="list-style-type: none"> Create a policy framework to facilitate patient data linkage in accordance with existing laws.

Functionality	Milestones
Collection of Participant-Provided Information	
<ul style="list-style-type: none"> ■ Support the development of tools and services that can be leveraged nationally and are not likely to be developed by the private sector. These tools and services will facilitate the collection and exchange of PPI, including national services for electronic capture and management of PPI and release of data for PCOR. 	
<ul style="list-style-type: none"> ■ Support the development of a core set of standards for the collection and integration of prevalent use cases of PPI for PCOR, by leveraging existing standards and filling gaps. 	
<ul style="list-style-type: none"> ■ Create policies and share best practices for collection and integration of prevalent use cases of PPI for PCOR. 	
Use of Enhanced Publicly Funded Data Systems for Research	
<ul style="list-style-type: none"> ■ Support the enhancement of strategic publicly funded data systems (including CMS data) to facilitate their access and use, and ease retrieval of data for research purposes. 	
<ul style="list-style-type: none"> ■ Support the further development of key federally initiated data systems for research. 	

Appendix B. Featured OS-PCORTF Awards

Table B-1. Awards Included in the Impact Report

Agency	Award Title
Agency for Healthcare Research and Quality (AHRQ)	
	<ul style="list-style-type: none"> ■ Capstone for Outcomes Measures Harmonization Project
	<ul style="list-style-type: none"> ■ Enhancing Patient-Centered Outcomes Research (PCOR): Creating a National Small-Area Social Determinants of Health (SDOH) Data Platform
	<ul style="list-style-type: none"> ■ Harmonization of Clinical Data Element Definitions for Outcome Measures in Registries
Centers for Disease Control and Prevention (CDC)	
	<ul style="list-style-type: none"> ■ Augmenting the National Hospital Care Survey (NHCS) Data through Linkages with Administrative Records: A Project
	<ul style="list-style-type: none"> ■ Childhood Obesity Data Initiative: Integrated Data for Patient-Centered Outcomes Research Project (CODI)
	<ul style="list-style-type: none"> ■ Enhancing Identification of Opioid-Involved Health Outcomes Using Linked Hospital Care and Mortality
	<ul style="list-style-type: none"> ■ Identifying Co-Occurring Disorders among Opioid Users Using Linked Hospital Care and Mortality Data: Capstone to an Existing FY18 PCORTF Project
	<ul style="list-style-type: none"> ■ Making Electronic Health Record (EHR) Data More Available for Research and Public Health
	<ul style="list-style-type: none"> ■ Strengthening the Data Infrastructure for Outcomes Research on Mortality Associated with Opioid Poisonings
	<ul style="list-style-type: none"> ■ Surveillance Network: Maternal, Infant, and Child Health Outcomes Following Treatment of Opioid Use Disorder During Pregnancy
Centers for Medicare and Medicaid Services (CMS)	
	<ul style="list-style-type: none"> ■ Improving Beneficiary Access to their Health Information through an Enhanced Blue Button Service (Blue Button 2.0)
Food and Drug Administration (FDA)	
	<ul style="list-style-type: none"> ■ Bridging the PCOR Infrastructure and Technology Innovation through Coordinated Registry Networks (CRN) Community of Practice (COP)
	<ul style="list-style-type: none"> ■ SHIELD—Standardization of Lab Data to Enhance Patient-Centered Outcomes Research and Value-based Care
	<ul style="list-style-type: none"> ■ Source Data Capture from Electronic Health Records: Using Standardized Clinical Research Data (OneSource)
	<ul style="list-style-type: none"> ■ Standardization and Querying of Data Quality Metrics and Characteristics for Electronic Health Data

Agency	Award Title
National Institutes of Health (NIH)	
	■ Creation of LOINC Equivalence Classes
	■ Emergency Medicine Opioid Data Infrastructure: Key Venue to Address Opioid Morbidity and Mortality
	■ 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
Office of the National Coordinator for Health Information Technology (ONC)	
	■ A Synthetic Health Data Generation Engine to Accelerate Patient-Centered Outcomes Research
	■ Training Data for Machine Learning to Enhance Patient-Centered Outcomes Research (PCOR) Data Infrastructure
Cross-Agency Funded Awards	
	■ Advancing the Collection and Use of Patient-Reported Outcomes through Health Information Technology [two awards dispersed: AHRQ (N=1), ONC (N=1)]
	■ Assessing and Predicting Medical Needs in a Disaster [two awards dispersed: AHRQ (N=1), ASPR (N=1)]
	■ Data Capacity for Patient-Centered Outcomes Research through Creation of an Electronic Care Plan for People with Multiple Chronic Conditions [two awards dispersed: AHRQ (N=1), National Institute of Diabetes and Digestive and Kidney Disease (NIDDK) (N=1)]
	■ Developing a Strategically Coordinated Registry Network to Support Research on Women's Health Technologies (WHT-CRN) [three awards dispersed: FDA (N=1), NLM (N=1), ONC (N=1)]
	■ Development of a Natural Language Processing Web Service for Public Health Use [two awards dispersed: CDC (N=1), FDA (N=1)]
	■ Enhancing Data Resources for Researching Patterns of Mortality in Patient Centered Outcomes Research: Projects 1–4 -four awards dispersed: CDC (N=2), CMS (N=1), FDA (N=1)]
	■ Harmonization of Various Common Data Models and Open Standards for Evidence Generation [five awards dispersed: FDA (N=1), NCI (N=1), NIH National Center for Advancing Translational Sciences (N=1), NLM (N=1), ONC (N=1)]
	■ Technologies for Donating Medicare Beneficiary Claims Data to Research Studies [two awards dispersed: CMS(N=1), NIH (N=1)]

Appendix C. 2017 Evaluation Gap Area Criteria

2017 Evaluation Results: Guidance on the five strategic areas of focus for the future of the portfolio was informed by stakeholder feedback. Based on the stakeholder feedback, the following criteria were used to assess how the OS-PCORTF awards address the strategic gap areas identified by the 2017.¹⁰⁸

- **Implementing Standards:** Develop best practices to develop, implement, and maintain data standards so that health care and research institutions can reduce the time and costs incurred when implementing and updating standards.
 - Awards that develop new content or information exchange standards
 - Awards that support access via repositories or portals to validated value sets and CDEs and current data and vocabulary standards
 - Awards that develop CDEs
 - Awards that harmonize CDMs
 - Awards that support structured data capture
 - Awards that result in standards frameworks and implementation guides
 - Awards that pilot-test standards and implementation guides
 - Awards that develop and/or pilot standards facilitate patient data donation for research
- **Enhancing Data Governance:** Additional effort is needed to address ongoing barriers to increased data capacity. Although this issue remains challenging, it is critical to the efficient use of the research-oriented data infrastructure across individual and organizations' boundaries of control and ownership.
 - Awards that develop policies or frameworks that specify who can access data (authorization) and for what purpose, based on existing law and regulations.
 - Awards that develop and/or pilot-test data provenance standards
- **Improving Data Quality:** Promote a focus on data quality and increase the quantity and accessibility of electronic health data to improve the efficiency and effectiveness of PCOR; also support core functions and improvements in data interoperability.
 - Awards that focus on improving data completeness
 - Awards that address issues around assessing the fitness-of-use of data (e.g., "fit-for-purpose" metadata standards); awards that address issues of data validity and reliability
 - Awards that develop and/or pilot approaches to analyzing unstructured data
 - Awards that support data linkages between clinical and claims data
- **Balancing Access with Enhancing Privacy and Security:** Spur strategies that enhance privacy and security and inform how research and health care entities can better balance data access with security. Strategies include employing innovative technologies that offer researchers access to data, securely and privately, as well as educating the public about the benefits of making available their anonymous health care data.
 - Awards that develop and/or pilot security standards

