2025

ANNUAL PORTFOLIO REPORT

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

The Office of the Secretary Patient-Centered Outcomes Research Trust Fund (OS-PCORTF) works to build data capacity for patient-centered outcomes research (PCOR). It does so through a portfolio of projects developed, implemented, and coordinated in partnership with United States (U.S.) Department of Health and Human Services (HHS) agencies. OS-PCORTF projects advance HHS policy priorities by strengthening the data infrastructure that makes it easier for researchers and private sector stakeholders to collect, link, and analyze data that can lead to more patient-centered care.

OS PCOR TF
OFFICE OF THE SECRETARY
PATIENT-CENTERED OUTCOMES
RESEARCH TRUST FUND

Visit the OS-PCORTF website at https://aspe.hhs.gov/collaborations-committees-advisory-groups/os-pcortf.

Explore the OS-PCORTF project portfolio at https://aspe.hhs.gov/collaborations-committees-advisory-groups/os-pcortf/explore-portfolio.

Mission

Build and strengthen data capacity for patient-centered outcomes research through coordination across agencies and federal programs.

Vision

Better data for patient-centered outcomes research to improve evidence generation, decision-making, and health outcomes for all Americans.

In Fiscal Year 2025, OS-PCORTF projects and other OS-PCORTF initiatives collectively enhanced the data infrastructure available for patient-centered and comparative clinical effectiveness research (CER) across national and HHS health objectives. OS-PCORTF projects are tackling critical health challenges that affect the wellbeing of all Americans, including chronic disease prevention and management, health impacts of environmental factors, and the ongoing mental health crisis. In 2019, Congress passed legislation that was signed by the President to reauthorize OS-PCORTF through 2029. The reauthorization specified maternal mortality, intellectual and developmental disabilities (ID/DD), and economic burden outcomes as research priorities.[1] The OS-PCORTF also advances gold standard science by promoting innovations in rigorous and transparent research methods and the use of high-quality data to support informed decision-making based on sound evidence.

This report describes the activities and accomplishments of 37 active OS-PCORTF projects during Fiscal Year 2025 in four key areas noted below:

Improving Data Capacity for National Health Needs

Enhancing Data Linkages for Longitudinal Research

Leveraging Technology Solutions to Innovate in Research

Advancing Person-Centeredness

More information about the 37 Fiscal Year 2025 OS-PCORTF projects, including detailed profiles of recently completed and newly funded projects, can be found in the accompanying 2025 Annual Portfolio Report Appendices. The OS-PCORTF website provides information on the entire OS-PCORTF portfolio, including past projects.

The report is organized in the following sections:

The Breadth and Scope of OS-PCORTF Projects 3

Highlights project engagement activities, data infrastructure accomplishments, and topics of national importance.

- HHS Interagency Partnerships and Stakeholder Engagement
- Accomplishments in Data Infrastructure

Improving Data Capacity for National Health Needs 10

Describes how OS-PCORTF projects and ASPE-led activities address national and HHS health needs related to building PCOR data capacity.

- Preventing and Managing Chronic Health Diseases
- Maternal and Pediatric Health
- Intellectual and Developmental Disabilities
- Responding to Health Emergencies

Enhancing Data Linkages for Longitudinal Research 16

Highlights OS-PCORTF projects focused on providing more relevant and comprehensive data to support longitudinal patient-centered outcomes research.

- Improving Data Linkage Efficiency
- Enhancing Data Transparency & Reproducibility
- Enabling Mortality Data Linkages

Leveraging Technology Solutions to Innovate in Research 21

Summarizes project activities that leverage technology to enhance analytic resources and improve data accessibility to patient-centered outcomes research.

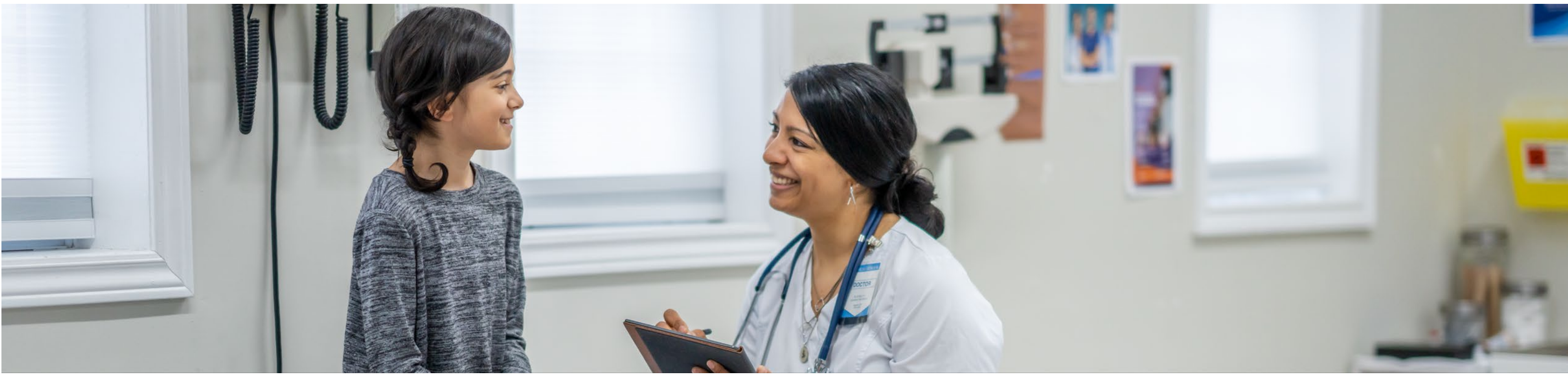
- Artificial Intelligence & Synthetic Data Creation
- Improving Interoperability for Real-Time Data Exchange

Advancing Person-Centeredness 24

Describes how OS-PCORTF projects expand the collection and analysis of data to advance person-centeredness in research.

- Addressing Non-Medical Factors
- Gathering Patient-Reported Data
- Facilitating Research for Medicaid Beneficiaries

The Breadth and Scope of OS-PCORTF Projects



The OS-PCORTF brings together HHS agency leaders, informaticians, data scientists, patients, and researchers around the shared interest in increasing data capacity for patient-centered outcomes research.

37 Projects in Fiscal Year

15 Completed Projects

19 Continued Projects

3 New Projects

HHS Interagency Partnerships and Stakeholder Engagement

Through interagency partnerships and stakeholder engagement, OS-PCORTF projects ensure that the perspectives of PCOR stakeholders are captured across the full lifecycle of product design, use, and uptake. These engagements ensure meaningful patient involvement in project work, support efficiencies across HHS activities, and spur innovation within HHS as well as in the private sector.

Capturing patient voices ensures that data products are responsive to patient needs and preferences. OS-PCORTF projects use many mechanisms to engage patients in the design and development of data infrastructure resources. For example, patients have been included as partners in research via steering committees, technical expert panels (TEPs), working groups, and task forces to provide insights during project planning and implementation. One project, which is linking administrative, assessment, and survey data for individuals with ID/DD, solicited input from stakeholders, including individuals with lived experience, to identify potential challenges with linking state data, and to build community buy-in for the dataset that is being created. Patients have also been engaged as research participants via surveys and user testing. For example, a project engaged patients to test the usability of a patient-facing care planning app and to ensure measures important to patients, such as goals and person-centered outcomes, were appropriately captured at the point of care.

End-User Engagement

28 Fiscal Year 2025 projects engaged data product end-users.

Patient Engagement

9 Fiscal Year 2025 projects specifically engaged patients.

Interagency collaborations support the development and implementation of transparent and more efficient data infrastructure across HHS agencies. Several project teams convened cross-agency TEPs and working groups to identify more efficient strategies for collecting and sharing data, reduce redundancy in data infrastructure initiatives, and promote transparency across HHS. These cross-agency collaborations reduce duplicative efforts by identifying opportunities to mobilize and leverage existing resources available within and outside of HHS. For example, one project's TEP helped identify existing disability measurement initiatives from federal and national partners that could be leveraged by the project. Cross-agency collaboration also helps inform more comprehensive processes for sharing linked datasets and other data products. For example, one project will convene a task force to modify data linkage contracts to accommodate interagency data access and sharing. Another project will work with federal partners to design a process that simplifies sharing of a new linked maternal health dataset. These OS-PCORTF collaborations have been particularly helpful for addressing research areas of national interest, such as research on maternal health and the ID/DD community, given that multiple agencies are conducting work related to these communities.

