



The 2022 Annual Portfolio Report

Building Data Capacity for the Office of the Secretary
Patient-Centered Outcomes Research

Patient-centered outcomes research (PCOR) aims to generate high-quality evidence about the effectiveness of treatments, services, and other health care interventions on the full range of outcomes that patients, caregivers, clinicians, policymakers, and other stakeholders have identified as important.

As part of the 2010 Patient Protection and Affordable Care Act, the Office of the Assistant Secretary for Planning and Evaluation (ASPE) administers the U.S. Department of Health and Human Services' (HHS) Office of the Secretary's Patient-Centered Outcomes Research Trust Fund (OS-PCORTF).¹ ASPE, in partnership with other HHS agencies, has helped develop, coordinate, and implement over 100 projects aimed at building data infrastructure capacity to strengthen patient-centered outcomes research and support the respective missions, statutory authorities, and annual priorities of HHS agencies and the Department as a whole.

The OS-PCORTF Strategic Plan (2020-2029) charts a course for improving capacity for collecting, linking, and analyzing data with respect to four goals: 1) data capacity for national health priorities; 2) data standards and linkages for longitudinal research; 3) technology solutions to advance research; and 4) person-centeredness, inclusion, and equity.²

This 2022 Portfolio Report highlights the accomplishments of the OS-PCORTF during Fiscal Year (FY) 2022. The report is organized into the following sections:

- Overview: 32 active project contributions to the Strategic Plan goals in FY 2022.
- Summary: Notable programmatic accomplishments and activities in advancing PCOR data capacity.
- Appendices A-C: Additional information on projects concluding, beginning, or continuing in FY 2022.



ASPE
ASSISTANT SECRETARY FOR
PLANNING AND EVALUATION

OFFICE OF
HEALTH POLICY



OFFICE OF THE SECRETARY
**PATIENT-CENTERED OUTCOMES
RESEARCH TRUST FUND**

FY 2022 Project Contributions to the OS-PCORTF Strategic Plan

Each of the 32 OS-PCORTF projects active in FY 2022 make a distinct contribution to, and collectively advance the work of, the OS-PCORTF portfolio in achieving the Strategic Plan goals. Projects that address more than one goal are noted with an asterisk. A full project list is also available in Appendix B.



Goal 1. Data Capacity for National Health Priorities. Eleven projects build data capacity for patient-centered outcomes research that informs the needs of federal health programs, clinicians, and patients. Projects address data infrastructure gaps in HHS priority areas—including maternal health, intellectual and developmental disabilities (ID/DD), COVID-19, and the opioid overdose crisis—by building more comprehensive, accessible datasets, guides, and software and engaging researchers and data stewards to create collaborative data exchange networks.

- [Bridging the PCOR Infrastructure and Technology Innovation through Coordinated Registry Networks \(CRN\) Community of Practice \(COP\)* \(FDA\)](#)
- [Building Infrastructure and Evidence for COVID-19 Related Research, Using Integrated Data* \(CDC/NCHS\)](#)
- [CURE ID: Aggregating and Analyzing COVID-19 Treatment from EHRs and Registries* \(FDA\)](#)
- [Dataset on Intellectual and Developmental Disabilities: Linking Data to Enhance Person-Centered Outcomes Research* \(ASPE\)](#)
- [Enhancing Surveillance of Maternal Health Clinical Practices and Outcomes with Federally Qualified Health Centers' Electronic Health Records Visit Data* \(CDC/NCHS\)](#)
- [Linking State Medicaid and Child Welfare Data for Outcomes Research on Treatment for Opioid Use Disorder and Other Behavioral Health Issues* \(ASPE, ACF\)](#)
- [Making Medicaid Data More Accessible Through Common Data Models and FHIR APIs \(FDA, NIH/NLM\)](#)
- [MAT-LINK: MATernal and Infant Network to Understand Outcomes Associated with Treatment for Opioid Use Disorder during Pregnancy* \(CDC\)](#)
- [MAT-LINK2: Expansion of MATernal and Infant Network to Understand Outcomes Associated with Treatment for Opioid Use Disorder during Pregnancy* \(CDC\)](#)
- [NIDA's AMNET: An Addiction Medicine Network to Address the United States Opioid Crisis* \(NIDA\)](#)
- [Understanding COVID-19 Trajectory and Outcomes in the Context of Multiple Chronic Conditions \(MCC\) through e-Care Plan Development* \(AHRQ, NIH/NDDK\)](#)



Goal 2. Data Standards and Linkages for Longitudinal Research. Twelve projects expand longitudinal data resources to collect and facilitate the use of data about people over time and across geographic boundaries and care settings. Multiple projects are developing or evaluating enhanced linkage methods to build longitudinal datasets for PCOR studies, particularly for research on long-term outcomes that matter to people and for which evidence is most often lacking. These linkages enable richer patient-centered outcomes research.

- [Augmenting the National Hospital Care Survey \(NHCS\) Data through Linkages with Administrative Records \(CDC/NCHS\)](#)
- [Creating a Federal COVID-19 Longitudinal Patient Outcomes Research Database Linked to Health Systems and Clinical Data* \(NIH/NCATS\)](#)
- [Data Linkage: Evaluating Preserving Privacy Methodology and Augmenting the National Hospital Care Survey with Medicaid Administrative Records \(CDC/NCHS\)](#)
- [Developing a Multi-State Network of Linked Pregnancy Risk Assessment Monitoring System \(PRAMS\) and Clinical Outcomes Data for Patient-Centered Outcomes Research* \(CDC/DRH\)](#)

- [Digitizing Consent and Regulatory Metadata Towards Streamlining Governance of Pediatric COVID-19 Research Data Linkages* \(NIH/NICHHD\)](#)
- [Enhancing Identification of Opioid-Involved Health Outcomes Using Linked Hospital Care and Mortality Data* \(CDC/NCHS\)](#)
- [Evaluation of Privacy-Preserving Record Linkage Solutions to Broaden Linkage Capabilities in Support of Patient-Centered Outcomes Research Objectives \(CDC/NCHS\)](#)
- [Human-Centered Design Study on Federal Data Access and Acquisition Processes of CMS Data \(CMS\)](#)
- [Identifying Co-Occurring Disorders among Opioid Users Using Linked Hospital Care and Mortality Data: Capstone to a FY18 PCORTF Project* \(CDC/NCHS\)](#)
- [Multistate Emergency Medical Services \(EMS\) and Medicaid Dataset \(MEMD\): A Linked Dataset for PCOR* \(ASPE\)](#)
- [Severe Maternal Morbidity and Mortality EHR Data Infrastructure* \(NIH\)](#)
- [SHIELD – Standardization of Lab Data to Enhance Patient-Centered Outcomes Research and Value-Based Care \(FDA\)](#)



Goal 3. Technology Solutions to Advance Research. Four projects leverage technology solutions to improve data standardization for patient-centered outcomes research. To better incorporate real-world data (RWD) from health care delivery systems, the OS-PCORTF portfolio supports efforts to develop exchange standards like Fast Healthcare Interoperability Resources (FHIR®) application programming interfaces (APIs) that can extract electronic health record (EHR) data for research purposes. The FY 2022 portfolio also supports the use of methods like machine learning and natural language processing (NLP) to perform complex analytic tasks, such as improving the collection, extraction, and identification of new and effective treatments for specified conditions. By improving the accessibility of RWD, these tools have the potential to both improve clinical decision-making and inform PCOR efforts.

- [A Synthetic Health Data Generation Engine to Accelerate Patient-Centered Outcomes Research \(ONC\)](#)
- [Making Electronic Health Record \(EHR\) Data More Available for Research and Public Health \(CDC\)](#)
- [Training Data for Machine Learning to Enhance Patient-Centered Outcomes Research \(PCOR\) Data Infrastructure \(ONC, NIH/NLM\)](#)
- [Using Machine Learning Techniques to Enable Health Information Exchange \(HIE\) Data Sharing to Support COVID-19-focused PCOR \(ONC\)](#)



Goal 4. Person-Centeredness, Inclusion, and Equity. Five projects expand the collection and analysis of socioeconomic and environmental data to improve individuals' access to tailored evidence when making health care decisions. The FY 2022 portfolio supports development of resources like databases created through novel data linkages and extraction tools that provide a more comprehensive view of health and health-related outcomes to identify and address disparities affecting underserved and underrepresented populations.

- [Childhood Obesity Data Initiative \(CODI\) 2.0: Integrated Data for Patient-Centered Outcomes Research Project* \(CDC\)](#)
- [Data Capacity for Patient-Centered Outcomes Research through Creation of an Electronic Care Plan for People with Multiple Chronic Conditions* \(AHRQ, NIH/NIDDK\)](#)
- [Data Capacity for Patient-Centered Outcomes Research through Creation of an Electronic Care Plan for People with Multiple Chronic Conditions 2.0: Development of the Patient-facing Application* \(AHRQ, NIH/NIDDK\)](#)
- [Enhancing Patient-Centered Outcomes Research \(PCOR\): Creating a National Small-Area Social Determinants of Health \(SDOH\) Data Platform* \(AHRQ\)](#)
- [Validating and Expanding Claims-Based Algorithms of Frailty and Functional Disability for Value-based Care and Payment* \(AHRQ, ASPE\)](#)

OS-PCORTF Portfolio Accomplishments and Activities

This section highlights notable programmatic accomplishments and activities of the OS-PCORTF in advancing PCOR data capacity.

2020-2029 Strategic Plan

[OS-PCORTF Strategic Plan for 2020-2029](#)³

The Strategic Plan outlines the four goals and outcomes HHS will prioritize to advance data capacity for patient-centered outcomes research through the OS-PCORTF portfolio.

[Building Data Capacity for Patient-Centered Outcomes Research: Priorities for the Next Decade](#)⁴

Produced by the National Academies of Sciences, Engineering, and Medicine, this report identifies and shares recommendations for strengthening PCOR data infrastructure.

COVID-19

[COVID-19 Collaborative – Collaboration on Data for Evidence \(CoDE\)](#)

This Collaborative brings together seven OS-PCORTF projects working to build infrastructure that supports current and future COVID-19 research, with a particular focus on SDOH data and data linkages.

[Scoping Review Report: Data Elements for Research on the Role of Social Determinants of Health in Coronavirus Disease 2019 Infection and Outcomes in the U.S.](#)⁵

This report presents the results of a scoping review that identified SDOH associated with COVID-19 infection risk and outcomes and describes the wide variability in SDOH measures and the need for standardized SDOH data elements.

