



ASPE
ASSISTANT SECRETARY FOR
PLANNING AND EVALUATION

OPRE

FINAL REPORT

CHILD AND CAREGIVER OUTCOMES USING LINKED DATA

Emily Madden and Robin Ghertner

Office of the Assistant Secretary for Planning and Evaluation

U.S. Department of Health and Human Services

Valeria Butler

Office of Planning, Research, and Evaluation

Administration for Children and Families

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.

Office of Planning, Research and Evaluation

The Office of Planning, Research, and Evaluation (OPRE) is part of the Administration for Children and Families (ACF) within HHS. OPRE serves as the principal advisor to the Assistant Secretary for Children and Families on improving the effectiveness and efficiency of programs designed to make measurable improvements in the economic and social well-being of children and families. OPRE builds and disseminates knowledge about effective approaches to helping children and families with low incomes through rigorous research and evaluation projects, including evaluations of existing programs and innovative approaches to help children and families, descriptive and exploratory studies, research syntheses, and more. OPRE also supports ACF programs in the responsible management and use of data to improve the effectiveness and efficiency of human services programs.

Suggested Citation

Madden, Emily, Valeria Butler, and Robin Ghertner. 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, and Office of Planning, Research and Evaluation, Administration for Children and Families. December 2023.

Acknowledgements

The authors would like to thank Tami Mark, Melissa Dolan, and the project team from RTI International, Susan Lumsden from ASPE, and Melissa Prosper from OPRE.

This communication was printed, published, or produced and disseminated at U.S. taxpayer expense.

Table of Contents

I. Executive Summary.....	4
II. Introduction	7
III. Background	8
IV. Objectives and Approach.....	11
IV. Project Accomplishments	18
VI. Lessons Learned for Projects Linking Child Welfare and Medicaid Data.....	25
VII. Directions for Future Medicaid and Child Welfare Data Infrastructure	29
VIII. Discussion	31

I. Executive Summary

This final report describes the objectives, approach, outcomes, and lessons learned from the project, Child and Caregiver Outcomes Using Linked Data (CCOULD). The Office of the Secretary Patient-Centered Outcomes Research Trust Fund ([OS-PCORTF](#)) funded the project, a collaboration between the Office of the Assistant Secretary for Planning and Evaluation (ASPE) and the Office of Planning, Research, and Evaluation (OPRE) within the Administration for Children and Families (ACF). OPRE and ACF implemented the project from 2019 to 2023.

Child maltreatment and involvement in child welfare systems are well-known adverse childhood experiences associated with poor health outcomes during childhood and later in life. Very little is known about the health needs of, and services used by families involved in child welfare systems, particularly those services funded by Medicaid. The lack of research is in large part due to the limited availability of large-scale datasets on Medicaid use of these families.

CCOULD (2019-2023) had four objectives:

Objective 1: Link data across State Medicaid and child welfare systems.

Objective 2: Prepare a research-use dataset and develop process for external researchers to access the data.

Objective 3: Develop an approach for states to have an ongoing link between Medicaid and child welfare data systems. Create a road map for other states to follow that documents the process and lends insight into lessons learned, challenges and successes.

Objective 4: Design, conduct, and encourage analyses on the linked dataset.

In partnership with OS-PCORTF, COULD developed several innovative products:

- Multistate datasets, deidentified and archived as restricted access files with a data center or repository that makes the data available to researchers to conduct secondary analyses.
- Road map for other states providing lessons learned on data sharing and data linkages between state Medicaid and child welfare agencies, including sample data agreements and advice about data structure, analysis, and overcoming organizational barriers.
- Analysis of outcomes in select states on patients receiving treatment under Medicaid and how receipt of (and adherence to) treatment relates to outcomes for their children involved in the child welfare system, as well as the possible effects of child welfare services on treatment success.

A team composed of federal leads, a contractor (RTI International), staff from participating states, and a technical expert panel implemented the project. The expert panel consisted of experts from across HHS in child welfare and Medicaid data, in data linking and data security, and in policy and program issues related to substance use and involvement in the child welfare system. In implementing the project, one of the first steps was to recruit states where data could be linked, and the infrastructure and staff capacity were sufficient for a high likelihood of success. This involved designing rigorous selection criteria to

identify factors predicting successful participation and a comprehensive strategy to recruit states to participate. The team selected Florida and Kentucky as the two participating states.

The team then developed a common data model (CDM) to guide the development of the final research dataset. Because each state collects its own data based on internal decisions, the team had to harmonize data extracted from state databases to make them useful for research purposes. In accordance with memoranda of understanding signed by each state, the CCOULD project team obtained data from each state's child welfare and Medicaid data systems. After conducting quality review, the team linked the data using deterministic linking methods. The team archived the final research-use dataset, along with a user's guide, codebook, and other documentation, at the National Data Archive on Child Abuse and Neglect at Cornell University, available free of charge.

CCOULD resulted in several key accomplishments. First, this project developed a comprehensive CDM to harmonize records across child welfare and Medicaid data systems, and to map variables together for a single, longitudinal, research-use dataset. Second, CCOULD developed a longitudinal, deidentified, research-use dataset, linking records from child welfare data systems with Medicaid enrollment and claims data, for records in Florida and Kentucky. Third, this project provided lessons learned to other states interested in developing their own research-use linked data across child welfare and Medicaid data systems. Fourth, CCOULD informed the federal and research community of the availability and utility of the linked dataset through research manuscripts, briefs, and presentations.