Cross-Agency Collaboration

34 Fiscal Year 2025 projects involve interagency partnerships.

OS-PCORTF projects accelerate private sector adoption by convening stakeholders to develop standards and resources that enable seamless data sharing across health systems and industry partners. Common data elements (CDEs), Health Level Seven International (HL7®) implementation guides, and Fast Healthcare Interoperability Resources (FHIR®) standards provide a common language and rules that electronic health record (EHR) developers can use to foster interoperability between data systems. These efforts to create standardized solutions reduce the burden for developers and support efficiency by eliminating the need for custom interfaces between systems. Several OS-PCORTF projects engaged with data standard stakeholders through participation in HL7 Connectathons as well as undergoing the formal HL7 "balloting" process to gather community input and enhance development of FHIR implementation guides.

Industry Involvement

8 Fiscal Year 2025 projects engaged industry stakeholders in project planning or execution.

Two OS-PCORTF projects developed implementation guides that have been published as “Standard for Trial Use” after HL7 balloting, allowing industry stakeholders to test and refine them within production environments. Other OS-PCORTF projects collaborate with health information exchange (HIE) organizations to develop methodologies for data exchange and standards processing. For example, one project will work with two HIE organizations to validate data element mapping for a real-world dataset that has been harmonized to a common data model. Another project is implementing and testing OS-PCORTF-developed FHIR standards with the HL7 Gravity Project,[2] a national collaborative representing private and public stakeholders. This project is working to harmonize concepts and use cases related to non-medical factors within the context of shared care planning.

OS-PCORTF data products are being enhanced and adopted across HHS and applied to novel research topics. Several OS-PCORTF projects have indicated that their data products are being used by other HHS initiatives or adapted in new ways by agencies to study various research topics. For example, projects will explore linkages of mortality datasets (e.g., National Death Index [NDI]) that had been enhanced through earlier OS-PCORTF efforts with other data sources such as the Healthcare Cost and Utilization Project (HCUP) and Centers for Medicare and Medicaid Services (CMS) claims data. These linkages will enable more granular research, such as investigating causes and patterns of mortality at the individual level. Another dataset created by an OS-PCORTF project that linked National COVID Cohort Collaborative (N3C) clinical data with CMS claims data has been used by 42 distinct research projects to explore critical patient-centered outcomes, including the impact of chronic diseases on mortality and morbidity, and health care utilization patterns across different populations. The longitudinal dataset captures over 10 million people and supports rigorous analyses comparing the effectiveness of COVID-19 treatments, such as steroids and antivirals, for different populations, accounting for patients’ health histories, demographic characteristics, and service utilization. Additionally, an OS-PCORTF project team developed a common data model (CDM)-formatted Transformed Medicaid Statistical Information System (T-MSIS) dataset which has been incorporated into the Food and Drug Administration (FDA) Sentinel System, and is now regularly used for post-market safety surveillance of medical products. The project’s Medicaid transformation code is also being applied by a health care insurer to claims data to assess the safety and effectiveness of medication use during pregnancy. Finally, OS-PCORTF project teams are collaborating with the United States Core Data for Interoperability Plus (USCDI+)[3] to include additional data elements related to care planning, maternal health, and sickle cell disease, which will provide essential guidance for data standardization for patient-centered outcomes research.

Spotlight | Engaging with the Private Sector to Promote Adoption of OS-PCORTF Products

Code Map Services for Interoperability of Common Data Models and Data Standards. This project developed common data elements (CDEs) and software tools that automatically map how the data fields and structures of one common data model (CDM) map to other CDMs. These tools support cross-research network interoperability and data sharing for patient-centered outcomes research. Several private sector data standard organizations, including the Clinical Data Interchange Standards Consortium (CDISC), OpenAPI, and Stanford’s Center for Expanded Data Annotation and Retrieval (CEDAR) Metadata Workbench, were involved in the development and promotion of the project’s data products. The endorsement of private sector organizations will help these products’ acceptability and use within the research community. The project team is in discussion with two state health information exchange (HIE) organizations and three data integration vendors to validate and test the standardized data element mapping products for researchers. These partnerships promote long-term impact and viability of the data products.

Accomplishments in Data Infrastructure

PCOR data infrastructure includes the methods, tools, and services that support access to high-quality data. Project impacts were evaluated across the four domains described below.

OS-PCORTF projects collectively advanced the ability of researchers to conduct patient-centered outcomes research by:

% of Projects

Improving Data Access and Use

Through efforts that support more efficient data access and timelier conduct of PCOR studies

Providing More Relevant, Comprehensive Data

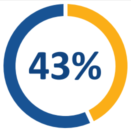
Through collection and integration of patient-generated health data, inclusion of new variables, and data linkages

Improving Data Quality[4]

By addressing the completeness, accuracy, consistency, timeliness, and granularity of data that affect research conclusions for patient-centered outcomes research and CER

Enhancing Analytic Resources

By improving or creating analytic resources for PCOR studies



Spotlight | Facilitating Community and Clinical Data Integration in Multiple States

Childhood Obesity Data Initiative (CODI): Integrated Data for Patient-Centered Outcomes Research Project 2.0. The Centers for Disease Control and Prevention (CDC) led the creation of the Childhood Obesity Data Initiative (CODI) project in 2018 to improve the utility of the data needed to evaluate and compare the effectiveness of interventions for childhood obesity. The original project developed and piloted tools and services including a common data model (CDM), data cleaning tools, and privacy-preserving record linkage (PPRL) techniques to securely link longitudinal pediatric data originating from the Colorado Health Observation Regional Data Services (CHORDS), a distributed data network. A subsequent OS-PCORTF-funded project, CODI 2.0, expands implementation of these tools to North Carolina's statewide health and human services coordinated care network, NCCARE360. CODI 2.0 supplements the existing CODI infrastructure by producing additional implementation guides to promote adoption of the CODI CDM among health systems and community organizations. The project also incorporates lifestyle factor data to capture the impact of non-medical factors on chronic disease prevention and management. In 2023, the Administration for Community Living (ACL) funded a data linkage project that leveraged CODI tools to examine outcomes for older adults receiving services from community-based organizations (CBOs) in Maryland. This project enabled innovative analyses of the effects of community-based programs on chronic disease outcomes, securely combining data from over 30,000 people served by local CBOs with the state's health information system. The linkage was successful, with nearly three-quarters of CBO records matched with clinical data in the health information exchange (HIE). The adoption and expansion of CODI infrastructure in Colorado, North Carolina, and Maryland demonstrates its adaptability across different health systems and its potential to advance integrated, person-centered research nationwide.

The projects completed in Fiscal Year 2025 produced **34** data and technical products and **19** reports and publications.

These products enhance researchers' **capacity to conduct robust patient-centered outcomes research** to address critical questions across areas related to **preventing and managing chronic diseases and maternal and pediatric health.**

Project Reports & Publications	User Implementation Tools & Guides	Databases & Datasets	Software & Analytic Services	Data Governance	Health IT Standards	Data Element & Informational Model
19	16	8	5	3	1	1

Advancing Gold Standard Science

The OS-PCORTF enables researchers to align their work with the nine principles of gold standard science: 1) reproducible, 2) transparent, 3) communicative of error and uncertainty, 4) collaborative and interdisciplinary, 5) skeptical of its findings and assumptions, 6) structured for falsifiability of hypotheses, 7) subject to unbiased peer review, 8) accepting of negative results as positive outcomes, and 9) without conflicts of interest.[5] The OS-PCORTF impact domains align with the principles of gold standard science in the following ways:

Scientifically rigorous, patient-centered research depends on high-quality data that are objective, reliable, and systematically collected. Communicating errors and uncertainty around study findings is crucial for understanding the limits of the evidence, leading to more informed decision-making. High-quality data supports use of advanced statistical techniques to test hypotheses and enables transparent reporting of uncertainty (e.g., margins of error and confidence intervals). In contrast, low-quality data, such as incomplete or inconsistent data with anomalies, can obscure uncertainty estimates, undermine statistical validity, and limit the reliability of conclusions.