Maternal Health

[Maternal Health Consortium](#)

ASPE created this consortium to facilitate collaboration across projects designed to improve EHR data for longitudinal research on maternal health and pregnancy-related conditions, outcomes, and procedures, and linking mothers with their infants. In addition to the project leads representing the National Institutes of Health (NIH) National Institute of Child Health and Human Development (NICHD) and the Centers for Disease Control and Prevention (CDC) National Center for Health Statistics (NCHS), the Consortium includes representation from the Office of the National Coordinator for Health Information Technology (ONC) and it is chaired by NICHD.

[Maternal Health Day of Action: Getting the Information Needed to Drive Change](#)⁶

This blog post describes the efforts of the Lantana Consulting Group in support of the Maternal Health Consortium to facilitate the development of a standardized approach to linking EHR data, thus supporting longitudinal PCOR studies on maternal and infant health.

[Improving Data Capacity for Maternal Health Research Through EHR Data Standards](#)⁷

This blog post discusses how data standardization and interoperability are critical for conducting high-quality patient-centered outcomes research focused on maternal and infant health, and also describes the contributions of OS-PCORTF projects in this area.

[Linking Medicaid Claims, Birth Certificates, and Other Sources to Advance Maternal and Infant Health](#)⁸

This report describes the results of projects to better understand the status of linking data from multiple data sources to conduct complex PCOR maternal and infant health studies and identifies next steps for advancing work in this area.

[Enhancing Data Infrastructure to Improve Women’s Health Outcomes](#)⁹

This vignette describes OS-PCORTF projects focused on improving women’s health by enhancing women’s health data for research, adding EHRs to the suite of maternal health data for research, and linking maternal health survey data with other health data sources.

[Addressing the Maternal Health Crisis Through Improved Data Infrastructure: Guiding Principles for Progress](#)¹⁰

This *Health Affairs* blog post discusses work from across the OS-PCORTF portfolio that advances PCOR data infrastructure to improve maternal health and describes key principles to advance this work over time.

Economic Outcomes

[Medicare Data Linkages for Conducting Patient-Centered Outcomes Research on Economic Outcomes](#)¹¹

This review highlights linked, federally funded administrative and survey data sources that could support patient-centered outcomes research on economic outcomes for Medicare fee-for-service beneficiaries, and identifies strategies for addressing persistent data gaps.

[Symposium on Building Data Capacity to Study Economic Outcomes for Patient-Centered Outcomes Research](#)¹²

Hosted by ASPE and the RAND Corporation, this scientific symposium focused on advancing data capacity for economic outcomes in PCOR studies.

Intellectual and Developmental Disabilities

[Improving Data Infrastructure for Patient-Centered Outcomes Research for People with Intellectual and Developmental Disabilities](#)¹³

This report presents the results of a listening session and key informant interviews that yielded insights on short-term opportunities for strengthening data capacity for patient-centered outcomes research related to individuals with ID/DD.

Opioid Epidemic

[Addressing the Opioid Overdose Crisis Through Better Data](#)¹⁴

This vignette presents OS-PCORTF projects that aim to address the opioid overdose epidemic through critical data-strengthening work, which includes improving the timeliness and quality of clinical outcomes data, building capacity for the collection of patient-reported outcomes, and creating linkages to address co-morbidities.

Other Initiatives Across the Portfolio

[Identifying Environmental Data, Barriers, and Opportunities for Health Services Research and Patient-Centered Outcomes Research: An Environmental Scan](#)¹⁵

This report presents the results of a scan conducted to identify data sources on environmental factors—such as air quality, chemical exposures, and lead levels—suitable for health services research and patient-centered outcomes research and identifies opportunities to strengthen environmental data infrastructure.

CMS Data Access and Linkage Workgroup

This Workgroup is a collaborative venue for various HHS Operating and Staff Divisions to discuss and collectively implement sustainable data infrastructure to support improved access and linkage of CMS data assets across the Department.

Conclusion

Since 2011, ASPE has coordinated across federal agencies to build and strengthen data capacity for patient-centered outcomes research through the OS-PCORTF portfolio. Improving the ability to collect, link, and analyze data supports PCOR studies that generate the strong, high-quality evidence needed to inform decisions by patients, providers, and policymakers and improve outcomes.

The OS-PCORTF portfolio helps address the nation's most significant health challenges including the COVID-19 pandemic, substance use disorder crisis, and maternal mortality. An accompanying impact case study of the portfolio will be forthcoming later this year. The OS-PCORTF portfolio reflects the commitment of HHS to responding to evolving data infrastructure needs, key HHS priorities, and other relevant developments, including legislative or policy changes and advances in health care, data science, and the needs of decision makers who use the findings from PCOR studies. These projects include a range of efforts and products needed to improve the availability, quality, interoperability, security, and accessibility of data for patient-centered outcomes research.



Appendices

Appendix A. Key Impacts of OS-PCORTF Projects Completed in FY 2022

This section includes key impacts and reports of nine projects that completed in FY 2022. A list of final reports for all projects are listed on the ASPE website.

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

Project Lead Agency
ONC

Point of Contact
Papia Paul, MS, MPA
Papia.Paul@hhs.gov

Project Profile
[ASPE Website Project Profile](#)

Strategic Plan
Goals Addressed



Goal 3:
Technology Solutions
to Advance Research

Clinical datasets are a key source of robust patient-level data that are essential to advancing patient-centered outcomes research. However, these datasets are often not available due to security concerns, legal restrictions, or high costs, and while some researchers instead utilize anonymized datasets, these still carry privacy risks. As a result, researchers often turn to synthetic health data, which can provide a more detailed, lower-risk source of data to support research.

This ONC project strengthened PCOR infrastructure by enhancing Synthea™, an open-source synthetic health data engine that produces realistic—but fictional—patient EHR data. The project increased the availability of synthetic health records in Synthea in three HHS priority health areas: 1) patients with complex care needs; 2) opioid use; and 3) pediatric populations. To help enhance the Synthea-generated data, and demonstrate utility of the synthetic data, the project engaged external researchers, developers, and innovators in a contest to develop novel innovations using Synthea-generated data—the Synthetic Health Data Challenge.

“The availability of reliable and robust synthetic data generation tools safeguard patient privacy because they support appropriate stewardship practices in which real patient data is only accessed and used when necessary.” – ONC Synthetic Health Data Challenge Winners Press Release

Key Impact

This project successfully developed five new Synthea modules. These critical data sources strengthen PCOR capabilities for rigorous research and analyses, all while protecting patients’ privacy. A demonstration study showed that Synthea can be successfully utilized in simulation studies as well as for software testing and study planning. Further, the Synthetic Health Data Challenge raised awareness about the availability of the Synthea-generated health data for research purposes by engaging a broad community of research stakeholders to develop practical enhancements to the open-source synthetic data generator.

In turn, the Challenge increased the amount and variety of synthetic data for research. The first place Challenge project was a pediatric asthma medication diversification tool that improved the diversity of medication orders in Synthea to better reflect real-world clinical practice. The ONC project team also identified several key opportunities to enhance Synthea and facilitate its utilization, including expanding Synthea to incorporate data on health care provided in non-clinical settings; adding modules focused on SDOH; enabling Synthea to simulate digital images, like X-rays; and creating a website to host user guidance and support direct communication with Synthea developers.

Key and Final Products



Synthea Disease-specific Modules and Accompanying Companion Guides. The project developed five Synthea modules on cerebral palsy, prescribing opioids for chronic pain and the treatment of opioid use disorder, sepsis, spina bifida, and acute myeloid leukemia. For each module, the project also produced a companion guide describing the module’s scope and intended use.¹⁶



Case Report: Evaluation of an Open-Source Synthetic Data Platform for Simulation Studies. This peer-reviewed evaluation of Synthea found that, with modifications, Synthea can successfully be used for certain types of simulation studies.¹⁷



Final Project Report. This final report summarizes project goals, activities, key findings, and lessons learned. Additionally, the report outlines several considerations for enhancing Synthea to further support patient-centered outcomes research.¹⁸

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

Project Lead Agency
CDC/NCHS

Point of Contact
Lisa Mirel, MS
Cordell Golden, MS
Christine Cox, MA
zav8@cdc.gov; cdg4@cdc.gov;
csc3@cdc.gov

Project Profile
[ASPE Website Project Profile](#)

Strategic Plan Goals Addressed



Goal 2:
Data Standards and Linkages
for Longitudinal Research

Although researchers can obtain hospital utilization and clinical data for PCOR studies, few datasets also integrate SDOH data, like health insurance status and information on access to safe and affordable housing. These limitations inhibit researchers’ ability to study relationships between SDOH and health outcomes. To expand data capacity for such research, this CDC project linked inpatient and emergency department claims and EHR data from the 2014 and 2016 NHCS to two administrative data sources: 1) the Centers for Medicare & Medicaid Services (CMS) Medicare Data Files, which contain information on Medicare enrollment and health care utilization; and 2) U.S. Department of Housing and Urban Development (HUD) data on individuals participating in HUD housing assistance programs.

This project was part of a broader effort at the NCHS to link NCHS survey data to federal administrative data, known as the [NCHS Data Linkage Program](#).¹⁹

“The linked data have been used by NCHS researchers to assess opioid-involved emergency department visits with subsequent respiratory illness and mortality, and for state-of-the-art linkage methodology research in utilizing machine learning for blocking in record linkage and synthetic data creation. There are currently 10 active agreements in place to analyze these linked data sources to support PCOR research goals.” – Lisa Mirel, MS, Project Lead

Key Impact

The project data files provide a new resource to study a wide range of patient health outcomes, including efficacy of treatment protocols and drugs, disparities among populations underrepresented in biomedical research, and the role of non-clinical and mental health and substance use programs on health outcomes. The resulting data files have also been an important recruitment tool to encourage hospital participation in the NHCS, as the facility-level mortality data are shared back with participating hospitals. As of November 2022, five publications have analyzed the NHCS linked data files that were created with OS-PCORTF support,^{20,21,22,23,24} with one peer-reviewed journal article, titled “Using Synthetic Data to Replace Linkage Derived Elements: A Case Study,” winning the 2022 CDC Statistical Science Awards for best theoretical paper. Within NCHS, the enhanced linkage algorithms used in this project will translate to improved efficiency and accuracy in future linkage efforts. Already, methods developed under this project have been applied in two subsequent OS-PCORTF projects.^{25,26}

Key and Final Products



Linked 2016 NHCS/2016-2017 CMS Medicare Data Files. The NCHS Data Linkage website includes the final 2016 NHCS data linked to 2016-2017 CMS Medicare enrollment, claims/encounters, and patient assessment data records, as well as a report on the linkage methods, analytic considerations, and codebooks.²⁷



CMS Data Linkage Methodology and Analytic Considerations Report. This report includes a brief overview of the data sources, a description of the novel methods used to link NHCS and CMS Medicare datasets leveraging a machine learning technique used in blocking to improve linkage efficiency, and analytic considerations to assist researchers in using the files.²⁸



Linked 2014 NHCS-2013/2015 HUD and 2016 NHCS-2015/2017 HUD Data Files. The NCHS Data Linkage website includes the final NHCS/HUD data files and a report on the linkage methods, analytic considerations, and codebooks.²⁹



HUD Data Linkage Methodology and Analytic Considerations Report. This report summarizes the data sources and the linkage methods used and provides analytic guidance for researchers.³⁰



Using Supervised Machine Learning to Identify Efficient Blocking Schemes for Record Linkage.

