OFFICE OF THE SECRETARY
PATIENT-CENTERED OUTCOMES RESEARCH TRUST FUND

Final Report

Creating Effective Medicaid Data Linkages for Health Outcomes Research

Prepared for
The Office of the Assistant Secretary for Planning and Evaluation (ASPE)
at the U.S. Department of Health and Human Services

By
NORC at the University of Chicago

May 2026

OFFICE OF THE ASSISTANT SECRETARY FOR PLANNING AND EVALUATION

The Assistant Secretary for Planning and Evaluation (ASPE) advises the Secretary of the U.S. Department of Health and Human Services (HHS) on policy development in health, disability, human services, data, and science; and provides advice and analysis on economic policy. ASPE leads special initiatives; coordinates the Department's evaluation, research, and demonstration activities; and manages cross-Department planning activities such as strategic planning, legislative planning, and review of regulations. Integral to this role, ASPE conducts research and evaluation studies; develops policy analyses; and estimates the cost and benefits of policy alternatives under consideration by the Department or Congress.

THE OFFICE OF HEALTH POLICY

The Office of Health Policy (HP) provides a cross-cutting policy perspective that bridges Departmental programs, public and private sector activities, and the research community, in order to develop, analyze, coordinate and provide leadership on health policy issues for the Secretary. HP carries out this mission by conducting policy, economic and budget analyses, assisting in the development and review of regulations, assisting in the development and formulation of budgets and legislation, and assisting in survey design efforts, as well as conducting and coordinating research, evaluation, and information dissemination on issues relating to health policy.

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

The Office of the Secretary Patient-Centered Outcomes Research Trust Fund (OS-PCORTF) was established as part of the 2010 Patient Protection and Affordable Care Act and is charged to build data capacity for patient-centered outcomes research. Coordinated by ASPE on behalf of the Department, OS-PCORTF has funded a rich portfolio of projects to meet emerging U.S. Department of Health and Human Services policy priorities and fill gaps in data infrastructure to enhance capabilities to collect, link, and analyze data for patient-centered outcomes research. For more information, visit https://aspe.hhs.gov/collaborations-committees-advisory-groups/os-pcortf

This report was funded by the Office of the Secretary Patient-Centered Outcomes Research Trust Fund (OS-PCORTF) under Contract Number HHSP233201500048II of the HHS Office of the Assistant Secretary for Planning and Evaluation (ASPE). The work was carried out by NORC at the University of Chicago (NORC) and ASPE. The authors are solely responsible for this document's contents, findings, and conclusions, which do not necessarily represent the views of HHS, ASPE, or NORC. Readers should not interpret any statement in this product as an official position of ASPE or of HHS.

Suggested Citation: Srinivasan M, Heaney-Huls K, Gauthreaux N, Shams R, Couture S, Wiltz J, and Dullabh P. Creating Effective Medicaid Data Linkages for Health Outcomes Research. Washington, D.C.: Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services. May 2026.

CONTRIBUTING AUTHORS

Mithuna Srinivasan, PhD, Principal Research Scientist, NORC
Krysta Heaney-Huls, MPH, Research Scientist, NORC
Nicole Gauthreaux, MPH, Senior Research Associate II, NORC
Rida Shams, MSPH, Senior Research Associate I, NORC
Sara Couture, MPH, Social Science Analyst, ASPE
Jennifer Wiltz, MD, MPH, Chief Medical Officer, ASPE
Prashila Dullabh, MD, Vice President and Senior Fellow, NORC

PROJECT OFFICERS AND PROJECT LEADERSHIP

Sara Couture (ASPE)
Jennifer Wiltz (ASPE)
Marcos Trevino (ASPE)
Prashila Dullabh (NORC)
Mithuna Srinivasan (NORC)

ACKNOWLEDGEMENTS

The authors would like to acknowledge the OS-PCORTF Project Leads who participated in key informant interviews and the ASPE staff who provided feedback on previous versions of this report including Marissa Abbott, Valeria Butler, Meredith Dost, Cordell Golden, “Chip” Haltermann III, Andrew Kelly, Vetisha McClair, Emma Plourde, Olivia Sappenfield, Andrew Shatto, and Brian Ward.

Table of Contents

Executive Summary

Patients, clinicians, decision makers, and other stakeholders depend on robust, evidence-based information to make informed health care decisions. Linking Medicaid data to other sources enables a deeper understanding of how Medicaid services and interventions influence health outcomes. Sharing insights and best practices from existing data linkage projects can support future linkage efforts that facilitate health outcomes research.

This report qualitatively assesses 11 Medicaid data linkage projects funded by the U.S. Department of Health and Human Services Office of the Secretary Patient-Centered Outcomes Research Trust Fund (OS-PCORTF) to understand data linkage approaches and share challenges and best practices for data linkage. The report also describes the impact of the linked datasets and identifies opportunities to strengthen data linkage infrastructure.

The project team reviewed documents produced by the 11 Medicaid data linkage projects, including scopes of work, progress and final reports, and manuscripts. We then conducted virtual key informant interviews in June 2025 with project teams and conducted a thematic analysis of the document review and interviews.

Approaches for Effective Medicaid Data Linkages

The OS-PCORTF projects described in this report linked a range of data sources, including clinical data, population-based survey data, utilization data, and vital records, to Medicaid data to create more comprehensive data to study the causes of disease and improve effective service delivery. Most projects (n=7) used Transformed Medicaid Statistical Information System data, while the rest used Medicaid data sourced from individual states. To link disparate data sources, project teams employed a range of linkage methods, including deterministic, probabilistic, and privacy-preserving record linkage (PPRL), based on the characteristics of the data sources and the resources available to perform the linkage. Three project teams used common data models to harmonize data before linkage, creating more efficient ways to link data in the future and improve interoperability.

Notably, the OS-PCORTF project teams highlighted challenges related to data quality, administrative hurdles to data access, and limited internal capacity. These challenges underscore the need for a more robust Medicaid data linkage infrastructure. Project teams identified several key opportunities to strengthen data capacity: (1) streamlining data governance and administrative processes; (2) documenting linkage procedures to support replication; (3) engaging key stakeholders—including patients—throughout the linkage process; (4) promoting linked datasets and linkage infrastructure as foundational resources for studying a range of populations; (5) providing multiple types of data files to meet varied research needs; (6) advancing secure PPRL techniques; and (7) building state and federal capacity for sustainable data linkage infrastructure.

Impact of Medicaid Data Linkages on Health Outcomes Research

The linked Medicaid datasets reviewed in this report provide more comprehensive data on patient outcomes related to behavioral health, chronic diseases, and maternal and child health. Over 100 studies have used these Medicaid-linked datasets for research that provides more complete longitudinal profiles of patients; captures care received across different providers and systems; evaluates the effectiveness of interventions or policy changes; and enables advanced statistical methodologies. Further, these data can be analyzed for health services utilization and quality improvement to inform the design of administrative programs.

1. Introduction

The U.S. Department of Health and Human Services (HHS) Office of the Secretary Patient-Centered Outcomes Research Trust Fund (OS-PCORTF)[1] funds intradepartmental projects that build data capacity for the collection, linkage, and analysis of data to conduct rigorous health outcomes research. This research generates high-quality evidence about the effectiveness of treatments, services, and other health care interventions on the full range of outcomes identified by patients, caregivers, clinicians, policymakers, and other stakeholders as important. [2] Longitudinal data enables researchers to observe changes and outcomes over time and across different health care and social services settings. Longitudinal datasets can be constructed by linking multiple data sources, creating more comprehensive data to study the causes of disease and improve effective service delivery.

Data linkages integrate disparate data sources, including claims data and electronic health record (EHR) data, across federal, state, and non-governmental data sources managed by different entities, creating more comprehensive data spanning multiple geographic regions and health care settings, as well as social services. Data linkages also provide the foundation for more effective data infrastructure, enabling greater transparency, more robust research, and informed policy decisions. As emphasized in the May 2025 *Make Our Children Healthy Again* report,[3] improved data linkages will be vital to advancing effective health services research and policy development to understand the root causes of disease, as well as enhancing data sharing and interagency collaboration. However, barriers to linking Medicaid data effectively with other sources stem from complexities of data acquisition, governance for sharing and accessing data, and data quality issues (e.g., missing or nonstandard data), which can significantly impact the accuracy of linked data and the use of these data for health outcomes research.[4,5]

The HHS Assistant Secretary for Planning and Evaluation (ASPE) has long recognized the importance of linking administrative Medicaid data—whether from the Transformed Medicaid Statistical Information System (T-MSIS) or state-level sources—with other data sources to analyze health outcomes and health care utilization patterns among Medicaid beneficiaries. Since 2017, the OS-PCORTF has supported 11 HHS projects that link Medicaid data with other data sources to increase the data available for rigorous health outcomes research. These projects were at different phases of development and employed different methods to link Medicaid data with other data sources such as birth certificates,[6] National Center for Health Statistics (NCHS) surveys,[7] and the National Death Index (NDI).[8] Using information gathered through document review and key informant interviews with project leads, this report summarizes the OS-PCORTF-funded Medicaid data linkage projects and related linkage methodologies, challenges, and lessons learned. The findings identify opportunities to replicate and scale successful data linkage strategies to leverage more comprehensive data for health outcomes research. **Appendix A** provides an overview of the projects included in this report. Information on the datasets developed by OS-PCORTF projects, the types of PCOR studies that use these datasets, and opportunities to enhance future data linkages is presented in a separate **infographic** available on the ASPE website **https://aspe.hhs.gov/**.