Several OS-PCORTF projects have enhanced data integrity by applying rigorous assessment and validation processes to ensure data consistency and standardization. For example, some projects harmonized data across environments by using CDEs, enabling end users to aggregate and analyze information in a consistent format across data contributors. Standardization efforts also improved consistency in how data are collected in different formats, such as EHR data and claims data, supporting more comprehensive analyses of clinical care histories and chronic diseases. Additionally, one project evaluated the standardization processes used to link survey and mortality data to ensure data accuracy and strengthen confidence in the dataset’s utility for research and public health surveillance.

Open-access data outputs and research methodologies support reproducible and transparent research and enhance generalizability and trust. Core tenets of gold standard science, reproducibility and transparency, are essential to patient-centered outcomes research, as they ensure that methods are accurate, adaptable across areas, and openly communicated throughout the research process. To achieve reproducibility, it is essential to employ rigorous experimental designs—from hypothesis generation through results reporting—that can be independently replicated and cross-validated. Transparency requires public sharing of information about each stage of the research process, including methodologies, validation procedures (e.g., peer review and disposition of comments), final outputs, and user documentation. Without these practices, research risks limited credibility, potential bias, and lower scientific integrity.

Several OS-PCORTF projects advance research accessibility and transparency by creating public-use data files from their standardization and linkage efforts, including synthetic datasets reflective of data in restricted-use files. Availability of these public-use files expands researcher access to high-quality federal and health system data, enabling wider uptake by researchers and encouraging their application to new research questions. In addition, some projects have produced public methodology reports, open-source codes hosted on GitHub, user guides that describe sampling strategies, data transformation and standardization techniques, and techniques for generating synthetic data that preserve both patient privacy and the integrity of the original data. These resources ensure that methods are transparent, reproducible, and adaptable across different contexts.

The **15** completed projects produced **26** data and technical products publicly available for research use.

26 new data resources are publicly accessible for research use

6 new data resources are open source and allow for modification and sharing

5 datasets are available upon request through a Research Data Center

Collaborative and interdisciplinary research encourages the engagement of various stakeholders, including interagency partnerships, end-users, and patients, in the development, deployment, and dissemination of research products. This approach enhances data capacity through multidisciplinary expertise and informs research outputs that are meaningful to end users and patients, thereby supporting broader uptake of evidence. Collaborative research brings together agencies and organizations with distinct missions, experts across disciplines (e.g., biomedical researchers and epidemiologists), and public-private sector partners. In contrast, siloed research efforts risk producing incomplete or biased findings due to limited scope and perspective.

Collaborative and interdisciplinary research is a cornerstone of OS-PCORTF projects, fostering engagement across federal agencies and among stakeholders. Multiple OS-PCORTF projects completed this year illustrate the impact of cross-agency collaboration, highlighting the value of shared expertise and coordinated efforts in advancing patient-centered outcomes research. These collaborations have also extended to researchers, decision-makers, patients, and advocacy groups through stakeholder meetings, tribal consultations, Connectathons, and patient-centered design processes. For example, end users, including researchers, helped prioritize data elements that comprehensively assess morbidity (e.g., hospitalization rates, chronic disease outcomes) and mortality to inform the development of synthetic datasets. Another project's research activities were informed by input from the NIH NCATS Advisory Council, which includes patient advocates. Additional projects directly involved patients in usability testing and app design to ensure that tools are accessible, user-friendly, and understandable, ultimately supporting broader adoption and scalability across different contexts and patient populations.

Improving Data Capacity for National Health Needs

The OS-PCORTF aims to improve PCOR data capacity for key national needs and HHS objectives, including chronic disease prevention and management, maternal and pediatric health, ID/DD, and responding to national emergencies. In this section we describe OS-PCORTF achievements in these areas.

Number of Fiscal Year 2025 Projects that Address National Needs

Preventing & Managing Chronic Diseases
11 Projects

Maternal & Pediatric Health
10 Projects

Intellectual & Developmental Disabilities
2 Projects

Responding to Health Emergencies
3 Projects

Preventing and Managing Chronic Diseases

Chronic diseases, including cancer, contribute to five of the top 10 leading causes of death in the U.S.[6] As a driver of morbidity and mortality as well as health care costs,[7] preventing chronic diseases—particularly pediatric chronic diseases—is a priority across federal agencies.[8] These prevention efforts include modernizing data infrastructure to understand chronic disease trends and outcomes.[9]

OS-PCORTF Project Efforts in Preventing and Managing Chronic Diseases

Eleven projects enhance the data infrastructure for evidence-based chronic disease prevention and management research – including for chronic kidney disease, sickle cell disease, cancer, and multiple chronic diseases. These projects **improve the quality and accessibility of data** by expanding data models and open-source tools for data linkages. For example, projects are harmonizing and standardizing chronic disease data elements to enable data exchange, including for chronic kidney disease and sickle cell disease. Two projects are also developing data sharing agreements and a governance framework to support secure and appropriate sharing of patient data between federal agencies.

Projects are improving the utility and richness of data on chronic diseases by linking separate data sources—such as cancer and transplant registry data, surveillance data, environmental factors data, and EHR data—to create new research datasets. These datasets bring together previously fragmented data sources to explore and compare relationships between service provision, social, economic, and environmental factors, and health outcomes in people with chronic diseases or at risk of developing a chronic disease. For example, one linked dataset enables research that identifies which community-based interventions are most effective in reducing childhood obesity for specific pediatric populations. Another dataset is enabling studies that compare effectiveness of different cancer treatment and medication protocols on improving patient health outcomes.

Finally, projects have developed tools, such as care planning apps to facilitate the collection and use of patient- and caregiver-reported data (see *New PCOR Resource for Preventing and Managing Chronic Diseases* box) and a patient-facing questionnaire to capture economic outcomes data directly from individuals and families. These efforts make patient-reported data regarding goals, treatments, economic burdens, and outcomes for chronic diseases more easily available to researchers and clinical care teams.

New PCOR Resource for Preventing and Managing Chronic Diseases

Project: ***Understanding Long-Term Outcomes in COVID-19 Survivors with Multiple Chronic Conditions (MCC) through eCare Plan Development***

EXPLORE PROJECT

The clinician-facing **eCarePlanner** application is open-source, EHR-agnostic, and FHIR-based. Health systems can implement this and the patient-facing version—**MyCarePlanner**—to improve care coordination for people with MCC. These apps collect and share data across home, community, clinical, and research settings. This project built on prior OS-PCORTF work by supporting app use for patients with Long COVID, in addition to existing capabilities for patients with chronic kidney disease, cardiovascular disease, diabetes, and/or chronic pain. This project also added a caregiver-facing component to the patient-facing app, allowing authorized caregivers access to share patient medical information for support care management. Caregivers can also share information on their burden and support needs with the patients' care team which can be used to tailor care plans.

Other OS-PCORTF Initiatives in Preventing and Managing Chronic Diseases

Exploring Data Infrastructure Availability and Expansion Opportunities for Health Outcomes Research on Sickle Cell Disease[10]

Sickle cell disease affects over 100,000 people in the U.S., causing severe health complications and significantly reducing life expectancy.[11, 12] While existing treatments offer promise, long-term data on their effectiveness are limited.[13, 14] Efforts to improve data collection are ongoing, but inconsistencies in state-level reporting of administrative, EHR, and vital records data contribute to gaps in understanding patient outcomes and health care utilization.[15] The report and inventory highlight existing individual-level and aggregate sickle cell disease data sources and outline key considerations for strengthening data collection and use to improve health outcomes research.

Maternal and Pediatric Health

The maternal care crisis is more severe in the U.S., driving higher rates of maternal and infant mortality than other developed countries, as over a third of U.S. counties in 2024 lacked adequate preventive, prenatal, and postpartum care.[16] Maternal health outcomes greatly affect the physical and mental health of children. Strengthening data infrastructure for studying maternal and child health outcomes is a key research priority specified in the PCORTF reauthorization.[17]

OS-PCORTF Project Efforts in Maternal and Pediatric Health

Over a quarter (n=10) of Fiscal Year 2025 projects address maternal and pediatric health research needs through **enhanced approaches to collect, harmonize, and link data** on mother and child health care and outcomes, and **providing higher-quality and more comprehensive data accessible to researchers**.

