Appendix A: Key Impacts of OS-PCORTF Projects Completed in Fiscal Year 2025

This appendix highlights key impacts of the Office of the Secretary Patient-Centered Outcomes Research Trust Fund (OS-PCORTF) projects that ended in Fiscal Year 2025.

The completed Fiscal Year 2025 OS-PCORTF projects made important contributions to improving data infrastructure for patient-centered outcomes research across four domains: 1) improving data quality; 2) providing more relevant, comprehensive data; 3) improving data access and use; and 4) enhancing analytic resources.

Improving Data Quality

Improving Data Access and Use

Providing More Relevant, Comprehensive Data

Enhancing Analytic Resources

Completed Project Resources

The completed Fiscal Year 2025 projects produced 34 data and technical resources. Of these resources, **76 percent** are open-source or publicly available to researchers, industry, and other PCOR stakeholders. The following two icons are used throughout this appendix to indicate the accessibility of the key project resources.

Icon indicates key resources that are open access/public

Icon indicates key resources that are restricted access

Ensuring Data Infrastructure Resources Meet End User Needs

73 percent of completed Fiscal Year 2025 projects engaged end users in the development and testing of resources.

Engaging end users can ensure their needs, values, and perspectives are adequately reflected in data infrastructure improvement efforts. End users participated in comprehensive user testing of data products, provided perspectives on research needs and appropriate use cases, and were engaged in the human-centered design activities for patient-facing applications. To improve the utility of OS-PCORTF data products, end users contributed to the development of linked datasets, tested implementation guides at Health Level Seven (HL7®) Fast Healthcare Interoperability Resources (FHIR®) Connectathons, and participated in cognitive testing of data collection tools. End users also engaged in technical expert panels, working groups, steering committees, and interviews.

Fiscal Year 2025 Completed Project Snapshots

Building Infrastructure and Evidence for COVID-19 Related Research, Using Integrated Data from National Center for Health Statistics (NCHS) Data Linkage Program

This project expanded the National Center for Health Statistics (NCHS) data infrastructure by improving access to datasets that link NCHS survey data with other federal administrative data sources. The project leveraged linkage methodologies used in other projects within the NCHS Data Linkage Program,[1] including those employed in the OS-PCORTF project, *Data Linkage: Evaluating Privacy Preserving Record Linkage Methodology and Augmenting the National Hospital Care Survey with Medicaid Administrative Records*. The current project developed two publicly accessible synthetic data files linking data from 1) the National Health Interview Survey (NHIS), Centers for Medicare and Medicaid Services (CMS), the U.S. Department of Housing and Urban Development (HUD), and 2) the National Hospital Care Survey (NHCS) and the National Death Index (NDI). The linked NHIS-HUD-CMS dataset is now available, and the NHCS-NDI linked dataset will be available in 2026. These synthetic linked data allow researchers to study outcomes related to risk factors for chronic diseases, other factors that impact health outcomes, and health care utilization, all while safeguarding patient privacy.

Challenge Addressed

To protect the confidentiality of NCHS survey participants, access to NCHS linked datasets is restricted to secure research centers, limiting their usability.[2,3] Additionally, stakeholders have expressed a strong preference for public use data. This project responded by developing publicly available synthetic datasets that are derived from linked NHIS, CMS, HUD, NHCS, and NDI data. The synthetic datasets maintain some statistical properties of the original data while preserving survey participant confidentiality.

Project Impact on Expanding Data Capacity for Patient-Centered Outcomes Research

The project team developed a novel verification process that allows users to compare statistical results using synthetic data against restricted use data. This approach helps researchers assess the reliability of synthetic data for certain analyses and determine when access to original data may be necessary.

The project developed a scalable and adaptable methodology that preserves statistical reliability while protecting privacy. The methodology generated publicly available synthetic linked data files that protect the privacy of survey and program participants and retain some of the statistical properties of in the underlying source data.

The project team expanded access to previously restricted NCHS data by creating synthetic datasets. The synthetic linked data files can be accessed on the NCHS data linkage website and used for modeling and statistical inference for selected analyses.

Key Resources

- **Synthetic Linked NHIS-HUD-CMS Dataset**. This synthetic linked file includes selected variables from the 2018 NHIS with selected linked data from Medicare health insurance records and HUD federal housing assistance programs.
- **Synthetic Linked NHIS-HUD-CMS Data Dictionary**. This data dictionary describes the 51 data elements included in the linked NHIS-HUD-CMS dataset.
- **Synthetic Linked NHIS-HUD-CMS Dataset User Documentation.** This technical documentation describes the methodology for developing the synthetic linked dataset.
- **NHCS-NDI Synthetic Linked Dataset, Data Dictionary, and User Documentation.** This synthetic linked file includes selected variables from the 2016 NHCS with selected linked data from the NDI. The dataset along with data dictionary and user documentation will be available in 2026.
- **Project Final Report.** This final report describes the project's key accomplishments, outcomes, and challenges they addressed. The report will be available in 2026.

Long-Term Impacts on Interagency Partnerships and Sustainability

This project laid the groundwork for broader adoption of synthetic data methods across the U.S. Department of Health and Human Services (HHS) by developing NCHS's first fully synthetic linked datasets. The development and verification approach can guide future HHS standards for synthetic data generation, enabling scalable and secure access to integrated health and social datasets that support patient-centered research. Early engagement with stakeholders ensured the datasets reflected high-value elements, increasing their usefulness to researchers. Strong initial uptake—over 800 webpage views within two months of launch—demonstrated clear demand for accessible, privacy-protected data. In recognition of these contributions, the project team received an NCHS Director's Merit Award in 2024 for their achievements and significant scientific advancement that position NCHS at the leading edge of synthetic data dissemination across the federal statistical system.

Implementing Agency
CDC/NCHS

Project Profile
ASPE Website Project Profile

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

To facilitate the assessment of survival and mortality in PCOR studies, this project developed a public report and database of data access and use policies for linkable mortality data from each of the 57 mortality-reporting jurisdictions. This database makes it easier, less expensive, and more efficient for researchers to determine the best ways to obtain the survival and mortality data needed for their studies. The mortality database can assist researchers in assembling more comprehensive, multijurisdictional samples by providing information on which states and territories have similar requirements to collect, utilize, and link mortality data. It will also allow researchers to group jurisdictions with similar data use characteristics to assemble larger samples, increasing efficiency and utility of the derived data.

Challenge Addressed

Access to mortality data is essential for exploring questions related to survival and risk factors for death.[4] Currently, linkable mortality data in the U.S. are owned by 57 separate jurisdictions with varying access and use policies.[5] There is no single publicly available source of information on the respective jurisdictions' requirements for accessing and using these data, and individual jurisdictions are not standardized in how they make this information available to data users. This information gap limits researchers' ability to assess survival, as they need to spend considerable time and resources to find and interpret mortality data access and use policies for each jurisdiction.

Project Impact on Expanding Data Capacity for Patient-Centered Outcomes Research

To facilitate researchers' access to information on mortality data for patient-centered outcomes research, this project conducted a systematic search and review of policies pertaining to the access and use of linkable mortality data. The project developed a publicly available report and database that describes and compares access policies along six dimensions: 1) the available data; 2) information about the application and review process; 3) legal, ethical, and privacy review requirements; 4) permitted and disallowed uses of the data; 5) data protection requirements; and 6) data linkage policies across the 57 mortality-reporting jurisdictions.

Key Resources

Mortality Data Linkages for Research Report. This publicly available final project report summarizes the process for developing a database of mortality policies and describes findings from the 57 mortality-reporting jurisdictions across the six policy dimensions. It highlights best practices for mortality data access policies across jurisdictions and shares lessons learned from compiling the database.

Compendium of Policies on Use of Linkable Mortality Data. This electronic database includes filterable information on access and use policies for mortality data from the 57 mortality-reporting jurisdictions across the six policy dimensions. The final policy database includes 137 statutes, 80 regulations, and 190 website URLs. The database will be made available in 2026.