- Awards that develop and/or pilot standards for securely accessing clinical data (e.g., APIs, data access standards)
- Awards that develop and/or pilot standards facilitate patient data donation for research
- Awards that develop and/or pilot standards to facilitate patient access to data through standards
- **Disseminating Research Findings:** Improve mechanisms for dissemination of OS-PCORTF-sponsored research so that stakeholders within and outside of HHS can better gauge federal efforts to build data capacity for PCOR.
 - Awards that have a defined plan for publishing the results of their work
 - Awards that partner with non-federal organizations to disseminate the findings from their work
 - Awards that support workforce or researcher education

Appendix D. Technical Expert Panel

Table D-1. Technical Expert Panel Members

TEP Member	Organizational Affiliation
Tom Carton	Principal Investigator, REACHnet
Charles Friedman	University of Michigan Medical School
John Glaser	Cerner Corporation
Michael Kahn	University of Colorado
Abel Kho	Feinberg School of Medicine, Northwestern University
Julia Skapik	Cognitive Medical Systems
Adam Wilcox	University of Washington School Medicine

Appendix E. HHS Data Strategy

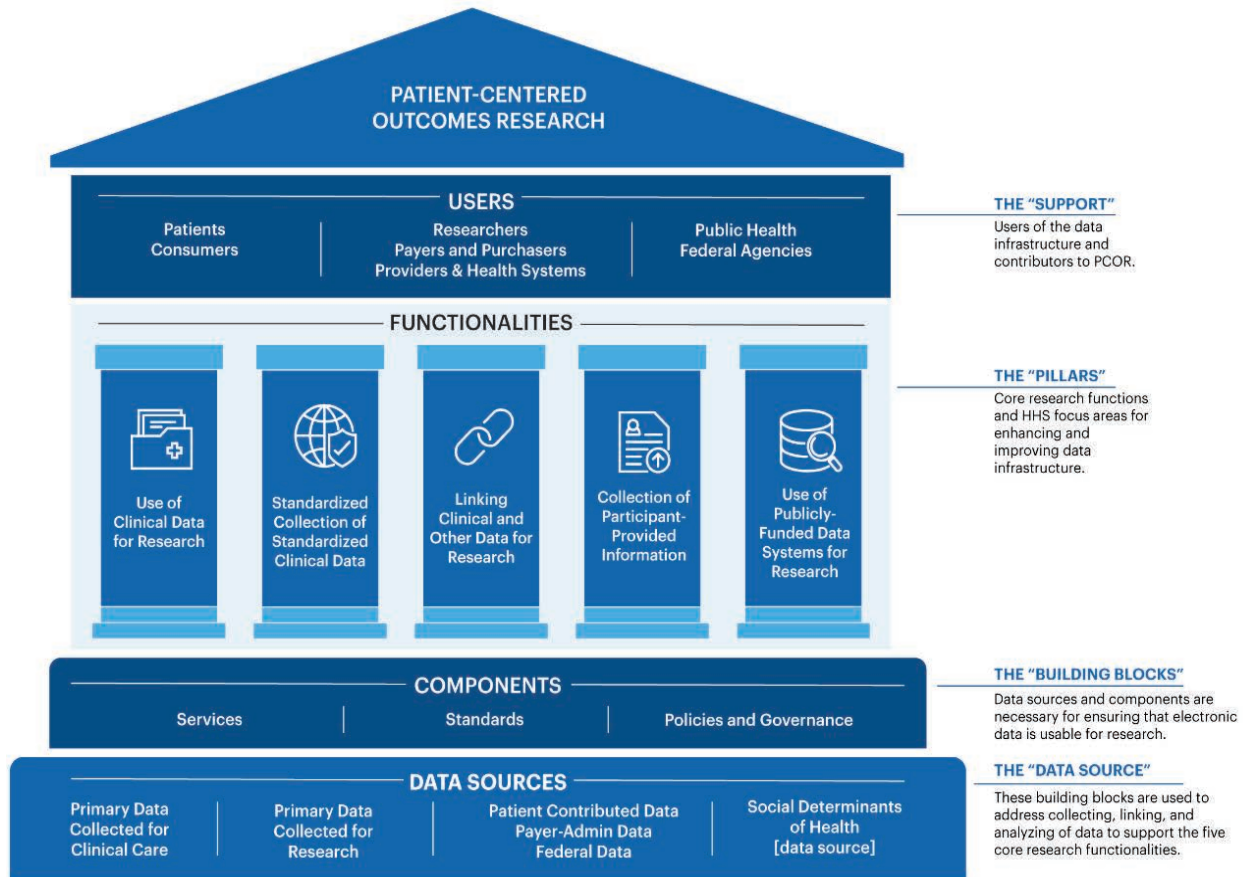
Table E-1. HHS Data Strategy

Priority 1 – Improving Access to HHS Data	
Opportunity 1: Increase the accessibility of HHS data to internal and external users while ensuring that the information is used responsibly	<ul style="list-style-type: none"> • <i>Strategy 1:</i> Streamline processes for accessing data
Opportunity 2: Increase awareness within the Department about available HHS data resources and research	<ul style="list-style-type: none"> • <i>Strategy 1:</i> Develop and implement a framework for a catalog of HHS Data resources • <i>Strategy 2:</i> Establish a process to coordinate the dissemination of major new data releases and research briefs across HHS
Priority 2 – Enhancing Administrative Data for Research	
Opportunity 1: Expand the use of administrative data in the Department	<ul style="list-style-type: none"> • <i>Strategy 1:</i> Improve the documentation and curation of HHS administrative data
Opportunity 2: Improve the quality of administrative data for research	<ul style="list-style-type: none"> • <i>Strategy 1:</i> Develop quality frameworks for administrative data collection • <i>Strategy 2:</i> Create procedures to benchmark big data for program evaluation
Priority 3 – Increasing Data Linkages across Diverse Data Assets	
Opportunity 1: Apply existing departmental knowledge and lessons learned from data linkages	<ul style="list-style-type: none"> • <i>Strategy 1:</i> Develop an HHS Data Linkage Repository that includes information on linkage strategies, barriers, and opportunities
Opportunity 2: Improve the capacity to link HHS data internally and with other data sources	<ul style="list-style-type: none"> • <i>Strategy 1:</i> Promote data linkage between HHS agencies and between HHS and other federal agencies to address departmental priorities • <i>Strategy 2:</i> Promote data linkage to nonfederal data
Priority 4 – Modernizing Privacy Protections	
Opportunity 1: Increase data-sharing without eroding privacy protections through better communication and coordination with experts	<ul style="list-style-type: none"> • <i>Strategy 1:</i> Use data intermediaries to facilitate HHS data sharing in accordance with privacy laws • <i>Strategy 2:</i> Increase the use of disclosure review boards or data disclosure boards and provide guidance on best practices for de-identification
Opportunity 2: Assist in standardization of departmental privacy policy practices	<ul style="list-style-type: none"> • <i>Strategy 1:</i> Support the Department’s development of more streamlined data sharing processes, including model enterprise-wide DUAs and inter-agency agreements (IAA) • <i>Strategy 2:</i> Address privacy and legal concerns about the use of data for policy research, statistical purposes, and program evaluation