To help researchers get a full picture of the health of mothers and children, projects are transforming data into consistent formats using data standards and CDMs, enabling cross-system data sharing and integration. These efforts enhance the completeness, accuracy, and consistency of data, and support more reliable and generalizable patient-centered outcomes research and CER analyses across settings and jurisdictions. One project is also modernizing data collection processes for the National Ambulatory Medical Care Survey (NAMCS) by automating processes to collect EHR data using an HL7 implementation guide in order to increase the representativeness of data collected from health centers delivering maternal health care (see *New PCOR Resource for Maternal and Pediatric Health Research* box).

Several OS-PCORTF projects improve the robustness and analytic utility of data used for maternal and pediatric health research by developing linked datasets that bring together disparate data on risk factors, outcomes, and other contextual information. These resources provide longitudinal data that support more detailed analyses of the effectiveness of current practices, programs, and treatments on improving outcomes related to childhood obesity, infant health, and maternal health care access. For example, one project is linking the Boston Birth Cohort, the Pregnancy to

Early Life Longitudinal Data System, and the Agency for Healthcare Research and Quality (AHRQ) Social Determinants of Health Database to create a large-scale, longitudinal dataset of health outcomes from mother-child pairs. Using a life course perspective, the dataset will support inquiries into how community and environmental factors, health care utilization, nutrition, genetic factors, and use of government-sponsored programs affect subsequent maternal and intergenerational health outcomes. It will also enable analyses that compare the effectiveness of different maternal health interventions. Another project's efforts to create a mother-child record linkage within T-MSIS data will help researchers assess the impact of medical product use or other exposures or events that occur during pregnancy on maternal and infant outcomes. By creating public-use versions of linked datasets, OS-PCORTF projects improve researchers' and decision makers' access to data that had previously been siloed and time consuming to link for analysis. Projects also support longitudinal linkages of these data sources by creating standardized linkage processes that researchers can leverage.

New PCOR Resource for Maternal and Pediatric Health Research

Project: ***Enhancing Surveillance of Maternal Health Clinical Practices and Outcomes with Federally Qualified Health Centers' Electronic Health Records Visit Data***

EXPLORE PROJECT

The project produced a nationally representative dataset of clinical care, including maternal health care, provided at federally qualified health centers (FQHCs) by enhancing data collection processes for the 2022 NAMCS Health Center Component. The project also linked EHR data from participating FQHCs to mortality data from the NDI and U.S. Housing and Urban Development Administrative (HUD) data. Researchers can analyze the linked data to understand maternal health outcomes and care received at FQHCs. Public and restricted-use files for the 2022 NAMCS Health Center Component data are available. To access the NAMCS-NDI Linked Data File and NAMCS-HUD Linked Data File, researchers need to apply through the National Center for Health Statistics (NCHS) RDC.

Other OS-PCORTF Maternal and Pediatric Health Initiatives

Data Sources for Conducting Research on the Effectiveness of Doula Services on Maternal and Infant Outcomes[18]

Strengthening the systems of support provided to women during pregnancy has potential to improve maternal health outcomes in the U.S.[19] Doulas are trained to provide emotional, physical, and informational guidance to mothers throughout the birthing process.[20] To advance understanding of how doula services influence health outcomes, the report and accompanying data inventory describe and assess data sources that could inform a comprehensive database on doula services and related outcomes. Researchers can use this resource to identify datasets suitable for analyzing patient-centered outcomes associated with doula services and their impact on maternal and infant health.

Better Data for Maternal Health Research: HHS Efforts to Strengthen Data Infrastructure for a Healthy America

The OS-PCORTF funds efforts to build maternal health data capacity to support research on the risks and factors that improve outcomes related to prenatal, perinatal, and postpartum care. This vignette highlights 10 projects that enabled new research on key maternal health outcomes and their potential to improve health care for pregnant women. It explores project impacts to reduce preventable pregnancy-related deaths and improve maternity care, study substance use treatment outcomes for pregnant women, and improve data representativeness for understanding maternal health outcomes.

Intellectual and Developmental Disabilities

People with ID/DD have historically been excluded from health research and have faced greater barriers to accessing high-quality care,[21] contributing to higher rates of chronic disease and maternal mortality.[22] Since the 2019 PCORTF reauthorization, HHS has prioritized addressing data infrastructure gaps in ID/DD health outcomes research and including people with ID/DD in research to strengthen the evidence generated to address the health needs of people with disabilities.[1]

OS-PCORTF Project Efforts in Intellectual and Developmental Disabilities

Two projects support the OS-PCORTF objective to **improve the quality and availability of clinical and administrative datasets** that support research for individuals with ID/DD. For example, one project is working with states to integrate and harmonize data from multiple state data sources, including the Support Intensity Scale – Adult Version, Medicaid claims, and the National Core Indicators – Intellectual and Developmental Disabilities (NCI-IDD) In-Person Survey. This linkage effort will generate a richer understanding of individual needs and enable research into the connections between assessed needs, service use (including the use of home and community-based services), Medicaid spending, and person-centered outcomes and delivery system performance for the ID/DD population.

Other OS-PCORTF Initiatives in Intellectual and Developmental Disabilities

Advancing an Operational Definition of Intellectual and Developmental Disabilities for Use in Research

CMS' Chronic Conditions Data Warehouse (CCW) uses diagnostic code-based flags to identify ID/DD in claims. However, some conditions (e.g., certain chromosomal abnormalities) appear in researchers' code lists but not in the CCW flags. This gap creates inconsistencies between federal and non-federal research, limiting comparability and generalizability. This OS-PCORTF funded initiative conducted exploratory analysis of Medicare and Medicaid administrative data and engaged ID/DD experts (including those with lived experience) to identify key considerations and lay the empirical groundwork for a future update of the CCW flags. Such updates would enable a more comprehensive, standardized claims-based definition for researchers to adapt to analytic goals and evolving clinical knowledge.

Responding to Health Emergencies

The OS-PCORTF builds capacity for research addressing national health emergencies and emerging threats. In response to national public health concerns, the OS-PCORTF has funded many projects enhancing the collection, use, and accessibility of data needed for research on outcomes for people with substance use disorder (SUD) and mental health issues. The OS-PCORTF also funds advances in environmental health data accessibility and interoperability needed to study environmental impacts on health outcomes.

OS-PCORTF Project Efforts in Responding to Health Emergencies

Three projects **improve the availability and representativeness** of data needed for research on patient-centered outcomes of the opioid epidemic and environmental threats. One project employs natural language processing (NLP) techniques to more easily analyze structured and unstructured hospital data on stimulant-associated health outcomes, making identification of substance-involved hospital encounters more efficient and accurate. Projects are also creating linked datasets of disparate data (e.g., administrative data, hospital survey data, and environmental data) for environmental and behavioral health emergencies. During and after a health emergency, researchers can leverage these OS-PCORTF data products to explore relationships between non-medical factors, the environment, service utilization, and patient outcomes. For example, one project is integrating data from HCUP with geospatial environmental data, including the Hazard Mapping System Fire and Smoke data from the National Oceanic and Atmospheric Administration (NOAA), to explore the impact of wildfire smoke on asthma-related health care utilization in Oregon. This project is also creating publicly available catalogs of curated data resources to facilitate linkage of data on environmental factors and exposures with health outcomes data.

Other OS-PCORTF Initiatives in Responding to Health Emergencies

Assessing the Feasibility of Creating a National Behavioral Health Workforce Database[23]

The behavioral health workforce has significant labor shortages and unequal geographic and specialty provider distribution, impacting health outcomes, and access to and quality of care. This report assesses the feasibility of creating a database of individual behavioral health providers (including clinicians and allied health professionals) and practices. Such a database would inform research aiming to address behavioral health workforce challenges, such as provider distribution, gaps in access to care, and the relationship between provider supply and patient outcomes.