Long-Term Impacts on Interagency Partnerships and Sustainability

This project's public report highlights jurisdictions with exemplary data access and use policies that can inform improvements to mortality data accessibility. Other states and jurisdictions can learn from and consider emulating the exemplar policies highlighted in the report to facilitate research using linked mortality data. Decision-makers can consider leveraging this report to consider how to make jurisdiction-level mortality data more available for researchers. National systems for accessing linkable mortality data for health outcomes research must reflect the range of jurisdictional policies that are represented. Therefore, by describing and comparing jurisdictions' data use policies in one place and highlighting best practices, the project's report can directly inform efforts to improve the usability of national mortality data systems.

Implementing Agency
ASPE/BHDAP

Project Profile
ASPE Website Project Profile

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

The National COVID Cohort Collaborative (N3C), led by the National Institutes of Health (NIH)/National Center for Advancing Translational Sciences (NCATS), is the largest U.S. platform of real-world electronic health record (EHR) data created to more efficiently study emergent health threats. This project linked CMS claims data with EHR data within the N3C's secure, privacy-preserving cloud enclave. To facilitate this linkage, the project team used resources from the previously funded OS-PCORTF project, ***Harmonization of Various Common Data Models and Open Standards for Evidence Generation*** to harmonize CMS data to the Observational Medical Outcomes Partnership (OMOP) common data model (CDM). The linked data provide a more complete view of patients' health journeys, moving beyond fragmented hospital encounters to reveal how health system interventions impact individual outcomes, as well as improve the timeliness and specificity of data used to identify effective treatment options. The longitudinal dataset captures over 10 million people and supports rigorous analyses comparing COVID-19 treatment effectiveness for different populations, taking patients' health history, characteristics, and service utilization into account.

Challenge Addressed

During the COVID-19 pandemic, researchers encountered significant barriers to conducting health outcomes research due to delays in accessing timely, real-world data from fragmented sources.[6] This challenge underscored the urgent need for a centralized, near real-time longitudinal research dataset to support evidence generation and inform decision-making.

Project Impact on Expanding Data Capacity for Patient-Centered Outcomes Research

The N3C team developed tools to harmonize and integrate different data sources using the OMOP CDM. This integration enables researchers to cross-validate key data elements across sources, improving the reliability of analyses. Additionally, it reduces the effort required for researchers to standardize and link these datasets.

The team created a dataset that combines longitudinal CMS claims data with clinical EHR data from 84 health systems. This linkage resulted in three data files on patients' clinical care history and outcomes (e.g., therapeutics, lab test and imaging results, diagnoses, health care utilization, and mortality) to help researchers answer complex questions related to disease progression and treatment outcomes.

The N3C team made the linked CMS-N3C dataset available in the centralized N3C data enclave. This centralized environment provides researchers with a single, standardized data asset and a common suite of analytic tools, reducing administrative burden by requiring only one data use request for access.

The N3C team used a privacy-preserving record linkage (PPRL) method to conduct data linkages between N3C data and CMS claims. This method uses tokenization and hashing to match and link patient records without revealing direct patient identifiers.

Key Resources

Linked CMS-N3C Dataset. The dataset contains billing data from CMS, formatted into the OMOP CDM for N3C researcher use. Researchers can request access to three file types (limited dataset, de-identified dataset, and synthetic dataset). Access can be requested through the N3C Data Enclave for approval by the Data Access Committee.

N3C Report on PPRL Methodologies. This methodology report describes the PPRL technique used to link N3C and CMS data.

N3C PPRL Data Governance. This document, intended for N3C data contributors, describes the PPRL process used in N3C, including linkage and deduplication, and how data will be used.

N3C PPRL Enrichment Dashboard. This 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 Final Report. This final report describes the project's key accomplishments, outcomes, and challenges they addressed.

Long-Term Impacts on Interagency Partnerships and Sustainability

The N3C-CMS linked dataset is being leveraged for additional research initiatives. The linked N3C-CMS dataset is being actively used by over 40 research teams for training, data harmonization, governance, and collaborative analyses, with 42 distinct projects led by federal and non-federal investigators focused on evaluating outcomes of various health interventions. These projects span high-priority topic areas such as multiple chronic diseases, pain management, and chronic disease prevention, among others. Building on this momentum, the OS-PCORTF is supporting NCATS in new studies focused on cancer and renal disorders.

Implementing Agency
NIH

Project Profile
ASPE Website Project Profile

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

Understanding COVID-19 Trajectory and Outcomes in the Context of Multiple Chronic Conditions through eCare Plan Development

The OS-PCORTF supported two complementary projects that significantly broadened the scope of interoperable electronic care (eCare) planning tools. The first 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***, developed a patient-facing eCare Plan app and associated FHIR implementation guide. The project implemented and evaluated the tools in multiple clinical settings among individuals with multiple chronic conditions (MCC). The second project, ***Understanding COVID-19 Trajectory and Outcomes in the Context of Multiple Chronic Conditions through eCare Plan Development***, built upon the clinician- and patient-facing apps by expanding functionality to include additional chronic disease and COVID-19-related data elements, as well as incorporating a caregiver-facing interface in the patient-facing app. Both projects leveraged the architecture established in the original eCare Plan 1.0 initiative, ***Data Capacity for Patient-Centered Outcomes Research through Creation of an Electronic Care Plan for People with Multiple Chronic Conditions***, which developed a clinician-facing app and implementation guide to enhance the interoperability of care planning resources for individuals with MCC.

Challenge Addressed

Researchers are increasingly seeking approaches to conduct chronic disease research that reflects real-world care. However, limited interoperability across EHRs creates barriers to accessing complete, consistent data, especially for individuals with MCC whose care can span multiple settings and providers.[7] To address this challenge, the Agency of Healthcare Research and Quality (AHRQ) and the NIH/National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) launched joint initiatives to enhance interoperable eCare plan tools. These tools enable the aggregation and exchange of patient data across home, community, clinical, and research environments, supporting more effective research on longitudinal health outcomes of individuals with chronic diseases.

Project Impact on Expanding Data Capacity for Patient-Centered Outcomes Research

The project teams developed consistently defined variables to standardize data collection and analysis. By implementing uniform data elements and value sets, they ensure consistent data capture across different patient populations and health systems, thereby strengthening data validity.

The project teams developed value sets to easily aggregate data across different settings. These value sets facilitate standardized data exchange among environments that operate on disparate systems—including home, community, clinical, and research contexts—ultimately supporting the integration of data for a more comprehensive eCare Plan.

The project teams developed, tested, and published FHIR-based eCare Plan applications for patients, clinicians, and caregivers, with an accompanying implementation guide. These open-source tools integrate with different EHR systems and include chronic disease and COVID-19-specific data elements, enabling users to view, contribute, and share comprehensive care plan information across diverse care settings. Patient testing enhanced application usability, supporting broader real-world adoption. The implementation guide underwent formal HL7 balloting, reinforcing credibility and alignment with industry standards.

Key Resources

- **Patient-Facing/Caregiver-Facing eCare Plan Application.** This app source code integrates with the clinician-facing version to support caregivers and individuals with MCC, including chronic kidney disease (CKD), diabetes, cardiovascular disease (CVD), chronic pain, and Long COVID.
- **Expanded Final eCare Plan Application HL7 FHIR Implementation Guide.** This implementation guide specifies the use of data standards and FHIR resources to support interoperable exchange of eCare plan data for CKD, type 2 diabetes, CVD, chronic pain, and Long COVID, as well as the coded content needed to implement the patient-facing and clinician-facing eCare plan applications.
- **Expanded Data Elements and Value Sets.** This resource contains over 1,100 data elements related to chronic diseases, including CKD, diabetes, CVD, chronic pain, and Long COVID, paired with value sets that utilize clinical terminologies (e.g., Systematized Nomenclature of Medicine—Clinical Terminology [SNOMED-CT], Logical Observation Identifiers Names and Codes [LOINC]) to support standardized implementation.
- **Expanded Clinician-Facing eCare Plan Application.** This freely available and open-source, EHR-agnostic, Substitutable Medical Applications, Reusable Technologies (SMART) on FHIR application can be used by clinicians to care for MCC populations.
- **Project Final Report**. This final report describes the development, implementation, and evaluation of the patient-facing/caregiver-facing and clinician-facing eCare plan applications with the expanded data elements and value sets to observe chronic disease impacts.