This peer-reviewed publication describes a case study using the Sequential Coverage Algorithm (SCA), a supervised machine learning algorithm developed by the project team. The SCA was used to develop the strategy for linking 2016 NHCS data with the CMS Enrollment Database,³¹ which improved the efficiency in conducting the linkage between extremely large data sources.



Final Project Report. The final report provides project background, major accomplishments, lessons learned, and a summary of key products and publications resulting from project activities.³²

The linked data files are available for research use through the NCHS federal Research Data Center (RDC). For more information on RDC access, please visit: <https://www.cdc.gov/rdc/>.

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

Project Lead Agency
FDA

Point of Contact
Danica Marinac-Dabic, MD, PhD, MSc
danica.marinac-dabiv@fda.hhs.gov

Project Profile
[ASPE Website Project Profile](#)

Strategic Plan
Goals Addressed



Goal 1:
Data Capacity for National
Health Priorities



Goal 4:
Person-Centeredness,
Inclusion, and Equity

Patient-centered outcomes research and patient registries often rely on data captured at the point of care, re-entered into clinical research systems, and then consolidated and transformed for analysis and research. The process is complex, labor-intensive, and expensive. As a result, both study designs and study infrastructure for using RWD are often limited. CRNs have emerged as a strategy for overcoming these challenges. CRNs are robust data ecosystems that generate real-world evidence (RWE) by linking data systems from multiple sources (e.g., EHR and administrative databases) to support patient care and quality improvement. CRNs streamline research by standardizing the capture of large amounts of RWD from diverse sources, such as EHRs, claims, and patients. Ultimately, CRNs increase access to high-quality data that allow for efficient, timely research assessing long-term outcomes.

This Food and Drug Administration (FDA) project enhanced existing registry infrastructure for studying and evaluating medical devices by harmonizing a minimum set of common data elements across registries, incorporating the collection of unique device identifiers (UDIs) and data reflecting the patient perspective, and facilitating data linkages between registries and other data sources across 12 clinical domains—from women’s health, to orthopedics, to prostate cancer.

Key Impact

This project strengthened the CRNs as a national RWD source of high-quality, reliable, actionable data for studying health outcomes for patients using medical devices, especially for technology affecting women’s health. The project increased the ability of the CRNs to capture standardized data.

For example, by including UDIs in the CRN minimum core dataset, the CRNs can routinely study device-specific questions. The FDA also conducted a pilot to test and refine the instrument used to capture patient preferences with registry data. This activity facilitates linking patient-contributed data to clinical data sources, supporting regulators in identifying how devices are performing for patients, and identifying patient-centric endpoints for future studies. The team also piloted and refined device-specific FHIR standards, with the goal of demonstrating FHIR’s ability to capture and exchange CRN data. Together, these activities support researchers focused on patient-centered outcomes research. Stakeholders have already applied this project’s findings; for instance, researchers used standardized CRN data in large-scale studies of demographic differences in medical device outcomes.^{33,34}

Key and Final Products



CRN Maturity Framework. This framework serves as a guide for researchers and clinicians to develop and advance high-quality CRNs. The framework describes proven approaches to seven core technical areas, including patient engagement, data quality, and efficiency, that contribute to the success of a CRN.^{35,36}



MDEpiNet Project Website. The Medical Device Epidemiology Network (MDEpiNet) Coordinating Center leads the CRN COP. The routinely updated MDEpiNet website summarizes the scope of activities and tools developed under each of the 12 CRNs.³⁷



MDEpiNet Supplement. Published in *BMJ Surgery, Interventions & Health Technologies*, this supplement includes an editorial and seven papers, all focused on advancements in and research applications of CRNs.³⁸



Nine Peer-reviewed Journal Manuscripts. Recent publications cover a wide range of studies utilizing the enhanced CRN data.^{39,40,41,42,43,44,45,46,47}

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

Project Lead Agency
AHRQ

Point of Contact
Patricia Keenan, PhD
Patricia.Keenan@ahrq.hhs.gov

Project Profile
[ASPE Website Project Profile](#)

Strategic Plan
Goals Addressed



Goal 2:
Data Standards and Linkages
for Longitudinal Research



Goal 4:
Person-Centeredness,
Inclusion, and Equity

Factors such as education, employment, transportation access, environment, and housing contribute significantly to overall health.^{48,49} By understanding how social factors impact health and health care use, policymakers and program developers can design interventions tailored to meet the needs of individuals given their unique social context. While incorporating SDOH into community and policy interventions can help address disparities in health care access and quality, the data needed to study SDOH are often not centrally stored or standardized.

This Agency for Healthcare Research and Quality (AHRQ) project fills the need for easily accessible community-level SDOH data by developing a national longitudinal database of valid and reliable SDOH factors at small-area geographic levels including county, zip code, and U.S. census tract. To develop the SDOH Database, AHRQ identified and linked SDOH variables derived from 44 federal datasets and publicly available data sources to create 1,366 harmonized SDOH variables. The increased precision afforded by small-area geographic levels supports localized public health surveillance and population health management efforts. Access to more specific geographic contextual factors facilitates patient-centered outcomes research, leading to more tailored research and programs that improve individual health.

“The purpose of the SDOH Database is to provide ‘one-stop’ access to publicly available community-level SDOH data to facilitate data linkages, research, and analysis to inform efforts to improve equity and health outcomes. Examples of outcomes of interest in recent studies using the SDOH Database include mortality prediction after trauma;⁵⁰ tracheostomy and mortality in bronchopulmonary dysplasia;⁵¹ cardiovascular disease mortality;⁵² and improvement in function among ventilated patients.”⁵³ – Patricia Keenan, PhD, Project Lead

Key Impact

The SDOH Research Database files are standardized, easy-to-use, publicly available data that facilitate research on SDOH, aligning with the HHS priority to address and improve health equity. Providing researchers access to a wide range of standardized and linkable SDOH data aims to improve the efficiency with which PCOR studies can be designed and conducted by reducing the need for repeated efforts to create longitudinal community SDOH files project-by-project or institution-by-institution. The database is also intended to enhance the robustness of the evidence that is generated by making it possible to leverage a more comprehensive set of variables. External researchers can link the data to other datasets for more detailed analyses for patient-centered outcomes research and health services research, or they can directly analyze the data for findings on SDOH-related characteristics across small area geographies, as demonstrated by the AHRQ visualization illustrating county-level patterns of poverty and access to internet.⁵⁴

Key and Final Products



SDOH Database. Updated July 2022, the current version of the Research Database is publicly available via AHRQ’s website. The Research Database includes data at the county and census tract levels for years 2009-2020 and zip code-level data from 2011-2020 across five key SDOH domains: social context, economic context, education, physical infrastructure, and health care context.⁵⁵ The current version (Version 1) builds on the initial version of the SDOH Database that was made publicly available in late December 2020.



Poverty and Access to Internet, by County Data Visualization. AHRQ created this data visualization to demonstrate potential applications of the SDOH Research Database for research.⁵⁶



Internet Access as a Social Determinants of Health. This summary of the SDOH data visualization was published in the July 2021 edition of the *Journal of the American Medical Association*.⁵⁷



SDOH Environmental Scan. AHRQ conducted an environmental scan of SDOH data sources at the zip code, county, and state levels to inform the development of AHRQ’s beta SDOH database (since archived). The scan identified existing SDOH-related data sources and variables for researchers and features filtering capabilities that allow the user to sort the variables by SDOH domain or geographic level.⁵⁸

Identifying Co-Occurring Disorders among Opioid Users Using Linked Hospital Care and Mortality Data: Capstone to a FY18 OS-PCORTF Project

Project Lead Agency
CDC/NCHS

Point of Contact
Carol DeFrances, PhD
cdefrances@cdc.gov

Project Profile
[ASPE Website Project Profile](#)

Strategic Plan
Goals Addressed



Goal 1:
Data Capacity for National Health Priorities



Goal 2:
Data Standards and Linkages for Longitudinal Research

Mental health issues increase the risk of opioid use and overdose,⁵⁹ but poor data availability limits research on the impact of co-occurring mental health issues and substance use disorders (SUDs) on opioid-involved hospitalizations. This project, led by the NCHS at the CDC, built upon existing methodologies to create and test an algorithm to identify co-occurring mental health issues and SUDs in health care data. The project team developed an algorithm to identify hospital visits in which use of opioids was documented and the patient had either or both an SUD and a mental health issue, termed the “co-occurring disorders algorithm.” The algorithm consists of two components to analyze all clinical data captured in the NHCS, including 1) a medical code-based algorithm for medical diagnoses and procedure code data, and 2) an efficient and innovative NLP algorithm to analyze unstructured data captured in EHR clinical notes.

The team validated both the co-occurring disorders algorithm and an opioid identification algorithm developed under another OS-PCORTF project using a sample of 900 validated encounters from nine hospitals. The validated algorithms were applied to the 2016 NHCS Enhanced Opioid Identification Dataset,⁶⁰ developed under the previous OS-PCORTF project, to link the 2016 NHCS to the National Death Index (NDI)⁶¹ and National Vital Statistics System restricted mortality data, drug-specific information⁶² files of the same year. This application allowed the project team to calculate the prevalence of co-occurring disorders among opioid users in 2016 and to produce a new NHCS dataset that flags opioid use and co-occurring opioid and mental health disorders within the 2016 NHCS data.

“The code component of the co-occurring disorders algorithm was used in the [NHCS Drug Dashboard](#)⁶³ to provide recent trend data on hospital encounters with drug use-associated hospital visits with co-occurring disorders. This information on drug use and co-occurring disorders is not available in other government reporting systems on hospital utilization related to drug use.” – Carol DeFrances, PhD, Project Lead

Key Impact

This project positions the NHCS as an important tool to leverage for public health surveillance of morbidity and mortality of substance use and severe mental illness co-occurrence. NCHS will use the products and procedures developed by this project to ensure ongoing improvement in the identification of opioid-involved encounters with co-occurring disorders in future NCHS products concerning opioid-involved hospital utilization and care. The co-occurring disorder algorithm will continue to be updated as additional years of NHCS data become available, leading to the production of new data resources relevant to patient-centered outcomes research for SUDs.