CCOULD generated important lessons for future efforts to link child welfare and Medicaid data. These lessons could also be informative for efforts to link Medicaid data with data from other human services programs. These lessons include the following:

- Recruiting states requires time and resources, engaging multiple stakeholders, and adapting to changing circumstances. The project achieved success by continually identifying how data linking supports the operational goals of the state agencies involved. These goals changed over the course of the project. Although state agencies consider data projects valuable, the projects often fall in priority due to limited resources and competing demands related to casework and program administration needs.
- States need support in navigating data governance. Federal and state rules on privacy and security govern state child welfare and Medicaid data systems, and states must navigate these rules successfully to share data across systems. On their own, states might not have the capacity to navigate the complex process of successfully establishing data-sharing agreements and maintaining data-sharing partnerships necessary for linking data across different state agencies.
- Linking projects need to identify and plan for issues and gaps in data quality and documentation. Teams managing data linking projects need to ensure they have a clear understanding of the data quality issues that state data systems have, particularly on key variables needed to execute linking. These issues can include data availability, consistency, validity, and reliability. The success of linkage hinges on the quality of the linking variables.
- States need support in strengthening and harmonizing data infrastructure on child welfare services. Many state child welfare systems provide services through different mechanisms and a

heterogenous network of providers. Consequently, the data systems used to collect information on families receiving these services are often not harmonized. For CCOULD, detailed information was not yet available on foster care prevention services provided to children and families, limiting the ability of researchers to use the dataset to understand how these services relate to health and child welfare outcomes. Medicaid data also need improvement around adding important variables, such as family identifiers.

- States and other entities undertaking linking projects would benefit from increased availability of resources needed to use probabilistic matching to link data sources. Although CCOULD used deterministic matching to link child welfare and Medicaid records, probabilistic matching could have produced more linked records. The states in this study did not have the time or resources to pursue probabilistic matching. Future efforts should include resources to support states in incorporating these methods through, for example, technical assistance and tools to implement probabilistic matching approaches.

II. Introduction

This final report describes the objectives, approach, outcomes, and lessons learned from the project, Child and Caregiver Outcomes Using Linked Data (CCOULD). A collaboration between the Office of the Assistant Secretary for Planning and Evaluation (ASPE) within the U.S. Department of Health and Human Services (HHS) and the Office of Planning, Research, and Evaluation (OPRE) within the Administration for Children and Families (ACF), the Office of the Secretary Patient-Centered Outcomes Research Trust Fund (OS-PCORTF) funded the project, which ASPE and ACF implemented from 2019 to 2023.

Child maltreatment and involvement in child welfare systems are well-known adverse childhood experiences associated with poor health outcomes during childhood and later in life (Ahrens et al., 2014; Rebbe et al., 2018; Zlotnick et al., 2012). Very little is known about the health needs of, and services used by families involved in child welfare systems, particularly those services funded by Medicaid. Many families involved in child welfare, ranging from being reported to maltreatment to having a child removed from home and placed into foster care, have low incomes and are eligible for Medicaid. Additionally, all children in foster care are eligible for Medicaid.

The lack of research on this topic is in large part due to the limited availability of large-scale datasets on Medicaid use of families involved with child welfare systems. One study examined high levels of Medicaid use among children in child welfare systems in Florida and Tennessee, identifying characteristics of children using a high degree of services, and doing so required obtaining administrative records and using complex methods to link and analyze them (Weigensberg, 2018). Administrative records on Medicaid claims do not contain adequate information on child welfare system involvement, and records from child welfare systems do not typically contain information on Medicaid use. National surveys of health services do not identify child welfare system involvement; surveys of child welfare-involved families do not typically ask a broad range of health questions. One notable example is the National Survey of Child and Adolescent Well-being (NSCAW), which includes health questions and has been linked to Medicaid in the past to answer important questions (Raghavan et al, 2016). The primary aim of this project was also to address this gap by developing important data resources linking child welfare records with Medicaid claims and to illustrate key lessons for implementing similar efforts in the future.

There are a myriad of important questions to answer regarding Medicaid use of system-involved families. The motivating questions behind this project were around parental substance use disorder (SUD) among child welfare involved families. This is because SUD is one of the biggest drivers of system involvement—particularly since the emergence of the opioid crisis—and one of the most challenging barriers to positive child welfare outcomes. After more than a decade of sustained declines in the national foster care caseload, the number of children entering foster care began to rise in 2012 (Radel et al., 2018). From 2012 to 2016, the number of children in foster care nationally rose by 10 percent, from 397,600 to 437,500, and areas most heavily affected by SUDs had the greatest increases in placements.

The project had four objectives:

Objective 1: Link data across State Medicaid and child welfare systems.

Objective 2: Prepare a research-use dataset and develop process for external researchers to access the data.

Objective 3: Develop an approach for states to have an ongoing link between Medicaid and child welfare data systems. Create a road map for other states to follow that documents the process and lends insight into lessons learned, challenges, and successes.

Objective 4: Design, conduct, and encourage analyses on the linked dataset.