The report is intended for decision makers and external researchers interested in linking Medicaid data to other sources to track and assess outcomes among Medicaid beneficiaries. Specifically, the report explores the following questions:

1. What approaches and methods have the Medicaid data linkage projects in the OS-PCORTF portfolio used for data linkages?
2. What limitations/challenges in approaches and/or methods have OS-PCORTF-funded Medicaid data linkage projects encountered?
3. What opportunities for building data capacities in Medicaid data linkages do these projects present for future research?
4. To what extent have these projects been used by stakeholders (e.g., researchers, decision makers) and how have project teams measured the impact of data linkage projects?

2. Methods

To identify the data linkage methods used by projects, understand the challenges and opportunities for strengthening data infrastructure, and explore how projects assess impact, we triangulated information from document reviews and key informant interviews with project teams.

2.1 Selection of Projects

We identified projects within the OS-PCORTF portfolio that used Medicaid data, such as T-MSIS data or state Medicaid data, to establish linkages with other data sources. Project selection included completed projects, those in progress, and projects in early stages. In total, we identified 11 projects for inclusion in the report.

2.2 Document Review

For each selected project, the NORC team reviewed information such as progress reports, ASPE website profiles, and final project reports when available. We reviewed 46 documents and created a spreadsheet inventory to abstract information from the available documents. Abstraction fields related to project information; data sources for linkage; linkage dataset description; data linkage methods and analytic tools; linked dataset access and use policies; health outcomes research data analyses supported by the linked dataset; data linkage challenges and lessons learned; and opportunities to build Medicaid data linkage capacity.

2.3 Key Informant Interviews

Of the 11 Medicaid data linkage projects, NORC completed interviews with representatives of eight projects. We conducted 60-minute, semi-structured discussions with project teams to gain a deeper understanding of the projects and their impacts. To prepare for these interviews, NORC developed an interview protocol informed by the key objectives of the report and the information gleaned from data abstraction. See **Appendix B** for the interview protocol. The interview protocol was designed to:

- Answer the key questions and objectives of the project, describing approaches and methodologies used to develop Medicaid datasets.
- Identify stakeholder usage and project impact measurements (as perceived by the project leads).
- Identify gaps and limitations of data linkage approaches used in these projects and, consequently, opportunities and guidelines for future work by health outcomes researchers.

NORC then tailored the interview protocol to each Medicaid data linkage project using information from the document review, probing for any identifiable informational gaps. A NORC facilitator and notetaker were present for each interview. We developed transcript-style notes for every interview and conducted evidence-based thematic analyses to answer the research questions, highlighting emerging concepts and themes from interview data and notable examples.[9] Results from both the document review and thematic analysis informed the report.

3. Findings

In this section we present cross-cutting themes and illustrative examples from the 11 Medicaid data linkage projects (Exhibit 1). We describe the data sources linked across projects (Section 3.1), the methodological approaches used to link disparate data sources (Section 3.2), and data access considerations (Section 3.3). We also discuss limitations and challenges to data linkage efforts (Section 3.4), identify best practices and opportunities to build Medicaid data linkage capacity (Section 3.5), and conclude by highlighting the utility of the new datasets for research (Section 3.6) and for policy insights (Section 3.7).

Exhibit 1. HHS Medicaid Data Linkage Projects Assessed

Creating a Federal COVID-19 Longitudinal Patient Outcomes Research Database Linked to Health Systems and Clinical Data *(National Institutes of Health [NIH]/National Center for Advancing Translational Sciences [NCATS]).* This project built a national longitudinal data infrastructure on COVID-19 and improved comprehensiveness of the National Clinical Cohort Collaborative (N3C) dataset through linkages between Centers for Medicare & Medicaid Services (CMS) claims data and EHR data in the N3C Data Enclave.

Data Linkage: Evaluating Privacy Preserving Record Linkage Methodology and Augmenting the National Hospital Care Survey (NHCS) with Medicaid Administrative Records *(Centers for Disease Control and Prevention [CDC]/National Center for Health Statistics [NCHS]).* This project evaluated privacy-preserving record linkage methodologies and conducted patient-level record linkages of 2014 and 2016 NHCS hospital administrative claims and EHR data to T-MSIS data from 2014-2017.

Data Infrastructure Supporting Research on Refugee Medicaid Service Utilization and Outcomes *(ASPE Office of Human Services Policy [HSP] and the Administration for Children and Families [ACF]).* This project planned to create a linked dataset of Refugee Arrivals Data System and T-MSIS data from 2016-2019 to provide more comprehensive data on refugee service utilization and health outcomes, strengthening the data infrastructure for conducting health outcomes research studies with refugee populations.

Dataset on Intellectual and Developmental Disabilities: Linking Data to Enhance Person-Centered Outcome Research *(ASPE - Office of Behavioral Health, Disability, and Aging Policy [BHDAP]).* To provide more relevant comprehensive data for research on person-centered outcomes for the intellectual and developmental disabilities (ID/DD) population, this project will link Support Intensity Scale scores, Medicaid claims, National Core Indicators survey data, National Core Indicators COVID-19 supplement survey data, and other relevant state-level data.

Developing a Multi-State Network of Linked Pregnancy Risk Assessment Monitoring System (PRAMS) and Clinical Outcomes Data *(CDC).* This project linked PRAMS to clinical outcomes data and other data to create a research database that included a wide range of factors related to maternal and infant health outcomes.

Enhancing Data Resources for Researching Patterns of Mortality in Patient-Centered Outcomes Research (Project 3) *(CMS).* This project linked NDI data to other sources, allowing researchers to develop national estimates of cause-specific death rates following emergency department visits and/or hospital stays for specific conditions.
Linking Federally Qualified Health Center (FQHC) EHR and Medicaid Data to Understand Maternal Health Care *(CDC/NCHS).* The project linked FQHC EHR data to T-MSIS 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 National Survey of Children's Health Longitudinal Cohort (NSCH-LC) and Medicaid Claims Data *(Health Resources and Services Administration, Maternal and Child Health Bureau [HRSA/MCHB]).* This project will link NSCH-LC data to T-MSIS to provide more comprehensive data on the long-term impacts of COVID-19 on children and families.
Linking State Medicaid and Child Welfare Data for Outcomes Research on Treatment for Opioid Use Disorder and Other Behavioral Health Issues *(ASPE HSP and ACF – Office of Planning Research and Evaluation [OPRE]).* This project aimed to enhance data infrastructure and increase data availability for research on a set of parents that have children in the child welfare system and are in need of treatment for opioid use disorder, other substance use disorders, or behavioral health issues.
Multistate Emergency Medical Services (EMS) and Medicaid Dataset (MEMD): A Linked Dataset for Patient-Centered Outcomes Research *(ASPE/BHDAP).* This project planned to create a publicly available dataset linking Medicaid and EMS records in five states, providing more comprehensive data regarding health outcomes among individuals receiving EMS.
Quality of Care and Outcomes Data for Pregnant Medicaid Beneficiaries and Newborn Infants *(CMS).* This project planned to link vital records and birth certificates from up to 20 states and for the calendar years 2016 – 2020 with T-MSIS Analytic Files data, thereby providing researchers with accessible, linked, longitudinal data on pregnant women and their infants.

At the time of report publication, the projects were at varying stages of development. Six projects successfully produced a linked dataset, two projects were in the planning or early execution phases, and three projects did not complete their linkage efforts. See **Appendix A** for an overview of these projects.

3.1 Data Sources Used in Linkage Efforts

Linking Medicaid data to other data sources enables researchers to study a range of interventions and outcomes that would otherwise be challenging to fully assess. These data linkages result in more comprehensive data to conduct effectiveness studies that can enhance clinical decision-making and policymaking. For example, projects linked Medicaid service utilization data with clinical outcomes data to more efficiently study treatment efficacy for emerging health threats, assess the causes and risk factors of different diseases and conditions, and study the relationship between inpatient and outpatient care. Other projects focused on linking Medicaid to other health and human services utilization data to improve maternal and infant health outcomes associated with chronic conditions, mental health, substance use disorder, and the use of health and social services.

For the primary Medicaid data source, most projects (n=7) used **T-MSIS data**, while the remaining four relied on **Medicaid data provided by individual states**. Projects linked a range of different data sources to Medicaid data (see Exhibit 2). Despite the availability of data from all 50 states, the District of Columbia, Puerto Rico, and the U.S. Virgin Islands, some projects only linked T-MSIS data for a subset of states.

Most projects linked Medicaid data to one other data source; two linked Medicaid data to more than one data source. The auxiliary data sources linked to Medicaid data, as shown in Exhibit 2, include clinical data (e.g., EHR data, case reports); federal and state population-based survey data; federal and state administrative sources (e.g., health care utilization, services, visit data); and vital records data (e.g., births and deaths). Predictably, most linkages to T-MSIS were with federal data sources, while most linkages to state Medicaid data were with state-based data sources.