Enhancing Data Linkages for Longitudinal Research

In Fiscal Year 2025, 15 projects worked to improve the efficiency, transparency, and reliability of data linkages for longitudinal patient-centered outcomes research and CER. These projects improve the utility and accessibility of longitudinal data needed to generate comprehensive comparisons of program or treatment outcomes. They also create more efficient and reproducible linkage procedures through streamlined and standardized linkage processes, governance frameworks, and policy documentation. Several key linkage efforts focused on enhancing mortality data to enable research on factors associated with increased risk of mortality.

Number of Fiscal Year 2025 Projects that Improve Data Linkage Efficiency Across Three Key Areas

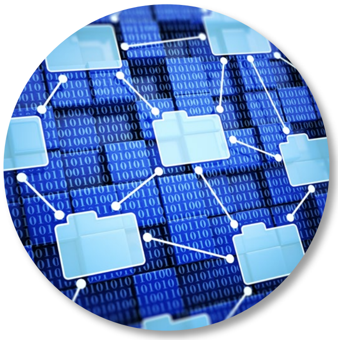
Improving Data Linkage Efficiency
6 Projects

Enhancing Data Transparency & Reproducibility
5 Projects

Enabling Mortality Data Linkages
4 Projects

Improving Data Linkage Efficiency

Fragmented data hinder research by limiting access to comprehensive intervention and outcome information needed to answer key questions. Preparing data sources for linkages and developing linkage methodologies are often time-intensive for researchers.



However, efficiencies can be created by leveraging and adapting existing linkage tools and processes created and tested by others. OS-PCORTF projects are addressing this by harmonizing disparate data sources and creating standardized linkage processes, streamlining the steps that researchers need to take and enhancing the overall efficiency of patient-centered outcomes research.[24]

OS-PCORTF Project Efforts in Improving Data Linkage Efficiency

Six projects are creating publicly accessible tools and resources that support researchers in linking disparate data sources containing administrative, EHR, and survey data. These projects create processes to **more efficiently link data** and increase **availability of longitudinal data** for research on non-clinical factors that can influence health. Fiscal Year 2025 projects enhance the data linkage infrastructure by improving PPRL techniques and creating methodological resources to standardize linkage processes for datasets commonly used in health outcomes research (e.g., Medicaid claims, EHR, and survey data). Projects are also leveraging CDMs to harmonize disparate data, including CMS data, across states and years (see *New PCOR Resources for Efficiency of Data Linkage* box).

The OS-PCORTF is helping to improve the overall security of processes that utilize protected health information to link datasets. For example, one project is advancing data linkage techniques that preserve individuals' privacy by evaluating the accuracy and completeness of data linked through an open-source PPRL tool compared to data linked through existing PPRL gold standard methodologies (see *New PCOR Resources for Efficiency of Data Linkage* box). Another project is developing and implementing a generalizable set of technical requirements and agreements to support broader data sharing across federal and non-federal organizations. These efforts are designed to reduce barriers to data sharing, aggregation, and linkage, thereby improving researchers' ability to access and use high-quality data for evidence generation.

Longitudinal data allow researchers to study interventions that are dynamic, comparing how changes impact outcomes over time. Furthermore, leveraging longitudinal datasets that include information on non-clinical factors provides a more comprehensive understanding of how interventions work under different circumstances. For example, one OS-PCORTF project is developing state-based databases that integrate data from disparate sources (e.g., claims data, provider-level data, community-level data) to give a more complete picture of patient factors that affect health outcomes, beyond care delivery. Another project is linking datasets to capture the full trajectory of cancer care, which enables comparative assessments of cancer treatments for distinct patient cohorts. Such linked data support research on the impact of treatments on outcomes, while controlling for non-medical factors.

New PCOR Resources for Efficiency of Data Linkage

Project: *Creating a Federal COVID-19 Longitudinal Patient Outcomes Research Database Linked to Health Systems and Clinical Data*

EXPLORE PROJECT

This project harmonized Medicare and Medicaid claims data to the Observational Medical Outcomes Partnership (OMOP) CDM and linked these data to clinical EHR data from 84 health systems that contribute data to NIH's National COVID Cohort Collaborative (N3C) data enclave of real-world EHR data. Researchers can request access to three file types (limited dataset, de-identified dataset, and synthetic dataset) of the **Linked CMS Medicare and Medicaid Claims-N3C Dataset**. The **N3C PPRL Enrichment Dashboard** displays key characteristics of the CMS-N3C linked data, including how data for patients in the N3C Data Enclave have been enriched through PPRL with CMS claims data.

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

EXPLORE PROJECT

This project evaluated the use of Annonlink, a Python-based open-source PPRL tool, to link national population survey data (National Hospital Care Survey [NHCS]) with NDI data. A manuscript summarizing the assessment findings is forthcoming and will help researchers understand the strengths and limitations of Annonlink and its applicability for their own studies.

Enhancing Data Transparency and Reproducibility

Transparent and reproducible data are key characteristics of gold standard science as such data allow others to validate the accuracy of study findings.[25] This accountability fosters trust and ensures the integrity of the evidence generated by studies. Several OS-PCORTF projects are developing resources to enhance researchers' understanding of the strengths, limitations, and appropriate research applications of linked data, how linked data are created, and how to reliably reproduce data linkages.

OS-PCORTF Project Efforts in Enhancing Data Transparency and Reproducibility

Five projects improve the transparency of data used for patient-centered outcomes research by developing tools and resources that inform end users about how data are collected, used, and should be managed. One project is developing machine-readable metadata standards that provide researchers with information on the characteristics, conditions, and usage requirements for high-priority HHS datasets, enabling their effective use and linkage to other datasets for research. Projects also **leverage standardized data formats to improve alignment of data across disparate sources**, enabling researchers to more readily use, retrieve, link, aggregate, and analyze data for patient-centered outcomes research. For example, one project formatted T-MSIS data into two different CDMs, and made the formatting code open source so it can be adapted for use in other research that requires T-MSIS linkage (see *New PCOR Resource for Data Transparency and Reproducibility* box).

OS-PCORTF projects are disseminating technical assistance tools and frameworks to assist researchers in applying consistent linkage approaches that can be reproduced. Outputs include a **data linkage roadmap** for replicating linkages between clinical, administrative, and survey data, as well as a **legal and governance framework** detailing a common data linkage governance process for a federated linked data system. Researchers can leverage these resources in their own work to securely and efficiently link datasets for new use cases.

New PCOR Resource for Data Transparency and Reproducibility

Project: ***Making Medicaid Data More Accessible Through Common Data Models and FHIR APIs***

EXPLORE PROJECT

To support the use of T-MSIS data for research, this project developed open-source code to standardize T-MSIS data into **Sentinel** and **OMOP** CDM formats. By standardizing T-MSIS data across years and states into a consistent format, researchers can perform more reliable multistate analyses. Additionally, this project produced a **public dataset of T-MSIS data formatted in the Sentinel CDM** to enable researchers to more readily retrieve, link, and analyze formatted T-MSIS data with other data sources formatted in the Sentinel CDM.

Enabling Mortality Data Linkages

Premature mortality is a fundamental outcome in patient-centered outcomes research, cutting across diseases and conditions. Linking mortality data to other sources of clinical- and health-related information can support research on factors that increase the risk of mortality and protect against premature death.

OS-PCORTF Project Efforts in Enabling Linkage of Mortality Data

Four projects address the need for greater standardization of mortality data to support health outcomes research and CER. Projects are developing standardized methodologies to link mortality data to clinical and administrative datasets, **providing more comprehensive datasets for longitudinal analyses of mortality outcomes** across different populations and care settings. Additionally, one project is increasing the **accessibility of timely mortality data** by enabling more frequent releases of NDI data and modernizing the NDI's infrastructure through cloud migration and development of PPRL methods. Finally, a project compiled a forthcoming database of requirements for using mortality data from each of the 57 reporting jurisdictions, which can help inform future efforts to standardize approaches for accessing these data (see *New PCOR Resource for Linked Mortality Data* box).