The Office of the Secretary Patient-Centered Outcomes Research Trust Fund ([OS-PCORTF](#)) funded the project in 2019. 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. Patient-centered outcomes research aims to generate high-quality evidence about the effectiveness of treatments, services, and other health care interventions on the full range of outcomes that patients, caregivers, clinicians, policymakers, and other stakeholders have identified as important.

In partnership with OS-PCORTF, CCOULD developed innovative products that met its objectives:

- Multistate datasets, deidentified and archived as restricted-access files with a data center or repository that makes the data available to researchers to conduct secondary analyses.
- Road map for other states providing lessons learned on data sharing and data linkages between state Medicaid and child welfare agencies, including sample data agreements and advice about data structure, analysis, and overcoming organizational barriers.
- Analysis of outcomes in select states on patients receiving treatment under Medicaid and how receipt of (and adherence to) treatment relates to outcomes for their children involved in the child welfare system, as well as the possible effects of child welfare services on treatment success.

While contributing to enhancing data sharing and data linkages for research purposes, this work was groundbreaking in the innovative ways it used existing data to gain new insights and promoting its use to the broader research community, with minimal burden to state agencies. It supports the HHS goal to become better informed using all the data available to find better outcomes for our nation struggling with a crisis in opioids and SUDs.

III. Background

Child Welfare and Medicaid Data

State child welfare agencies generally track outcomes of children involved in their systems, including measures such as permanency outcomes (such as reunification with family or adoption or guardianship), and regularly monitor outcomes as youth transition to adulthood. Agencies rarely track health outcomes regularly and consistently, however, and this impedes our ability to understand how SUDs affect these disadvantaged populations differently from other children.

Many states link Medicaid and child welfare case records for children in the foster care system, which is required for new Comprehensive Child Welfare Information Systems (CCWIS) currently under development in most states because of regulations published in 2016. However, before this project, the project team could not identify any states that linked parents' records. Very little is known about use of Medicaid for SUD or mental health treatment by parents in child welfare systems, and how that treatment might be associated with child outcomes (including several social determinants of health such as repeated maltreatment, family reunification, and later health and well-being outcomes). In addition, the project team is not aware of any state with the capacity to monitor SUD treatment outcomes of parents with children in child welfare systems, as states rarely if ever collect adequate data on parents.

The timing of this project was ideal because ACF had recently issued grants to states to modernize data collection systems. States are in the position to make significant changes to their information technology (IT) infrastructure enabling improved data collection on parents and enhanced linkages to Medicaid data for both parent and child records. The research and documentation resulting from this project can support states as they invest in data systems enhancements. Aside from developing data infrastructure to enable research, the project served to inform states as they develop their new data systems.

Child Welfare System Involvement and SUDs

As stated above, the number of children entering foster care began to rise in 2012 following a decade of declines. One of the most consistent reasons given for a child's removal is parental alcohol abuse or other drug use, and the percentage of cases that identified alcohol or drug abuse as a contributing factor has increased nationally from 12 percent in 2000 to 34 percent in 2016 (Radel et al., 2018). ASPE research has shown that areas heavily affected by the opioid crisis have higher rates of foster care placement, after accounting for other factors.

One family member's SUD affects the entire family. Child well-being is inextricably linked to family well-being and parents' capacity to care safely for and nurture their children. Parental SUD can influence a child's likelihood of developing SUD (Leventhal et al., 2011). A parent's SUD affects his or her ability to effectively carry out parental tasks and responsibilities (Arria et al., 2014). Child welfare-involved parents with opioid use disorders (OUD) are an especially vulnerable population. They face additional barriers beyond those faced by others with OUD, which are important to understand to provide successful treatment and sustain recovery (Clary et al., 2020; Ghertner et al, 2020; Sieger et al., 2023). In addition, the treatment approach affects the entire family, and it is important to examine how treatment affects substance-using parents and their children.

No single system or agency has the resources needed to effectively address this problem. Collaboration across systems can help ensure programs identify parents in need of substance abuse treatment and they receive appropriate treatment in a timely manner; agencies also must address children's intervention needs. A family's involvement in the child welfare system can be an opportunity to connect them to integrated, evidenced-based treatment and services to support their path to recovery.

The Family First Prevention Services Act¹ (FFPSA) signed into law February 2018, permits states to access Social Security Act Title IV-E federally matched dollars for foster care prevention services, including treatment for substance use and mental health disorders for parents and youth, intended to provide enhanced support to children and families and prevent foster care placements. FFPSA requires an independent systematic review of evidence to designate programs and services as “promising,” “supported,” and “well-supported” practices. To meet these requirements, ACF established the Title IV–E Prevention Services Clearinghouse (the Clearinghouse). While programs and services are being reviewed and rated by the Clearinghouse on an ongoing basis, there is a need to continue to build the evidence of effectiveness for programs and services. Thus, for states to be able to access these funds, more evaluation is necessary to test the plethora of interventions being used by states and counties. This project developed data infrastructure that can support evidence-building activities.

The Role of Medicaid in SUD Treatment for Families Involved in Child Welfare Systems

Some of the major barriers to receiving SUD treatment for parents with children in the child welfare system include (1) ability to pay for treatment, (2) adherence to treatment and the inability of courts to monitor adherence effectively, and (3) inadequate recovery supports for people receiving treatments and their families.