Exhibit 2. Data Sources for OS-PCORTF Medicaid Data Linkage Projects

Medicaid Data Source	Linked Auxiliary Data Source(s)
T-MSIS	Birth certificates
	EHR data within the National COVID Cohort Collaborative Data Enclave
	National Ambulatory Medical Care Survey
	National Death Index
	National Hospital Care Survey
	National Survey of Children’s Health Longitudinal Cohort
	Office of Refugee Arrivals Data System
	Vital records
State Medicaid Data	Comprehensive Child Welfare Information Systems
	Emergency Medical Services electronic patient care reports
	National Core Indicators – Intellectual and Developmental Disabilities (NCI-IDD) In-Person Survey
	NCI-IDD In-Person Survey COVID-19 Supplement
	Pregnancy Risk Assessment Monitoring System
	Supports Intensity Scale – Adult Version

3.2 Data Linkage Process

Record linkage creates more comprehensive data for health outcomes research by combining information from different data sources related to the same person. Although data linkage can be complex and requires methodological expertise, linkage algorithms are well established and routinely implemented.[10] A critical first step in the data linkage process is selecting a method that effectively protects data privacy and ensures security. To create a linked dataset, most projects followed four steps: (1) selecting identifiers for linkage, (2) preparing data for linkage, (3) implementing the linkage, and (4) assessing the quality of the resulting linkage.

Selecting identifiers for linkage. Assessing the quality and completeness of available linkage identifiers helps data linkage teams select the correct identifiers for linkage. Many projects linked records using a combination (at least two) of the following identifiers: the individual's name, date of birth, Social Security number (SSN), sex, ZIP code, state of residence, birth certificate number, or data source specific identifiers.

Projects used one or more of the following Identifiers to link data sources

- **Personally identifiable information (PII)** that directly identifies an individual (ex: first and last name; Social Security Number) (n=8 projects).
- **Other identifiers** such as official numbers or codes assigned by entities like state agencies (ex: Medicaid ID) (n=8 projects).
- **De-identified tokens** that are unique, encrypted identifiers created from PII to facilitate privacy-preserving record linkage (n=1 project).

SSN was the most frequently used identifier to link Medicaid data to other data sources. Of the 11 projects, five used or planned to use SSNs to link data. While SSNs are known to provide high-quality linkages, they are not always available in all records or across data sources, and the quality of SSNs can vary.[11]

To address inconsistent availability of SSNs in the auxiliary data sources linked to state Medicaid data, **one project** developed a project-specific identifier that concatenated—or linked in a series or chain—the first three letters of the last name, plus the first two letters of the first name, plus the date of birth, plus the sex.

> *"There were about 400 different variables. They did quality checks on outliers, missing values, out of range values, [and] logic inconsistencies. [For example,] removing date days from their data so they just report month and year. They did all these harmonization things and looked for consistency in their variables in terms of names, and formats, and value ranges."*

Several projects leveraged unique Medicaid beneficiary IDs, either the state-assigned ID or the T-MSIS ID (MSIS ID), to facilitate linkage. However, using Medicaid IDs can present challenges, including the same MSIS IDs used for different beneficiaries in different states; multiple MSIS IDs assigned to the same beneficiary; and missing MSIS IDs. These challenges can make it difficult to link records for the same individual across time and across states. For example, one project that linked T-MSIS data to NDI data from multiple states encountered issues for beneficiaries who moved in and out of Medicaid coverage; beneficiaries who moved to a different state; and for infants assigned their mothers' MSIS ID while awaiting SSN assignment. These types of inconsistencies in beneficiaries' identifiers (e.g., new IDs assigned to the same person when they leave and rejoin Medicaid) can make it difficult to identify and follow a person's health care over time.

To address this challenge, and improve the accuracy of longitudinal Medicaid data linkages, CMS created the Eligible Identifiers Segment in 2020.[12] The Eligible Identifiers Segment records all a Medicaid and Children's Health Insurance Program beneficiary's different IDs, tracks when those identifiers where active, and provides an explanation for why an ID change (e.g., re-enrollment). This record segment helps CMS keep track of changes in beneficiary identifiers over time, making it easier for CMS and researchers to follow individuals consistently even when their IDs change.

Preparing data for linkage. To enable high-quality data linkages, project teams undertook measures to clean and prepare the data for linkage by understanding the data formats and structure of the data sources. This includes ensuring that data are formatted consistently across sources, particularly for data elements that may vary widely in format, such as dates (e.g., MM/DD/YYYY versus YYYY/MM/DD) or the order of first and last names (e.g., First Name, Last Name versus Last Name, First Name).

In some instances, researchers must harmonize the data by verifying comparable linkage variables across data sources. This step involves checking for consistent meaning of terms and concepts—known as semantic harmonization—and checking compatibility of data structure and formats—known as syntactic harmonization. One harmonization approach is using a **common data model (CDM)** for formatting and structuring data so information from different sources and applications can be exchanged, integrated, and analyzed consistently.

The CCOULD project team developed a CDM that maps variables from child welfare reports, foster care files, and Title IV-E services including:

- Child and caregiver identifiers
- Demographic variables
- Child welfare event and maltreatment variables
- Service receipt and intervention variables
- Foster care episode and placement variables
- Medicaid enrollment and claims data
- Child and caregiver risk factors
- Maltreatment perpetrator identification

To achieve semantic and syntactic data harmonization, three projects planned to or used CDMs to enable more effective and consistent linkages. The Dataset on Intellectual and Development Disabilities project is currently developing a CDM, while two projects leveraged CDMs to standardize state-level auxiliary data before linkage to Medicaid data. Of the two completed projects, the Child and Caregiver Outcomes Using Linked Data (CCOULD) project—linking child welfare records with Medicaid data in Florida and Kentucky—developed a CDM[13] to address the substantial heterogeneity in the underlying state-level data sources, which would have made it challenging to create a unified dataset absent the CDM. The project published the CDM to enable other states to "map," or code and structure, their data to the CCOULD CDM. This mapping supports future integration of additional states' data into the database of linked CCOULD data. The NC3 project mapped Medicaid data to the Observational Medical Outcomes Partnership[14] (OMOP) CDM to harmonize the claims data to the clinical data already harmonized to that CDM.

The N3C project team mapped Medicaid data to the following data elements found in the OMOP CDM:

- Demographic variables
- Observation period
- Visit occurrence (inpatient, outpatient, emergency, etc.)
- Diagnoses and medical conditions
- Drug exposure
- Procedures performed
- Clinical measurements (e.g., blood pressure)
- Death
- Device exposure

Project teams that leveraged a CDM described two key benefits. First, CDMs are useful to standardize highly heterogeneous data. For example, state programmatic data can vary widely across states (e.g., structures, formats, quality), in part due to different goals, information management systems, requirements, and staff resources. Second, CDMs support scalability by providing a "map" that others can use to help reproduce the linkage. For example, the CDM developed by the CCOULD project is being leveraged within HHS by another data linkage initiative. This effort is using the CCOULD CDM to promote data consistency for child welfare data that will be linked to Medicaid data to enhance service delivery, monitor spending, and conduct research on service outcomes.

> *"I don't want to undersell the multi-state component of this … states are very heterogeneous in terms of … the data systems that they have and also the organizational components of their individual states are all completely different … that's the real challenge with this."*

Implementing the linkage. Three types of linkage approaches were used across the projects—deterministic linkage, probabilistic linkage, and privacy-preserving record linkage (PPRL). Of the projects specifying a linkage method, four used a combination of deterministic and probabilistic matching; one used deterministic matching; one used probabilistic matching; and one used PPRL. Additionally, two projects incorporated machine learning techniques into the deterministic and probabilistic linkage process to improve linkage efficiency and accuracy.

Deterministic linkage is used when valid, high-quality unique identifiers are available, and this approach relies on exact or rules-based matching of specified information or fields between records to establish matches.[15] Probabilistic methods are useful when the data contain errors, missing values, or no unique identifiers. Probabilistic linkage estimates the likelihood that two records match through statistical theory.[16] The one project using probabilistic matching will leverage a unique approach developed by the U.S. Census Bureau called the Person Identification Validation System (PVS) to link data from the National Survey of Children's Health - Longitudinal Cohort to T-MSIS data. The PVS uses probabilistic matching based on sensitive PII, such as name, date of birth, SSN, and address to apply a unique person identifier that the Census Bureau uses to facilitate linkages across and within files for federal, commercial, census, and survey data.[17] This approach is effective because the Census Bureau has access to more PII than CMS and can link across datasets that do not share PII elements, which enhances the accuracy of data linkages.

Projects using a combination of deterministic and probabilistic approaches applied a phased approach to increase the total number of matched records by first conducting deterministic matching, followed by a round of probabilistic matching. For instance, in **the project** linking hospital survey data with T-MSIS claims data, the team began with deterministic linkage requiring an exact match using SSN and validating the linked pairs by comparing other identifying fields (i.e., first name, last name, date of birth, etc.).[18] Next, the team performed probabilistic matching using the other identifying fields besides SSN, retaining only those records that met the established match probability threshold.[19] Multiple probability thresholds were tested, and the threshold that minimized total linkage error was selected as the cutoff. Pairs were selected that were believed to represent the same individual between data sources. In another instance, the data quality assessment conducted by the project team to link NDI and T-MSIS data confirmed that a combination of deterministic and probabilistic linkage methods was necessary to address data gaps, particularly missing SSNs.