Priority 5 – Increasing Data Policy Coordination and Information-sharing Across the Departments	
<p>Opportunity 1: Increase coordination in the Department regarding data collection, system and software investments, and data management and governance</p>	<ul style="list-style-type: none"> • <i>Strategy 1:</i> Identify high-priority data and information policy issues that require OS-level coordination • <i>Strategy 2:</i> Increase communication between the Data Council, OCIO, ONC, and the CTO
<p>Opportunity 2: Inform policymakers and researchers about the value and uses of HHS data</p>	<ul style="list-style-type: none"> • <i>Strategy 1:</i> Communicate the value of HHS data collections and systems for policymaking • <i>Strategy 2:</i> Inform internal stakeholders about potential tradeoffs between timeliness and quality and how this impacts the usefulness of data for policy purposes
Priority 6 – Building a 21st Century Data-Oriented Workforce	
<p>Opportunity 1: Enhance the data science capability of the current HHS data workforce</p>	<ul style="list-style-type: none"> • <i>Strategy 1:</i> Increase data science and statistical training opportunities • <i>Strategy 2:</i> Promote multidisciplinary data science teams and increased cross-program collaboration • <i>Strategy 3:</i> Promote awareness and education of data ethics in the Department
<p>Opportunity 2: Reinforce capacity to explore the application of data science and alternative data to HHS research and program evaluation</p>	<ul style="list-style-type: none"> • <i>Strategy 1:</i> Develop capacity to investigate new or more-blended statistics for health and human services • <i>Strategy 2:</i> Develop the capacity to coordinate the evaluation of alternative data sources • <i>Strategy 3:</i> Explore the NIH Strategic Plan for Data Science as a tool to support data science across the Department
<p>Opportunity 3: Invest in the future of data science</p>	<ul style="list-style-type: none"> • <i>Strategy 1:</i> Increase the number of new data scientists • <i>Strategy 2:</i> Ensure that staff have the expertise to explore the coordinated implementation of technology or software that facilitates ethical data-sharing and use for data science capabilities

Appendix F. High-Level Overview of Agency-Specific Frameworks and Data Strategies

Exhibit F-1. OS-PCORTF Strategic Framework for Building PCOR Data Infrastructure



NIH – Strategic Plan for Data Science

GOAL 1 - Support a Highly Efficient and Effective Biomedical Research Data Infrastructure

- Objective 1-1 | Optimize Data Storage and Security
- Objective 1-2 | Connect NIH Data Systems

GOAL 2 - Promote Modernization of the Data-Resources Ecosystem

- Objective 2-1 | Modernize the Data Repository Ecosystem
- Objective 2-2 | Support the Storage and Sharing of Individual Datasets
- Objective 2-3 | Leverage Ongoing Initiatives to Better Integrate Clinical and Observational Data into Biomedical Data Science

GOAL 3 - Support the Development and Dissemination of Advanced Data Management, Analytics, and Visualization Tools

- Objective 3-1 | Support Useful, Generalizable, and Accessible Tools and Workflows
- Objective 3-2 | Broaden Utility, Usability, and Accessibility of Specialized Tools
- Objective 3-3 | Improve Discovery and Cataloging Resources

GOAL 4 - Enhance Workforce Development for Biomedical Data Science

- Objective 4-1 | Enhance the NIH Data-Science Workforce
- Objective 4-2 | Expand the National Research Workforce
- Objective 4-3 | Engage a Broader Community

GOAL 5 - Enact Appropriate Policies to Promote Stewardship and Sustainability

- Objective 5-1 | Develop Policies for a FAIR Data Ecosystem
- Objective 5-2 | Enhance Stewardship

Exhibit F-3. National Library of Medicine (NLM) – A Platform for Biomedical Discovery and Data-Powered Health: National Library of Medicine Strategic Plan 2017–2027¹¹⁰

Goal 1	Goal 2	Goal 3
Accelerate discovery and advance health through data-driven research	Reach more people in more ways through enhanced dissemination and engagement	Build a workforce for data-driven research and health
1.1 Connect the resources of a digital research enterprise	2.1 Know NLM users and engage with persistence	3.1 Expand and enhance research training for biomedical informatics and data science
1.2 Advance research and development in biomedical informatics and data science	2.2 Foster distinctiveness of NLM as a reliable, trustable source of health information and biomedical data	3.2 Assure data science and open science proficiency
1.3 Foster open science policies and practices	2.3 Support research in biomedical and health information access methods and information dissemination strategies	3.3 Increase workforce diversity
1.4 Create a sustainable institutional, physical, and computational infrastructure	2.4 Enhance information delivery	3.4 Engage the next generation and promote data literacy

Exhibit F-4. Food and Drug Administration (FDA) – Framework for FDA’s Real-World Evidence Program¹¹¹

FDA – Framework for Evaluating RWD/ RWE for Use in Regulatory Decisions
Using Trials or Studies with RWD/RWE for Effectiveness Decisions <ul style="list-style-type: none"> • Whether the RWD are fit for use • Whether the trial or study design used to generate RWE can provide adequate scientific evidence to answer or help answer the regulatory question • Whether the study conduct meets FDA regulatory requirements (e.g., for study monitoring and data collection)
Assessing Fitness of RWD for Use in Regulatory Decisions <ul style="list-style-type: none"> • Assessing Data Reliability (Data Accrual and Data Assurance) and Relevance • Addressing Gaps in RWD Sources
Potential for Study Designs Using RWD to Support Effectiveness <ul style="list-style-type: none"> • Randomized Designs Using RWD • Non-randomized, Single-Arm Trials with External RWD Control Observational Studies
Regulatory Considerations for Study Designs Using RWD <ul style="list-style-type: none"> • Use of Electronic Source Data • Regulatory Considerations for Clinical Studies Generating RWE

FDA – Framework for Evaluating RWD/ RWE for Use in Regulatory Decisions

Data Standards — Appropriate Data Standards for Integration and Submission to FDA

- Identify data standards and implementation considerations that apply to proposed uses of RWD/RWE at FDA
- Review existing RWD/RWE-driven work, both internally and with external stakeholders, to identify gaps that need to be addressed
- Collaborate with internal and external stakeholders to adapt or develop standards and implementation strategies for RWD/RWE-driven solutions at FDA
- Integrate RWD/RWE-driven solutions with existing FDA systems

Exhibit F-5. Food and Drug Administration (FDA) – Sentinel System Five-Year Strategy 2019–2023¹¹²