As many of these parents are eligible for Medicaid, Medicaid has a role in reducing or removing each of these barriers. However, we know little about how this population uses Medicaid for SUD treatment, and whether treatment accessed through Medicaid leads to successful treatment and recovery, positive child welfare outcomes (such as family reunification and prevention of foster care placement, reduced repeat maltreatment, and shorter time in foster care), and child health outcomes. Little is known about how services provided by child welfare agencies align with those paid for by Medicaid. In addition to linking data strictly for research purposes, Medicaid plays an important role for states to fulfill expectations for interagency collaboration among public health, child welfare and other entities in Child Abuse Prevention and Treatment Act grant programs and Title IV-B.

Linked Medicaid and child welfare data can address other relatively recent policy changes. In particular, the FFPSA allows child welfare agencies to use foster care funds to pay for services to children staying with their parents in family-friendly residential treatment facilities, substance abuse treatment for caregivers, and other behavioral health services. The Substance Use Disorder Prevention that Promotes Opioid Recovery and Treatment for Patients and Communities (SUPPORT) Act, enacted on October 24, 2018, created a state option for Medicaid to pay for up to 30 days of care in a year for patients with SUDs in institutions for mental disease (IMDs), including residential treatment facilities. This presents new opportunities for parents in need of SUD treatment services, though the complexity of navigating the payments necessitates research and data sharing to understand how to implement it effectively.

¹<https://www.congress.gov/bill/115th-congress/house-bill/1892/text#toc-HBF8A6BEC2EC643E6993C8D8B06A01F02>

IV. Objectives and Approach

This section describes the project’s objectives and the associated deliverables, as well as the general approach taken. Exhibit 1 outlines the project’s objectives and the final products associated with them. More details on the deliverables are available in the Accomplishments section (on page 19).

Exhibit 1. Project Objectives and Deliverables

Objective	Deliverables
1: Develop a linked dataset across State Medicaid and child welfare systems.	Datasets for each state that contain linked Medicaid eligibility, enrollment, and claims data with Child Welfare Records; these datasets will primarily be available only to the state child welfare and Medicaid agencies and the state’s specified research partner. The project team will submit documented affirmation from the state and the contractor.
2: Prepare a research use dataset and develop a process for external researchers to access the data.	Deidentified, research-use dataset and accompanying documentation, available at the National Data Archive on Child Abuse and Neglect: https://www.ndacan.acf.hhs.gov/datasets/dataset-details.cfm?ID=272
3: Develop an approach for states to have an ongoing link between Medicaid and child welfare data systems. Create a road map for other states to follow that documents the process and lends insight into lessons learned, challenges, and successes.	Published a report, “ Linking Child Welfare and Medicaid Data: Lessons Learned from Two States ,” that provided guidance that will help other states navigate how to obtain data-sharing agreements between state Medicaid and child welfare agencies, among other lessons.
4: Design, conduct, and encourage analyses on the linked dataset.	Manuscripts submitted to peer-reviewed journals: The Child and Caregiver Outcomes Using Linked Data (CCOULD): A New Data Resource from Two States to Identify Ways to Improve the Welfare of Children and Their Families, and Linked Child Welfare and Medicaid Data in Kentucky and Florida Highlights Racial Disparities in Access to Care. HHS report that contains examples for additional analyses and guidance for completing those analyses. Presentations at academic conferences and webinars for HHS and external audiences.

Project Team and Participating States

In implementing the project, the team consisted of federal staff from ASPE and OPRE, a contractor (RTI International), and staff from the participating states and a university. Over the course of the project, the team consulted with a technical expert panel (TEP). The TEP consisted of experts from across HHS in child welfare and Medicaid data, in data linking and data security, and in policy and program issues related to substance use and child welfare system involvement. Members of the TEP came from the following agencies: ASPE, OPRE, the Children’s Bureau, the Centers for Medicare & Medicaid Services (CMS), the National Institute on Drug Abuse, the Substance Abuse and Mental Health Services Administration, and the Agency for Healthcare Research and Quality.

After an initial kickoff meeting, the project team involved members of the TEP at critical points during the project, depending on their expertise. This included selecting and recruiting states, legal and data security issues when developing memoranda of understanding with participating states, design of the common data model, and identifying topics to research with the dataset.

The team engaged with two states to carry out data linking for this project: Florida and Kentucky. Each state had a unique internal structure, with Kentucky keeping project efforts internal to the state and Florida teaming with a research partner, The University of Florida, to carry out all analytical and technical tasks related to data linking.

State Selection and Recruitment

State participation was critical for the success of CCOULD. One of the first steps in CCOULD was to recruit states to be sites capable of linking data. This involved designing rigorous selection criteria to identify factors predicting successful participation and a comprehensive strategy to recruit states to participate. The team developed four broad selection criteria to identify factors that would predict a given state’s likelihood of success in meeting the project’s objectives. The team identified these criteria based on the subject matter expertise of the project team and the TEP, as well as review of current and past related ASPE and OPRE projects. Exhibit 2 shows how the team developed the main selection criteria and put those criteria into operation.

Exhibit 2. Project Selection Criteria

Criterion	Elements
Interest and buy-in: The extent to which the state is interested in the project and has buy-in from key leadership and relevant agencies	Political and career leadership in both child welfare and Medicaid agencies are brought in and support the project’s goals. State has interest in accessing Title IV-E funds for prevention services under FFPSA.