PPRL promotes increased privacy for records by transforming PII (e.g., names, birthdates) into encrypted tokens that can be matched across datasets without sharing PII. A key advantage of this method is the trust it builds among data contributors. PPRL involves a neutral third party (e.g., a linkage honest broker) that compares the encrypted (i.e., hashed data) data across sources to identify matches but is unable to view the PII.[20] One project aimed to enhance health outcomes research relevant to emergent health threats by linking claims data from Medicaid and Medicare with deidentified clinical EHR data in the National COVID Cohort Collaborative (N3C), a data enclave of more than 90 institutional data contributors across the nation. The project used PPRL techniques to enable N3C data contributors to retain control of their data—similar to when they contribute other data to N3C. Finally, while it did not use PPRL to create a linked dataset for public release, one project team explored using a commercial PPRL solution to link data and compared the PPRL results to results from a prior established linkage conducted using clear-text (unencrypted) PII that was considered to be a benchmark. The results of that project showing comparable results can increase opportunities and inform enhancements to the linkage infrastructure across HHS.

Assessing the quality of the resulting linkage. Of the six projects resulting in linked datasets, one project estimated Type I errors (when a true nonmatch is classified as a match) and Type II errors (when a true match is classified as a nonmatch) to evaluate the quality of matches, applying a cutoff threshold to optimize linkage accuracy. The validation process can also be performed in an iterative manner (rather than at the end of the linkage) to identify enhancements to the matching algorithm. For example, to validate the PPRL accuracy, the N3C project used a linkage honest broker to complete data validation for a 1% sample of CMS data before completing the entire linkage. Specifically, the broker compared a subset of reidentified hashed patients with identified patients in the state's health information exchange. The project also performed a second level of validation to assess the accuracy of using PPRL to supplement missing data from an individual's clinical information in the N3C. The quality of the PPRL was assessed on the basis of age, sex, date of death, within-site uniqueness, and cluster sizes.

3.3 Access to Linked Data

Five of the six linked datasets produced at the time of this report's publication were made available to HHS and external researchers for health outcomes research in a restricted-access data center, such as the NCHS Research Data Center (RDC) and the CMS Virtual Research Data Center (VRDC). Such centers provide secure access to data while ensuring data confidentiality. They can also support more efficient and timely data access compared to requesting physical data. For example, in the CMS VRDC, researchers can directly access approved datasets, perform analyses within the platform, and download aggregated results. However, project teams stated that these restricted-use environments are costly. In 2024, each CMS VRDC annual seat license cost $20,000 to $22,000, while project fees assessed per data use agreement (DUA) ranged from $15,000 to $18,000 annually.[21] The sixth completed **project** restricts access to its linked dataset—based on state Medicaid data and Pregnancy Risk Assessment Monitoring System (PRAMS) data—to state use for programmatic and surveillance activities. While none of the projects produced public-use files, one completed project produced a synthetic dataset using patient information from the restricted-access dataset

Many projects planned or have produced associated documentation for the linked dataset, such as data dictionaries, codebooks, user guides, technical documentation, or presentations to aid researchers in using the dataset. These resources can be found in **Appendix A**.

Two of the not-yet-completed projects also plan to make data available for restricted use by researchers. One plans to host the aggregate, deidentified dataset in a third-party restricted access data repository, for example the Inter-University Consortium for Political and Social Research.

3.4 Limitations or Challenges to Implementing Data Linkages

The challenges OS-PCORTF project teams highlighted focused specifically on issues related to data quality, administrative hurdles to data access, and limited staff capacity.

> *"The administrative challenges are so much with these linkages. That's the hard part. The technical stuff is … doable."*

Data quality and consistency issues. Because Medicaid is administered at the state level, variability in state data collection and reporting impact both federal data sources like T-MSIS and state Medicaid data that is submitted to T-MSIS. This poses challenges to harmonizing data within T-MSIS and across states. Additionally, differences in available data exist depending on when states migrated from MSIS submissions to T-MSIS. In T-MSIS, there are inconsistencies in available identifiers across states and within a state over time. For example, some states do not collect and report enrollee demographics when Medicaid covers only specific limited services, and states use different methods to report labor and delivery services provided to women and newborns.[22] One project noted that the time to release of final T-MSIS data compared to preliminary data had downstream impacts on its project timeline. Project teams also reported data quality (e.g., data missingness) challenges within the auxiliary data sources linked to Medicaid data.

Difficulty navigating data access. Project teams described legal barriers navigating various regulatory requirements and DUAs to securely access data, which can take significant time and effort. Even when existing data exchange agreements are in place with CMS, executing project-specific DUAs can take several months. This challenge is amplified when linking Medicaid data across states or between different state agencies. Each data owner's legal analysis must be interpreted and aligned with others, requiring agreement among all parties to move forward. Five projects identified administrative and legal challenges as the most significant, substantially impacting the overall project timeline. In terms of data governance, project teams noted that setting up DUA and memorandums of understanding (MOU) can take anywhere from a few months to a few years. Federal agencies familiar with the data access process (e.g., NCHS Data Linkage Program project teams)—and those with existing DUAs that only need extensions—typically moved faster. However, requests for nonstandard T-MSIS data can extend timelines. Similarly, projects linking state-level data to Medicaid data (including both state-based and T-MSIS data) often face longer timelines. Project teams advised accounting for additional time to navigate the data request process.

> *"Privacy laws and data sharing was really a challenge. Just getting to a place of establishing a data use agreement … took us 2 years with [data owner], from the initial call to the finalization of the DUA. Similarly, with [data owner], almost about 2 years. So, it was a long road to get to a data sharing agreement. Once we had it, the linking was fast."*

Siloed data governance and data infrastructures. The data linked to Medicaid data are often stored in fragmented, disparate systems, each with their own governance structures. They also require personnel with specialized knowledge of the data source to explain data elements and infrastructure to the data linkage team. Project teams noted that individual state agencies often lacked existing relationships and familiarity with each other's program data, complicating coordination and data sharing. Additionally, the information technology underlying state programs vary in sophistication, which can make it technically difficult to access and extract data to support ongoing data linkage efforts.[23]

Limited state-level capacity for conducting linkages. Multiple project teams discussed challenges related to limited state staff capacity for data linkage efforts. For example, one project wanted to use probabilistic matching to yield more linked data but was limited by the state's resources and time span for the project.[24] Another project team encountered staff capacity challenges because auxiliary data owners had limited experience with Medicaid data, making it difficult to assess linkage feasibility based on the data fields available, which impeded linkage efforts.[25] Staff turnover also contributed to knowledge continuity issues and an insufficient workforce with dedicated time available for the linkage project.[26]

3.5 Best Practices to Facilitate Data Linkage and Opportunities to Build Medicaid Data Linkage Capacity

The assessment identified several best practices for data linkage efforts that future project teams can follow, as well as future opportunities to advance Medicaid data linkage capacity.

Streamline governance processes for data sharing and acquisition. Data sharing across multiple entities requires a significant amount of time and collaboration with data source owners. Data sharing can be enhanced by encouraging data owners to improve their DUA processes to minimize time and resource barriers. Many project teams recommended factoring in enough time before data linkage, particularly when working with multiple actors that need to review and approve legal agreements. This administrative work includes time for transferring funding to partnering agencies, identifying and gaining buy-in from the appropriate contacts for each data source (factoring in possible changes in leadership), establishing DUAs and MOUs, and acquiring the data. When working with state data, it is also essential to engage a stakeholder or team member who understands the specific state laws and regulations for data sharing in addition to federal laws.

> *"If somebody uses [tool], somebody else uses [tool], and so on, then I have to put in 20 PPRL tools and that's probably $200,000 a pop, which makes the data cost prohibitive and people can't request it…Ideally, there's some way the government could develop…a tool internally for the government that other people could then use and that would be the better solution instead of some commercial or proprietary product."*

Some project teams were working with federal agencies to find better models for data sharing for their unique requests, such as acquiring updated files annually. The Research Data Assistance Center's DUA instructions were highlighted as an exemplary model for providing instructions to data requesters. New pricing models for accessing and acquiring T-MSIS data could also be particularly helpful. One project considered using T-MSIS data instead of individual state data but did not have the budget to do so, and it was unclear whether the team would be able to create a public-use file of the linked dataset. One project team also advocated for establishing universal data agreements or other mechanisms to promote sharing data at a national level.[27]

Advance secure PPRL techniques. While robust PPRL algorithms are available and can be leveraged for a range of linkage efforts, project teams also identified opportunities to fully realize their potential and support broader, scalable implementation. For example, one project team noted that internally managing the tokenization process rather than relying on an external partner could improve efficiency and control. Another project team suggested a future OS-PCORTF project to develop a standardized tokenization tool for cross-agency linkages, reducing the costs of relying on disparate commercial solutions and promoting consistency across federal efforts.

Document linkage processes to support replication. Many project teams recommended maintaining records or notes and developing technical assistance resources to improve the efficiency of future projects. For instance, to facilitate the transfer of funds between multiple agencies, one project team documented prior processes to share with financial staff unfamiliar with the process. Project teams can also develop a data linkage roadmap or user guide to outline the steps and considerations for implementing or replicating the data linkage process. For instance, one project team plans to create a roadmap of approaches for identifying and partnering with other states that use National Core Indicators data; establishing DUAs and navigating regulatory requirements; securely transferring and storing data; and maintaining the data. Another project created a framework to guide other projects in linking PRAMS with administrative data such as Medicaid claims.[30] Future linkage endeavors can also leverage the CDMs developed by some projects to harmonize data in preparation for linkage.