1. Enhance and expand the Sentinel System’s foundation.
 - a. Expand data sources and linkages so that the Sentinel System can answer a broader range of questions (expanding the breadth, granularity, and duration of data)
 - b. Enhance data infrastructure and methods to improve the utility of available data sources
 - c. Improve operational efficiency to help FDA staff continue to engage more effectively with Sentinel System partners
2. Further enhance safety analysis capabilities using advances in data science and signal detection.
 - a. Increase the Active Post-market Risk Identification and Analysis (ARIA) sufficiency
 - b. Enhance active signal detection in the post-market setting by leverage advances in data science and signal detection
3. Use the Sentinel System to accelerate access to and broaden the use of real-world data (RWD) for RWE.
 - a. Establish standards for high-quality RWD and evaluation RWE applications
4. Create a national resource by broadening the Sentinel System’s ecosystem of stakeholders (user base) to pursue the vision of a national resource.
 - a. Improving operations and procedures for accessing tools, methods, and results to create a broader ecosystem not reliant on FDA-sponsored Sentinel infrastructure
5. Disseminate knowledge and advance regulatory science to encourage innovation and meet the Agency’s scientific needs.
 - a. Convene public working sessions with active and potential Sentinel System users to better understand how they interact with the Sentinel System
 - b. Shape the discussion and participate as a thought leader in the learning healthcare community
 - c. Communicate Sentinel System results and insights

Exhibit F-6. Food and Drug Administration (FDA) – Digital Health Innovation Action Plan¹¹³

Strategic Aims

1. Issue guidance to provide clarity on the medical software provisions of the 21st Century Cures legislation;
2. Launch an **innovative pilot precertification program** (Pre-Cert for Software) to develop a new approach to digital health technology oversight; and
3. Build FDA's bench strength and expertise in CDRH's digital health unit.

Exhibit F-7. Centers for Disease Control and Prevention (CDC) – Surveillance Strategy¹¹⁴

A strategy for improving the Centers for Disease Control and Prevention's activities in public health surveillance (2014–2018)

- Goal 1. Enhance the accountability, resource use, workforce and innovation for surveillance at CDC and in support of STLT agencies.
 - By April 2014, establish a CDC Surveillance Leadership Board charged with reviewing, guiding and overseeing the evolution of CDC surveillance systems in accordance with principles established in this strategy document.
 - By September 2014, develop a federal and STLT workforce training and support plan that integrates CDC's strategy for improving surveillance systems and technological considerations practitioners will face with CDC workforce investments.
 - By May 2014, create a CDC Health Information Innovation Consortium that fosters and promotes creative solutions to surveillance challenges implemented through innovative projects in CDC programs and STLT agencies. This
- Goal 2. Accelerate the utilization of emerging tools and approaches to improve the availability, quality, and timeliness of surveillance data.
 - CDC has established two senior-level positions in the Office of Public Health Scientific Services (OPHSS) to improve Health IT policy engagement and awareness: the Chief Public Health Informatics Officer and the Senior Policy Advisor for Public Health Scientific Services.
 - By May 2014, CDC will develop a forum to systematically engage vendors regarding informatics technologies and tools that can advance surveillance practice and systems.
 - Beginning in FY 2014, OPHSS will provide funding and technical support for small project awards generated through the CHIC to advance specific innovations in the major areas of interest. Projects to be funded should: 1) advance innovation on a specific area related to data collection, transport, storage, analysis, visualization, or availability; 2) if successful, provide insights or tools that can be generalized
- Goal 3. Through cross-cutting agency initiatives, improve surveillance by addressing data availability, system usability, redundancies, and incorporation of new information technologies in major systems or activities.
 - Initiative 1. National Notifiable Diseases Surveillance System (NNDSS) Modernization Initiative
 - Initiative 2. BioSense Enhancement Initiative:
 - Initiative 3. Accelerating Electronic Laboratory Reporting:
 - Initiative 4. Mortality Surveillance-Related initiatives with the National Vital Statistics System

The proposed rule focuses on the following areas:

- **Patient Access Through APIs.** CMS is proposing to require Medicare Advantage (MA) organizations, state Medicaid and CHIP FFS programs, Medicaid managed care plans, CHIP managed care entities, and QHP issuers in FFEs to implement, test, and monitor an openly-published Health Level Seven (HL7®) Fast Healthcare Interoperability Resources (FHIR®)-based APIs to make patient claims and other health information available to patients through third-party applications and developers.
- **Health Information Exchange and Care Coordination across Payers.** The rule proposed that as patients move through the health system, including payer to payer, patients should be able to maintain access to their healthcare information. CMS proposes to require MA organizations, Medicaid managed care plans, CHIP managed care entities, and QHP issuers in the FFEs to support electronic exchange of data for transitions of care as patients move between these plan types. This data includes information about diagnoses, procedures, tests and providers seen and provides insights into a beneficiary's health and healthcare utilization.
- **API Access to Published Provider Directory Data.** Health plan provider directories help patients find in-network providers and allow healthcare professionals to locate other providers for access to medical records, referrals, transitions of care, and care coordination. To ensure patients and providers have easy access to this information, CMS proposes to require MA organizations, state Medicaid and CHIP FFS programs, Medicaid managed care plans, and CHIP managed care entities to make their provider networks available to enrollees and prospective enrollees through API technology. CMS is not proposing these requirements for QHP issuers at this time.
- **Care Coordination through Trusted Exchange Networks.** Exchanging health information on the internet requires a reliable "trust framework" that verifies the security and identity of participants. To expand participation, CMS proposes that payers in CMS programs be able to participate in a trusted exchange network. The goal would be to enable the information to flow securely and privately between plans and providers throughout the healthcare system. CMS proposes requiring MA organizations (including MA-PD plans), Medicaid managed care plans, CHIP managed care entities, and QHP issuers in the FFEs to participate in trust networks to improve interoperability.
- **Improving the Dual-Eligible Experience by Increasing Frequency of Federal-State Data Exchanges.** CMS proposes an update on the frequency with which states are required to exchange certain Medicare/Medicaid data on dually eligible beneficiaries from a monthly exchange to a daily exchange to improve benefit coordination for the dual eligible population. The data exchanged include files of all eligible Medicaid beneficiaries by state, as well as "buy-in" data, or information about beneficiaries states are using Medicaid funds to "buy-in" Medicare services.
- **Public Reporting and Prevention of Information Blocking.** CMS states that it would benefit patients and caregivers to know if individual clinicians, hospitals, and critical access hospitals (CAHs) have submitted a "no" response to any of the three attestation statements regarding the prevention of information blocking in the Promoting Interoperability Programs. Making this information publicly available may motivate clinicians, hospitals, and CAHs to refrain from information blocking.

- **Provider Digital Contact Information.** The 21st Century Cures Act required the Secretary to create a provider digital contact information index, and as of June 2018, the National Plan and Provider Enumeration System (NPPES) has been updated to include one or more pieces of digital contact information that can be used to facilitate secure sharing of health information. To ensure that the NPPES is updated with this information, CMS is proposing to publicly report the names and National Provider Identifiers (NPIs) of those providers who have not added digital contact information to their entries in the NPPES system beginning in the second half of 2020.
- **Revisions to the Conditions of Participation (CoPs) for Hospitals and Critical Access Hospitals.** The CoPs for hospitals and CAHs set basic health and safety standards for how effective care transitions for discharged patients should occur. CMS proposes requiring Medicare-participating hospitals, psychiatric hospitals, and CAHs to send electronic notifications when a patient is admitted, discharged or transferred.
- **Advancing Interoperability in Innovative Models.** The Center for Medicare and Medicaid Innovation (“Innovation Center”) models are an important lever to advance interoperability. The Innovation Center is seeking public comment on promoting interoperability among model participants and other healthcare providers as part of the design and testing of innovative payment and service delivery models.