Long-Term Impacts on Interagency Partnerships and Sustainability

The interoperable care planning tools developed by the eCare plan projects are being leveraged across federal and industry initiatives. The FHIR implementation guide, now published as "Standard for Trial Use," following successful balloting and reconciliation, is positioned for broader implementation across health care systems. Ongoing collaboration with the Assistant Secretary for Technology Policy/Office of the National Coordinator for Health Information Technology (ASTP/ONC) to include eCare plan elements in the United States Core Data for Interoperability (USCDI) and leverage them in health information exchanges (HIEs) promotes the adoption of consistent EHR data across health care systems, ultimately enhancing the quality and scalability of data for both care and research. The project's tools are already being leveraged in federally-funded research and workforce programs, including studies by the NIH/National Institute on Aging (NIA), the Health Services and Research Administration's (HRSA) Geriatric Workforce Enhancement Program, and a **2022 ASTP/ONC LEAP in Health IT award**. Additionally, partnerships with the National Committee for Quality Assurance (NCQA) and the Oregon Health & Science University (OHSU) are advancing goal attainment data standards and expanding use cases into behavioral health, demonstrating the projects' influence on different areas of patient-centered outcomes research.

Implementing Agency
AHRQ & NIH/NIDDK

Project Profiles
ASPE Website Project Profile
(Development of the Patient-facing Application)

ASPE Website Project Profile
(Understanding COVID-19 Trajectory and Outcomes and Caregiver-facing Component)

Data Infrastructure Supporting Research on Refugee Medicaid Service Utilization and Outcomes

The Refugee Arrivals Dataset (RADS) contains information on refugees and asylees who access services provided through the Administration for Children and Families (ACF) Office of Refugee Resettlement (ORR) in their first few years of entry to the U.S. This project explored the feasibility of linking RAD data to Transformed Medicaid Statistical Information System (T-MSIS) Analytic Files (TAF) data from 2016-2022, which includes information on Medicaid beneficiaries' health care utilization. The project planned to link these data to understand Medicaid service utilization among people receiving asylum and refugee services and identify how participation in ORR programs impacts access to Medicaid benefits and outcomes, with a focus on refugees' long-term health trajectories and interventions to increase access to preventive care.

Challenge Addressed

Refugees arriving to the U.S. can receive various education and employment benefits and services through the ACF ORR refugee resettlement program.[8] To ensure that ORR programs meet the needs of the people they serve, decision-makers need comprehensive data that provide insight into 1) how programs support refugees' health-related needs and 2) where there are gaps. However, there is limited health-related data available on refugee populations, including data on health outcomes, access to and utilization of health services, and environmental and non-medical factors.[9] Additionally, given the lack of unique identifiers shared between RADS and T-MSIS, it can be labor-intensive and time-consuming for researchers to use the two datasets together to understand the health care utilization and outcomes of people who participate in ORR's refugee programs.

Project Impact on Expanding Data Capacity for Patient-Centered Outcomes Research

To provide more relevant, comprehensive data, the project assessed the feasibility of linking RAD and T-MSIS data, determining the linkage is feasible for T-MSIS data from 16 states. The feasibility assessment found linkages are less feasible in the remaining states and for data prior to 2019 due to variations in T-MSIS data quality for the variables needed to perform the linkage. These findings can support future efforts to link RADS and T-MSIS data.

Key Resources

Refugee and Asylee Data on the Utilization of Medicaid. This brief describes the project's objectives and value of linking data from agencies managing refugees and Medicaid services.

Project Final Report. This final report describes the project's key accomplishments, outcomes, and challenges addressed. It highlights lessons learned from the project activities that are relevant for other administrative data linkage efforts, in particular linkages using T-MSIS. This report will be released in 2026.

This project was sunset prior to producing the linked dataset, but valuable key lessons learned from the project's linkage feasibility assessment are noted below.

Lessons Learned

Through the RADS-T-MSIS linkage feasibility assessment, the project team concluded that a probabilistic, rather than a deterministic, linkage approach would result in a more accurate and complete dataset. This linkage approach can inform other efforts that link data lacking social security numbers by demonstrating the strengths and limitations of such a probabilistic linkage methodology. Additionally, other researchers can benefit from the project's T-MSIS data quality analyses, which found low missingness for key variables that could be used as individual identifiers to support linkage. In addition to lessons learned related to the feasibility of the RAD-T-MSIS linkage, ACF surfaced several considerations that could benefit others seeking to work with CMS data. The project team had started navigating the data use agreement (DUA) process with CMS, revealing opportunities to improve and streamline processes for obtaining data across federal institutions. For example, projects seeking to upload non-CMS data to CMS's Virtual Research Data Center (VRDC)—to perform linkages to CMS data—do not need to involve CMS in DUA approval processes for obtaining access to data originating or managed outside of CMS, although CMS approval is needed for use of the VRDC.

Implementing Agency
ACF

Project Profile
ASPE Website Project Profile

Engaging People with Intellectual and Developmental Disabilities to Enhance Functional Disability Representation in Point of Care Settings

The aim of this project was to develop a functional disability demographic checklist for use at the point of care. The proposed checklist would facilitate representation of people with intellectual and development disabilities (ID/DD) in administrative data systems and allow for collection of ID/DD data that were not previously available. Access to these data would facilitate researchers' ability to conduct research on health outcomes associated with different services and supports designed to improve the health and well-being of people with ID/DD and their quality of care. To develop the checklist, the project team planned to leverage already validated and widely accepted questions from existing ID/DD surveys (e.g., American Community Survey [ACS], Washington Group Short Set on Functioning). To ensure the patient-centeredness of the checklist, the project planned to conduct interviews and cognitive testing with individuals with ID/DD as part of checklist development.

Challenge Addressed

People with ID/DD experience differences in health care access and outcomes compared to people without ID/DD.[10] To understand the impact of differences in care experience, service utilization data are necessary for identifying and addressing differences in health outcomes faced by people living with ID/DD.[11] However, there are challenges to collecting these data consistently, particularly given the heterogeneity of diagnoses and the wide range of impacts that ID/DD have on daily living and functioning.[12] Existing surveys that do elicit functional disability status have not been adapted for clinical use, highlighting an opportunity for translating these tools for clinical settings.[13] Researchers need more standardized and accessible data to understand how different services and treatments affect outcomes. These findings can then be used to improve the quality of care for the ID/DD community.

Project Impact on Expanding Data Capacity for Patient-Centered Outcomes Research

To improve the capture of data relevant to the ID/DD population, the project drafted supplemental questions across a range of disability domains based on existing survey questions and emergent areas identified by people with ID/DD. The project team conducted qualitative interviews with people living with ID/DD to gather suggestions for possible adaptations to the proposed questions for inclusion in the functional disability checklist. A methodological plan and interview guide was developed for cognitive testing of the draft checklist questions.

Key Resources

Project Final Report. This final report describes the project's key accomplishments, outcomes, and challenges addressed. This report will be released in 2026.

External complications prevented the project from producing the disability checklist and accompanying resources.

Implementing Agency
CDC/NCBDD

Project Profile
ASPE Website Project Profile

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

The National Ambulatory Medical Care Survey (NAMCS) collects visit data on outpatient medical services provided in clinics, physician offices, or health center settings, also called ambulatory care, and the patients who receive those services. NAMCS is the leading source of information on the care provided at Federally Qualified Health Centers (FQHCs). This project improved the representativeness of NAMCS data on maternal health care provided by FQHCs by expanding and modernizing NAMCS data collection. As part of the modernization efforts, the project supported an expansion of the transition from manual data collection to EHR data transmission for clinical visit data. Starting in 2021, health centers that participate in NAMCS submit EHR data, which are processed and made available to researchers at the visit level. Additionally, the project linked the expanded NAMCS data to the NDI and administrative data from HUD to provide additional information on post-FQHC visit mortality and understand associations between housing needs and health outcomes. The project leveraged existing data linkage algorithms from a Fiscal Year 2017 OS-PCORTF project, ***Enhancing Data Resources for Researching Patterns of Mortality in Patient Centered Outcomes Research*** and a Fiscal Year 2019 OS-PCORTF project, ***Augmenting the National Hospital Care Survey (NHCS) Data through Linkages with Administrative Records***. These prior algorithms informed the development of the NAMCS data linkage algorithm used in the current project. The Centers for Disease Control and Prevention (CDC) team evaluated the quality of these linkages and assessed the usability of the linked NAMCS datasets for generating accurate statistical estimates on maternal health visits at FQHCs.