	State is willing to invest its own resources (staff and budget) to support the project.
Partnerships: The strength of partnerships and collaboration among key stakeholders needed to successfully implement the project	<p>State has experience working with one or more research partners.</p> <p>State Medicaid and child welfare agencies have experience collaborating.</p>
Infrastructure and capacity: The extent of the existing technical infrastructure to link datasets, as well as conduct necessary data quality reviews and analysis; the extent of existing staffing and budgetary capacity to support the linking and analysis	<p>State is in the process of updating or has updated its CCWIS system with funding from ACF’s Children’s Bureau.</p> <p>State Medicaid office has used existing federal and state resources to update Medicaid infrastructure.</p> <p>State child welfare and Medicaid agency has dedicated staff to support research.</p> <p>State has standardized SUD and behavioral health measures in Medicaid data.</p>
Experience: Experience the state has in implementing linking projects	<p>State and/or research partners have linked child records across Medicaid and child welfare systems or have linked other sources of data to Medicaid or child welfare.</p> <p>State has experience addressing data-sharing concerns related to privacy and experience with data use agreements.</p>

The team considered several other factors, though these were not critical for the project’s success:

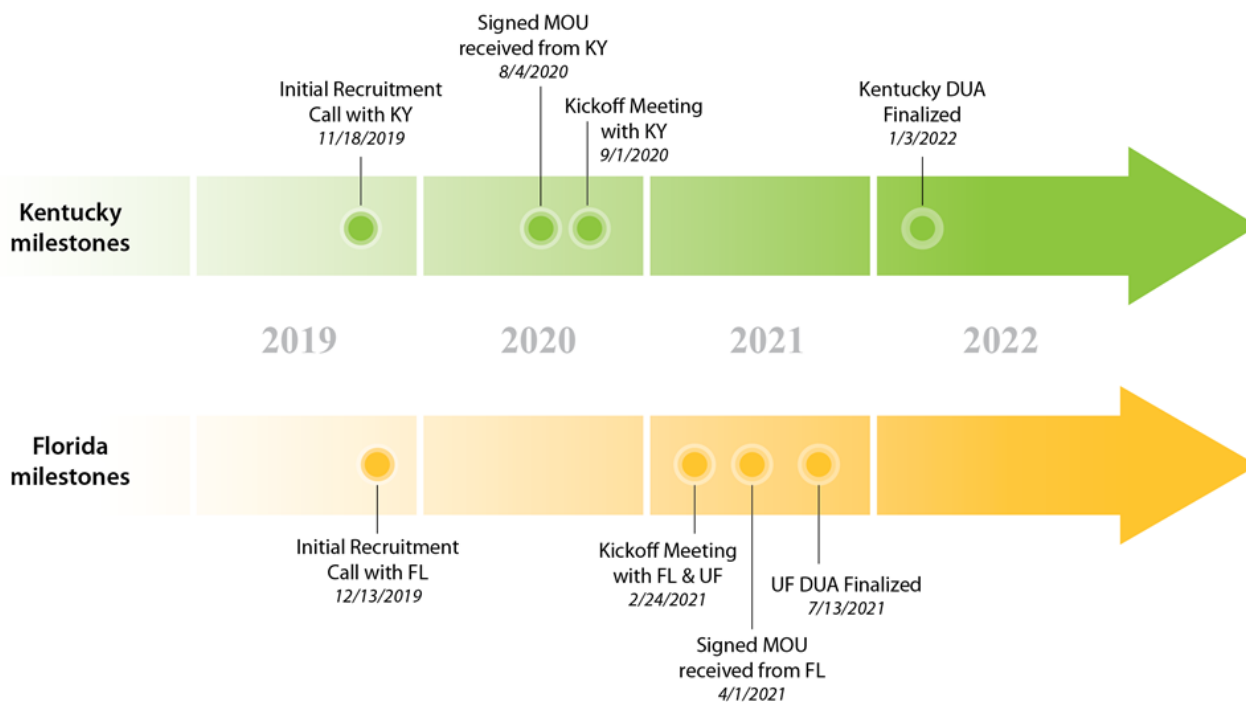
1. **Foster care caseloads.** Selected states ideally would have sufficient caseloads to result in a large sample size to provide statistical power for subgroup analyses. In addition, the caseload should have large enough sample to avoid identification of cases.
2. **Medicaid enrollment and eligibility.** Selected states ideally would have sufficient enrollment of relevant subgroups in Medicaid to provide statistical power and avoid identification.
3. **The prevalence** of opioid use disorders and other SUDs in the state.
4. **Additional relevant experience** the state has engaged in that would enhance its contribution to the success of this project.
5. **Consideration for the final selection** of states to include states with different data capacity and approaches to their own data infrastructure.

Recruiting states to participate in data-linking projects is a multistage process that is often not straightforward. After developing the selection criteria outlined before, the team took several steps to recruit states. First, the team relied on expert consultation, holding discussions with key federal and

nonfederal stakeholders with knowledge about state data capacity and infrastructure related to this project’s objectives. In the discussions, the project team laid out the objectives and selection criteria, then asked experts to provide their opinion on the degree to which states they were familiar with met or did not meet the criteria. Second, the team conducted outreach through the extensive state network maintained by ACF’s Children’s Bureau. The team sent out information of the project to data-related leadership and staff in all 50 states and the District of Columbia, including an invitation to a webinar about the project. The team hosted a webinar with interested states, with more than 60 attendees. After the webinar, the team held targeted follow-up discussions with states that had demonstrated interest and met the selection criteria.