Key and Final Products



Co-occurring Disorders Algorithm Methodology Report. This report details the methodology used to develop the co-occurring opioids disorder algorithm and the findings from applying the algorithm to linked 2016 NHCS data.⁶⁴



2016 Enhanced NHCS Inpatient and Emergency Department Data File. This report describes the 2016 NHCS enhanced data file identifying opioid-involved encounters with mention of a mental health illness or SUD, including a brief description of the source datasets, a description of the mental health issues and SUD definitions, a data dictionary, and instructions on how to access the data in the federal and NCHS RDC.⁶⁵



Co-occurring Disorders ICD-10-CM and/or Procedure Code-Based Algorithm. The code used to develop the co-occurring disorders algorithm component that analyzes medical diagnosis data is available via the CDC GitHub website.⁶⁶



Co-occurring Disorders NLP-based Algorithm. The NLP component of the co-occurring disorders algorithm used to analyze unstructured EHR data is available via the CDC GitHub website.⁶⁷



Final Project Report. The final report summarizes critical background information, major project accomplishments, lessons learned, and a summary of key products and publications resulting from project activities.⁶⁸

SHIELD – Standardization of Lab Data to Enhance Patient-Centered Outcomes Research and Value-Based Care

Project Lead Agency

FDA

Point of Contact

Keith Campbell, MD, PhD
keith.campbell@fda.hhs.gov

Project Profile

[ASPE Website Project Profile](#)

Strategic Plan
Goals Addressed



Goal 2:
Data Standards and Linkages
for Longitudinal Research

Laboratory in vitro diagnostic (IVD) tests—tests performed on samples taken from the human body—are important sources of patient health information and are frequently used by clinicians to inform care decisions.

However, IVD test results are often encoded using different terminology standards within and between health care organizations, leading to heterogeneous data that can be exchanged between systems, but not easily utilized because of difficulties in aggregation and analysis. As a result, manual and error-prone techniques are required to determine equivalence of test results in reporting systems. This lack of standardization in the data hinders interoperability and the provision of high-quality patient care, as well as large-scale public health research.

To address this challenge, researchers have called for more standardized use of Logical Observation Identifiers Names and Codes (LOINC[®]), the most commonly used data standard for identifying health measurements, observations, and documents important for laboratory tests and orders.⁶⁹ This FDA-based project helped to develop LOINC code mapping manuals which are intended to provide clearer guidance to laboratories in assigning the correct code to each IVD test type. In parallel, the FDA project team sought to better understand the integrity of laboratory data in its current state by conducting a pilot study with five high-profile academic medical center laboratories. This study evaluated the implementation of LOINC-to-In Vitro Diagnostic (LIVD) coding specifications and focused on COVID-19 test reporting required by the Coronavirus Aid, Relief, and Economic Security (CARES) Act.

“The importance of improving the integrity and agility of laboratory data was on clear display during the rapidly changing environment that was the COVID-19 pandemic. The work of this project to provide practical improvements to laboratories’ workflows, as well as quantify, for the first time, the integrity of laboratory data as it moves between laboratory device and information systems, will serve as the foundation for a number of larger FDA efforts to create a safer, more efficient laboratory ecosystem.” – Keith E. Campbell, MD, PhD, Project Lead

Key Impact

This project’s findings have been leveraged to support the national response to the COVID-19 pandemic. The pilot study found significant variability (41%) in how laboratory test manufacturers and health care organizations curate and categorize laboratory data, and, ultimately, that LIVD mappings alone were not enough to promote comprehensive data interoperability. Strengthening data interoperability has been identified as a critical task to facilitate an effective continued response to COVID-19, making the project’s efforts highly relevant.⁷⁰ Further, the findings from this project can help determine future directions for promoting the use of standardized IVD data.

Key and Final Products



Encoding Laboratory Testing Data: Case Studies of the National Implementation of HHS Requirements and Related Standards in Five Laboratories. This peer-reviewed journal article in the *Journal of the American Medical Informatics Association* (JAMIA) provides case studies on the implementation of LIVD in five pilot sites.⁷¹



LOINC Mapping Guides. These pilot-tested and vetted resources provide guidance for using LOINC terms across six laboratory domains: allergy, cell markers, chemistry, drug and toxicology, hematology and serology, and molecular pathology.⁷²



LIVD Test Code Mapping Tools. These tools define LOINC codes to support the standardized reporting of SARS-CoV-2 test results to public health agencies.⁷³

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

Project Lead Agency
ONC & NIH/NLM

Point of Contact
Matthew Rahn (ONC)
Stefan Jaeger, PhD (NIH/NLM)
Matthew.Rahn@hhs.gov
Stefan.Jaeger@nih.gov

Project Profile
[ASPE Website Project Profile](#)

Strategic Plan
Goals Addressed



Goal 3:
Technology Solutions to
Advance Research

Artificial intelligence (AI) is widely recognized as a set of powerful methods and tools with the potential to advance medical research and support high-quality, evidence-based health care.^{74,75} AI techniques, such as machine learning, rapidly analyze large clinical datasets to identify new, effective treatments, which can then be used to support clinical comparative effectiveness research and patient-centered clinical decision-making. However, ongoing challenges, such as the limited availability of training data used to improve pattern recognition, hinder the use of these important tools. This ONC and National Institutes of Health/National Library of Medicine (NIH/NLM) project worked to overcome these persistent barriers by building high-quality training datasets and machine learning models focused on chronic kidney disease and tuberculosis (TB), which were selected based on evidence gaps on treatments for these conditions.

“The project addressed the goal of building and testing high-quality training datasets for a kidney disease use case that can potentially be utilized for AI and machine learning applications, including joint clinician-patient informed decision-making. PCOR researchers can build off the foundational work completed through this project and extend the application of these methods to a wider array of use cases and advance the application of machine learning to enhance PCOR infrastructure.” — Adam Wong, MPP, ONC

“This project is the first major study combining radiographical, clinical, and genomic pathogen data for TB drug resistance analysis with machine learning. The developed training data and software will enable PCOR researchers to predict treatment outcomes and find potentially undiscovered relationships.” — Stefan Jaeger, PhD, NIH/NLM Project Lead

Key Impact

This project developed and validated high-quality training datasets that can be used to support patient-centered care and shared decision-making regarding patient preferences and treatment decisions. NIH/NLM also used these training datasets to develop, train, and improve algorithm performance for detecting TB drug resistance in radiographs, while ONC demonstrated how these training datasets can potentially be utilized in AI and machine learning applications for patient-centered outcomes research using chronic kidney disease as a use case. The project documented lessons learned, best practices, and recommendations, which can support future applications of machine learning to patient-centered outcomes research, including the development of an industry-wide strategy to facilitate the timely access of data and ensuring that a standard set of widely used data elements are fully and accurately captured in national data registries. Ultimately, the PCOR community can build upon the foundational work completed by this project by applying these methods to new use cases, expanding the ability of machine learning to support patient-centered health care.

Key and Final Products



Training Datasets and Validated Machine Learning Algorithms from the United States Renal Data System (USRDS). These training datasets and validated machine learning algorithms use data from the USRDS. The training dataset includes 188 elements, including demographics, prior care, clinical variables, comorbidities, and patient education, and contains over one million individual patients. Researchers can request access to the datasets via the USRDS website.⁷⁶



Training Dataset and Validated Machine Learning Algorithms for the TB Drug Resistance Use Case. These training data can be used by researchers to develop models using machine learning. These data are available for download on the NIH National Institute of Allergy and Infectious Diseases (NIH/NIAIDS) TB Portals.⁷⁷



Training Data for Machine Learning to Enhance PCOR Project Overview and Implementation Guide. This implementation guide provides the methodology for and lessons learned from developing effective, high-quality training datasets and machine learning models.^{78,79}



Machine Classifiers Training Implementation Guide. This implementation guide describes the data acquired during this project and the software tools used for machine learning. It shows how to train machine classifiers and test these classifiers on unseen data for each of the classification problems addressed in the project.⁸⁰



Data Dictionary. A data dictionary that provides information on the features included in the training dataset is available on the project's website.⁸¹



Chronic Kidney Disease Use Case Open-Source Codes. The training dataset and machine learning algorithm codes are available on the ONC GitHub page.⁸²



Drug Resistant and Drug Sensitive TB Use Case Open-Source Codes. The codes for segmentation and classifying drug resistance and drug sensitivity and TB classification using lung images are available on the NIH/NLM GitHub page.⁸³



Six Peer-reviewed Manuscripts. The project’s publications cover a range of topics relevant to PCOR researchers, including applications of machine learning in diagnosing and detecting diseases such as TB.^{84,85,86,87,88,89}



ONC Final Project Report. This report summarizes project activities performed by ONC, including methodology employed to build the training dataset, machine learning models, and use case selection, as well as recommendations to support ongoing use of the resources for patient-centered outcomes research.⁹⁰



NIH/NLM Final Project Report. The final report provides background information about TB drug resistance and its relevance to patient-centered outcomes research. It explains the classification and prediction problems addressed in this project. Furthermore, it describes the lessons learned and proposes directions for future PCOR research.⁹¹

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

Project Lead Agency
ASPE & AHRQ

Point of Contact

Lok Wong Samson, PhD, MHS (ASPE)
Arlene Bierman, MD, MS (AHRQ)
Janey Hsiao, PhD (AHRQ)
Lokwong.Samson@hhs.gov
Arlene.Bierman@ahrq.hhs.gov
Janey.Hsiao@ahrq.hhs.gov

Project Profile

[ASPE Website Project Profile](#)

Strategic Plan
Goals Addressed



Goal 3:
Technology Solutions to
Advance Research



Goal 4:
Person-Centeredness,
Inclusion, and Equity

Older adults who are frail and persons with functional disabilities (e.g., vision impairment, deaf or hard of hearing, difficulties with mobility) are at increased risk for poor health outcomes. This project, led by ASPE’s Office of Health Policy, in collaboration with AHRQ, used functional risk factors such as frailty, predictors of functional dependence, mental and behavioral health disorders, and other disabling conditions to improve risk-adjustment models used in value-based care programs and patient-centered outcomes research for older adults and persons with functional disabilities. The goal of this project was to develop or identify one or more measures of frailty that could be added to the CMS Chronic Conditions Data Warehouse (CCW)⁹² so that these data would be available for use by researchers, health care providers, measure developers, and other stakeholders to identify populations with or at-risk for functional disability and poor health outcomes (e.g., hospital readmission and mortality). The project also assessed the use of EHR-based frailty measures for risk stratification using linked claims and EHR data from several large health systems.