Project teams produced resources that can be leveraged by other projects to facilitate data sharing, such as a **legal framework** that provides guidance for sharing data between state child welfare agencies and state Medicaid agencies[28] and **data governance documentation** for the entities involved in a PPRL project.[29]

Provide datasets in multiple file formats to support various research needs. Most linked Medicaid datasets that were produced are restricted-use files that require an application process and DUA for data access and use. The identifying variables in these restricted-use files are often necessary to conduct researchers' analyses. However, aggregate and deidentified data can also be used to answer health outcomes research questions. Recognizing different research needs, some projects planned for different levels of access to the linked data. One project made three file types available to researchers—a limited dataset, a deidentified dataset, and a synthetic dataset. Although access requires an application, this approach offers researchers flexibility in data access and use. Additionally, one project was considering producing a public-use file, and another project was considering developing a publicly available summary dashboard but encountered cost constraints. These alternate data files, if feasible, may increase the utilization of linked datasets.

Expand the utility of linked datasets as foundational resources for studying a range of populations. One promising direction is using the N3C dataset to expand the use cases for data linkages. The team developed the dataset originally for research on emergent health threats and was planning to leverage the existing infrastructure to other clinical domains, such as renal disease and cancer. Moreover, the OS-PCORTF has funded three additional studies using N3C EHR data that will serve as reference architecture for data linkages. Additionally, one project identified a strong infrastructure for child welfare data linkages at the state level, with two-thirds of states already conducting linkages between child maltreatment data and Medicaid eligibility files. This finding prompted a new project to link data in additional states and develop more lasting infrastructure. Following these examples, other linked datasets could potentially leverage existing infrastructure to broaden their scope to other areas.

> *"There's usually not stakeholder engagement sessions at the jump, and I think particularly with populations like intellectual and developmental disabilities where there's been misuse of data related to the disability population, that is really critical to get that buy-in and get that reassurance and credibility."*

Engage key stakeholders, including patients, early and throughout the data linkage process. Project teams described the importance of engaging key stakeholders in data linkage efforts. These stakeholders include institutional review board staff, communications staff, staff with experience establishing DUAs, finance staff with experience transferring funds to appropriate parties, a statistical programmer for the data linkage, a statistician who understands probabilistic models, and a subject matter expert for the data sources used for the linkage. Project teams also emphasized the importance of gaining buy-in and credibility among the population that the dataset will serve. Several projects engaged technical expert panels (TEPs) of federal and nonfederal stakeholders, and they recommended communicating with TEP members periodically between convenings. One project team held stakeholder engagement sessions with targeted end users (e.g., state decision-makers, people with ID/DD, their caregivers, and researchers) at the beginning of the project to increase awareness of the project, listen to community concerns, and improve the likelihood of state participation. Another project team was exploring the possibility of creating a portal to enable patients to see which datasets their data are used in, and to either consent or remove their data from the datasets. Such practices can help improve patient awareness and trust in research datasets

Continue to expand capacity for data linkage infrastructure at the state and federal levels. To address the capacity limitations described by project teams, investments need to be made to modernize technology and train state staff with specialized skills and knowledge. For instance, one project had to rely on deterministic linkage methods due to the team's capabilities. However, training internal staff on probabilistic matching could improve linkage efficiency and linkage match rates. To build internal knowledge, project teams encouraged more collaboration and information sharing across federal and nonfederal stakeholders, such as through communities of practice or consortia to share best practices and lessons learned. In addition to staffing, project teams learned that it was important to financially compensate data partners such as state agencies to drive participation in linkage projects, while acknowledging that the funding may not be sufficient to compensate for linkage activities and that participation is primarily motivated by the dataset's utility for informing programmatic decisions.

> *"Finding ways to just leverage that experience and knowledge is really key because this is a really complicated effort...So even thinking about developing a community of practice of some sort where some of these experiences and some of these helpful tips and forms of guidance can be housed would be extremely valuable for those that are looking to do this work."*

3.6 Impact of Linked Medicaid Data for Health Outcomes Research

The six project teams that completed a linked dataset used several approaches to measure uptake and dissemination of the linked datasets, including tracking citations via Zotero, Altmetric scores, and data access requests through the hosting data center. Across five of the projects, key informants noted that **over 100 studies** have used these Medicaid-linked datasets for research that provides more complete longitudinal profiles of patients; captures care received across different providers and systems; evaluates the effectiveness of interventions or policy changes; and enables advanced statistical methodologies.

Project teams also described the value of the linked datasets in terms of additional health outcomes research questions that can be studied. They described how Medicaid claims data alone can provide information on the types of services received but not additional context to understand possible causes of disease or effective preventive strategies to inform health care decisions. By linking Medicaid to survey data, patient-reported outcomes, clinical data, and other administrative datasets, researchers can explore a wide array of innovative research questions (Exhibit 3).

Exhibit 3. New Health Outcomes Research Questions Enabled by Medicaid Data Linkages

Research Topic	New Information Provided
Health services and quality improvement	Prior to linkages, there was little understanding of which Medicaid funding allocations produced better outcomes for Medicaid beneficiaries. Linking health and human services data reveals how individuals interact with multiple systems, highlighting gaps and redundancies in care. For example, the linkage to ID/DD surveys will answer service level questions regarding ID/DD cost containment and quality improvement, and the linkage to child welfare data can assess which Title IV-E prevention services are used most often among child welfare-involved families and how those services are related to child health outcomes.
Maternal and infant health	Medicaid claims often lack detailed information on pre- and post-natal care trajectories and, importantly, the relationship between maternal and infant outcomes. Linking Medicaid to vital records, EHR data, and survey data offers insights into the effectiveness of Medicaid coverage for pregnant women and information on parent-child dyads that was previously separated. For example, linking pregnancy risk assessments with clinical outcomes data can answer questions such the percentage of women who self-report symptoms of postpartum depression and received mental health services, or the pre-birth familial stressors and injury-related health care visits occurring during the first two years of the infant’s life.
Methodological advancements	Not only do linked datasets improve the scientific validity of data, resulting in better consistency and effect estimates compared to individual datasets, they also drive advances in research methods, such as developing validated prognostic risk models. Large-scale datasets also enable the use of advanced analytics, such as machine learning, to generate new insights.
Rare diseases	Aggregating data from multiple sources increases the sample size for rare conditions, which facilitates research that would not be possible with siloed datasets, enabling better surveillance.
Social and behavioral factors	Integrating health data with information on social services, education, and environment provides a more nuanced understanding of how social factors drive health outcomes. Researchers can examine longitudinal data on child development, answering questions such as causal pathways between health services and treatment and educational outcomes. They can also study the relationship between caregiver substance use disorders or mental health conditions and child welfare engagement.

3.7 Using Linked Medicaid Datasets for Policy Insights

Medicaid data linked with other data sources enhances the ability of researchers and decision-makers to address critical national policy priorities, such as chronic disease prevention and management, pregnancy outcomes, and youth behavioral health. For example, studies using these linked datasets have examined anxiety, depression, and trauma-related symptoms among children; state actions to prevent and mitigate adverse childhood experiences; the effectiveness of treatments such as steroids and antivirals; and diabetes and prediabetes surveillance and prevention efforts.

Linked Medicaid datasets can also enable policy-relevant analyses of the relationship between Medicaid enrollment, healthcare utilization and beneficiary health outcomes. For example, one study leveraging the CCOULD dataset examined the timing of children’s Medicaid enrollment relative to their engagement in child welfare systems to identify program enhancements and ensure families receive the supports they need.[31] Another CCOULD study explored data on children entering foster care through custody relinquishment to help states and localities identify prevention practices to better serve children's behavioral health and disability needs without relying on foster care .[32] In addition, the University of Wisconsin-Madison Institute for Research on Poverty has funded several extramural policy-relevant research studies using the CCOULD dataset, such as examining opportunities to support families with substance use disorder and understanding long-term service trajectories of children in residential care.[33]

These linked datasets can also facilitate economic analyses related to service utilization, spending, economic burden to patients, and workplace productivity. Projects have leveraged the NDI-Medicaid dataset to assess the impact of changes to the child tax credit on the behavioral health of adolescents enrolled in Medicaid, and to identify the economic burden of Parkinson's disease. Researchers interested in using the CCOULD dataset, researchers will be able to explore questions like: "How much does Medicaid spend on substance use disorder treatment among parents with children in foster care?" or "Does spending on parent substance use disorder treatment result in subsequent savings to Medicaid?"[34]

4. Conclusion

Patients, caregivers, clinicians, and decision makers rely on high-quality evidence to assess relationships between interventions and outcomes to make informed health care decisions. However, individual datasets often lack complete, longitudinal information on individuals due to fragmented and siloed data systems. Linked Medicaid datasets have a profound impact by enabling richer, more accurate health services research and evidence-based policy development. The purpose of this project was to qualitatively investigate the OS-PCORTF Medicaid data linkage projects to document their linkage approaches, understand challenges encountered, and identify best practices. Looking ahead, the Medicaid linkage data infrastructure could be improved by streamlining data sharing and governance processes, advancing tokenization tools to securely exchange PII across agencies, broadening the scope of existing linked datasets, and building state and federal capacity to conduct data linkages.

Appendix A: Overview of OS-PCORTF Linked Medicaid Datasets

This table summarizes information on the 11 projects described in this report.