The project had budgeted for participation of two to four states, and the recruitment process resulted in two states—Florida and Kentucky—with real interest and capacity to participate. The next step was to develop and finalize agreements with these states. Exhibit 3 outlines the timeline for finalizing these agreements. This was a lengthy process, requiring multiple meetings with interested parties in different agencies and organizations (including universities) and at different levels of responsibility in each state. Each stakeholder had different perspectives and interests and, as a result, the project team had to tailor its approach to pitching the project depending on the stakeholder.

Exhibit 3. Timeline for State Recruitment and Finalization of Agreements



DUA = data use agreement; FL = Florida; KY = Kentucky; MOU = memorandum of understanding; UF = University of Florida.

Collaboration with Participating States

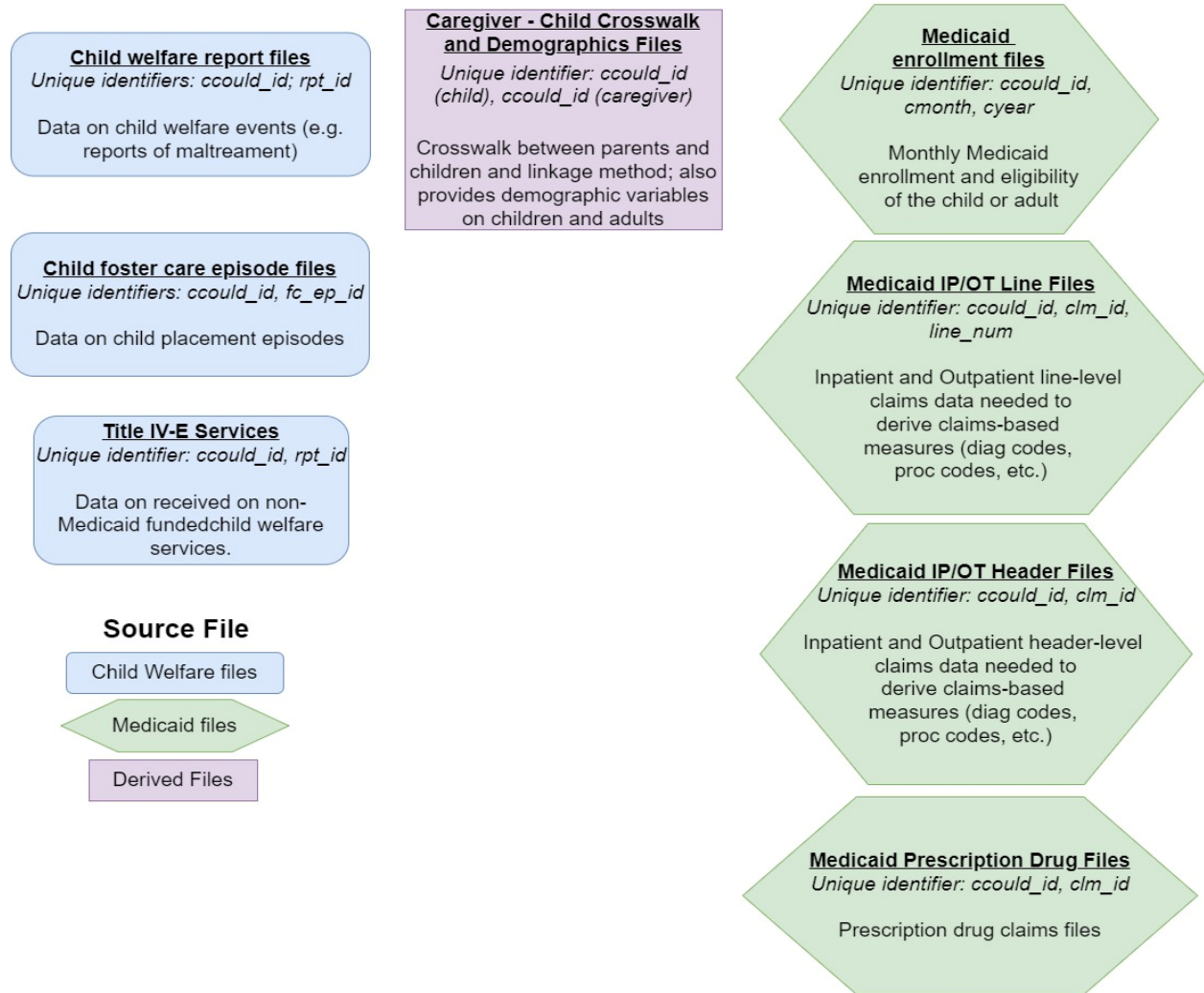
Throughout the project, the team engaged regularly with state partners to provide extensive technical assistance from start to finish. This began with in-depth work with state staff to work through any legal, privacy, and ethical issues and questions. The project team worked with legal staff from both states to review any relevant state laws and ensure they met all Health Insurance Portability and Accountability Act (HIPAA) requirements. The project team also worked closely with states to develop comprehensive data use agreements that were critical to ensure the success of any data-linking efforts. Project staff frequently served as liaisons between state child welfare and Medicaid agencies to help strength the partnership between these two agencies within each state. As the project progressed, technical assistance focused on remediating data quality concerns, developing data-linking methodologies and executing the data linkage, and cleaning and harmonizing the final linked dataset. These tasks ranged from creating a method to link records that lacked Social Security numbers (SSNs, correcting mapping and extract errors, and reconciling data with the states' original records.