“This project confirms claims-based frailty algorithms can identify and predict persons at risk for frailty and poor health outcomes, such as hospitalizations, nursing home stays, and fewer healthy days at home. Frailty scores, to be added to the CMS Chronic Conditions Warehouse, will help health systems and providers proactively target those at risk of frailty for early interventions and help researchers and measure developers assess or account for frailty in evaluating patient-centered outcomes.” – Lok Wong Samson, PhD, MHS, ASPE Project Lead

“Electronic frailty index (EFI) scores from electronic health records had acceptable concordance with the claims-based EFI and can be used by health systems, practices, and PCOR researchers to assess frailty risk to improve the management of high-risk individuals and populations. There is opportunity to improve EHR-based indices as EHR data improves and by incorporating additional methods such as natural language processing.” – Arlene Bierman, MD, MS, AHRQ Project Lead

Key Impact

The project tested and validated the use of claims-based algorithms of frailty and functional disability using a sample of over 35 million Medicare fee-for-service (FFS) beneficiaries. Based on the results of this testing, the project team recommended use of an existing algorithm—the Kim algorithm⁹³—to calculate a claims-based frailty index (CFI) score to identify Medicare beneficiaries at risk for functional impairments. The CFI scores will be included in the CCW. The project also demonstrated the predictive power of EFI scores to identify current and future health care utilization outcomes, such as inpatient hospitalization, emergency department admission, and nursing home admission. Together the CFI and EFI scores can be used by health systems for risk adjustment or for tracking quality of care and utilization for at-risk populations by stratifying measures. Researchers might also consider including these scores as controls in evaluations of policies or using them to study the way(s) in which frailty may impact different interventions in this population.

Key and Final Products



Electronic Frailty Index. The EFI used to support frailty validation using EHR data is available on GitHub.⁹⁴



Validation of the CFI Developed. The claims-based frailty algorithm developed by researcher Dae Kim had the best overall performance at predicting claims-based outcomes of interest for most metrics and subpopulations tested. Work is ongoing to include the CFI scores derived from the Kim algorithm in the CCW.



AHRQ Final Project Report. This report focuses on the methods, results, and discussion related to linking and comparing claims data against the EHR, as well as evaluating the accuracy and concordance of claims versus EHR-based measures of frailty.⁹⁵

Appendix B. OS-PCORTF Project Portfolio⁹⁶

The list below includes 32 new, complete, and active projects in FY 2022.

[Digitizing Consent and Regulatory Metadata Towards Streamlining Governance of Pediatric COVID-19 Research Data Linkages](#)

Lead Agency
NIH/NICHHD

Project Status
New

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

Lead Agency
FDA

Project Status
Complete

[Evaluation of Privacy-Preserving Record Linkage Solutions to Broaden Linkage Capabilities in Support of Patient-Centered Outcomes Research Objectives](#)

Lead Agency
CDC/NCHS

Project Status
New

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

Lead Agency
AHRQ

Project Status
Complete

[Human-Centered Design Study on Federal Data Access and Acquisition Processes of CMS Data](#)

Lead Agency
CMS

Project Status
New

[Identifying Co-Occurring Disorders among Opioid Users Using Linked Hospital Care and Mortality Data: Capstone to a FY18 PCORTF Project](#)

Lead Agency
CDC/NCHS

Project Status
Complete

[A Synthetic Health Data Generation Engine to Accelerate Patient-Centered Outcomes Research](#)

Lead Agency
ONC

Project Status
Complete

[SHIELD – Standardization of Lab Data to Enhance Patient-Centered Outcomes Research and Value-Based Care](#)

Lead Agency
FDA

Project Status
Complete

[Augmenting the National Hospital Care Survey \(NHCS\) Data through Linkages with Administrative Records](#)

Lead Agency
CDC/NCHS

Project Status
Complete

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

Lead Agency
ONC, NIH/NLM

Project Status
Complete

[Validating and Expanding Claims-Based Algorithms of Frailty and Functional Disability for Value-Based Care and Payment](#)

Lead Agency
AHRQ, ASPE

Project Status
Complete

[Building Infrastructure and Evidence for COVID-19 Related Research, Using Integrated Data](#)

Lead Agency
CDC/NCHS

Project Status
Active

[Childhood Obesity Data Initiative \(CODI\) 2.0: Integrated Data for Patient-Centered Outcomes Research Project](#)

Lead Agency
CDC

Project Status
Active

[Creating a Federal COVID-19 Longitudinal Patient Outcomes Research Database Linked to Health Systems and Clinical Data](#)

Lead Agency
NIH/NCATS

Project Status
Active

[CURE ID: Aggregating and Analyzing COVID-19 Treatment from EHRs and Registries](#)

Lead Agency
FDA

Project Status
Active

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

Lead Agency
AHRQ, NIH/NIDDK

Project Status
Active

[Data Capacity for Patient-Centered Outcomes Research through Creation of an Electronic Care Plan for People with Multiple Chronic Conditions 2.0: Development of the Patient-facing Application](#)

Lead Agency
AHRQ, NIH/NIDDK

Project Status
Active

[Data Linkage: Evaluating Preserving Privacy Methodology and Augmenting the National Hospital Care Survey with Medicaid Administrative Records](#)

Lead Agency
CDC/NCHS

Project Status
Active

[Dataset on Intellectual and Developmental Disabilities: Linking Data to Enhance Person-Centered Outcomes Research](#)

Lead Agency
ASPE

Project Status
Active

[Developing a Multi-State Network of Linked Pregnancy Risk Assessment Monitoring System \(PRAMS\) and Clinical Outcomes Data for Patient-Centered Outcomes Research](#)

Lead Agency
CDC

Project Status
Active

[Enhancing Identification of Opioid-Involved Health Outcomes Using Linked Hospital Care and Mortality Data](#)

Lead Agency
CDC/NCHS

Project Status
Active

[Enhancing Surveillance of Maternal Health Clinical Practices and Outcomes with Federally Qualified Health Centers' Electronic Health Records Visit Data](#)

Lead Agency CDC/NCHS	Project Status Active
--------------------------------	---------------------------------

[Linking State Medicaid and Child Welfare Data for Outcomes Research on Treatment for Opioid Use Disorder and Other Behavioral Health Issues](#)

Lead Agency ASPE, ACF	Project Status Active
---------------------------------	---------------------------------

[Making Electronic Health Record \(EHR\) Data More Available for Research and Public Health](#)

Lead Agency CDC	Project Status Active
---------------------------	---------------------------------

[Making Medicaid Data More Accessible Through Common Data Models and FHIR APIs](#)

Lead Agency FDA, NIH/NLM	Project Status Active
------------------------------------	---------------------------------

[MAT-LINK: MATernal and Infant Network to Understand Outcomes Associated with Treatment for Opioid Use Disorder During Pregnancy](#)

Lead Agency CDC	Project Status Active
---------------------------	---------------------------------

[MAT-LINK2: Expansion of MATernal and Infant Network to Understand Outcomes Associated with Treatment for Opioid Use Disorder during Pregnancy](#)

Lead Agency CDC	Project Status Active
---------------------------	---------------------------------

[Multistate Emergency Medical Services \(EMS\) and Medicaid Dataset \(MEMD\): A Linked Dataset for PCOR](#)

Lead Agency ASPE	Project Status Active
----------------------------	---------------------------------

[NIDA's AMNET: An Addiction Medicine Network to Address the United States Opioid Crisis](#)

Lead Agency NIDA	Project Status Active
----------------------------	---------------------------------

[Severe Maternal Morbidity and Mortality-EHR Data Infrastructure](#)

Lead Agency NIH	Project Status Active
---------------------------	---------------------------------

[Understanding COVID-19 Trajectory and Outcomes in the Context of Multiple Chronic Conditions \(MCC\) through e-Care Plan Development](#)

Lead Agency AHRQ, NIH/NIDDK	Project Status Active
---------------------------------------	---------------------------------

[Using Machine Learning Techniques to Enable Health Information Exchange \(HIE\) Data Sharing to Support COVID-19-focused PCOR](#)

Lead Agency ONC	Project Status Active
---------------------------	---------------------------------

Appendix C. FY 2022 New Projects

Digitizing Consent and Regulatory Metadata Towards Streamlining Governance of Pediatric COVID-19 Research Data Linkages

Project Lead Agency

NIH National Institute of Child and Human Development (NICHD)

Point of Contact

Rebecca Rosen, PhD
rebecca.rosen@nih.gov

Project Profile

[ASPE Website Project Profile](#)

Strategic Plan Goals Addressed



Goal 1:
Data Capacity for
National Health Priorities



Goal 2:
Data Standards and Linkages
for Longitudinal Research

Successful clinical and administrative dataset linkages require robust privacy and confidentiality policies to ensure that individuals' sensitive information is protected. Although privacy-preserving record linkage (PPRL) is a set of methods that can address privacy concerns, there is no standardized governance approach to determine how to address informed consent, regulatory, or any other requirements associated with linking datasets and the access to and use of linked datasets. Due to the ambiguity in governance practices, researchers experience significant difficulties performing data linkages in a timely and appropriate manner.

To improve the timeliness, accessibility, and security of high-priority HHS datasets on pediatric COVID-19 data, the NIH/NICHD team will develop a standard data linkage governance approach. This governance approach will include detailed linkage requirements for pediatric COVID-19 datasets and a generalizable metadata schema that captures consent and regulatory-based elements relevant to data linkage governance. The project will also develop and test the use of the governance metadata schema in a data collection tool to demonstrate streamlined decision-making for data linkage and use of linked datasets for research.

Evaluation of Privacy-Preserving Record Linkage Solutions to Broaden Linkage Capabilities in Support of Patient-Centered Outcomes Research Objectives

Project Lead Agency

CDC/NCHS

Point of Contact

Lisa Mirel, MS
zav8@cdc.gov

Project Profile

[ASPE Website Project Profile](#)

Strategic Plan Goals Addressed



Goal 2:
Data Standards and Linkages
for Longitudinal Research

PPRL is a linking methodology that works to mitigate privacy concerns when linking person-level data across data sources. PPRL allows for additional privacy protections that encrypt or mask personally identifiable information (PII). However, researchers may choose to utilize a variety of PPRL tools, which may produce different linkage results and subsequently different PCOR study findings. To maximize the potential of PPRL, it is important to understand the strengths and limitations, including accuracy, of existing approaches.

Using datasets that link the National Health Information Survey, National Health and Nutrition Examination Survey, and National Hospital Care Survey data to other federal sources like HUD, this OS-PCORTF funded project will compare the linkage results obtained from three different PPRL tools currently in use or in development within HHS to a reference linked dataset s developed using gold standard linkage methods. The project will assess a variety of scenarios that may impact the accuracy of the linkage, including PII that is incomplete (e.g., missing Social Security Numbers), and of varying levels of quality. The analysis will also consider the security and re-identification risks of the three PPRL tools to join records across multiple data

Human-Centered Design Study on Federal Data Access and Acquisition Processes of CMS Data

Project Lead Agency
CMS

Point of Contact
Kari Gaare, JD, MHA
kari.gaare@cms.hhs.gov

Project Profile
[ASPE Website Project Profile](#)

Strategic Plan
Goals Addressed



Goal 2:
Data Standards and Linkages
for Longitudinal Research

Federal data linkage across HHS agencies allows for the creation of robust PCOR data. CMS makes data that contain PII and personal health information available for research purposes. CMS established a standard process all researchers (including federal agencies) follow to request CMS data, including entering into a Data Use Agreement with CMS. In addition, federal agencies often must complete supplemental agreements. CMS experienced a significant increase in the number and complexity of data requests during the COVID-19 pandemic and anticipates an increase in data requests from other federal agencies to facilitate data linkages and support patient-centered outcomes research.