Project Title (hyperlink to project page on ASPE website)	Agency	Project Status	Linked Dataset Description	Data Sources	Linkage Method	Data Linkage Identifiers	Dataset Access Description	Publicly Available Resources for Researchers
Creating a Federal COVID-19 Longitudinal Patient Outcomes Research Database Linked to Health Systems and Clinical Data	National Institutes of Health (NIH)/National Center for Advancing Translational Sciences (NCATS)	Completed with a linked dataset	The PPRL-CMS national dataset contains longitudinal, de-identified billing data from CMS and clinical electronic health record (EHR) data from the NCATS National COVID Cohort Collaborative (N3C) for N3C researcher use.	▪ Transformed Medicaid Statistical Information System (T-MSIS) Analytic File (TAF) including inpatient, long-term care, other service and prescription claims ▪ N3C Data Enclave	The project team used privacy-preserving record linkage (PPRL).	▪ NC3 person identifier ▪ Global Person identifier for patients represented in multiple N3C sites ▪ Data partner identifiers ▪ Other variables from CMS records are matched to the Observational Medical Outcomes Partnership (OMOP) domain table	**Restricted Use:** Researchers can request access to three file types (limited dataset, deidentified dataset, and synthetic dataset). Access can be requested through the N3C Data Enclave for approval by the Data Access Committee.	▪ **Final Report** ▪ **N3C PPRL-CMS dataset** ▪ **PPRL Enrichment Dashboard** ▪ Methodology Report: **Augmenting the N3C Dataset with CMS Data, Secure and Deidentified Clinical Dataset**

Project Title (hyperlink to project page on ASPE website)	Agency	Project Status	Linked Dataset Description	Data Sources	Linkage Method	Data Linkage Identifiers	Dataset Access Description	Publicly Available Resources for Researchers
Data Linkage: Evaluating Privacy Preserving Record Linkage Methodology and Augmenting the National Hospital Care Survey (NHCS) with Medicaid Administrative Records	Centers for Disease Control and Prevention (CDC)/National Center for Health Statistics (NCHS)	Completed with a linked dataset	The linked NHCS-CMS Medicaid dataset includes longitudinal, patient-level record linkages of NHCS hospital administrative claims and EHR data to T-MSIS data.	▪ T-MSIS data ▪ NHCS	The project team used deterministic and probabilistic record linkage incorporating machine learning techniques.	▪ Social Security number (SSN) ▪ First and last name, middle initial ▪ Date of birth (month, day, year) ▪ Sex ▪ Zip code ▪ State of residence	**Restricted Use**: Researchers can request access through the NCHS and Federal Statistical Research Data Centers (RDCs)	▪ Final Report ▪ NHCS-CMS Linked Dataset ▪ NHCS-CMS T-MSIS Data Dictionaries ▪ Methodology Report: The Linkage of the 2016 National Hospital Care Survey to 2015–2017 Centers for Medicare & Medicaid Services Transformed Medicaid Statistical Information System Claims Data: Matching Methodology and Analytic Considerations

Project Title (hyperlink to project page on ASPE website)	Agency	Project Status	Linked Dataset Description	Data Sources	Linkage Method	Data Linkage Identifiers	Dataset Access Description	Publicly Available Resources for Researchers
Developing a Multi-State Network of Linked Pregnancy Risk Assessment Monitoring System (PRAMS) and Clinical Outcomes Data	CDC	Completed with a linked dataset	Three of the PRAMS Learning Community teams (Alaska, Montana, South Dakota) linked PRAMS data to state Medicaid claims and clinical outcomes data. The Alaska dataset is longitudinal.	**Alaska:** ■ Medicaid child injury claim data ■ Alaska Child Abuse Linkage Project (ALCANLink) **Montana:** ■ Montana's Program for Automating and Transforming Healthcare (MPATH) ■ Vital records (Birth Certificate and Death Certificate) ■ Newborn screening ■ Child maltreatment reports ■ Home visiting services ■ Children's special health services **South Dakota:** ■ Medicaid claims data ■ PRAMS	The state project teams used the following linkage methods: **Alaska:** Deterministic and probabilistic **Montana:** Deterministic and probabilistic **South Dakota:** Deterministic	**Alaska**: ■ First name ■ Last name ■ Middle name ■ Date of birth ■ Sex ■ Year of birth (exact match needed) **Montana**: ■ Birth certificate number ■ Infant first name ■ infant last name ■ Infant date of birth ■ Infant sex ■ Mother maiden/ last name **South Dakota**: ■ Mother SSN ■ Mother first name ■ Mother last name	**Restricted Use:** None of the resulting datasets are available to external researchers	■ **Final Report** ■ **A Framework for Linking PRAMS with Administrative Data**

Project Title (hyperlink to project page on ASPE website)	Agency	Project Status	Linked Dataset Description	Data Sources	Linkage Method	Data Linkage Identifiers	Dataset Access Description	Publicly Available Resources for Researchers
Enhancing Data Resources for Researching Patterns of Mortality in Patient-Centered Outcomes Research (Project 3)	Centers for Medicare & Medicaid Services (CMS)	Completed with a linked dataset	The Medicaid Enrollee Supplemental File (MESF): National Death Index (NDI) Segment includes longitudinal data on cause and manner of death from death certificates linked with Medicaid decedents data from 2007-2015.	▪ Medicaid enrollment and claims data ▪ NDI	The project team used deterministic and probabilistic record linkage.	All NDI data fields except SSN	**Restricted Use:** The Restricted Research Identifiable File (RIF) is only available through the Virtual Research Data Center (VRDC). Researchers can request access through the Research Data Assistance Center (ResDAC).	▪ **Final Report** ▪ **MESF-NDI Linked Dataset** ▪ **CMS Bridge File**, information for which is available in the TAF User Guide, section 5A, footnote 29

Project Title (hyperlink to project page on ASPE website)	Agency	Project Status	Linked Dataset Description	Data Sources	Linkage Method	Data Linkage Identifiers	Dataset Access Description	Publicly Available Resources for Researchers
Linking State Medicaid and Child Welfare Data for Outcomes Research on Treatment for Opioid Use Disorder and Other Behavioral Health Issues	Office of the Assistant Secretary for Planning and Evaluation (ASPE) Office of Human Services Policy (HSP) and the Administration for Children and Families - Office of Planning Research and Evaluation (ACF/OPRE)	Completed with a linked dataset	The Child and Caregiver Outcomes Using Linked Data (CCOULD) is a longitudinal, deidentified dataset linking records from child welfare data systems with Medicaid enrollment and claims data in Florida and Kentucky.	▪ State Medicaid data: enrollment files, inpatient claims, outpatient and other service claims, and prescription drug claims ▪ Comprehensive Child Welfare Information Systems data on child welfare and foster care episode data and Title IV-E services provided	The project team used deterministic record linkage.	▪ SSN ▪ LongID: concatenation of the first three letters of child or caregiver's last name plus the first two letters of the child or caregiver's first name plus the child or caregiver's date of birth (as DDMMYYYY) plus sex (M/F)	**Restricted Use**: The dataset and accompanying documentation are available through the National Data Archive on Child Abuse and Neglect at no cost.	▪ Final Report ▪ CCOULD Linked Dataset ▪ CCOULD Dataset User Guide ▪ CCOULD Project Common Data Model ▪ Linking Child Welfare and Medicaid Data: Lessons Learned from Two States ▪ Toolkit: Data Sharing for Child Welfare Agencies and Medicaid ▪ Issue Brief: Advancing Research on Intersections of Child Welfare and Medicaid Using Linked Data from the CCOULD Project

Project Title (hyperlink to project page on ASPE website)	Agency	Project Status	Linked Dataset Description	Data Sources	Linkage Method	Data Linkage Identifiers	Dataset Access Description	Publicly Available Resources for Researchers
Linking Federally Qualified Health Center (FQHC) EHR and Medicaid Data to Understand Maternal Health Care	CDC/NCHS	Completed with a linked dataset	This project will develop and evaluate a linked longitudinal dataset of CY2021 National Ambulatory Medical Care Survey (NAMCS) and EHR data from FQHCs with the T-MSIS administrative data.	▪ T-MSIS ▪ NAMCS FQHC EHR data	The project team plans to use deterministic and probabilistic record linkage incorporating machine learning techniques.	▪ SSN ▪ Medicaid IDs ▪ Medicare Beneficiary Identifier	**Restricted Use:** The resulting dataset will be available through the NCHS and Federal Statistical RDCs.	▪ **NAMCS Health Center Component-CMS Medicaid Linked Dataset** ▪ **Technical Report:** Linkage of the 2021 NAMCS HC Component to 2020-2022 T-MSIS Data: Linkage Methodology and Analytic Considerations ▪ **Data Dictionaries:** Match Status, Demographic and Eligibility, Inpatient, Long-term Care, Pharmacy, Other Services

Project Title (hyperlink to project page on ASPE website)	Agency	Project Status	Linked Dataset Description	Data Sources	Linkage Method	Data Linkage Identifiers	Dataset Access Description	Publicly Available Resources for Researchers
Data Infrastructure Supporting Research on Refugee Medicaid Service Utilization and Outcomes	ASPE HSP and ACF	Ended	This project planned to create a longitudinal dataset with new linkages between the Office of Refugee Arrivals Data System (RADS) and T-MSIS data.	■ TAF Annual Demographic and Eligibility file ■ RADS, containing information on refugee demographics and services provided	The project team planned to use probabilistic record linkage.	■ Date of birth ■ Gender ■ Immigration status ■ Enrollment year ■ State ■ Race and Ethnicity Other demographic information	The project did not result in a linked dataset.	None available at the time this report was written.
Multistate Emergency Medical Services (EMS) and Medicaid Dataset (MEMD): A Linked Dataset for Patient-Centered Outcomes Research	ASPE- Office of Behavioral Health, Disability, and Aging Policy (ASPE/BHDAP)	Ended	This project aimed to develop MEMD, a longitudinal dataset containing linked Medicaid and EMS records from up to five states.	■ State Medicaid data ■ EMS electronic patient care reports	*Not specified in the project information collected*	*Not specified in the project information collected*	The project did not result in a linked dataset.	None available at the time this report was written.