New PCOR Resource for Linked Mortality Data

Project: ***Compendium of Policies on Use of Linkable Mortality Data for PCOR in the Nation's 57 Mortality-Reporting Jurisdictions***

EXPLORE PROJECT

The **Mortality Data Linkages for Research report** summarizes the Compendium of Policies on Use of Linkable Mortality Data—a database detailing polices on accessing and using mortality data from all 57 mortality-reporting jurisdictions in the U.S. The report identifies key features of mortality data policies to help researchers determine which jurisdictions' data would be appropriate to use for their study. The report describes five jurisdictions (KY, NC, OR, PA, and WA) as exemplar jurisdictions that meet the essential data and policy needs for patient-centered outcomes research. The report also highlights data policy best practices for decision makers to consider to facilitate the broader use of mortality data for research. The Compendium is a filterable database that enables users to further explore mortality data policies across jurisdictions, categorizing these policies across six dimensions including data description; application and review process; legal, ethical, and privacy review requirements; permitted and disallowed use and users; data protection requirements; and data linkage. The Compendium will be published in 2026.

Looking Ahead

In Fiscal Year 2025, OS-PCORTF funded two new projects that enhance mortality data linkages. These projects seek to enhance existing federal infrastructure by designing, developing, and implementing methods for linking mortality data from the NDI with various data sources. One project is split into two phases—the initial phase seeks to design and develop methods for linking mortality data from the NDI with AHRQ HCUP data. Methods may be adapted for linking mortality data from the NDI with Medicare and Medicaid data. The second phase will apply the methodology to link NDI data with AHRQ HCUP data in collaboration with HCUP Partners and the NCHS and disseminate the linked NDI data for patient-centered outcomes research. Additionally, one of these projects will offer training resources on how to use the linked NDI data for patient-focused research.

Leveraging Technology Solutions to Innovate in Research

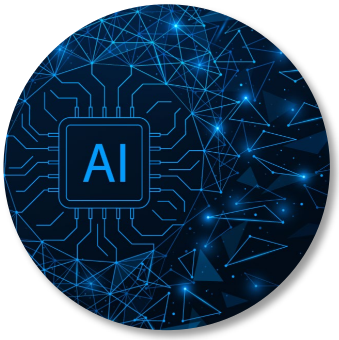
Ten OS-PCORTF projects leverage cutting-edge technology solutions to enhance analytic resources and improve the quality and accessibility of PCOR data. OS-PCORTF funding supports work to develop and test new technologies and methods for clinical data interoperability and automated exchange of real-world data for research. Additionally, projects leverage artificial intelligence (AI)-enabled tools to identify and collate relevant EHR data for research and develop statistical techniques to create synthetic datasets that reflect trends in real-world data. These efforts support more widespread and efficient access to data that provide a holistic view of individuals' health while preserving their privacy.

Number of Fiscal Year 2025 Projects that Leverage Two Types of Cutting-Edge Technology Solutions

Artificial Intelligence & Synthetic Data Creation
3 Projects

Improving Interoperability for Real-Time Data Exchange
7 Projects

Artificial Intelligence and Synthetic Data Creation



AI and synthetic data are transforming health outcomes research. AI enables faster and scalable analyses of large datasets, which helps to improve data integration, accelerate analytical workflows, and preserve individual privacy.[26] AI can also address data limitations common in health outcomes studies such as timeliness, quality, and usability.[27] Synthetic data, which is generated to reflect real-world patterns without revealing personal information, also facilitates secure data sharing and enhances researchers' access to timely, high-quality data. Synthetic data enables research on topics where real-world data are scarce, such as rare diseases.[28] Additionally, growing availability of synthetic data supports the application of AI for health outcomes research, as it can be used to train or evaluate AI models, especially when use of real-world data are constrained by privacy or regulatory policies.

OS-PCORTF Project Efforts in Artificial Intelligence and Synthetic Data Creation

Three projects are **expanding access to clinical EHR data while safeguarding privacy**. Two projects leverage AI, including split learning predictive modeling and NLP, to 1) facilitate EHR data exchange between health information exchange (HIE) organizations while preserving patient privacy, and 2) extract data from structured and unstructured EHR and administrative data to identify hospital encounters involving stimulants. Collectively, these tools will help to improve the interoperability of EHR data from three HIEs, and enable linkages of hospital EHR and NCHS survey data to NDI data to study SUD outcomes. A third project developed two synthetic linked datasets to increase access to data resources previously available only to approved users via secure data enclaves (see *New PCOR Resource for Synthetic Data* box). The synthetic data generation infrastructure created under this project can be applied to future linked datasets, supporting broader use of federal data for patient-centered outcomes research.

New PCOR Resource for Synthetic Data

Project: ***Building Infrastructure and Evidence for COVID-19 Related Research, Using Integrated Data***

EXPLORE PROJECT

This project created two synthetic linked datasets to preserve individual privacy while improving the public availability of NCHS data that has traditionally been restricted to approved users. The first synthetic dataset is based on linked 2018 National Health Interview Survey (NHIS) data, CMS Medicare claims data, and HUD administrative data. The second linked dataset derives from 2016 NHCS data linked to NDI data, and is expected to be released in 2026. Researchers can access the public-use synthetic **2018 NHIS-HUD-CMS dataset** for studying the effect of housing assistance on hospitalizations and other types of healthcare use. To provide end users with feedback on the similarity of results from synthetic and confidential data analyses, the NCHS team has developed a process for end users to verify selected synthetic data results against the original data upon which it is based. The verification process can provide synthetic data users with feedback on analyses where the restricted-use data might be more suitable.

Improving Interoperability to Support Real-Time Data Exchange

Data interoperability is essential for system information exchange that supports integrated clinical care and research on the full scope of outcomes meaningful to patients, clinicians, and decision makers. Tools that transform data into a common format and structure enable efficient exchange and use of data across traditionally siloed systems. Data standardization and harmonization tools are key to producing comparable and interoperable data.

OS-PCORTF Project Efforts in Improving Interoperability to Support Real-Time Data Exchange

Seven projects are developing tools and resources to **enhance data interoperability and facilitate more efficient exchange** of comprehensive data needed for health outcomes research and CER. These projects are creating technical data standards and leveraging CDMs to transform and harmonize disparate datasets into common formats, enabling integration and linkage with external data. Projects are producing FHIR implementation guides that define and structure EHR data for research to support consistent transmission and exchange when implemented by research networks, health systems, and other organizations. Specifically, they are creating and pilot testing FHIR implementation guides that support research on risk factors of and effective interventions for childhood obesity, care coordination for people with multiple chronic diseases, and research on the full range of factors that affect patients with sickle cell disease. Two projects are also leveraging the USCDI+, which is an Assistant Secretary for Technology Policy/Office of the National Coordinator for Health IT (ASTP) initiative to define, standardize, and harmonize data elements for broad health-related areas as well as specific federal priority programs and use cases. These projects are using USCDI+ to identify gaps and create data element sets for research on maternal health and sickle cell disease. Two other projects are using CDMs to develop innovative tools that expand researcher access to clinical data across a range of populations and settings. One project is building an automated CDM mapping tool that enables cross-network and cross-CDM interoperability and data sharing. Another project is creating a CDM to standardize, exchange, and link data submitted by organizations participating in a clinical research network using PPRL techniques. Converting disparate datasets into CDMs improves their interoperability, allowing them to be shared and integrated across applications and organizations. This, in turn, allows researchers to leverage data that represent different patients, leading to findings that are more generalizable to broader populations.

Looking Ahead

In Fiscal Year 2025, OS-PCORTF funded one new project that enhances data interoperability. This project aims to define the infrastructure needed to develop a clinical performance registry for the Indian Health Service (IHS). The registry seeks to improve clinical performance measurement across IHS providers by integrating and harmonizing data from multiple EHR systems. These clinical performance metrics intend to inform decision-making related to effective clinical interventions and quality improvement efforts to improve patient outcomes.

Advancing Person-Centeredness

The OS-PCORTF advances person-centered outcomes research by funding 15 projects that enhance the collection and analysis of robust, relevant data for rigorous studies. These projects are developing resources that enhance the availability of key data, including data on relevant and measurable non-medical factors and patient-reported data. OS-PCORTF projects also provide higher quality health outcomes data for Medicaid beneficiaries through improved data collection efforts, data linkages, and database development. These efforts ensure availability of representative data that supports research that is responsive to patient needs, prioritizes patient outcomes, and leads to actionable evidence for improving health.