In response to the increased demand and complexity of federal agency research requests, CMS will be conducting a human-centered design study to understand and identify the current gaps in the federal agency request process. Under this project, CMS will evaluate user experiences with the current data request process and use the findings to inform the design and testing of a more user-friendly process. The goal is to design a more streamlined process that meets federal agency partners' needs, while ensuring compliance with relevant laws and CMS policies.

Appendix D. Acknowledgments

We thank the following individuals for their contributions to this report:

OS-PCORTF Project Leads

Arlene Bierman (AHRQ)	Kari Gaare (CMS)	Shin Kim (CDC)	Matt Rahn (ONC)
Carlos Blanco (NIDA)	Ken Gersing (NIH/NCATS)	Danica Marinac-Dabic (FDA)	Rebecca Rosen (NIH/NICHHD)
Rachael Boicourt (AHRQ)	Aly Goodman (CDC)	Maria Michaels (CDC)	Lok Wong Samson (ASPE)
Keith Campbell (FDA)	Robin Ghertner (ASPE)	Lisa Mirel (CDC)	Daniel Schwartz (ASPE)
Nahida Chakhtoura (NIH)	Janey Hsiao (AHRQ)	James Mork (NIH/NLM)	Heather Stone (FDA)
Shanna Cox (CDC/DRH)	Stefan Jaeger (NIH/NLM)	Jenna Norton (NIH/NIDDK)	Brian Ward (CDC/NCHS)
Carol DeFrances (CDC/NCHS)	Geoff Jackson (CDC/NCHS)	Papia Paul (ONC)	Jaime Zimmerman (AHRQ)
Sarah Dutcher (FDA)	Patricia Keenan (AHRQ)	Emma Plourde (ASPE)	

ASPE Office of Health Policy

Nancy DeLew, Associate Deputy Assistant Secretary for Health Policy

Division of Health Care Quality and Outcomes

Scott R. Smith, Division Director

Cagla Ada-Anderson	Idia Enogieru
Michelle Bohnes	Madjid Karimi
Sara Couture	Euny Lee
Jessica Elosa	Susan Lumsden
Emily Evans	Marcos Trevino
Aldren Gonzales	Sara Wei
Violanda Grigorescu	

References

- 1 Patient Protection and Affordable Care Act, Publ. L. No. 111-148, 124 Stat. 119(2010). <https://www.govinfo.gov/content/pkg/PLAW-111publ148/html/PLAW-111publ148.htm>
- 2 The Office of the Assistant Secretary for Planning and Evaluation. Building data capacity for patient-centered outcomes research: OS-PCORTF Strategic Plan for 2020-2029. ASPE. September 2022. Accessed January 31, 2023. Retrieved from <https://aspe.hhs.gov/sites/default/files/documents/b363671a6256c6b7f26dec4990c2506a/aspe-os-pcortf-2020-2029-strategic-plan.pdf>.
- 3 The Office of the Assistant Secretary for Planning and Evaluation. Building data capacity for patient-centered outcomes research: OS-PCORTF Strategic Plan, 2020-2029. ASPE. September 2022. Accessed January 31, 2023. Retrieved from <https://aspe.hhs.gov/sites/default/files/documents/b363671a6256c6b7f26dec4990c2506a/aspe-os-pcortf-2020-2029-strategic-plan.pdf>
- 4 National Academies of Sciences, Engineering, and Medicine. Building data capacity for patient-centered outcomes research: Priorities for the next decade. Washington, DC: The National Academies Press. 2022. Accessed April 28, 2023. Retrieved from <https://nap.nationalacademies.org/catalog/26489/building-data-capacity-for-patient-centered-outcomes-research-priorities-for>
- 5 Kim KK, Qudsi H, Halter HM, et al. Scoping review report: Data elements for research on the role of social determinants of health in coronavirus disease 2019 infection and outcomes in the U.S. Office of the Assistant Secretary for Planning and Evaluation. March 7, 2022. Accessed April 28, 2023. Retrieved from <https://aspe.hhs.gov/sites/default/files/documents/53fc356858b67f5e9810d107f09ce169/sdoh-covid-19-scoping-review-report.pdf>
- 6 Lantana Consulting Group. Maternal Health Day of Action: Getting the Information Needed to Drive Change. December 20, 2021. Accessed January 31, 2023. Retrieved from <https://www.lantanagroup.com/maternal-health-day-of-action-getting-the-information-needed-to-drive-change/>
- 7 Grigorescu V. Improving Data Capacity for Maternal Health Research Through EHR Data Standards. Healthcare IT News. May 31, 2022. Accessed January 31, 2023. Retrieved from <https://www.healthcareitnews.com/news/improving-data-capacity-maternal-health-research-through-ehr-data-standards>
- 8 Heins S, Predmore Z, Hoch E, Baxi S. Linking Medicaid claims, birth certificates, and other sources to advance maternal and infant health: Final report. Office of the Assistant Secretary for Planning and Evaluation. September 2022. Accessed April 28, 2023. Retrieved from <https://aspe.hhs.gov/sites/default/files/documents/db89f9777de0142dc5b6562692f67032/linking-medicaid-claims-other-data.pdf>
- 9 The Office of the Assistant Secretary for Planning and Evaluation. Enhancing Data Infrastructure to Improve Women's Health Outcomes. ASPE. November 2022. Accessed April 28, 2023. Retrieved from <https://aspe.hhs.gov/reports/maternal-health-vignette>
- 10 Chappel A, DeLew N, Grigorescu V, Smith SR. Addressing the maternal health crisis through improved data infrastructure: Guiding principles for progress. Health Affairs Blog. August 11, 2021. Accessed April 28, 2023. Retrieved from <https://www.healthaffairs.org/doi/10.1377/forefront.20210729.265068/>
- 11 Brown D, Srinivasan M, Zott C, Wilson K, Dullabh P, Smith S. Medicare data linkages for conducting patient-centered outcomes research on economic outcomes. Office of the Assistant Secretary for Planning and Evaluation.. September 2022. Accessed April 28, 2023. Retrieved from <https://aspe.hhs.gov/sites/default/files/documents/66b598da75b717c6b2181ffe63d5830d/medicare-linkages-white-paper.pdf>

- 12 RAND Corporation. Symposium on Building Data Capacity to Study Economic Outcomes for Patient-Centered Outcomes Research. RAND. Accessed January 31, 2023. Retrieved from <https://www.rand.org/health-care/projects/pcor/hhs-symposium/agenda.html>
- 13 Dhopeswarkar R, Rotondo C, Jiménez F, et al. Improving data infrastructure for patient-centered outcomes research for people with intellectual and developmental disabilities. Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services. July 26, 2022. Accessed April 28, 2023. Retrieved from <https://aspe.hhs.gov/reports/improving-data-infrastructure-pcor-people-iddd>
- 14 The Office of the Assistant Secretary for Planning and Evaluation. Addressing the Opioid Overdose Crisis Through Better Data. ASPE. November 2022. Accessed April 28, 2023. Retrieved from <https://aspe.hhs.gov/reports/addressing-opioid-overdose-crisis-through-better-data-opioids-vignette>
- 15 Redman S, Wagstaff L, Flanagan E, Trevino M. Identifying environmental data, barriers, and opportunities for health services research (HSR) and patient-centered outcomes research (PCOR): An environmental scan. Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services. September 2022. Accessed April 28, 2023. Retrieved from <https://aspe.hhs.gov/sites/default/files/documents/5d6538254418204062ec50460083c290/Final-Environmental-Data-Report-508.pdf>
- 16 The Office of the National Coordinator for Health Information Technology. Synthetic Health Data Generation to Accelerate Patient-Centered Outcomes Research: Example Modules and Companion Guides. HealthIT.gov. 2022. Accessed December 1, 2022. Retrieved from <https://www.healthit.gov/topic/scientific-initiatives/pcor/synthetic-health-data-generation-accelerate-patient-centered-outcomes>
- 17 Meeker D, Kallem C, Heras Y, Garcia S, Thompson C. Case report: evaluation of an open-source synthetic data platform for simulation studies. *JAMIA Open*. 2022;5(3):ooac067. doi:10.1093/jamiaopen/ooac067
- 18 Garcia S, Kallem C, Thompson C, Tan M, Heras Y, Meeker D. . Synthetic health data generation to accelerate patient-centered outcomes research (PCOR): Final report. The Office of the National Coordinator for Health Information Technology. March 2022. Accessed February 2, 2023. Retrieved from https://www.healthit.gov/sites/default/files/page/2022-03/20220314_Synthetic%20Data%20Final%20Report_508.pdf
- 19 National Center for Health Statistics. NCHS Data Linkage Activities. Centers for Disease Control and Prevention. Accessed November 11, 2022. Retrieved from <https://www.cdc.gov/nchs/data-linkage/index.htm>
- 20 Jackson G, Brown A, DeFrances C. Opioid-involved emergency department visits in the National Hospital Care Survey and the National Hospital Ambulatory Medical Care Survey. *Natl Health Stat Report*. 2020;(149):1-15. PMID: 33395383
- 21 Ashman J, Cairns C, DeFrances C, Schwartzman A. Respiratory illness emergency department visits in the National Hospital Care Survey and the National Hospital Ambulatory Medical Care Survey. *Natl Health Stat Report*. 2021;(151):1-18. PMID: 33541514
- 22 Resnick D, Cox C, Mirel L. Using synthetic data to replicate linkage derived elements: A case study. *Health Serv Outcomes Res Methodol*. 2021; 21: 389-406. doi:10.1007/s10742-021-00241-z
- 23 Campbell S, Resnick D, Cox C. Using supervised machine learning to identify efficient blocking schemes for record linkage. *Stat J IAOS*. 2021; 37(2): 673-680. doi:10.3233/sji-200779
- 24 Spencer MR, Flagg LA, Jackson G, DeFrances C, Hedegaard H. National Hospital Care Survey demonstration projects: Opioid-involved emergency department visits, hospitalizations, and deaths. *Natl Health Stat Report*. 2020;(141):1-19. PMID: 32600515
- 25 Evaluation of Privacy-Preserving Record Linkage Solutions to Broaden Linkage Capabilities in Support of Patient-Centered Outcomes Research Objectives. ASPE. Accessed January 26, 2023. Retrieved from <https://aspe.hhs.gov/evaluation-privacy-preserving-record-linkage-solutions-broaden-linkage-capabilities>