Project Title (hyperlink to project page on ASPE website)	Agency	Project Status	Linked Dataset Description	Data Sources	Linkage Method	Data Linkage Identifiers	Dataset Access Description	Publicly Available Resources for Researchers
Quality of Care and Outcomes Data for Pregnant Medicaid Beneficiaries and Newborn Infants	CMS	Ended	The project planned to create a longitudinal dataset with linked vital records and birth certificates for Medicaid covered pregnancies and births with claims data on mothers and infants for four states.	▪ TAF ▪ Vital records birth certificates	The project team planned to use probabilistic record linkage.	SSN where available. If missing, researchers planned to use mother-infant identifiers from birth certificate data.	The project did not result in a linked dataset.	None available at the time this report was written.

Project Title (hyperlink to project page on ASPE website)	Agency	Project Status	Linked Dataset Description	Data Sources	Linkage Method	Data Linkage Identifiers	Dataset Access Description	Publicly Available Resources for Researchers
Dataset on Intellectual and Developmental Disabilities: Linking Data to Enhance Person-Centered Outcome Research	ASPE/BHDAP	Ongoing	This project will produce an integrated Dataset on Intellectual and Developmental Disabilities (DIDD), a multi-state dataset enabling researchers to analyze relationships between support needs, service utilization, Medicaid expenditures, and person-centered outcomes for the ID/DD population.	▪ State Medicaid data ▪ State-level data sources, including Support Intensity Scale – Adult Version scores, National Core Indicators – Intellectual and Developmental Disabilities (NCI-IDD) In-Person survey data, and NCI-IDD In-Person survey COVID-19 Supplement	The project team plans to use deterministic record linkage.	▪ SSN ▪ First name ▪ Last name	**Restricted Use:** A de-identified dataset is expected to be available to researchers via a third-party data repository, which has not been selected yet.	None available at the time this report was written.

Project Title (hyperlink to project page on ASPE website)	Agency	Project Status	Linked Dataset Description	Data Sources	Linkage Method	Data Linkage Identifiers	Dataset Access Description	Publicly Available Resources for Researchers
Linking the National Survey of Children's Health Longitudinal Cohort (NSCH-LC) and Medicaid Claims Data	Health Resources and Services Administration, Maternal and Child Health Bureau (HRSA/MCHB)	Ongoing	This project will create a longitudinal dataset that provides a source of pediatric health services and treatment data linked to longitudinal data on child development, academic performance, school disruptions, mental health, caregiver health, and social factors.	- T-MSIS - NSCH-LC Study	The project team plans to use deterministic and probabilistic record linkage	- Protected Identification Key (PIK), which uniquely identifies a record based on PII such as name, date of birth, SSN and address - If neither are available, then T-MSIS IDs will be used	**Restricted Use:** Information for researchers interested in the dataset will be made available when the file is published.	None available at the time this report was written.

Appendix B: Key Informant Interview Questions

The following interview protocol was tailored to each Medicaid data linkage project using information from the document review.

1. To begin, could you please tell us your name and position, and provide a quick overview of your [PROJECT NAME]?
    a. [Ask only if unclear from data abstraction] Please specify which Medicaid data files are being (will be) linked.
    b. [Ask only if unclear from data abstraction] Please clarify whether the data linkage generates (will generate) *longitudinal* data.

I want to now ask you a few questions about the data and methods used for the data linkages.

2. [For projects using state-level Medicaid data] Why did you go the route of requesting data from individual states instead of using data from T-MSIS? What do you see are the advantages or disadvantages of acquiring data from individual states versus T-MSIS?
   [For projects using T-MSIS] We understand that some researchers request data from individual states even though that same data is available within T-MSIS. What governed your decision to use T-MSIS? What do you see are the advantages or disadvantages of acquiring data from T-MSIS versus individual states?
3. Please describe the approach/methods used to link [DATASET NAME].
4. Please describe any data quality or validation checks the team performed to optimize linkage accuracy and precision?
    a. Were any specific analytic tools used in data quality checks (e.g., algorithms)? If yes, please describe.
    b. Were there any specific limitations of these tools? If yes, please discuss.
    c. [If project results in longitudinal data] Are there specific considerations about accuracy and precision for the longitudinal nature of these linked data?
    d. How were findings from data quality checks used to refine the linkage methodology?

Moving on to challenges the project team encountered to linking Medicaid data.

5. What, if any, challenges did the project team encounter while linking data?
    a. To what extent are these challenges specific to the linkage method used or the specific datasets used, versus apply more broadly to any linkage method/datasets?
    b. [If using T-MSIS data for linkage]
        1. What data governance-related challenges did your team face when accessing T-MSIS data?
        2. Did you encounter restrictions on how T-MSIS data could be linked with other datasets?
        3. Were there any unexpected requirements or administrative hurdles to requesting access to T-MSIS data from CMS?
            - Were state-based approvals required prior to the linkage?
            - What strategies did your team use to navigate these hurdles?
    c. [If using state Medicaid data for linkage]
        1. What governance-related challenges did your team face when requesting state-level Medicaid data?

        2. Was there variation in the approvals, DUAs, or governance structures across states that complicated the data acquisition and/or linkage process?
        3. What strategies did your team use to navigate state-level data access restrictions?
    d. To what extent did these challenges make you change your original approach or affect the scope of the final or planned linked dataset? For example, did you have to adjust your matching criteria or modify your analytic plans?
6. What solutions did you implement to address the data linkage challenges you encountered?
    e. To what extent are these solutions reproducible by other researchers interested in linking Medicaid data to other data sources?
    f. [If project results in longitudinal data] Are there specific considerations (on these solutions) for future longitudinal linkages?

Thinking now about analytic documentation that was (will be) produced from this project.

7. We gathered from our review of project materials that other researchers will have access to [describe data/methods documentation/artifacts or tools]. Would you like to add anything to this list?
   [If we didn't find any information about documentation] What documentation on the data and methods, or data artifacts/tools, will other researchers have access to, from this project?
8. What analysis has (will) the project team performed with the linked dataset?
9. How can researchers access the dataset?
    a. To what extent, that you know of, has the dataset been used by stakeholders, and how (e.g., researchers, policymakers)?
    b. We are interested in documenting research use cases for the dataset. Is there an opportunity for our team to receive a summary of received and approved data access requests?
    c. Is this dataset included on data.gov?

We would now like to understand your project's impacts or your perceived impacts of the project/resulting data.

10. Can you talk about the value of the linked dataset and the kind of PCOR questions that can be addressed that perhaps were not possible or difficult to assess prior to the linkage?
11. What has been the broader impact of the linked dataset? For example, enhanced data infrastructure or improved collaboration?
12. From an economic perspective, what cost savings or efficiency gains do you anticipate this linked dataset could provide for future linkage efforts?
13. How has (can) the project team measured impact and use of the resulting linked data and other artifacts?

Finally, I would like to ask you about lessons learned and future recommendations related to linking Medicaid data.

14. What strategies and opportunities to strengthen data capacities in Medicaid data linkages did your project identify?
    a. What incentives, or facilitating factors, need to be in place for encouraging other researchers to invest in such linkage efforts?
    b. Based on the work done in your project, what future work do you think the OS-PCORTF can support in this area?
15. Beyond this specific project, what opportunities do you see for leveraging this Medicaid linkage methodology with different datasets or for different analytical purposes? (e.g., social service data, nutrition data)?

    a. What types of linkages with Medicaid data do you think researchers should prioritize, more generally?

16. In general, what were the most important lessons learned, best practices, or resources that other researchers could consider when linking to Medicaid data?
17. And finally, did this project engage/will this project engage with patients in any way? If yes, please describe the patient engagement activities.

References

[1] Office of the Assistant Secretary for Planning and Evaluation. Office of the Secretary's Patient-Centered Outcomes Research Trust Fund (OS-PCORTF): Strengthening Data Infrastructure for Patient-Centered Outcomes Research (PCOR). U.S. Department of Health and Human Services. **https://aspe.hhs.gov/patient-centered-outcomes-research-trust-fund**.

[2] Office of the Assistant Secretary for Planning and Evaluation. Office of the Secretary Patient-Centered Outcomes Research Trust Fund (OS-PCORTF). **https://aspe.hhs.gov/collaborations-committees-advisory-groups/os-pcortf**

[3] The White House, Make America Healthy Again Commission. (2025, February 13). *Make Our Children Healthy Again: Assessment*. The White House. **https://www.whitehouse.gov/wp-content/uploads/2025/05/MAHA-Report-The-White-House.pdf**

[4] Predmore Z, Heins S, Hoch E, Baxi S, Grigorescu V, Smith S. State Experiences Linking Medicaid Data With Birth Certificates and Other Data Sources. Med Care. 2023 Jun 1;61(6):353-359. doi: 10.1097/MLR.0000000000001853. Epub 2023 Apr 12. PMID: 37068021; PMCID: PMC10168110.

[5] Dost M, Ghertner R, Madden E. Linking Health and Human Services Data Can Empower Patient Decisions and Increase Efficiency. Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services; March 2025. **https://aspe.hhs.gov/sites/default/files/documents/a4aef2b535dfc8a5a5571912e35c51a6/aspe-brief-human-services-health-data-linking.pdf**.