Challenge Addressed

FQHCs play a key role in providing maternal health care to women in the U.S. NAMCS is the only source of nationally representative visit-level data on ambulatory health services, including maternal health care provided at FQHCs. Prior to 2021, data collection procedures for the NAMCS Health Center Component relied on in-person, manual abstraction of patient records from participating FQHCs.[14] These outdated data collection processes were time-consuming and burdensome for health centers and introduced potential for human error and data inaccuracy in the final data. The new processes both modernize and expand NAMCS data collection procedures by using an HL7 implementation guide to facilitate FQHC's automated and standardized EHR data submission, expanding the time frame of data collected at each interval, and collecting data from a larger number of centers.

Project Impact on Expanding Data Capacity for Patient-Centered Outcomes Research

The CDC team improved the representativeness, completeness, consistency, and timeliness of NAMCS data through several updates to modernize the EHR data collection process. Among the most impactful updates, the project expanded the number of FQHC sites sampled from 50 to 104 in the primary sample for 2022. As a result of these efforts, the NAMCS is now the only nationally representative dataset of FQHC visits. The increased sample size and amount of data collected will allow for more robust and reliable estimates of maternal health visits at FQHCs with future rounds of NAMCS data collection.

The team improved the comprehensiveness of PCOR data available for maternal health care and outcomes by linking the expanded 2022 NAMCS data to other federal datasets on mortality outcomes and housing assistance program participation. The resulting datasets link NAMCS FQHC data to NDI and HUD for the first time ever, bringing together information on clinical visits, health care utilization, patient characteristics, and mortality outcomes to support more robust inquiries into the risk factors for poor maternal health outcomes.

The project enhanced the timeliness with which data are made available and developed publicly accessible data resources for researchers. The project made a de-identified version of the 2022 NAMCS data file publicly available for research use. By including a variable flagging whether a visit qualified as maternal care-related, the dataset supports inquiries into maternal health care received at FQHCs. A public dashboard facilitates the timely release of data by disseminating preliminary national estimates of data within six months of collection.

The project evaluated and adapted linkage algorithms developed by prior OS-PCORTF projects and assessed their feasibility for linking NAMCS FQHC data. From their evaluation, the CDC team determined that such data linkages were feasible. These findings advance the capabilities of CDC to perform future linkages with NAMCS FQHC data as they provided proof-of-concept that NAMCS data could be linked to other federal administrative datasets.

Key Resources

Expanded NAMCS Health Center Component 2022 Data File. The nationally representative dataset of visit-level patient data from FQHCs from the 2022 NAMCS contains data on over 5.6 million patient visits. Restricted use and public-use (personally identifiable information [PII] removed) versions of the dataset are available. To access the restricted use dataset, researchers need to apply through the NCHS Research Data Center (RDC).

NAMCS Health Center Component 2022 Restricted Use File Data Dictionary. The data dictionary defines variables included in the restricted use file and provides additional information to support end users working with the data.

Linked NAMCS 2021 and NDI Data File. Clinical EHR data from the 2021 NAMCS Health Center (HC) Component was linked to 2021-2022 NDI data. The restricted-used dataset is accessible through the NCHS RDC.

Linked NAMCS 2021 and HUD Data File. Clinical data from the 2021 NAMCS HC Component was linked to 2020-2022 HUD administrative data on federal housing assistance program participation. The restricted-used dataset is accessible through the NCHS RDC.

Preliminary Estimates of Visits to Health Centers in the United States Dashboard. This interactive dashboard disseminates preliminary bi-annual estimates of medical diagnoses and health center service utilization from NAMCS participants. Estimates can be stratified by patient characteristics.

Sampling Procedures for the Collection of Electronic Health Record Data From Federally Qualified Health Centers, 2021–2022 National Ambulatory Medical Care Survey. This methodology report provides details on the changes to the sampling methods and procedures for the 2022 NCHS-NAMCS HC component performed under the project. This report serves as the project’s final report.

The Linkage of the 2021 National Ambulatory Medical Care Survey (NAMCS) Health Center (HC) Component to 2021-2022 National Death Index: Linkage Methodology and Analytic Considerations. This methodology report describes the methods used to link NAMCS and NDI data and provides analytic guidance for researchers when using the linked dataset.

The Linkage of the 2021 National Ambulatory Medical Care Survey (NAMCS) Health Center (HC) Component to 2020-2022 U.S. Department of Housing and Urban Development Administrative Data: Linkage Methodology and Analytic Considerations. This methodology report describes the methods used to link NAMCS and HUD data and provides analytic guidance for researchers when using the linked dataset.

Long-Term Impacts on Interagency Partnerships and Sustainability

The new sampling approach established by the project is expected to be used in future years whenever the NAMCS HC sample is updated or replaced, enabling sustained data quality improvements of the NAMCS for future years. By establishing that linkages of NAMCS data to external federal data sources are feasible, the project lays the foundation for future NAMCS linkages. These future linkages can enhance the analytic capabilities of NAMCS EHR data for patient-centered outcomes research.

Implementing Agency
CDC

Project Profile
ASPE Website Project Profile

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

To support patient-centered outcomes research, health policy evaluation, and public health monitoring, the NCHS Data Linkage Program[1] has performed over 70 linkages of their population-based survey data to federal administrative data, including from the NDI, CMS claims, Social Security Administration (SSA), HUD and the Department of Veterans Affairs (VA).[15] Traditionally, these linkages, using a combination of probabilistic and deterministic techniques, have been performed using clear-text PII and have produced highly accurate results. To address concerns for data privacy and security and increase potential linkage opportunities, the NCHS has explored PPRL methods to link data without sharing PII.

To ensure that PPRL results are consistent with high-quality results generated using traditional linkage methods, this project evaluated the performance of an open-source PPRL tool, Anonlink. The CDC previously used AnonLink to link data in two other OS-PCORTF-funded projects, the ***Childhood Obesity Data Initiative (CODI)*** and ***Childhood Obesity Data Initiative (CODI): Integrated Data for Patient-Centered Outcomes Research Project 2.0 (CODI 2.0)***. This project also builds on a prior Fiscal Year 2020 OS-PCORTF project led by NCHS, ***Assessment of Privacy Preserving Record Linkage Using the Linked National Hospital Care Survey (NHCS) and the National Death Index (NDI) as the Gold Standard***, that focused on evaluating results produced using a commercial PPRL solution and provided a foundation for conducting these types of PPRL evaluations. Under the current project, the NCHS team sought to extend the findings from the Fiscal Year 2020 project by assessing the performance of an open-source PPRL solution using data from NHCS and mortality data from NDI.

Challenge Addressed

The project's evaluation sought to address manifold needs in developing data infrastructure capacity for PPRL tools within public health and PCOR studies. First, growing concerns about data privacy and security across the federal government and health research community have highlighted the need for data linkage solutions that protect individual privacy through secure linkage techniques.[16,17,18] Recent advancements in PPRL methods have resulted in more sophisticated tools that can more effectively link large datasets. However, these tools are still evolving and their feasibility for use with administrative datasets has not been widely tested.[19] Second, patient-centered outcomes research often requires the use of a range of data sources, including clinical data, administrative data, programmatic data, survey data, and patient-reported data, which can vary in the quality of PII. Given such variation, it is important to understand how different PPRL tools perform not only with data sources containing high-quality PII, but also data sources containing PII data with more variable quality (e.g., missing unique identifiers, inconsistent or informal names).

Project Impact on Expanding Data Capacity for Patient-Centered Outcomes Research

This project assessed the quality of linked datasets produced using an open-source PPRL tool, compared to gold standard data linkage techniques (i.e., clear-text linkage). By comparing Anonlink's linkage accuracy to clear-text linkage approaches, the evaluation results can help maximize linkage accuracy when using PPRL tools, increase confidence in the scientific validity of linked data resources created with PPRL, and expand linkage opportunities across the federal data ecosystem.