- 26 Data Linkage: Evaluating Privacy Preserving Record Linkage Methodology and Augmenting the National Hospital Care Survey with Medicaid Administrative Records. ASPE. Accessed January 26, 2023. Retrieved from <https://aspe.hhs.gov/data-linkage-evaluating-privacy-preserving-record-linkage-methodology-augmenting-national-hospital>
- 27 National Center for Health Statistics. Linked NHCS-CMS Medicare Data. Centers for Disease Control and Prevention. Accessed November 28, 2022. Retrieved from <https://www.cdc.gov/nchs/data-linkage/CMS-Medicare-Restricted.htm>
- 28 National Center for Health Statistics, Division of Analysis and Epidemiology. The linkage of the 2016 National Hospital Care Survey to the 2016/2017 Centers for Medicare & Medicaid Services Medicare enrollment, claims/encounters and assessment data: Matching methodology and analytic considerations. Centers for Disease Control and Prevention. September 2020. Accessed November 28, 2022. Retrieved from <https://www.cdc.gov/nchs/data/datalinkage/2016-nhcs-cms-linkage-methodology.pdf>
- 29 National Center for Health Statistics. Linked NHCS-HUD Housing Assistance Program Data. Centers for Disease Control and Prevention. Accessed November 28, 2022. Retrieved from <https://www.cdc.gov/nchs/data-linkage/nhcs-hud.htm>
- 30 National Center for Health Statistics, Division of Analysis and Epidemiology. The Linkage of the National Hospital Care Survey to U.S. Department of Housing and Urban Development Administrative Data: Matching Methodology and Analytic Considerations. Centers for Disease Control and Prevention. August 2021. Accessed November 28, 2022. Retrieved from <https://www.cdc.gov/nchs/data-linkage/index.htm>
- 31 Campbell SR, Resnick DM, Cox CS, Mirel LB. Using supervised machine learning to identify efficient blocking schemes for record linkage. *Stat J IAOS*. 2021;37(2):673-680.
- 32 National Center for Health Statistics. Augmenting the National Hospital Care Survey (NHCS) data through linkages with administrative records: Linking the National Hospital Care Survey to Centers for Medicare & Medicaid Services and Housing and Urban Development data: Final report. Centers for Disease Control and Prevention. Published April 2022. Accessed February 2, 2023. Retrieved from <https://aspe.hhs.gov/sites/default/files/documents/34696d3d740fcecb2b950e21e602994/aspe-final-report-fy19-pcortf.pdf>
- 33 Behrendt CA, Sedrakyan A, Katsanos K, et al. Sex disparities in long-term mortality after paclitaxel exposure in patients with peripheral artery disease: A nationwide claims-based cohort study. *J Clin Med*. 2021 Jul 2;10(13):2978. doi:10.3390/jcm10132978
- 34 Chen A, Paxton L, Zheng X, et al. Association of sex with risk of 2-year revision among patients undergoing total hip arthroplasty. *JAMA Network Open*. 2021;4(6), e2110687. doi: 10.1001/jamanetworkopen.2021.10687
- 35 Sedrakyan A, Aryal S. Maturity framework and select approaches for developing Coordinated Registry Networks (CRNs): Medical Device Epidemiology Network (MDEpiNet) supplement. *BMJ Surg Interv Health Technol*. 2022;4(Suppl 1):e000148. Published 2022 Nov 11. doi:10.1136/bmjst-2022-000148
- 36 Sedrakyan A, Marinac-Dabic D, Campbell B, et al. Advancing the real-world evidence for medical devices through coordinated registry networks. *BMJ Surg Interv Health Technol*. 2022;4(Suppl 1):e000123. Published 2022 Nov 11. doi:10.1136/bmjst-2021-000123
- 37 The Medical Device Epidemiology Network. MDEpiNet. Accessed December 1, 2022. Retrieved from <https://www.mdepinet.net/>
- 38 Sedrakyan A, Aryal S. Maturity framework and select approaches for developing Coordinated Registry Networks (CRNs): Medical Device Epidemiology Network (MDEpiNet) supplement. *BMJ Surg Interv Health Technol*. 2022;4(Suppl 1):e000148. doi:10.1136/bmjst-2022-000148

- 39 Stoddard MD, Zheng X, Mao J, Te A, Sedrakyan A, Chughtai B. Safety and efficacy of outpatient surgical procedures for the treatment of benign prostatic enlargement in New York State and California (2005-2016). *J Urol*. 2021 Mar;205(3):848-854. doi: 10.1097/JU.0000000000001401. Epub 2020 Oct 7.
- 40 Chughtai B, Mao J, Matheny ME, Mauer E, Banerjee S, Sedrakyan A. Long-Term safety with sling mesh implants for stress incontinence. *J Urol*. 2021 Jan;205(1):183-190. doi: 10.1097/JU.0000000000001312. Epub 2020 Aug 4.
- 41 Bouhadana D, Nguyen DD, Zhang X, et al. Safety and efficacy of TURP vs. laser prostatectomy for the treatment of benign prostatic hyperplasia in multi-morbid and elderly individuals aged ≥ 75 . *World J Urol*. 2021 Dec;39(12):4405-4412. doi:10.1007/s00345-021-03779-2
- 42 Bertges DJ, Eldrup-Jorgensen J, Robbins S. Vascular Quality Initiative surveillance of femoropopliteal artery paclitaxel devices. *JACC Cardiovasc Interv*. 2021 Dec 13;14(23):2598-2609. doi: 10.1016/j.jcin.2021.08.058
- 43 Mao J, Moore KO, Columbo JA, Mehta KS, Goodney PP, Sedrakyan A. Validation of an indirect linkage algorithm to combine registry data with Medicare claims. *J Vasc Surg*. 2022;76(1):266-271.e2. doi:10.1016/j.jvs.2022.01.132
- 44 Chughtai B, Cho A, Simonyan V, et al. Development and usability testing of a mobile application to monitor patient-reported outcomes after stress urinary incontinence surgery. *Urology*. 2022;159:66-71. doi:10.1016/j.urology.2021.10.011
- 45 Behrendt CA, Sedrakyan A, Katsanos K, et al. Sex disparities in long-term mortality after paclitaxel exposure in patients with peripheral artery disease: A nationwide claims-based cohort study. *J Clin Med*. 2021;10(13):2978. Published 2021 Jul 2. doi:10.3390/jcm10132978
- 46 Chen A, Paxton L, Zheng X, et al. Association of sex with risk of 2-year revision among patients undergoing total hip arthroplasty. *JAMA Netw Open*. 2021;4(6):e2110687. Published 2021 Jun 1. doi:10.1001/jamanetworkopen.2021.10687
- 47 Garland SK, Falster MO, Beiles CB, et al. Long-term outcomes following elective repair of intact abdominal aortic aneurysms: A comparison between open surgical and endovascular repair using linked administrative and clinical registry data. *Ann Surg*. 2023;277(4):e955-e962. doi:10.1097/SLA.0000000000005259
- 48 Office of Disease Prevention and Health Promotion, Office of the Assistant Secretary for Health. Social Determinants of Health. Healthy People 2030. Accessed November 7, 2022. Retrieved from <https://health.gov/healthypeople/priority-areas/social-determinants-health>
- 49 Post WS, Watson KE, Hansen S, et al. Racial and ethnic differences in all-cause and cardiovascular disease mortality: The MESA Study. *Circulation*. 2022;146(3):229-239. doi:10.1161/CIRCULATIONAHA.122.059174
- 50 Phelos HM, Kass NM, Deeb AP, Brown JB. Social determinants of health and patient-level mortality prediction after trauma. *J Trauma Acute Care Surg*. 2022;92(2):287-295. doi:10.1097/TA.0000000000003454
- 51 Smith MA, Steurer MA, Mahendra M, Zinter MS, Keller RL. Sociodemographic factors associated with tracheostomy and mortality in bronchopulmonary dysplasia. *Pediatr Pulmonol*. 2023;58(4):1237-1246. doi:10.1002/ppul.26328
- 52 Son H, Zhang D, Shen Y, et al. Social determinants of cardiovascular health: A longitudinal analysis of cardiovascular disease mortality in US counties from 2009 to 2018. *J Am Heart Assoc*. 2023;12(2):e026940. doi:10.1161/JAHA.122.026940
- 53 Jain S, Walkey AJ, Law AC, Ferrante LE, Lindenauer PK, Krumholz HM. Association between residential segregation and long-term acute care hospital performance on improvement in function among ventilated patients. *Ann Am Thorac Soc*. 2022;19(1):147-150. doi:10.1513/AnnalsATS.202107-796RL

- 54 Agency for Healthcare Research and Quality. Poverty and Access to Internet, by County. AHRQ. June 2021. Accessed November 28, 2022. Retrieved from <https://www.ahrq.gov/sdoh/data-analytics/sdoh-tech-poverty.html>
- 55 Agency for Healthcare Research and Quality. Social Determinants of Health Database. AHRQ. July 2022. Accessed November 28, 2022. Retrieved from <https://www.ahrq.gov/sdoh/data-analytics/sdoh-data.html>
- 56 Agency for Healthcare Research and Quality. Poverty and Access to Internet, by County. AHRQ. June 2021. Accessed November 28, 2022. Retrieved from <https://www.ahrq.gov/sdoh/data-analytics/sdoh-tech-poverty.html>
- 57 Rubin R. Internet access as a social determinant of health. *JAMA*. 2021;326(4):298. doi:10.1001/jama.2021.11733
- 58 Agency for Healthcare Research and Quality. Social Determinants of Health Environmental Scan. AHRQ. July 2020. Accessed November 28, 2022. Retrieved from <https://www.ahrq.gov/sites/default/files/wysiwyg/sdoh/sdoh-environmental-scan.xlsx>
- 59 Park TW, Lin LA, Hosanagar A, Kogowski A, Paige K, Bohnert AS. Understanding risk factors for opioid overdose in clinical populations to inform treatment and policy. *J Addict Med*. 2016;10(6):369-381. doi:10.1097/ADM.0000000000000245
- 60 National Center for Health Statistics Research Data Center. Linked data on hospitalizations, mortality, and drugs: Data from the National Hospital Care Survey 2016, National Death Index 2016-2017, and the Drug-Involved Mortality 2016-2017. October 2020. Accessed April 28, 2023. Retrieved from <https://www.cdc.gov/nchs/data/nhcs/Task-3-Doc-508.pdf>
- 61 National Center for Health Statistics. National Death Index. Centers for Disease Control and Prevention. Accessed November 11, 2022. Retrieved from <https://www.cdc.gov/nchs/ndi/index.htm>
- 62 National Center for Health Statistics. Restricted-Use Vital Statistics Data. Centers for Disease Control and Prevention. Accessed November 11, 2022. Retrieved from <https://www.cdc.gov/nchs/nvss/nvss-restricted-data.htm>
- 63 National Center for Health Statistics. Drug Use-associated Hospital Encounters Data from Selected Hospitals. Centers for Disease Control and Prevention. Accessed January 25, 2023. Retrieved from <https://www.cdc.gov/nchs/dhcs/drug-use/index.htm>
- 64 Brown AM, White DG, Adams NB, et al. Identifying co-occurring disorders among patients with an opioid involved hospital encounter using National Hospital Care Survey data. *Vital Health Stat 1*. 2022;(193):1-21.
- 65 National Center for Health Statistics Research Data Center. Identifying co-occurring disorders among opioid users using linked hospital care and mortality data: Capstone to an existing FY18 PCORTF project. June 2021. Accessed April 28, 2023. Retrieved from <https://www.cdc.gov/nchs/data/nhcs/FY19-RDC-2021-06-01-508.pdf>
- 66 Adams, N. Opioid_SUD_MHI_MedCodes. GitHub. March 2022. Accessed November 28, 2022. Retrieved from https://github.com/CDCgov/Opioid_SUD_MHI_MedCodes
- 67 Adams, N. Opioid_SUD_MHI_NLP. GitHub. December 2021. Accessed November 28, 2022. Retrieved from https://github.com/CDCGov/SUD_MHI_NLP
- 68 National Center for Health Statistics. Identifying co-occurring disorders among opioid users using linked hospital care and mortality data: Capstone to an existing FY18 PCORTF project: Final report. Centers for Disease Control and Prevention. October 2021. Accessed April 28, 2023. Retrieved from <https://aspe.hhs.gov/sites/default/files/documents/6b78783ff57c2d5baccce2e71e148f4b/fy19-nhcs-aspe-final-report-10-18-21.pdf>