[6] Office of the Assistant Secretary for Planning and Evaluation. Improving Quality Of Care & Outcomes Data For Pregnant Medicaid Beneficiaries & Newborn Infants. **https://aspe.hhs.gov/quality-care-outcomes-data-pregnant-medicaid-beneficiaries**

[7] Office of the Assistant Secretary for Planning and Evaluation. Data Linkage: Evaluating Privacy Preserving Record Linkage Methodology And Augmenting The National Hospital Care Survey With Medicaid Administrative Records. **https://aspe.hhs.gov/data-linkage-evaluating-privacy-preserving-record-linkage-methodology-augmenting-national-hospital**

[8] Office of the Assistant Secretary for Planning and Evaluation. Enhancing Data Resources For Researching Patterns Of Mortality In Patient-Centered Outcomes Research. **https://aspe.hhs.gov/enhancing-data-resources-researching-patterns-mortality-patient-centered-outcomes-research**

[9] Nowell LS, Norris JM, White DE, Moules NJ. Thematic Analysis: Striving to Meet the Trustworthiness Criteria. *Int J Qual Methods*. 2017:16(1). **https://doi.org/10.1177/1609406917733847**

[10] Dusetzina SB, Tyree S, Meyer AM, et al. Linking Data for Health Services Research: A Framework and Instructional Guide. Rockville (MD): Agency for Healthcare Research and Quality (US); 2014 Sep. 4, An Overview of Record Linkage Methods. **https://www.ncbi.nlm.nih.gov/books/NBK253312/**

[11] Mark T, Dolan M, Allaire B, Bradley C. Linking Child Welfare and Medicaid Data: Lessons Learned from Two States Child and Caregiver Outcomes Using Linked Data (CCOULD). Office of Planning, Research, and Evaluation, Administration for Children and Families; 2022. **https://acf.gov/sites/default/files/documents/opre/ccould_lessons_learned_march2023.pdf**

[12] Centers for Medicare & Medicaid Services. CMS Guidance: Reporting Eligible Identifiers in T-MSIS. Medicaid.gov. Published September 25, 2020. Accessed August 21, 2025. **https://www.medicaid.gov/tmsis/dataguide/v3/t-msis-coding-blog/cms-guidance-reporting-eligible-identifiers-in-t-msis**

[13] Child and Caregiver Outcomes Using Linked Data (CCOULD) Project Common Data Model (CDM). **https://acf.gov/sites/default/files/documents/opre/ccould_lessons_learned_appendix_a_cdm_jan2023_0.pdf**

[14] Observational Health Data Sciences and Informatics. (n.d.). *Data Standardization – OHDSI*. OHDSI. **https://www.ohdsi.org/data-standardization/**

[15] Dusetzina SB, Tyree S, Meyer AM, et al. Linking Data for Health Services Research: A Framework and Instructional Guide. Rockville (MD): Agency for Healthcare Research and Quality (US); 2014 Sep. 4, An Overview of Record Linkage Methods. **https://www.ncbi.nlm.nih.gov/books/NBK253312/**

[16] Dusetzina SB, Tyree S, Meyer AM, et al. Linking Data for Health Services Research: A Framework and Instructional Guide. Rockville (MD): Agency for Healthcare Research and Quality (US); 2014 Sep. 4, An Overview of Record Linkage Methods. **https://www.ncbi.nlm.nih.gov/books/NBK253312/**

[17] U.S. Census Bureau. Record Linkage and the Person Identification Validation System (PVS). Census.gov. 2021. **https://www.census.gov/about/adrm/linkage/projects/pvs.html**

[18] National Center for Health Statistics. Division of Analysis and Epidemiology. The Linkage of the 2016 National Hospital Care Survey to 2015–2017 Centers for Medicare & Medicaid Services Transformed Medicaid Statistical Information System Claims Data: Matching Methodology and Analytic Considerations. Centers for Disease Control and Prevention; 2022. **https://www.cdc.gov/nchs/data/datalinkage/2016-nhcs-cms-medicaid-linkage-methodology.pdf**.

[19] National Center for Health Statistics. Division of Analysis and Epidemiology. The Linkage of the 2016 National Hospital Care Survey to 2015–2017 Centers for Medicare & Medicaid Services Transformed Medicaid Statistical Information System Claims Data: Matching Methodology and Analytic Considerations. Centers for Disease Control and Prevention; 2022. **https://www.cdc.gov/nchs/data/datalinkage/2016-nhcs-cms-medicaid-linkage-methodology.pdf.**

[20] CDC Foundation. Data Linkage and Identity Management -Privacy Protecting Record Linkage (PPRL). 2023. **https://www.cdcfoundation.org/CDCFoundationPPRLSummary.pdf?inline**

[21] Research Data Assistance Center (ResDAC). CMS Fee List for Chronic Conditions Warehouse Virtual Research Data Center Cloud Environment. ResDAC. Published April 2024. Accessed July 25, 2025. **https://resdac.org/sites/datadocumentation.resdac.org/files/202404/CMS%20Fee%20List%20for%20CCW%20VRDC%20Cloud%20Environment.pdf**.

[22] National Center for Health Statistics. Division of Analysis and Epidemiology. The Linkage of the 2016 National Hospital Care Survey to 2015–2017 Centers for Medicare & Medicaid Services Transformed Medicaid Statistical Information System Claims Data: Matching Methodology and Analytic Considerations. Centers for Disease Control and Prevention; 2022. **https://www.cdc.gov/nchs/data/datalinkage/2016-nhcs-cms-medicaid-linkage-methodology.pdf**

[23] Madden E, Butler V, Ghertner R. Child and Caregiver Outcomes Using Linked Data: Final Report. Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health & Human Services; December 2023. **https://aspe.hhs.gov/sites/default/files/documents/aece720a686c0ab93f5f6ce68723582e/aspe-ccould-pcor-final-report.pdf**.

[24] Mark TL, Dolan M, Allaire B, Bradley C. Linking Child Welfare and Medicaid Data: Lessons Learned from Two States. Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services; March 2023. **https://acf.gov/sites/default/files/documents/opre/ccould_lessons_learned_march2023.pdf**.

[25] Vance S, Strahle S, Cox S, Kuruganti S, Kipling L. Multi-State Network of Linked Pregnancy Risk Assessment Monitoring System (PRAMS) and Clinical Outcomes Data for Patient-Centered Outcomes Research. Association of State and Territorial Health Officials; 2024. **https://www.astho.org/4900c1/globalassets/report/multi-state-network-of-linked-prams-and-clinical-outcomes-data-for-patient-centered-outcomes-research.pdf**.

[26] Gersing K. National COVID-19 Longitudinal Research Database Linked to Medicare and Medicaid Data: PCOR Final Report—CMS-N3C Linked Dataset and Use Cases. National Center for Advancing Translational Sciences, NCATS Division of Clinical Innovation; November 15, 2024. **https://aspe.hhs.gov/sites/default/files/documents/8dc3a838c7c9652d041c07c2cb966d0f/n3c-11-15-24-final.pdf**.

[27] Multi-State Network of Linked PRAMS and Clinical Outcomes Data for Patient-Centered Outcomes Research. Association of State and Territorial Health Officials (ASTHO). 2022. **https://www.astho.org/4a2688/globalassets/report/multi-state-network-of-linked-prams-and-clinical-outcomes-data-for-patient-centered-outcomes-research.pdf**

[28] Comprehensive Child Welfare Information System. Toolkit: Data Sharing for Child Welfare Agencies and Medicaid. Administration for Children and Families. **https://acf.gov/sites/default/files/documents/cb/data-sharing-and-medicaid-toolkit.pdf**.

[29] Gersing K. N3C Privacy-Preserving Record Linkage and Linked Data Governance. 2021. doi:10.5281/zenodo.5165212

[30] Parrish J, Strahle S, Vance S. A Framework for Linking PRAMS with Administrative Data: Enhancing PRAMS through Data Linkages. Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services; 2024. **https://aspe.hhs.gov/sites/default/files/documents/e43417753fc9c1afd3d798c16b938751/astho-prams-data-linkage-framework-final.pdf**.

[31] Greenfield B. Timing of Medicaid Enrollment for Children Engaged with Child Welfare Services in Two States. Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services. December 2024. **https://aspe.hhs.gov/sites/default/files/documents/ab3a421123cfbaca8804e8d3a60d8d0b/Medicaid-Enrollment-Child-Welfare.pdf**

[32] Gross M, Keating B, Colten J, Miller R, Radel L, Abbott M. Prevalence and Characteristics of Children Entering Foster Care to Receive Behavioral Health or Disability Services: An Analysis of Custody Relinquishment Using Administrative Data. Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services. January 2025. **https://aspe.hhs.gov/sites/default/files/documents/35ca9fc3ef5f522e2ed18ec2a277f99c/custody-relinquishment-prevalence-characteristics.pdf**

[33] 2024-2026 IRP Extramural Large Grants. Institute for Research on Poverty, University of Wisconsin-Madison. 2024. **https://www.irp.wisc.edu/irp-extramural-large-grants-2024-2026/**

[34] Greenfield B, Madden E, Ghertner R. Advancing Research on Intersections of Child Welfare and Medicaid Using Linked Data from the CCOULD Project. Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services. November 2023. **https://aspe.hhs.gov/sites/default/files/documents/e48f62bd4d5417db61b2642dd70436d2/ccould-research-topics.pdf**