This project improves access to federal data assets by determining best practices in PPRL linkage, allowing for accelerated linkages of federal data. These efforts can improve data sharing and access to research initiatives that are unable to conduct linkages using personal identifiers or PII.

This project improves understanding of the strengths and limitations of a PPRL tool regarding data security, linkage accuracy, and scientific validity of resulting data resources. The NCHS team's evaluation of the PPRL tool, Anonlink, provides insights into the application of PPRL for data linkage within HHS and across other federal agencies.

Key Resources

Project Final Report. This final report describes the project's key accomplishments, outcomes, and challenges addressed. The final report, as well as a manuscript describing PPRL assessment findings, will be available in 2026.

Long-Term Impacts on Interagency Partnerships and Sustainability

This project informs HHS and Operating Divisions' objectives to identify methodologies and opportunities to expand linkages with HHS data resources while protecting privacy. The results of this project have increased HHS confidence in both the reliability of utilizing PPRL tools and the scientific validity of the resulting linked data resources. While the project did not directly link HHS data, its evaluation of a PPRL method will inform future efforts to perform data linkages. Improving linkage processes can result in decreased administrative burden, improved data quality (e.g., standardized for linkage), and increased opportunities for stakeholders across the health ecosystem to link and access HHS data to improve research on patient-centered health outcomes. The project team also shared a manuscript describing the evaluation and security risk assessment results of AnonLink with colleagues in CDC and federal partners to inform data sharing efforts and implementation of PPRL technologies across federal data systems.

The project has helped position the NCHS Data Linkage program as PPRL experts within HHS who can provide technical assistance to other agencies on deploying PPRL tools.

Implementing Agency
CDC

Project Profile
ASPE Website Project Profile

Making Medicaid Data More Accessible Through Common Data Models and FHIR APIs

This project developed and published open-source code to transform T-MSIS Research Identifiable Files (RIFs) into standardized formats using the OMOP and Sentinel CDMs. To facilitate this transformation, data quality was characterized using the harmonized Data Quality Assessment Framework developed by a previously funded OS-PCORTF project, ***Standardization and Querying of Data Quality Metrics and Characteristics for Electronic Health Data***. To illustrate the benefits of the CDM transformation, the project used open-source tools to conduct linkages of transformed data to support maternal mortality studies and assessed the feasibility of using FHIR standards to link EHR data to Sentinel CDM-formatted T-MSIS claims data. The transformed T-MSIS dataset improves access to high-quality, standardized Medicaid data for researchers—an ideal data source to study service utilization and outcomes for underrepresented populations. This, in turn, will accelerate analyses and enable multi-database studies.

Challenge Addressed

The T-MSIS RIFs are a relatively new research-optimized national Medicaid dataset that includes rich information on enrollment, demographics, service utilization, and payment for Medicaid and Children's Health Insurance Program (CHIP) beneficiaries.[20] However, T-MSIS data can be complex and challenging to use due to variations in data quality across states and difficulties combining it with other data sources, making it an ideal target for efforts to make it more useful for analysis.[21] CDMs are increasingly used by researchers to standardize data structure across different data sources, which supports evidence generation by ensuring that databases can be developed and used with greater efficiency, versatility, consistency, and scalability.

Project Impact on Expanding Data Capacity for Patient-Centered Outcomes Research

The project improved the completeness, accuracy, consistency, and utility of Medicaid data quality by developing tools to standardize T-MSIS data into the Sentinel and OMOP CDM formats. This allows researchers to ensure consistency of T-MSIS data across states and years, enabling more reliable multistate analyses. The project also developed 30 data quality metrics to characterize the impact of CDM transformation and assess the data's fitness for research use.

The team expanded the availability of data to study maternal and infant outcomes by linking live birth deliveries to individual infants using the Sentinel CDM-formatted T-MSIS data. To inform future efforts to enrich T-MSIS with more detailed clinical data, a feasibility assessment of implementing FHIR application programming interfaces (APIs) to link T-MSIS data with EHR data was conducted.

The team made the Sentinel-formatted T-MSIS dataset publicly accessible to researchers. This enables researchers to more readily use, retrieve, aggregate, link, and analyze formatted T-MSIS data with other data formatted in the same CDM. Additionally, researchers using TAF RIF data will be able to format their data into Sentinel or OMOP using the open-source code, better enabling them to leverage these data infrastructures and their associated toolkits in their own research.

The project developed new code for transforming T-MSIS data into the two CDM formats, as well as new methodologies for linking and analyzing data through the mother-infant linkage. The open-source code and accompanying user guide used to transform T-MSIS files into the two CDMs and conduct the mother-infant linkage is available to the public.

Key Resources

T-MSIS Dataset Formatted in Sentinel CDM. This dataset contains T-MSIS Analytic Files (TAF) RIFs converted to the Sentinel CDM. The dataset contains Medicaid and CHIP data and is housed in the CMS VRDC.

Medicaid/CHIP Transformation User Documentation. This user guide contains information on how to transform five separate Medicaid TAF RIF files into the Sentinel CDM format.

Sentinel CDM Data Processing Code. This open-source code contains information researchers need to transform T-MSIS RIF files into the Sentinel CDM format.

Mother-Infant Linkage Code. This freely available code can be used by researchers to create mother-infant linkages for T-MSIS and EHR data in the Sentinel CDM.

OMOP CDM Data Processing Code. This open-source code, posted on GitHub, contains information researchers need to transform T-MSIS RIF files into the OMOP CDM format.

OMOP CDM Transformation User Guide. This user guide, posted on GitHub, describes the data transformation and mapping process of T-MSIS data into the OMOP CDM format.

Project Final Report. This final report describes the project's key accomplishments, outcomes, and challenges addressed.

Long-Term Impacts on Interagency Partnerships and Sustainability

The CDM-formatted T-MSIS dataset is being leveraged for new federal surveillance and private sector research activities. The Sentinel-formatted T-MSIS dataset has been incorporated into FDA's Sentinel Distributed Database[22] and is regularly used by FDA investigators for post-market safety surveillance of medical products. The results from these studies are used to inform regulatory decision-making and contribute to FDA actions such as updates to medical product labeling and drug safety communications. Additionally, the Harvard Pilgrim Health Care Institute is using the Sentinel CDM data process code to transform T-MSIS data for a NIH-funded research project studying the effects of medications taken during pregnancy on pregnancy and birth outcomes.[23]

Implementing Agency
FDA & NIH/NLM

Project Profile
ASPE Website Project Profile

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

The project aimed to create new data capabilities for federal and external researchers to longitudinally analyze the relationship between emergency medical services (EMS) provision, including treatment and transportation, and health outcomes among Medicaid beneficiaries. The project planned to develop the Multistate EMS and Medicaid Dataset (MEMD), a publicly available dataset containing Medicaid claims data linked to EMS records from up to five states. The project also proposed creating technical assistance materials to enable other states to link their own EMS and Medicaid data, which would facilitate research to inform clinical protocols, insurance coverage policies, and scope of practice laws for EMS.

Challenge Addressed

EMS are a major component of the acute care system, providing on-demand care for people experiencing medical emergencies.[24] The importance of EMS was highlighted during the COVID-19 pandemic due to higher rates of acute medical emergencies as well as the need for EMS to deliver care at the site of an emergency without transportation to minimize overcrowding in hospitals.[25] Despite the important role of EMS in the health care system, few data sources contain information on the outcomes of patients who receive these services. In the absence of a robust data infrastructure to support research on the outcomes of patients after their EMS encounter, there is limited ability to assess the impact of EMS service provision and identify policies that could improve these services.

Project Impact on Expanding Data Capacity for Patient-Centered Outcomes Research

To provide more relevant, comprehensive data, the project obtained informal agreements from two state Medicaid offices and secured Institutional Review Board (IRB) approval to access state Medicaid data for an EMS-Medicaid linkage.

Key Resources

Project Final Report. This final report describes the project's key accomplishments, outcomes, and challenges addressed.

This project was sunset prior to producing the linked MEMD dataset, but valuable key lessons learned from the work are noted below.