Exhibit F-9. Office of the National Coordinator of Health IT (ONC) – 21st Century Cures Act: Interoperability, Information Blocking, and the ONC Health IT Certification Program¹¹⁶¹¹⁷

Summary of the key points of the ONC rule:

- The 21st Century Cures Act (Section 4002) requires the Secretary of HHS to establish Conditions and Maintenance of Certification requirements for the ONC Health IT Certification Program. There are seven Conditions of Certification with accompanying Maintenance of Certification Requirements. They are:
 - Information Blocking
 - Assurances
 - Communications
 - Application Programming Interfaces (APIs)
 - Real World Testing
 - Attestation
 - Future EHR Reporting Criteria Submission
- The Rule developed 10 recommendations for the voluntary certification of health IT for pediatric care that does NOT include a separate certification program for pediatric care and practice settings.
- The Proposed Rule also identifies rules for information blocking and allowable exceptions (reasonable and necessary activities that do not constitute information blocking). They identified seven categories of activities that fall under the allowable exceptions:
 - Preventing harm
 - Promoting the privacy of EHI
 - Promoting the security of EHI
 - Recovering costs reasonably incurred

- Responding to requests that are infeasible
- Licensing of interoperability elements on reasonable and non-discriminatory terms
- Maintaining and improving Health IT performance
- In order to implement the Cures Act, ONC has proposed a set of certification requirements. These proposed requirements would improve interoperability by focusing on standardized, transparent, and pro-competitive API practices. This would further support the access, exchange, and use of EHI) by patients and providers. APIs are to access USCDI using FHIR, with a base set of 13 Resources and two specific data fields within the Patient Resource that must be supported.
 - The rules outline the allowable fees charged by API technology suppliers to support information exchange (but not to the patient to access their data)
- The Rule proposes to remove the CCDS definition and its references from the 2015 Edition and replace it with the United States Core Data for Interoperability (USCDI) standard. This will increase the minimum baseline of data classes that must be commonly available for interoperable exchange. The new U.S. Core Data for Interoperability (USCDI) adds provenance information, eight types of clinical notes, additional demographic information, and pediatric vital signs to the former Core Clinical Data Set.
- ONC's new proposal would require health IT developers to provide the capability to electronically export all EHI they produce and electronically manage in a computable format. ONC proposes to make this criterion part of the 2015 Edition Base EHR definition, and for providers and developers to implement this within 24 months of the final rule's effective date. This requirement would facilitate the ability to export data for a patient and for all patients (when a provider is changing EHRs)

Exhibit F-10. Office of the Chief Technology Officer (CTO) – The State of Data Sharing at the U.S. Department of Health and Human Services¹¹⁸

1. Process for Data Access HHS lacks consistent and standardized processes for one agency to request data from another agency.
 - a. Agencies are not accountable for their responses to requests for access to internal data. If access is inappropriately denied or if access is significantly and inappropriately delayed, there are no consequences.
2. Technology for Data Access & Analysis
 - a. The technical formats and approaches to sharing restricted and nonpublic data across agencies vary widely. In addition, the analytical tools to interpret data can be redundant. Agencies are tracking who has access to restricted and nonpublic data, but it can be challenging for agencies to successfully audit for misinterpretation and misuse.
3. Regulatory Environment
 - a. Each data collection effort has statutes, regulations, and policies that govern the collection of and access to the data. Some statutes limit access to data and its use. In order to increase access or broaden use, changes to the relevant statutes may be required.

Appendix G. Case Study Strategic Framework Functionalities

Table G-1. Functionalities Addressed by Each Case Study

Award Title	 Optimizing Clinical Data for Research	 Standardizing Data Collection	 Linking Data	 Collecting Participant-Provided Information	 Using Federal Databases for Research
Use of ADAPTABLE Trial to Strengthen Methods to Collect and Integrate Patient-reported Information with Other Data Sets and Assess Its Validity (NIH)	✓		✓	✓	
Advancing the Collection and Use of PROs through Health IT (AHRQ/ONC)	✓			✓	
Technologies for Donating Medicare Beneficiary Claims Data to Research Studies (CMS/NIH)	✓	✓	✓	✓	✓
Development of an NLP Web Service for Public Health Use (CDC/FDA)	✓			✓	
Standardization and Querying of Data Quality Metrics and Characteristics for Electronic Health Data (FDA)	✓	✓		✓	✓
Harmonization of Various Common Data Models and Open Standards for Evidence Generation (FDA/NCATS/NIH/NCI/ONC)	✓	✓	✓	✓	✓
Enhancing Data Resources for Studying Patterns and Correlates of Mortality in Patient-Centered Outcomes Research: Project 4 – NDI Workshop and Strategy Paper (CDC)					✓
Enhancing Data Resources for Studying Patterns and Correlates 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 (CDC)			✓	✓	✓
Enhancing Data Resources for Studying Patterns and Correlates of Mortality in Patient-Centered Outcomes Research: Project 2 - Pilot Linkage of NDI+ to Commercially and Publicly Insured Populations (FDA)			✓	✓	✓
Security and Privacy Standards for Patient Matching, Linking and Aggregation (ONC)	✓		✓	✓	
Developing a Strategically Coordinated Registry Network (CRN) for Women's Health Technologies (FDA/NLM/ONC)	✓	✓	✓		✓
Privacy and Security Blueprint, Legal Analysis and Ethics Framework for Data Use, & Use of Technology for Privacy (CDC/ONC)	✓		✓	✓	