Common Data Model

The team developed a common data model (CDM) to guide the development of the final research dataset, shown in Exhibit 5. Because each state collects its own data based on internal decisions, data extracted from state databases must be harmonized to make them useful for research purposes. CDMs standardize datasets across different applications and sources, with the goal of permitting the organization and exchange of information between them. To support the harmonization and standardization of data across child welfare and Medicaid data systems within each state, and across the two states, the CCOULD common data model mapped state variables and transformed them to make them comparable. The model's structure depended on existing federal data systems for Medicaid claims and child welfare information. For Medicaid, the model relied on the Transformed Medicaid Statistical Information System ([T-MSIS](#)), which collects data directly from all state Medicaid information systems. CMS operates the T-MSIS, which standardizes the disparate data systems used by states to permit cross-state analysis. For child welfare records, the CCOULD common data model used by the Children's Bureau: National Child Abuse and Neglect Data System ([NCANDS](#)) and the Adoption and Foster Care Reporting System ([AFCARS](#)). NCANDS is a voluntary data collection system that gathers information from all 50 states, the District of Columbia, and Puerto Rico about reports of child abuse and neglect. AFCARS collects data from states and tribes on case-level information on all children in foster care and children who have been adopted from foster care for agencies funded through Social Security Act Title IV-E.

Exhibit 4. CCOULD Common Data Model

CCOULD COMMON DATA MODEL STRUCTURE



Two files describe the CDM: a Microsoft Excel workbook and programming specifications. The Excel workbook contains worksheets coinciding with each data file. The worksheets list the variables contained in the data files, the source of the variables, the variable’s character type and length, and the variable label. An introduction sheet includes file structure information and research questions to answer with the data. The programming specifications provide detailed guidance on variable definition and how to derive each of the Medicaid-specific files in the CDM. The files in the CDM adhere to the structure laid out in Exhibit 4 (above). The complete CCOULD dataset comprises several smaller subsets of data identified and also described in Exhibit 4, which is a visual display of the structure of the database created for initial collection of CCOULD dataset components.

Data Linking

In accordance with the memorandum of understanding (MOU) signed by each state, the CCOULD project team obtained data from each state’s child welfare and Medicaid data systems. The team negotiated the scope of the data—such as number of years, which cases, and which variables—with each agency as part of the MOU and as required by the CDM. After obtaining the data, the team developed a set of data quality checks for states to assess and adjust their data as needed before submitting them to the project team for statistical disclosure analysis and harmonization. Examples of these quality checks include an examination of missing data, comparison of population counts to corresponding counts pulled directly from the states’ data systems, and tracing mapped variables to their components in each state system to check for fidelity.

Following the data quality review, the team linked the data. After consulting with the analysts in Florida and Kentucky who would support the linking efforts, the team decided to use deterministic linkage method. Probabilistic methods could have yielded different and perhaps better results, but the states did not have the time or resources to pursue it. Deterministic linking relied primarily on an individual’s Social Security Number (SSN). Using additional variables for deterministic linkage can increase specificity (correctly identifying individuals who should not be linked) at the cost of lowering the sensitivity (correctly identifying individuals across the datasets who should be linked); thus, the team opted to link using SSN when possible (Dusetzina et al., 2014). They linked records for children and caregivers without SSNs on an additional set of identifiers called the LongID. The team modeled the LongID after the SLK-581 identifier, which is straightforward to implement and has a low rate of incorrect links (Australian Government, n.d.; Coulson et al, 2021). The team specified the LongID for this use as the 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).

Florida and Kentucky linked 80 and 81 percent, respectively, of the children in the child welfare system with a report to a Medicaid record. Florida linked 70 percent via SSN and Kentucky linked 93 percent. Linkage rates for the LongID were much lower (30 percent for Florida and 7 percent for Kentucky). The caregiver linkage rate was not as high as the link for the children; 68 percent of caregivers were linked to Medicaid in Florida and 85 percent of caregivers in Kentucky were linked to Medicaid. Caregivers were almost exclusively linked via SSN; more than 96 percent in both states were linked with their SSNs. Almost all children linked to Medicaid had only one caregiver linked to Medicaid and among those who did, very few had one: 9 percent children in the Florida dataset also had a caregiver linked to Medicaid, whereas 21 percent in Kentucky had a least one caregiver linked to Medicaid.

Data Archiving

One of the primary goals of this project was to ensure outside researchers have access to the linked dataset created from the project’s efforts. The project team held discussions with multiple data archives to understand requirements for archiving, the level of assistance that the archive can provide to researchers accessing the data, the types of researchers who access data from the archive, and any data-related or technical requirements. The team selected the National Data Archive on Child Abuse and

Neglect (NDACAN) at Cornell University as the archive to house the data. In collaboration with the NDACAN, the project team developed access requirements and steps for how researchers could obtain the data. As a part of the archiving process, the team developed a user's guide, codebook, and other relevant data documentation that researchers who obtain the linked dataset will receive.