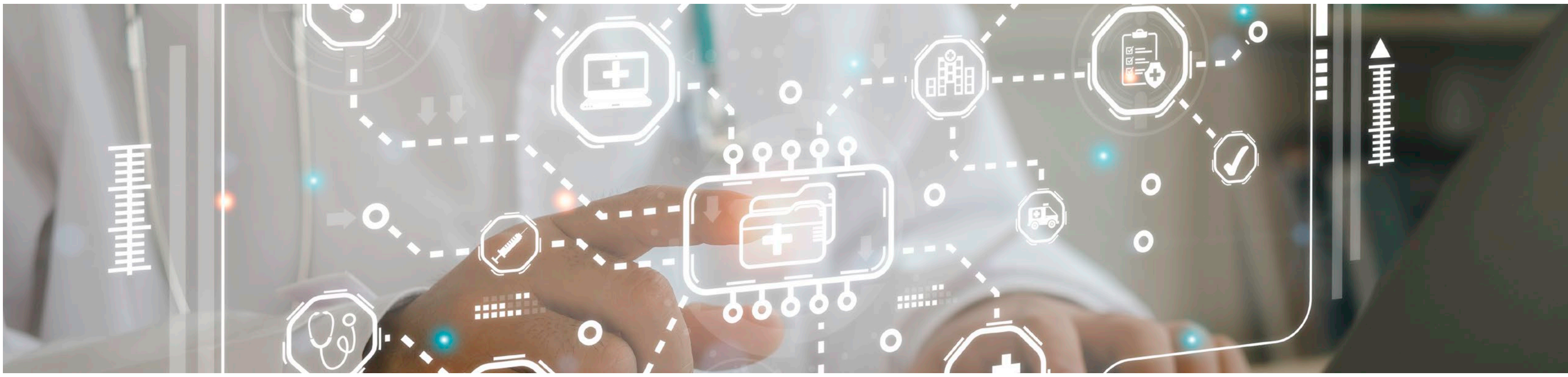
Lessons Learned

The project team successfully assessed the quality of states' Medicaid claims and electronic patient care report data and determined that 11 states had data quality sufficient for linkage. In addition, this project surfaced lessons learned related to enabling state participation and securing formal agreements with states for data sharing that can be addressed by future projects. The MEMD project received positive feedback from state officials and stakeholders for its potential to investigate outcomes of Medicaid beneficiaries who use EMS, underscoring that a future linkage of Medicaid and EMS data would be valuable to researchers and decision-makers.

Implementing Agency
ASPE/BHDAP

Project Profile
ASPE Website Project Profile

Appendix B: New Projects Funded in Fiscal Year 2025



This appendix provides an overview of the three new Office of the Secretary Patient-Centered Outcomes Research Trust Fund (OS-PCORTF) projects funded in Fiscal Year 2025. The three newly funded Fiscal Year 2025 OS-PCORTF projects strengthen data infrastructure for patient-centered outcomes research across four domains: 1) improving data quality; 2) providing more relevant, comprehensive data; 3) improving data access and use; and 4) enhancing analytic resources.

Improving Data Quality	Improving Data Access and Use	Providing More Relevant, Comprehensive Data	Enhancing Analytic Resources

Iterative Data Infrastructure

67%

of new projects funded in Fiscal Year 2025 build on prior OS-PCORTF projects.

The achievements of previous OS-PCORTF projects continue to inform and strengthen new initiatives, highlighting the lasting impact of the portfolio and the interconnected nature of the work. In Fiscal Year 2025, **two of the three** newly funded projects build directly on prior OS-PCORTF products and findings to further advance data infrastructure and promote data access and transparency.

One project builds off the ***Enhancing Data Resources for Researching Patterns of Mortality in Patient Centered Outcomes Research*** project by linking National Death Index (NDI) data to Centers for Medicare and Medicaid Services (CMS) data. This linkage will expand access to cause-of-death information previously unavailable in the CMS Virtual Research Data Center (VRDC), enhancing researchers' ability to study mortality-related outcomes.

One project will link data from the Agency of Healthcare Research and Quality (AHRQ) Healthcare Utilization Cost and Utilization Project (HCUP) to NDI data, building off four prior projects: (1) ***Improving the Mortality Data Infrastructure for Patient Centered Outcomes***, (2) ***Augmenting the National Hospital Care Survey (NHCS) Data through Linkages with***

Administrative Records, (3) ***Compendium of Policies on Use of Linkable Mortality Data for PCOR in the Nation's 57 Reporting Jurisdictions***, and (4) ***Evaluation of Privacy-Preserving Record Linkage Solutions to Broaden Linkage Capabilities in Support of PCOR Objectives***. The resulting linkage will expand researcher access to mortality and survival-related outcomes data on a state-by-state basis.

New Project Profiles

Enhancing All-Payer Data for Researching Patterns of Mortality in Patient-Centered Outcomes Research

This project will develop and implement approaches for linking data between the AHRQ HCUP—the only comprehensive all-payer billing database for hospital care in the U.S.[26]—with NDI records.

Although AHRQ's HCUP collects outcomes data related to hospital encounters and inpatient and emergency department care, its scope does not capture post-discharge mortality. This gap restricts research into premature mortality, particularly deaths related to hospital complications.[27] By linking HCUP and NDI data, this project will enhance data infrastructure to support more comprehensive research on mortality outcomes, helping to uncover drivers of premature death and identify opportunities to improve survival and quality of care. Importantly, the project will also address state-by-state legal, regulatory, and policy variations related to access and use of NDI data by developing flexible linkage approaches that accommodate jurisdiction-specific requirements with patient privacy while ensuring data validity and consistency.[28]

The team will link NDI data to existing HCUP state databases that contain annual inpatient and outpatient source data files for the years 2019-2022, expanding the dataset to a census-level scope and enabling comprehensive mortality analyses.

The team will explore and implement linkage methods, including privacy-preserving record linkages (PPRL), to enable connections between datasets that were previously not feasible. Linkage plans will be comprehensive and tailored to account for the specific requirements and policies of individual state jurisdictions.

To enhance access to NDI-linked HCUP data, the project team will create consistently formatted publicly available and restricted-access state databases in collaboration with 10 state partners.

Implementing Agency
AHRQ

Enhancing Data Resources for Studying Patterns and Correlates of Mortality in Patient-Centered Outcomes Research

This project will enhance the mortality data available in the CMS VRDC by linking mortality data for Medicare beneficiaries with descriptive data on the date, cause, and manner of death from the NDI. This linkage will improve researchers' ability to study mortality-related outcomes and generate more accurate estimates of survival and risk.

Currently, many population-based datasets, including the CMS Master Beneficiary Summary File (MBSF), do not collect comprehensive data on date or fact of death, which can result in misestimation of mortality risk in associated research.[29,30] The MBSF is a comprehensive dataset that contains person-level information on Medicare beneficiaries, including demographic characteristics, enrollment details, and indicators for medical conditions.[31] By integrating NDI data with the MBSF, this project will enhance the accuracy and completeness of mortality information, enabling more robust analyses to advance understanding and improvement of survival outcomes.

The project team will link individual-level NDI data with the 2021-2024 MBSF, allowing researchers to examine outcomes such as survival rates, cause-specific mortality, and the effectiveness of interventions across different Medicare beneficiaries.

The project team will promote the utilization and adoption of the new linked mortality dataset by developing free training resources for interested researchers, guiding them on how to use the CMS VRDC and how to incorporate these mortality data into their research studies.

Implementing Agency
CMS

Developing a Context-Relevant Indian Health Service Clinical Performance Registry to Support Data-Driven Improvements in Service Delivery

This project will create a detailed roadmap to guide the development of a Clinical Performance Registry at three Indian Health Service (IHS) clinics. The roadmap will guide the development of infrastructure to aggregate and harmonize data from multiple electronic health record (EHR) systems to support quality improvement efforts in key clinical areas including maternal health, chronic disease prevention and management, and nutrition.

American Indian/Alaska Native (AI/AN) populations experience poorer health outcomes in chronic disease, nutrition, and maternal morbidity and mortality, driven by socioeconomic factors and limited access to care.[32,33] Limited access to comprehensive clinical data about these outcomes hinders meaningful research and the ability to advance clinical quality improvement in health care delivery for these populations. The Clinical Performance Registry will harmonize and analyze data across disparate health systems, serving as a decision-making tool to inform policy and program development, clinical interventions, and quality improvement efforts.