Number of Fiscal Year 2025 Projects that Advance Person-Centeredness Across Three Key Areas

Addressing Non-Medical Factors
5 Projects

Gathering Patient-Reported Data
4 Projects

Facilitating Research for Medicaid Beneficiaries
8 Projects

Strengthening Data Infrastructure to Address Non-Medical Factors

Non-medical factors, such as the social, natural, and built environment where people live, can influence individual health and contribute to the risk of chronic disease and other poor health outcomes.[29] As research continues to surface evidence on the importance of these non-medical factors to health, researchers and decision makers require actionable insights on how to address these factors to improve health outcomes through effective treatments and programs.[30] Incorporating data from outside the health system can enhance the person-centeredness of the evidence by broadening the focus of health outcomes research to the whole individual, rather than focusing on clinical encounters alone. This approach helps ensure evidence incorporates the varied experiences of different individuals and can be used to tailor recommendations for treatments or interventions to the individual's circumstances or needs.

OS-PCORTF Project Efforts in Data Infrastructure to Address Non-Medical Factors

Five projects are improving the inclusion and use of data on relevant non-medical and environmental data to enable actionable, evidence-based insights to improve outcomes meaningful to patients, caregivers, and their families. Projects are **expanding the availability and utility of data on environmental exposures and other factors** known to influence health by collecting new data and linking state and federal administrative databases, survey data, and clinical data for research on childhood obesity, chronic health conditions, and morbidity and mortality. These projects provide **robust data to answer important questions** about how geographic location, environmental factors, socioeconomic status (e.g., income level), and health care coverage affect health care utilization and outcomes. The data tools and resources produced by the OS-PCORTF also support inquiries into whether and how programs or treatments improve health outcomes, and which interventions are most helpful to patients based on their own unique circumstances.

One project developed a FHIR-enabled app for patients, caregivers, and clinicians to support integration of non-medical factors into patients' care plans. The app presents patient health data and goals in a single place and documents actionable steps and information for achieving health and wellness goals and improving health outcomes.

Gathering Patient-Reported Data

A key tenet of person-centered research is that individuals are seen as more than just patients, and that decisions about their care and treatment reflect their preferences, goals, and values.[31] By collecting data directly from patients, researchers can incorporate their perspectives into PCOR studies and ensure that health outcomes research reflects the outcomes most important to patients, which can differ from those traditionally studied in a clinical setting.[32]

OS-PCORTF Project Efforts in Gathering Patient-Reported Data

Four projects are advancing the ability to collect information directly from patients for research by **improving the quality and accessibility of important patient-reported data.** For example, projects developed and tested a patient-facing, FHIR-based application that collects patient-generated health data for an electronic care plan (see *New PCOR Resource for Gathering Patient-Reported Data* box). Another project is creating and fielding a supplemental questionnaire to a national survey to collect information from patients and caregivers on the economic burden and impact of seeking and receiving health care. Together, these projects improve the infrastructure for collecting patient-reported data and enrich analyses by making new data available on outcomes that are important to patients and their families.

New PCOR Resource for Gathering Patient-Reported Data

Project: ***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***

EXPLORE PROJECT

Building on prior OS-PCORTF work, AHRQ and NIH project teams developed MyCarePlanner, a patient-facing version of the eCare Plan app. The app collects relevant data about five different chronic diseases (chronic kidney disease, type 2 diabetes, common cardiovascular disease, chronic pain, and long COVID) directly from patients, integrates it with EHR data, and shares this information with patients' clinicians across health care settings via the clinician-facing version of the app, eCarePlanner. The personalized patient-facing app simplifies the process for sharing health updates, improves concordance of patient and clinician perspectives and decision-making, and facilitates future health outcomes research. Open-source code is available for health system stakeholders to implement the apps in their own institutions.

Other OS-PCORTF Initiatives in Gathering Patient-Reported Data

Engaging Patients to Enhance Data for Health Outcomes Research[33]

This brief summarizes project activities and accomplishments of six completed OS-PCORTF projects that integrate patient input and improve the collection of data from patients about their care. These projects ensure evidence generation supports research that reflects what matters most to patients, promotes the use of patient-reported outcomes, and empowers patients to more actively engage in health-related decision-making about their care. The brief also offers recommendations for future directions to improve patient engagement in research.

Facilitating Research for Medicaid Beneficiaries

To ensure evidence is relevant and representative of all populations, it is essential to study outcomes among individuals who face significant barriers to accessing health care. Medicaid provides health care coverage for 25% of all people in the U.S. and 41% of live births.[34] As the single largest source of coverage in the U.S.,[35] it is important to have a robust data infrastructure for Medicaid beneficiaries that can provide broad insights on the impact of Medicaid coverage for patient-centered outcomes.

OS-PCORTF Project Efforts in Research for Medicaid Beneficiaries

Eight projects are **creating more comprehensive data for researchers to study the factors that contribute to health care utilization and health outcomes of Medicaid beneficiaries** by linking Medicaid data to relevant clinical, survey, and administrative data sources. For example, projects are linking Medicaid claims data (from either T-MSIS or directly from state Medicaid agencies) to EHR data, the National Core Indicators-In Person Survey, the NAMCS, and the National Survey of Children’s Health Longitudinal Cohort. The resulting datasets will enable researchers to study the outcomes and effects of policies and programs designed to increase access to and improve quality of health care services for Medicaid beneficiaries. These linked datasets may also support studies on the effects of interventions and programs on person-centered outcomes, such as effects of home- and community-based services use, as well as the impact of family- and patient-level factors on COVID-19 treatment success. OS-PCORTF projects are also **improving T-MSIS data quality and accessibility** by harmonizing T-MSIS data into standard formats using CDMs and by creating standard linkage processes to enable comparison of outcomes across different years and states. These efforts **support more efficient analyses of Medicaid data** by creating replicable processes that researchers can leverage for their own work to streamline data cleaning and linkage, and by creating CDM formatted T-MSIS files that can be more easily linked to other data sources formatted to the same CDM.

Other OS-PCORTF Initiatives in Facilitating Research for Medicaid Beneficiaries

Creating Effective Medicaid Data Linkages for Health Outcomes Research

This report qualitatively assesses 11 OS-PCORTF-funded Medicaid data linkage projects to understand Medicaid data linkage approaches and share data linkage challenges and best practices. The report also describes the impact of the linked datasets and identifies opportunities to strengthen Medicaid data linkage infrastructure.

Conclusion

Over the past year, the OS-PCORTF program has worked collaboratively across federal agencies and programs to enhance data capacity for patient-centered outcomes research and CER. This report highlights progress made by the OS-PCORTF in Fiscal Year 2025 towards improving health outcomes for all Americans. Ongoing efforts to promote the adoption of OS-PCORTF resources—and to apply them to emerging health challenges—underscore both the enduring value and long-term sustainability of the data infrastructure developed through this initiative. Looking ahead, OS-PCORTF projects will continue to build on the momentum established within the research and industry communities, driving innovation toward meaningful and positive health impacts.

Acknowledgements

The authors would like to acknowledge the OS-PCORTF Project Leads and the ASPE staff in the Division of Health Care Quality and Outcomes who provided feedback on this report.