References

- ¹ Office of the Assistant Secretary for Planning and Evaluation. Patient-Centered Outcomes Research Trust Fund. <https://aspe.hhs.gov/patient-centered-outcomes-research-trust-fund>
- ² Office of the Assistant Secretary for Planning and Evaluation. Patient-Centered Outcomes Research Trust Fund FAQs. <https://aspe.hhs.gov/patient-centered-outcomes-research-trust-fund-faqs>
- ³ Amoozegar, J., Blumenfield, B., et al. (2017). Building Data Capacity for Patient-Centered Outcomes Research in HHS: A Formative Evaluation of 2012-2016 Projects. Prepared by RTI International for the Division of Healthcare Quality and Outcomes, Office of Health Policy/ASPE/HHS. December 22, 2017. Retrieved from <https://aspe.hhs.gov/system/files/pdf/259016/ASPEPCORTEFEvaluation.pdf>
- ⁴ Office of the Assistant Secretary for Planning and Evaluation. Patient-Centered Outcomes Research Trust Fund FAQs. <https://aspe.hhs.gov/patient-centered-outcomes-research-trust-fund-faqs>
- ⁵ Office of the Assistant Secretary for Planning and Evaluation. Patient-Centered Outcomes Research Trust Fund FAQs. <https://aspe.hhs.gov/patient-centered-outcomes-research-trust-fund-faqs>
- ⁶ Amoozegar, J., Blumenfield, B., et al. (2017). Building Data Capacity for Patient-Centered Outcomes Research in HHS: A Formative Evaluation of 2012-2016 Projects. Prepared by RTI International for the Division of Healthcare Quality and Outcomes, Office of Health Policy/ASPE/HHS. Retrieved from <https://aspe.hhs.gov/system/files/pdf/259016/ASPEPCORTEFEvaluation.pdf>
- ⁷ Office of the Assistant Secretary for Planning and Evaluation. Patient-Centered Outcomes Research Trust Fund FAQs. Retrieved from <https://aspe.hhs.gov/patient-centered-outcomes-research-trust-fund-faqs>
- ⁸ Office of Health Policy, ASPE. Building the Data Capacity for Patient-Centered Outcomes Research: The 2018 Annual Report. <https://aspe.hhs.gov/system/files/pdf/259016/2018PortfolioReport.pdf>
- ⁹ Assistant Secretary for Planning and Evaluation. (2018). *2018 Planning guide for OS-PCORTF portfolio participants: Building data capacity for research on patient outcomes*. Unpublished Manuscript.
- ¹⁰ Secretary Priorities. HHS.gov. <https://www.hhs.gov/about/leadership/secretary/priorities/index.html>
- ¹¹ U.S. National Library of Medicine. NIH Common Data Element (CDE) Resource Portal. <https://www.nlm.nih.gov/cde/>
- ¹² U.S. National Library of Medicine. Glossary. <https://www.nlm.nih.gov/cde/glossary.html#cdedefinition>
- ¹³ U.S. National Library of Medicine. Guidance to Encourage the Use of CDEs. <https://www.nlm.nih.gov/cde/policyinformation.html>
- ¹⁴ Food and Drug Administration. (2009.) Guidance for Industry Patient Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Retrieved from <https://www.fda.gov/downloads/drugs/guidances/ucm193282.pdf>
- ¹⁵ Office of the National Coordinator for Health Information Technology. (2018). Conceptualizing a Data Infrastructure for the Capture, Use, and Sharing of Patient-Generated Health Data in Care Delivery and Research through 2024. (White paper.) Retrieved from https://www.healthit.gov/sites/default/files/onc_pghd_final_white_paper.pdf
- ¹⁶ Amoozegar, J., Blumenfield, B., et al. (2017). Building Data Capacity for Patient-Centered Outcomes Research in HHS: A Formative Evaluation of 2012-2016 Projects. Prepared by RTI International for the Division of Healthcare Quality and Outcomes, Office of Health Policy/ASPE/HHS. Retrieved from <https://aspe.hhs.gov/system/files/pdf/259016/ASPEPCORTEFEvaluation.pdf>
- ¹⁷ Food and Drug Administration. Electronic Source Data in Clinical Investigations: Guidance for Industry. Retrieved from <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM328691.pdf>

-
- ¹⁸ Food and Drug Administration. Use of Electronic Health Record Data in Clinical Investigations Guidance for Industry. Retrieved from <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/use-electronic-health-record-data-clinical-investigations-guidance-industry>
- ¹⁹ Hart, N., & Shaw, T. (2018). Congress Provides New Foundation for Evidence-based Policymaking. (Blog.) Bipartisan Policy Center. Retrieved from <https://bipartisanpolicy.org/blog/congress-provides-new-foundation-for-evidence-based-policymaking/>
- ²⁰ Commission on Evidence-Based Policymaking. (2017). The Promise of Evidence-Based Policymaking: Report of the Commission on Evidence-Based Policy Making. Retrieved from <https://www.cep.gov/content/dam/cep/report/cep-final-report.pdf>
- ²¹ Commission on Evidence-Based Policymaking. (2017). The Promise of Evidence-Based Policymaking: Report of the Commission on Evidence-Based Policy Making. Retrieved from [cep.gov/report/cep-final-report.pdf](https://www.cep.gov/report/cep-final-report.pdf)
- ²² Commission on Evidence-Based Policymaking. (2017). The Promise of Evidence-Based Policymaking: Report of the Commission on Evidence-Based Policy Making. Retrieved from [cep.gov/report/cep-final-report.pdf](https://www.cep.gov/report/cep-final-report.pdf)
- ²³ Congress.gov. H.R.4174 – Foundations for Evidence-Based Policymaking Act of 2018. Retrieved from <https://www.congress.gov/bill/115th-congress/house-bill/4174/text>
- ²⁴ Congress.gov. H.R.4174 – Foundations for Evidence-Based Policymaking Act of 2018. Retrieved from <https://www.congress.gov/bill/115th-congress/house-bill/4174/text>
- ²⁵ Rucker, D. (2018.) Achieving the Interoperability Promise Of 21st Century Cures. (Blog.) Health Affairs. Retrieved from <https://www.healthaffairs.org/doi/10.1377/hblog20180618.138568/full/>
- ²⁶ Congress.gov. H.R.6 - SUPPORT for Patients and Communities Act. <https://www.congress.gov/bill/115th-congress/house-bill/6>
- ²⁷ Dullabh P, Dhopeswarkar R, Heaney-Huls K, Hovey, L, Rajendran N, Moriarty E, Steiner C. Building the Data Capacity for Patient-Centered Outcomes Research: The 2018 Annual Report. Prepared under Contract No. HHSP233201600020I.
- ²⁸ U.S. Department of Health and Human Services. 2018 HHS Data Strategy: Enhancing the HHS Evidence-Based Portfolio. Retrieved from <https://aspe.hhs.gov/system/files/pdf/261591/2018HHSDataStrategy.pdf>
- ²⁹ Executive Office of the President, U.S. Department of Commerce, Small Business Administration. DRAFT 2019-2020 Federal Data Strategy Action Plan. <https://strategy.data.gov/assets/docs/draft-2019-2020-federal-data-strategy-action-plan.pdf>
- ³⁰ National Institutes of Health (NIH). About NIH Collaboratory. <https://rethinkingclinicaltrials.org/about-nih-collaboratory/>
- ³¹ Wilson, H., Anatchkova, M., Gelhorn, H. (2017). A Perspective on the 21ST Century Cures Act: Patient-Focused Drug Development. The Evidence Forum. Retrieved from https://www.evidera.com/wp-content/uploads/2017/11/04-A-Perspective-on-The-21st-Century-Cures-Act-Patient-Focused-Drug-Development_2017Nov-1.pdf
- ³² Centers for Medicare & Medicaid Services. Centers for Medicare & Medicaid Services: Innovation Center New Direction. <https://innovation.cms.gov/Files/x/newdirection-rfi.pdf>
- ³³ Centers for Medicare & Medicaid Services. CMS Drives Patient-Centered Care over Paperwork in Proposals to Modernize Medicare and Reduce Burden. <https://www.cms.gov/newsroom/press-releases/cms-drives-patient-centered-care-over-paperwork-proposals-modernize-medicareand-reduce-burden>
- ³⁴ National Institutes of Health. What We Do. <https://www.nih.gov/about-nih/what-we-do/mission-goals>
- ³⁵ National Institutes of Health. Strategic Plan for Data Science. https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf

-
- ³⁶ National Library of Medicine. A Platform for Biomedical Discovery and Data-Powered Health: National Library of Medicine Strategic Plan 2017–2027.
- ³⁷ National Library of Medicine. A Platform for Biomedical Discovery and Data-Powered Health: National Library of Medicine Strategic Plan 2017–2027.
- ³⁸ Food and Drug Administration. What We Do. <https://www.fda.gov/about-fda/what-we-do>
- ³⁹ Food and Drug Administration. Digital Health Innovation Action Plan. <https://www.fda.gov/media/106331/download>
- ⁴⁰ Food and Drug Administration. (2018). Framework for FDA’s Real-World Evidence Program. <https://www.fda.gov/media/120060/download>
- ⁴¹ Food and Drug Administration. Sentinel System Five-Year Strategy 2019-2023. <https://www.fda.gov/media/120333/download>
- ⁴² Centers for Disease Control and Prevention. Mission, Role, and Pledge. <https://www.cdc.gov/about/organization/mission.htm>
- ⁴³ Centers for Disease Control and Prevention. A Strategy for Improving the Centers for Disease Control and Prevention’s Activities in Public Health Surveillance (2014-2018). <https://www.cdc.gov/ddphss/docs/CDC-Surveillance-Strategy-Final.pdf>
- ⁴⁴ Centers for Medicare and Medicaid Services. <https://www.cms.gov/index.html>
- ⁴⁵ Centers for Medicare and Medicaid Services. Interoperability and Patient Access Proposed Rule. <https://www.cms.gov/Center/Special-Topic/Interoperability/CMS-9115-P.pdf>
- ⁴⁶ Office of the National Coordinator. About ONC. <https://www.healthit.gov/topic/about-onc>
- ⁴⁷ Office of the National Coordinator. Notice of Proposed Rulemaking to Improve the Interoperability of Health Information. <https://www.healthit.gov/topic/laws-regulation-and-policy/notice-proposed-rulemaking-improve-interoperability-health>
- ⁴⁸ Office of the Chief Technology Officer. Office of the CTO. <https://www.hhs.gov/cto/index.html>
- ⁴⁹ Office of the Chief Technology Officer. (2018). The State of Data Sharing at the U.S. Department of Health and Human Services. https://www.hhs.gov/sites/default/files/HHS_StateofDataSharing_0915.pdf
- ⁵⁰ Assistant Secretary of Planning and Evaluation. 2018 HHS Data Strategy. <https://aspe.hhs.gov/system/files/pdf/261591/2018HHSDataStrategy.pdf>
- ⁵¹ Food and Drug Administration. Framework for FDA’s Real-World Evidence Program. <https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence>
- ⁵² Food and Drug Administration. Digital Health Innovation Action Plan. <https://www.fda.gov/media/106331/download>
- ⁵³ NIH Collaboratory. Rethinking Clinical Trials - A Living Textbook of Pragmatic Clinical Trials. <https://rethinkingclinicaltrials.org/>
- ⁵⁴ HHS.gov. Overview: HHS Strategic Plan, FY 2018-2022. <https://www.hhs.gov/about/strategic-plan/overview/index.html>
- ⁵⁵ FHIR. Patient Reported Outcomes FHIR Implementation Guide. <https://build.fhir.org/ig/HL7/patient-reported-outcomes/>
- ⁵⁶ Oberd. Smart Outcomes Data Collection. <https://new.oberd.com/>
- ⁵⁷ AHRQ. (2019). Phase 2 Winner Announcement. <https://www.ahrq.gov/stepupappchallenge/phase2-winners.html>
- ⁵⁸ NIH Strategic Plan for Data Science. (2018). Retrieved from https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf

-
- ⁵⁹ NIH Strategic Plan for Data Science. (2018). Retrieved from https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf
- ⁶⁰ Sync for Science: Helping Patients Share EHR Data with Researchers. Retrieved from <http://syncfor.science/>
- ⁶¹ Sync for Science: Helping Patients Share EHR Data with Researchers. Retrieved from <http://syncfor.science/>
- ⁶² Federal Register. (2015). Medicare and Medicaid Programs; Electronic Health Record Incentive Program-Stage 3 and Modifications to Meaningful Use in 2015 Through 2017: A Rule by the Centers for Medicare & Medicaid Services. Retrieved from <https://www.federalregister.gov/documents/2015/10/16/2015-25595/medicare-and-medicaid-programs-electronic-health-record-incentive-program-stage-3-and-modifications>
- ⁶³ Federal Register. (2015.) 2015 Edition Health Information Technology (Health IT) Certification Criteria, 2015 Edition Base Electronic Health Record (EHR) Definition, and ONC Health IT Certification Program Modifications. Retrieved from <https://www.federalregister.gov/documents/2015/10/16/2015-25595/medicare-and-medicaid-programs-electronic-health-record-incentive-program-stage-3-and-modifications>
- ⁶⁴ NIH. All of Us Research Program. <https://allofus.nih.gov/>
- ⁶⁵ Apple. Apple Health Records. <https://www.apple.com/healthcare/health-records/>
- ⁶⁶ CMS Blue Button 2.0. Blue Button 2.0. <https://bluebutton.cms.gov/>
- ⁶⁷ HealthIT.gov. Inferno. <https://inferno.healthit.gov/>
- ⁶⁸ HHS. 21st Century Cures Act: Interoperability, Information Blocking, and the ONC Health IT Certification Program. Retrieved from <https://www.healthit.gov/sites/default/files/nprm/ONCCuresActNPRM.pdf>
- ⁶⁹ SAS. Natural Language Processing: What It Is and Why It Matters. Retrieved from https://www.sas.com/en_us/insights/analytics/what-is-natural-language-processing-nlp.html
- ⁷⁰ Kreimeyer K., Foster M., Pandey A., Arya N., Halford G., Jones S.F., Forshee R., Walderhaug M., Botsis T. (2017). Natural Language Processing Systems for Capturing and Standardizing Unstructured Clinical Information: A Systematic Review. *Journal of Biomedical Informatics*.
- ⁷¹ LAPPS Grid. The Language Application Grid. <https://www.lappsgrid.org/>
- ⁷² CDC. NLP Workbench Web Services Cancer Domain Pilot Project | CDC. <https://www.cdc.gov/cancer/npcr/informatics/nlp-workbench/pilot-cancer.htm>
- ⁷³ CDC. NLP Workbench Web Services Safety Surveillance Domain Pilot Project. <https://www.cdc.gov/cancer/npcr/informatics/nlp-workbench/pilot-safety-surveillance.htm>
- ⁷⁴ Foster M., Pandey A., Kreimeyer K., Botsis T. (2018). *Vaccine*, 36 (29): 4325-4330.
- ⁷⁵ GitHub. Food and Drug Administration. <https://github.com/FDA/>
- ⁷⁶ Kahn M.G., Callahan T.J., Barnard J., Bauck A.E., Brown J., Davidson B.N., et al. (2016). A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. eGEMs (Generating Evidence & Methods to improve patient outcomes). Retrieved from <https://egems.academyhealth.org/articles/abstract/10.13063/2327-9214.1244/>
- ⁷⁷ FDA. (2018). PCORTF Common Data Model Harmonization Project and the Oncology Use Case.
- ⁷⁸ CDC. Data Access - National Death Index. <https://www.cdc.gov/nchs/ndi/index.htm>
- ⁷⁹ CDC. NVSS - Mortality Data. <https://www.cdc.gov/nchs/nvss/deaths.htm>
- ⁸⁰ CDC. The Linkage of the 2014 National Hospital Care Survey to the 2014/2015 National Death Index: Methodology Overview and Analytic Considerations (2019, January 31). Retrieved July 12, 2019 from https://www.cdc.gov/nchs/data/datalinkage/NHCS14_NDI14_15_Methodology_Analytic_Consider.pdf
- ⁸¹ CDC. (2019). The Linkage of the 2016 National Hospital Care Survey to the 2016/2017 National Death Index: Methodology Overview and Analytic Considerations. Retrieved from https://www.cdc.gov/nchs/data/datalinkage/NHCS16_NDI16_17_Methodology_Analytic_Consider.pdf