The project team will work with a technical expert panel (TEP) to develop strategies to identify and harmonize measures to include in the Clinical Performance Registry, facilitating streamlined data integration across different EHRs.

The project team will collaborate with the TEP to create metadata for each measure in the Clinical Performance Registry, specifying the data elements needed to locate and extract relevant variables from the EHR. This will establish a consistent framework for data contributors, improving the consistency and reliability of data and measures.

To support long-term interoperability, the team will design metadata and variable structures that align with future EHR initiatives, including those planned under the **IHS Health IT Modernization Program**.[34]

Implementing Agency
IHS

Appendix C: In Progress Projects in Fiscal Year 2025

This appendix provides a brief overview of the 19 Office of the Secretary Patient-Centered Outcomes Research Trust Fund projects in progress during Fiscal Year 2025.

Projects in Progress in Fiscal Year 2025
Capturing the Full Trajectory of Patient-Centered Cancer Care Via Enhanced Data Linkages (NIH/NCI and NIH/NCATS). This project will create and evaluate longitudinal linked datasets from cancer registry and electronic health record (EHR) data that capture the full trajectory of cancer patients' care that can be used by researchers to understand and compare the impacts of cancer treatments on patient outcomes.
Childhood Obesity Data Initiative (CODI): Integrated Data for Patient-Centered Outcomes Research Project 2.0 (CDC). This project will refine CODI's existing data tools and services to promote data standardization, improving researchers' ability to combine data from pediatric patient-level EHRs, weight management intervention, and community (census) data to assess intervention effectiveness for different pediatric populations.
Code Map Services for Interoperability of Common Data Models and Data Standards (NIH/NCATS). This project will create an automated set of code mapping services for common data models (CDMs) to promote interoperability and sharing of real-world data for research.
Dataset on Intellectual and Developmental Disabilities: Linking Data to Enhance Person-Centered Outcome Research (ASPE/BHDAP). To provide more relevant comprehensive data for research on person-centered outcomes for people with intellectual and developmental disabilities (ID/DD), this project will link Adult Version scores of the Support Intensity Scale (SIS), Medicaid claims, data from the National Core Indicators – Intellectual and Developmental Disabilities (NCI-IDD) In-Person survey and COVID-19 Supplement, and other relevant state-level data.
Digitizing Consent and Regulatory Metadata Towards Streamlining Governance of Pediatric COVID-19 Research Data Linkages (NIH/NICHD). Based on pediatric COVID-19 studies, this project will develop a data governance approach that standardizes and digitizes consent metadata to help researchers determine the appropriate use of these data.

Projects in Progress in Fiscal Year 2025

Establishing the Governance, Legal and Analytical Framework for a Federated Linked Data System: Creating a New Data Research Environment (NIH/NCI). This project will enhance data capacity for cancer research by developing a legal and governance framework for a federated linked data system that will allow secure data linkages across federal agencies and care settings.

Expanding Climate Change and Health Data Infrastructure to Advance Health Interventions: Linking Health and Environmental Data to Improve Patient and Community Health (NIH/NIEHS). This project will provide researchers with accessible, timely, harmonized, interoperable, and linkable data by developing a publicly accessible catalog of curated data sources, tools, methods, and educational resources related to climate change and health as well as datasets to enable research on the health impact of environmental exposures.

Improving Maternal Health by Leveraging Existing Resources, Data Linkage and Standard Reporting for Patient-Centered Outcomes Research (ASTP/ONC). This project will improve the quality and accessibility of United States Core Data for Interoperability Plus (USCDI+) data standards for maternal health by expanding the current list of maternal health data elements and developing implementation standards to ensure data interoperability across health care delivery and data systems.

Improving the Availability of National Mortality Data by Accelerating Data Release and Laying the Groundwork for IT Modernization of the National Death Index (NDI) (CDC). This project will improve the timeliness and quality of mortality data in the NDI by shifting from annual end-of-year releases to more frequent, provisional monthly releases, and by upgrading the data infrastructure to a more flexible and secure cloud-based platform.

Leveraging Data Standards for Patient-Centered Outcomes Research in Sickle Cell Disease under the United States Core Data for Interoperability+ Program (USCDI+) (ASTP/ONC). This project will improve the quality and accessibility of standards for sickle cell disease by developing a minimum set of USCDI+ data elements and developing implementation specifications to ensure the interoperability of standards across care delivery settings.

Linking Federally Qualified Health Center (FQHC) EHR and Medicaid Data for Increased Data Capacity to Understand Maternal Health Care (CDC/NCHS). The project will link FQHC EHR data to the Transformed Medicaid Statistical Information System (T-MSIS) data to enhance researchers' ability to analyze data on maternal health care provided at FQHCs and subsequently gain insight into the relationships among patient characteristics, service utilization, and outcomes for maternal health visits.

Linking the Boston Birth Cohort and Pregnancy to Early Life Longitudinal Data System and Social Determinants of Health Databases: A Longitudinal Cohort of Mother-Child Dyads (HRSA/MCHB). This project will link data from the Boston Birth Cohort Pregnancy, Early Life Longitudinal Data System, and the AHRQ Social Determinants of Health (SDOH) Database to generate comprehensive data on mothers and children related to health service utilization, environmental impacts, and clinical and intergenerational health outcomes.

Linking the National Survey of Children's Health Longitudinal Cohort and Medicaid Claims Data (HRSA/MCBH) This project will link National Survey of Children's Health Longitudinal Cohort Study data to T-MSIS data to provide more comprehensive data on the long-term impacts of COVID-19 on children and families and effectiveness of treatments.

Medical Expenditure Panel Survey (MEPS) Enhancements on Burdens and Economic Impacts of Medical Care (AHRQ). This project will enhance the comprehensiveness of MEPS data by collecting and incorporating insights collected directly from patients and caregivers on economic burdens and outcomes. This project will also enable future linkages to the AHRQ SDOH Database.

Projects in Progress in Fiscal Year 2025

N3C Renal Freeport: A Collaborative Data Analytics and Data Sharing Model for Patient-Centered Outcomes Research (NIH/NCATS). This project will improve the comprehensiveness and accessibility of real-world clinical data for chronic kidney disease (CKD) and renal transplantation by developing a secure data-sharing infrastructure containing linked datasets of real-world evidence including de-identified EHR data, CMS claims, and Scientific Registry for Transplant Recipients data. Additionally, it will implement a governance structure and technical requirements for interagency data sharing and collaboration.

Panoramic View of Patient Care Through Data Innovations and Linkages Implementing (AHRQ). This project will link data across multiple sources, including non-medical data and community information, to capture a panoramic view of patient care in two-to-three states, providing richer, more complete data to answer PCOR and comparative effectiveness questions related to health care coverage and health outcomes.

Small-Area Community SDOH Data: Enhancements and Linkages to Inform Action (AHRQ). This project will improve the granularity and comprehensiveness of data available in the AHRQ SDOH Database by adding more recent years of data and adding new variables on non-medical factors that influence health, as well as standardize data in the database to enhance its usability for researchers.

Using Machine Learning Techniques to Enable Health Information Exchange (HIE) Data Sharing to Support COVID-19-focused PCOR (ASTP/ONC). This project will leverage USCDI, application programming interfaces, and privacy-preserving machine learning techniques to facilitate the more efficient use of EHR data from HIEs for research.

Utilizing Natural Language Processing and Machine Learning to Enhance the Identification of Stimulant and Opioid-Involved Health Outcomes in the National Hospital Care Survey (CDC/NCHS). To provide researchers access to more relevant, high-quality data on stimulant-involved health encounters, this project will develop an algorithm using natural language processing and machine learning to identify hospital encounters involving stimulants in both structured and unstructured hospital data.