- 69 Stram M, Gigliotti T, Hartman D, et al. Logical Observation Identifiers Names and Codes for laboratorians: potential solutions and challenges for interoperability. *Arch Pathol Lab Med*. 2020;144(2):229-239. doi:10.5858/arpa.2018-0477-RA
- 70 Galaitsi SE, Cegan JC, Volk K, Joyner M, Trump BD, Linkov I. The challenges of data usage for the United States' COVID-19 response. *Int J Inf Manage*. 2021;59. doi:10.1016/j.ijinfomgt.2021.102352
- 71 Cholan RA, Pappas G, Rehwoldt G, et al. Encoding laboratory testing data: Case studies of the national implementation of HHS requirements and related standards in five laboratories. *J Am Med Inform Assoc*. 2022;29(8):1372-1380. doi:10.1093/jamia/ocac072
- 72 LOINC from Regenstreif. LOINC Mapping Guides. LOINC. 2022. Accessed December 1, 2022. Retrieved from <https://loinc.org/guides/#in-development>
- 73 Centers for Disease Control and Prevention. LOINC in Vitro Diagnostic (LIVD) Test Code Mapping. CDC Division of Laboratory Systems. 2022. Accessed December 1, 2022. Retrieved from https://www.cdc.gov/csels/dls/livd-codes.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcsels%2Fdls%2Fsars-cov-2-livd-codes.html
- 74 Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Future Healthc J*. 2019;6(2):94-98. doi:10.7861/futurehosp.6-2-94
- 75 U.S. Department of Health and Human Services (HHS). Artificial Intelligence (AI) strategy. HHS. January 2021. Accessed April 28, 2023. Retrieved from <https://www.hhs.gov/sites/default/files/final-hhs-ai-strategy.pdf>
- 76 National Institute of Diabetes and Digestive and Kidney Diseases. United States Renal Data System. National Institutes of Health. Accessed December 1, 2022. Retrieved from <https://www.niddk.nih.gov/about-niddk/strategic-plans-reports/usrds>
- 77 National Institute of Allergy and Infectious Diseases. TB Portals. NIH. Accessed December 1, 2022. Retrieved from <https://tbportals.niaid.nih.gov/download-data>
- 78 The Office of the National Coordinator for Health Information Technology. Training data for machine learning to enhance PCOR data infrastructure: Implementation guidance. HealthIT.gov. August 2021. Accessed April 28, 2023. Retrieved from https://www.healthit.gov/sites/default/files/page/2021-09/ONC%20Training%20Data%20for%20ML_Implementation%20Guide.pdf
- 79 Booz Allen Hamilton. Training Data for machine learning (ML) to enhance patient-centered outcomes research (PCOR) data infrastructure: Implementation guide. HealthIT.gov. September 2021. Accessed April 28, 2023. Retrieved from https://www.healthit.gov/sites/default/files/page/2021-09/ONC%20Training%20Data%20for%20ML_Implementation_Guide%20Project%20Overview.pdf
- 80 Karki M, Kantipudi K, Haghghi, et al. Training data for machine learning to enhance patient-centered outcomes research (PCOR) data infrastructure — A case study in tuberculosis drug resistance: Implementation guide. National Library of Medicine, National Institutes of Health. March 7, 2023. Accessed April 28, 2023. Retrieved from https://github.com/NLMLHC/NLM-PCOR-TB/blob/main/Implementation_Guide_PCOR.pdf
- 81 The Office of the National Coordinator for Health Information Technology. Building Data Infrastructure to Support Patient-Centered Outcomes Research (PCOR) Data Infrastructure: Resources. HealthIT.gov. Accessed December 1, 2022. Retrieved from <https://www.healthit.gov/topic/scientific-initiatives/pcor/machine-learning>
- 82 The Office of the National Coordinator for Health Information Technology. Training Data for Machine Learning to Enhance PCOR Data Infrastructure. GitHub. 2022. Accessed December 1, 2022. Retrieved from <https://github.com/onc-healthit/2021PCOR-ML-AI>

- 83 National Library of Medicine, National Institutes of Health. NLM PCOR TB. GitHub. November 17, 2022. Accessed January 25, 2023. Retrieved from <https://github.com/NLMLHC/NLM-PCOR-TB>
- 84 Zheng Q, Lu Y, Lure F, Jaeger S, Lu P. Clinical and radiological features of novel coronavirus pneumonia. *J Xray Sci Technol*. 2020;28(3):391-404. doi:10.3233/XST-200687
- 85 Ma L, Wang Y, Guo L, et al. Developing and verifying automatic detection of active pulmonary tuberculosis from multi-slice spiral CT images based on deep learning. *J Xray Sci Technol*. 2020;28(5):939-951. doi:10.3233/XST-200662
- 86 Yang F, Yu H, Kantipudi K, et al. Differentiating between drug-sensitive and drug-resistant tuberculosis with machine learning for clinical and radiological features. *Quant Imaging Med Surg*. 2022;12(1):675-687. doi:10.21037/qims-21-290
- 87 Karki M, Kantipudi K, Yu H, et al. Identifying drug-resistant tuberculosis in chest radiographs: Evaluation of CNN architectures and training strategies. *Annu Int Conf IEEE Eng Med Biol Soc*. 2021;2021:2964-2967. doi:10.1109/EMBC46164.2021.9630189
- 88 Karki M, Kantipudi K, Yang F, et al. Generalization challenges in drug-resistant tuberculosis detection from chest X-rays. *Diagnostics (Basel)*. 2022;12(1):188. Published 2022 Jan 13. doi:10.3390/diagnostics12010188
- 89 Yang F, Yu H, Kantipudi K, et al. . Automated drug-resistant TB screening: Importance of demographic features and radiological findings in chest X-ray. 50th IEEE Applied Imagery Pattern Recognition Workshop (AIPR). 2021;1-4, doi: 10.1109/AIPR52630.2021.9762198.
- 90 Booz Allen Hamilton. Training data for machine learning (ML) to enhance patient-centered outcomes research (PCOR) data infrastructure: Final report. The Office of the National Coordinator for Health Information Technology. Published September 2021. Accessed February 2, 2023. Retrieved from https://www.healthit.gov/sites/default/files/page/2021-09/ONC_Training%20Data%20for%20ML_PCOR_Final%20Report.pdf
- 91 Karki M, Kantipudi K, Haghghi, et al. Training data for machine learning to enhance patient-centered outcomes research (PCOR) data infrastructure — A case study in tuberculosis drug resistance: Final Report. National Library of Medicine, National Institutes of Health. March 7, 2023. Accessed April 28, 2023. Retrieved from https://github.com/NLMLHC/NLM-PCOR-TB/blob/main/Final_PCOR_Report.pdf
- 92 Chronic Conditions Data Warehouse. Centers for Medicare and Medicaid Services. Accessed January 26, 2023. Retrieved from <https://www2.ccwdata.org/web/guest/home/>
- 93 Kim DH, Schneeweiss S, Glynn RJ, Lipsitz LA, Rockwood K, Avorn J. Measuring frailty in Medicare data: Development and validation of a claims-based frailty index. *J Gerontol A Biol Sci Med Sci*. 2018;73(7):980-987. doi:10.1093/gerona/glx229
- 94 The Office of the Assistant Secretary for Planning and Evaluation. Frailty Algorithm Codes. Github. 2022. Accessed December 12, 2022. Retrieved from <https://github.com/grimnr14/frailty>.
- 95 Kharrazi H, Roblin WD, Weiner JP. Developing and assessing the validity of claims-based indicators of frailty & functional disabilities in electronic health records: Final report. Agency for Healthcare Research and Quality. September 2022. AHRQ Publication No. 22-0058. Accessed April 28, 2023. Retrieved from <https://www.ahrq.gov/sites/default/files/wysiwyg/learning-health-systems/frailty-ehr.pdf>.
- 96 The Office of the Assistant Secretary for Planning and Evaluation. Explore the OS-PCORTF Project Profiles. ASPE. Accessed January 31, 2023. Retrieved from <https://aspe.hhs.gov/collaborations-committees-advisory-groups/os-pcortf/explore-portfolio>

Assistant Secretary for Planning and Evaluation
U.S. Department of Health and Human Services
200 Independence Avenue, SW
Washington, D.C. 20201
Room 415F
+1 202.690.7858

Publication Date: May 2023

This report was prepared under Contract No. HHSP233201500048I/75P00121F37020 between the Department of Health and Human Services' Office of the Assistant Secretary for Planning and Evaluation (ASPE) Office of Health Policy and NORC at the University of Chicago.

Suggested Citation: Dullabh P, Dhopeswarkar R, Heaney-Huls K, Desai P, Jimenez F, Peterson C, Kane R, Tarver D, Wilson K, Wei S, and Smith S. The 2022 Portfolio Report Building Data Capacity for Patient-Centered Outcomes Research. Washington, DC: Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services. May 2023.