-
- ⁸² CDC. (2019). The Linkage of the 2014 National Hospital Care Survey to the 2014/2015 Centers for Medicare & Medicaid Services Master Beneficiary Summary File: Methodology Overview and Analytic Considerations. Retrieved from <https://www.cdc.gov/nchs/data/datalinkage/NHCS-CMS-Medicare-Linkage-Methods-and-Analytic-Considerations.pdf>
- ⁸³ CDC. (2019). NCHS - Linkages with the National Hospital Care Survey Data. Retrieved from https://www.cdc.gov/nchs/data-linkage/nhcs_linkage.htm
- ⁸⁴ CDC. National Health Statistics Report. <https://www.cdc.gov/nchs/products/nhsr.htm>
- ⁸⁵ CDC. National Vital Statistics System – Mortality Data. <https://www.cdc.gov/nchs/nvss/deaths.htm>
- ⁸⁶ FDA. FDA's Sentinel Initiative. (2018). <https://www.fda.gov/safety/fdas-sentinel-initiative>
- ⁸⁷ CDC. National Center for Health Statistics: National Death Index. <https://www.cdc.gov/nchs/ndi/index.htm>
- ⁸⁸ Patient Matching Algorithm Challenge. Home. <https://www.patientmatchingchallenge.com/>
- ⁸⁹ Patient Demographic Data Quality Framework. (n.d.). Retrieved July 23, 2019, from <https://www.healthit.gov/playbook/pddq-framework/introduction/>
- ⁹⁰ ONC. Patient Demographic Data Quality Framework. <https://www.healthit.gov/playbook/pddq-framework/>
- ⁹¹ OpenID. HEART WG. <https://openid.net/wg/heart/>
- ⁹² HealthIT.gov. Testing and Utilities. https://www.healthit.gov/techlab/testing_and_utilities.html
- ⁹³ CC Innovation Center. (2017). Blockchain in Healthcare Code-A-Thon. <https://www.cccinnovationcenter.com/challenges/blockchain-in-healthcare-code-a-thon/>
- ⁹⁴ HealthIT.gov. Announcing the Blockchain Challenge. <https://www.healthit.gov/topic/grants-contracts/announcing-blockchain-challenge>
- ⁹⁵ CC Innovation Center. (2017). Blockchain in Healthcare Code-A-Thon. <https://www.cccinnovationcenter.com/challenges/blockchain-in-healthcare-code-a-thon/>
- ⁹⁶ HealthIT.gov. Announcing the Blockchain Challenge. <https://www.healthit.gov/topic/grants-contracts/announcing-blockchain-challenge>
- ⁹⁷ NPES NPI Registry. Search NPI Records. <https://npiregistry.cms.hhs.gov/>
- ⁹⁸ HealthIT.gov. (2018). Patient Matching, Aggregating and Linking (PMAL). <https://www.healthit.gov/topic/scientific-initiatives/pcor/patient-matching-aggregating-and-linking-pmal>
- ⁹⁹ HL7 FHIR. Women's Health Technologies (WHT) Coordinated Registry Network (CRN) FHIR Implementation Guide. <http://hl7.org/fhir/us/womens-health-registries/2019May/>
- ¹⁰⁰ HL7 FHIR. WHT CRN Overview. <http://hl7.org/fhir/us/womens-health-registries/2018Sep/crn-overview.html>
- ¹⁰¹ AUGS. ACQUIRE SUI Surgery Module. <https://www.augs.org/clinical-practice/acquire-sui-surgery-module/>
- ¹⁰² SMART. What Is SMART? (2018). <https://smarthealthit.org/an-app-platform-for-healthcare/about/>
- ¹⁰³ HealthIT.gov. Legal and Ethical Architecture for Patient-Centered Outcomes Research (PCOR) Data ("Architecture"). (2018). <https://www.healthit.gov/topic/scientific-initiatives/pcor/legal-and-ethical-architecture-patient-centered-outcomes-research-pcor-data-architecture>
- ¹⁰⁴ ONC. (2014, October 29). Electronic Consent Management: Landscape Assessment, Challenges, and Technology. Retrieved from https://www.healthit.gov/sites/default/files/privacy-security/ecm_finalreport_forrelease62415.pdf
- ¹⁰⁵ DeSalvo, K., & Savage, M. (2019). The Nationwide Interoperability Roadmap and the ONC's and CMS's Proposed New Regulations: Are We Halfway There Yet? (Blog.) <https://www.healthaffairs.org/doi/10.1377/hblog20190621.419379/full/>
- ¹⁰⁶ ASPE. E. Datasets Released by HHS. (2014). <https://aspe.hhs.gov/report/minimizing-disclosure-risk-hhs-open-data-initiatives/e-datasets-released-hhs>

-
- ¹⁰⁷ CDC. (2017). Legal and Ethical Framework to Use Centers for Disease Control and Prevention Data for Patient-Centered Outcomes Research. Retrieved from https://aspe.hhs.gov/system/files/pdf/259016/PCOR_Legal_508_2.pdf
- ¹⁰⁸ Division of Healthcare Quality and Outcomes, Office of Health Policy. (2017.) Building Data Capacity for Patient-Centered Outcomes Research in HHS: A Formative Evaluation of 2012-2016 Projects. Prepared by RTI International.
- ¹⁰⁹ NIH. Strategic Plan for Data Science. https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf
- ¹¹⁰ NLM. (2017). A Platform for Biomedical Discovery and Data-Powered Health: National Library of Medicine Strategic Plan 2017–2027. Retrieved from https://www.nlm.nih.gov/pubs/plan/lrp17/NLM_StrategicReport2017_2027.pdf
- ¹¹¹ FDA. (2018). Framework for FDA’s Real-World Evidence Program. Retrieved from <https://www.fda.gov/media/120060/download>
- ¹¹² FDA. (2019.) Sentinel System Five-Year Strategy 2019-2023. Retrieved from <https://www.fda.gov/media/120333/download>
- ¹¹³ FDA. Digital Health Innovation Action Plan. <https://www.fda.gov/media/106331/download>
- ¹¹⁴ CDC. A Strategy for Improving the Centers for Disease Control and Prevention’s Activities in Public Health Surveillance (2014-2018). <https://www.cdc.gov/ddphss/docs/CDC-Surveillance-Strategy-Final.pdf>
- ¹¹⁵ HHS. Interoperability and Patient Access Proposed Rule. <https://www.cms.gov/Center/Special-Topic/Interoperability/CMS-9115-P.pdf>
- ¹¹⁶ Regulations.gov. 21st Century Cures Act: Interoperability, Information Blocking, and the ONC Health IT Certification Program. <https://www.regulations.gov/document?D=HHS-ONC-2019-0002-0001>
- ¹¹⁷ HealthIT.gov. Notice of Proposed Rulemaking to Improve the Interoperability of Health Information. <https://www.healthit.gov/topic/laws-regulation-and-policy/notice-proposed-rulemaking-improve-interoperability-health>
- ¹¹⁸ Office of the Chief Technology Officer (CTO). (2018). The State of Data Sharing at the U.S. Department of Health and Human Services. Retrieved from https://www.hhs.gov/sites/default/files/HHS_StateofDataSharing_0915.pdf