References

[1] Centers for Disease Control and Prevention, National Center for Health Statistics. NCHS Data Linkage Activities. Content last reviewed May 19, 2025. **https://www.cdc.gov/nchs/data-linkage/index.htm**

[2] King JH, Hall MAK, Goodman RA, Posner SF. Life in Data Sets: Locating and Accessing Data on the Health of Americans Across the Life Span. J Public Health Manag Pract. 2021;27(3):E126-E142. doi:10.1097/PHH.0000000000001079

[3] U.S. Government Accountability Office. *Open Data: Additional Action Required for Full Public Access*. GAO-22-104574. Published December 16, 2021. Accessed October 12, 2025. **https://www.gao.gov/assets/720/718205.pdf**

[4] Parpia, Alyssa S., Isabel Martinez, Abdulrahman M. El-Sayed, Chad R. Wells, Lindsey Myers, Jeffrey Duncan, Jim Collins, Meagan C. Fitzpatrick, Alison P. Galvani, and Abhishek Pandey, "Racial Disparities in COVID-19 Mortality Across Michigan, United States," EClinicalMedicine, Vol. 33, February 26, 2021.

[5] Cisewski JA, Ahmad FB, Rossen LM. Timeliness of death certificate data by jurisdiction and selected causes of death: United States and Puerto Rico, 2023. Vital Statistics Rapid Release. 2025 Sep;(40):1–11. doi: https://dx.doi.org/10.15620/cdc/174624.

[6] Brown JS, Bastarache L, Weiner M. Aggregating Electronic Health Record Data for COVID-19 Research—Caveat Emptor. JAMA Network Open. 2021;4(7):e2117175. doi:10.1001/jamanetworkopen.2021.17175,

[7] Joo JY. Fragmented care and chronic illness patient outcomes: A systematic review. *Nurs Open*. 2023;10(6):3460-3473. doi:10.1002/nop2.1607

[8] Office of Refugee Resettlement, Office of the Administration for Children and Families. Refugee Resettlement Program. Accessed October 14, 2025. **https://acf.gov/orr/programs/refugees**

[9] Office of the Assistant Secretary for Planning and Evaluation. Refugee and Asylee Data on the Utilization of Medicaid. Published October 19, 2023. Accessed on October 14, 2025. **https://aspe.hhs.gov/reports/refugee-asylee-data-utilization-medicaid**

[10] 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. Intellect Dev Disabil. 2013;51(5):385-398. doi:10.1352/1934-9556-51.5.385

[11] Krahn GL, Fox MH. Health disparities of adults with intellectual disabilities: what do we know? What do we do?. *J Appl Res Intellect Disabil*. 2014;27(5):431-446. doi:10.1111/jar.12067

[12] Karimi M, Dhopeshwarkar R, Jiménez F, Ryan S, Plourde E. Improving Data Infrastructure for Person-Centered Outcomes Research on Intellectual and Developmental Disabilities. *Am J Intellect Dev Disabil*. 2024;129(3):231-241. doi:10.1352/1944-7558-129.3.231

[13] Hall JP, Goddard KS, Ipsen C, Myers A, Kurth NK. Counting everyone: evidence for inclusive measures of disability in federal surveys. *Health Aff Sch*. 2024;2(9):qxae106. Published 2024 Aug 21. doi:10.1093/haschl/qxae106

[14] Williams SN, Ukaigwe J, Ward BW, Okeyode T, Shimizu IM, National Center for Health Statistics. Sampling Procedures for the Collection of Electronic Health Record Data From Federally Qualified Health Centers, 2021–2022 National Ambulatory Medical Care Survey. Vital Health Stat. Series 2. 2023 Jun 26;(203). Accessed October 14, 2025. **https://stacks.cdc.gov/view/cdc/127730**

[15] Centers for Disease Control and Prevention, National Center for Health Statistics. NCHS Data Linkage Activities. Content last reviewed May 19, 2025. **https://www.cdc.gov/nchs/data-linkage/index.htm**

[16] U.S. Department of Health and Human Services. HHS Priorities. Content last reviewed September 30, 2025. **https://www.hhs.gov/about/priorities/index.html**

[17] U.S. Food and Drug Administration. Enterprise Modernization Action Plan (EMAP). 2022. Accessed October 14, 2025. **https://www.fda.gov/media/158206/download**

[18] H.R.4174 - 115th Congress (2017-2018): Foundations for Evidence-Based Policymaking Act of 2018. (2019, January 14). **https://www.congress.gov/bill/115th-congress/house-bill/4174**

[19] Pathak A, Serrer L, Zapata D, et al. Privacy preserving record linkage for public health action: opportunities and challenges. J Am Med Inform Assoc. 2024;31(11):2605-2612. doi:10.1093/jamia/ocae196

[20] Centers for Medicare and Medicaid Services. Medicaid and CHIP T-MSIS Analytic Files Data Release. November 7, 2019. Accessed October 14, 2025. **https://www.cms.gov/newsroom/fact-sheets/medicaid-and-chip-t-msis-analytic-files-data-release**

[21] Rai A, Maro JC, Dutcher S, Bright P, Toh S. Transparency, reproducibility, and replicability of pharmacoepidemiology studies in a distributed network environment. Pharmacoepidemiology and Drug Safety. 2024;33(6). doi:https://doi.org/10.1002/pds.5820

[22] U.S. Food and Drug Administration. FDA's Sentinel Initiative. March 8, 2024. Accessed October 14, 2025. **https://www.fda.gov/safety/fdas-sentinel-initiative**

[23] NIH RePORTER. Big data apprOaches fOr Safe Therapeutics in Healthy Pregnancies (BOOST-HP). Accessed October 14, 2025. **https://reporter.nih.gov/search/l-SFCiE4SEWfQo_fF1qDqg/project-details/11082300**

[24] Ems.gov. What is EMS? Accessed October 14, 2025. **https://www.ems.gov/what-is-ems/**

[25] Richter H, Schneider M, Eisenberger J, Jafari N, Haumann H, Häske D. Impact of the COVID-19 pandemic on prehospital emergency medical service: a scoping review. Front Public Health. 2025 Mar 19;13:1543150. doi: 10.3389/fpubh.2025.1543150.

[26] Agency for Healthcare Research and Quality. Healthcare Cost and Utilization Project (HCUP). Published 2025. Accessed October 12, 2025. **https://www.ahrq.gov/data/hcup/index.html**

[27] Davies SM, Geppert J, McClellan M, et al. Refinement of the HCUP Quality Indicators. Rockville (MD): Agency for Healthcare Research and Quality (US); 2001 May. (Technical Reviews, No. 4.) Available from: **https://www.ncbi.nlm.nih.gov/books/NBK43831/**

[28] da Graca B, Sorensen A, Ettlinger J, Priest EL. Evolving trends and challenges in national death data. Proc (Bayl Univ Med Cent). 2025 May 8;38(4):460-461. doi: 10.1080/08998280.2025.2499428. PMID: 40557196; PMCID: PMC12184195.

[29] Barberio J, Naimi AI, Patzer RE, et al. Influence of incomplete death information on cumulative risk estimates in US claims data. *Am J Epidemiol*. 2024;193(9):1281-1290. doi:10.1093/aje/kwae034

[30] Research Data Assistance Center. Death Information in the Research Identifiable Medicare Data. Published November 9, 2022. Accessed October 12, 2025. **https://resdac.org/articles/death-information-research-identifiable-medicare-data**

[31] Centers for Medicare & Medicaid Services. Master Beneficiary Summary File (MBSF) Limited Data Set (LDS). Published August 18, 2025. Accessed October 12, 2025. **https://www.cms.gov/data-research/files-for-order/limited-data-set-lds-files/master-beneficiary-summary-file-mbsf-lds**

[32] Singh GK, Lee H, Kim LH, Williams SD. Social Determinants of Health Among American Indians and Alaska Natives and Tribal Communities: Comparison with Other Major Racial and Ethnic Groups in the United States, 1990-2022. Int J MCH AIDS. 2024;13:e010. Published 2024 May 20. doi:10.25259/IJMA_10_2024

[33] U.S. Government Accountability Office. *Tribal Epidemiology Centers: HHS Actions Needed to Enhance Data Access*. GAO-22-104698. Washington, DC: U.S. Government Accountability Office; 2022. **https://www.gao.gov/assets/D22104698.pdf**

[34] Indian Health Service. Health Information Technology Modernization Program. Indian Health Service. Accessed October 15, 2025. **https://www.ihs.gov/hit/**