References

[1] Further Consolidated Appropriates Act, 2020, Publ. L. No. 116-94, 133 Stat. 2534 (2019). **https://www.congress.gov/116/plaws/publ94/PLAW-116publ94.pdf**

[2] HL7 Gravity Project. Introducing the Gravity Project. Accessed December 24, 2025. **https://thegravityproject.net/**

[3] Assistant Secretary for Technology Policy. USCDI+. HealthIT.gov. Published December 2023. **https://www.healthit.gov/topic/interoperability/uscdi-plus**

[4] Federal Committee on Statistical Methodology. *A Framework for Data Quality: FCSM-20-04*. Federal Committee on Statistical Methodology, National Center of Education Statistics. September 2020. **https://nces.ed.gov/fcsm/pdf/FCSM.20.04_A_Framework_for_Data_Quality.pdf**

[5] The White House. Restoring Gold Standard Science. Executive Order. May 23, 2025. Retrieved from: **https://www.whitehouse.gov/presidential-actions/2025/05/restoring-gold-standard-science/**

[6] Leading causes of death. Centers for Disease Control and Prevention. Updated January 23, 2023. **https://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm**

[7] National health expenditure data: historical. Center for Medicare & Medicaid Services. Updated December 18, 2024. **https://www.cms.gov/data-research/statistics-trends-and-reports/national-health-expenditure-data/historical**

[8] The White House. Make Our Children Healthy Again. Accessed September 12, 2025. **https://www.whitehouse.gov/maha/**

[9] Carney TJ, Wiltz JL, Davis K, Briss PA, Hacker K. Advancing Chronic Disease Practice Through the CDC Data Modernization Initiative. Prev Chronic Dis. 2023 Nov 30;20:E110. doi: 10.5888/pcd20.230120. **https://www.cdc.gov/pcd/issues/2023/23_0120.htm**

[10] Srinivasan M, Shah A, Peterson C, Aronoff A, Eloso J, and Dullabh P. Exploring Data Infrastructure Availability and Expansion Opportunities for Health Outcomes Research on Sickle Cell Disease. Washington, D.C.: Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services. November 2025.

[11] Centers for Disease Control and Prevention. Data and statistics on sickle cell disease. May 15, 2024. Accessed May 2, 2025. Retrieved from: **https://www.cdc.gov/sickle-cell/data/index.html**

[12] Jiao B, Johnson KM, Ramsey SD, Bender MA, Devine B, Basu A. Long-term survival with sickle cell disease: a nationwide cohort study of Medicare and Medicaid beneficiaries. *Blood Adv*. 2023;7(13):3276-3283. doi:10.1182/bloodadvances.2022009202

[13] National Heart, Lung, and Blood Institute. Sickle cell disease treatment. September 30, 2024. Accessed May 2, 2025. Retrieved from: **https://www.nhlbi.nih.gov/health/sickle-cell-disease/treatment**

[14] Desai P, George A, Ahmed F, et al. Access to gene therapy in sickle cell disease. *Blood*. 2024;144(1):7674. doi:10.1182/blood-2024-209630

[15] Smeltzer MP, Reeves SL, Cooper WO, et al. Common data model for sickle cell disease surveillance: considerations and implications. *JAMIA Open*. 2023;6(2):ooad036. doi:10.1093/jamiaopen/ooad036

[16] Stoneburner A, Lucas R, Fontenot J, Brigance C, Jones E, DeMaria AL. Nowhere to Go: Maternity Care Deserts Across the US. (Report No 4). March of Dimes. 2024. **https://www.marchofdimes.org/sites/default/files/2024-09/2024_MoD_MCD_Report.pdf**

[17] Further Consolidated Appropriates Act, 2020, Publ. L. No. 116-94, 133 Stat. 2534 (2019). **https://www.congress.gov/116/plaws/publ94/PLAW-116publ94.pdf**

[18] Srinivasan M, Fernandez J, Leaphart L, Ryan S, Byrne M, Shen K, Wiltz, J, and Dullabh P. Data Sources for Conducting Research on Doula Services and Related Outcomes. Washington, D.C.: Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services. October 2025.

[19] Knocke K, Chappel A, Sugar S, De Lew N, Sommers BD. Doula Care and Maternal Health: An Evidence Review. (Issue Brief No. HP-2022-24). Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services. December 2022.

[20] Kozhimannil KB, Attanasio LB, Jou J, Joarnt LK, Johnson PJ, Gjerdingen DK. Potential benefits of increased access to doula support during childbirth. *Am J Manag Care*. 2014;20(8):e340-e352. Published 2014 Aug 1.

[21] Anderson LL, Humphries K, McDermott S, Marks B, Sisirak J, Larson S. The State of the Science of Health and Wellness for Adults with Intellectual and Developmental Disabilities. Intellectual and Developmental Disabilities. 2013;51(5):385-398. doi:10.1352/1934-9556-51.5.385

[22] Pham HH, Benevides TW, Andresen M, et al. Advancing Health Policy and Outcomes for People With Intellectual or Developmental Disabilities: A Community-Led Agenda. JAMA Health Forum. 2024;5(8):e242201. doi:10.1001/jamahealthforum.2024.2201

[23] Jiménez F, Gauthreaux N, Segal C, Srinivasan M, Dullabh P, Steverman S, Dey J, Karimi M, and Creedon T. Assessing the Feasibility of Creating a National Behavioral Health Workforce Database. Washington, D.C.: Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services. November 2024.

[24] NORC at the University of Chicago. Challenges and Improvements for PCOR Data Infrastructure: Results from a Stakeholder Prioritization Activity. Office of the Assistant Secretary for Planning and Evaluation, U.S. Dept of Health and Human Services; November 10, 2020. **https://aspe.hhs.gov/sites/default/files/private/pdf/259016/PCOR-Data-Infrastructure.pdf**

[25] U.S. Department of Health and Human Services. Implementing Gold Standard Science: Report to the White House Office of Sciene and Technology Policy. August 22, 2025. **https://www.hhs.gov/sites/default/files/hhs-gold-standard-science-report-2025.pdf**

[26] Chen IY, Pierson E, Rose S, Joshi S, Ferryman K, Ghassemi M. Ethical Machine Learning in Healthcare. Annu Rev Biomed Data Sci. 2021;4:123-144. doi:10.1146/annurev-biodatasci-092820-114757

[27] Jarmakovica A. Machine learning-based strategies for improving healthcare data quality: an evaluation of accuracy, completeness, and reusability. *Front Artif Intell*. 2025;8:1621514. Published 2025 Jul 21. doi:10.3389/frai.2025.1621514

[28] Finetti R, Roncaglia B, Visibelli A, Spiga O, Santucci A. Data Augmentation and Synthetic Data Generation in Rare Disease Research: A Scoping Review. *Med Sci (Basel)*. 2025;13(4):260. Published 2025 Nov 6. doi:10.3390/medsci13040260

[29] Benavidez GA, Zahnd WE, Hung P, Eberth JM. Chronic Disease Prevalence in the US: Sociodemographic and Geographic Variations by Zip Code Tabulation Area. Prev Chronic Dis 2024;21:230267. doi: http://dx.doi.org/10.5888/pcd21.230267

[30] National Academies of Sciences, Engineering, and Medicine; Division on Engineering and Physical Sciences; Health and Medicine Division; Division of Behavioral and Social Sciences and Education; Computer Science and Telecommunications Board; Board on Health Care Services; Committee on National Statistics; Committee on Building Data Capacity for Patient-Centered Outcomes Research: An Agenda for 2021 to 2030. Building Data Capacity for Patient-Centered Outcomes Research: Priorities for the Next Decade. Washington (DC): National Academies Press (US); 2022 May 18. 3, Priority Areas for the PCOR Data Infrastructure. Available from: **https://www.ncbi.nlm.nih.gov/books/NBK584415/**

[31] Hsiao B, Fraenkel L. Incorporating the patient's perspective in outcomes research. Curr Opin Rheumatol. 2017 Mar;29(2):144-149. doi: 10.1097/BOR.0000000000000372.

[32] Snyder CF, Jensen RE, Segal JB, Wu AW. Patient-reported outcomes (PROs): putting the patient perspective in patient-centered outcomes research. Med Care. 2013;51(8 Suppl 3):S73-S79. doi:10.1097/MLR.0b013e31829b1d84

[33] Office of the Assistant Secretary for Planning and Evaluation. Engaging Patients to Enhance Data for Health Outcomes Research. July 28, 2025. **https://aspe.hhs.gov/reports/patient-engagement-vignette**

[34] Kaiser Family Foundation. 10 Things to Know About Medicaid. February 18, 2025. Accessed September 12, 2025. **https://www.kff.org/medicaid/10-things-to-know-about-medicaid/**

[35] Centers for Medicare and Medicaid Services. "October 2024, Medicaid and CHIP Eligibility Operations and Enrollment Snapshot" Available at **https://www.medicaid.gov/resources-for-states/downloads/eligib-oper-and-enrol-snap-october2024.pdf**.