Analysis of Linked Data

This project developed resources to inform PCOR researchers about the value of using the CCOULD dataset for research. The project produced a summary of the topics that CCOULD is designed to address. In addition, the project team produced two manuscripts using the newly created data to demonstrate the utility of the data for patient-centered outcomes research. The team is preparing to submit both manuscripts to peer-reviewed journals. The next section provides more details on these products.

IV. Project Accomplishments

This section describes the key accomplishments of CCOULD.

Common data model to link records from child welfare data systems and Medicaid claims

As described, this project developed a comprehensive CDM to harmonize records across child welfare and Medicaid data systems and to map variables together for a single, longitudinal, research-use dataset. Administrative data systems such as child welfare information systems and T-MSIS seek to manage programs, such as case management, reporting on operations, and fund reimbursement. They are not designed to interact with other systems, nor with research in mind. The CDM provides a framework for linking these complex systems so researchers can easily use them for analysis. Although the team developed the model specifically for data from Kentucky and Florida, they also designed it to apply to other states.

First research-use dataset linking records from child welfare data systems with Medicaid claims, for children and caregivers

CCOULD developed a longitudinal, deidentified, research-use dataset, linking records from child welfare data systems with Medicaid enrollment and claims data, for records in Florida and Kentucky (note, records across the two states were not linked). The dataset is available from the NDACAN for no cost.²

The archive includes accompanying data documentation; codebooks that define variables contained in the dataset and explain proper use and interpretation; and a user guide with documentation of deidentification process, masking procedures, standards for extracting data from child welfare and Medicaid systems, and steps taken to link the data. The dataset contains linked records on 1,087,763

² Data available at the National Data Archive on Child Abuse and Neglect. Child and Caregiver Outcomes Using Linked Data (CCOULD). <https://www.ndacan.acf.hhs.gov/datasets/dataset-details.cfm?ID=272>.

children and 89,871 caregivers and covers the periods of January 2017 through June 2021 in Florida, and January 2017 through 2020 in Kentucky.

There are eight file types contained in the CCOULD: Medicaid enrollment files, Medicaid inpatient (IP) claims files, Medicaid outpatient and other service (OT) claims files, prescription drug (RX) claim files, child welfare report files, foster care episode files, the Title IV-E services provided files, and the caregiver–child crosswalk and demographic files. The IP and OT files contain information on health care use and the enrollment file contain Medicaid monthly enrollment and eligibility information. Child welfare reports and foster care files contain critical information documented from each states’ child welfare system, primarily child welfare reports compiled by caseworkers and placement information about children in the foster care system.

The dataset also contains a comparison sample of Medicaid records from both states that child welfare data systems do not include. These comparison records represent a separate 10 percent random sample of Medicaid beneficiaries who did not have records in the child welfare system. The team assembled these data to allow for comparisons between children and adults with and without child welfare involvement.

Exhibit 5 reports some of the demographic characteristics of the dataset produced by this project, focused on the year 2020.

Exhibit 5. Characteristics of Caregivers with Children Investigated by Child Protective Services and Enrolled in Medicaid and Age and Gender Matched Adults in Medicaid, 2020

	Caregivers with Children Involved in Child Welfare		Age and Gender Matched Adults Without Children Involved with Child Welfare	
	Count	Percentage	Count	Percentage
Kentucky	33,786	58	33,786	58
Florida	24,765	42	24,765	42
Ages 18 to 25	8,342	14	8,342	14
Ages 26 to 40	38,124	65	38,124	65
Ages 41 to 55	10,959	19	10,959	19
Ages 56 to 64	863	1	863	1
Ages 65 and older	1	0	1	0
Age missing	262	0	262	0
Male	18,061	31	18,061	31
Female	40,489	69	40,489	69
Gender unknown*	1	0	1	0
American Indian or Alaska Native	122	0	95	0
Asian	164	0	677	1
Black or African American	11,740	20	11,025	19
Native Hawaiian/Other Pacific Islander	70	0	1,633	3
White	45,499	78	31,959	55
Race missing*	956	2	13,161	22
Hispanic or Latino	3,199	5	7,186	12
Not Hispanic or Latino	45,583	78	46,745	80
Ethnicity missing*	9,769	17	4,620	8

* Data elements were missing or unknown for some records in the administrative data systems that formed the basis of CCOULD data. Data may be missing for a variety of reasons, for example, caseworkers may not have collected the information, participants may not have provided the data, or there may be data entry errors.

Lessons learned to support other states, researchers, and organizations to conduct similar data linking in the future

In addition to the valuable research the CCOULD dataset can generate, this project also provided guidance to other states, other organizations, or researchers interested in developing their own research-use linked data across child welfare and Medicaid data systems. The project team published an ASPE report, "[Linking Child Welfare and Medicaid Data: Lessons Learned from Two States](#)," that provides guidance that will help other states navigate how to obtain data sharing agreements between state Medicaid and child welfare agencies, guidance and best practices for linking Medicaid and child welfare data, how to establish needed infrastructure to properly house and work with the data, common barriers to linking Medicaid and child welfare data, and suggestions on how to overcome them. This documents also outlines important considerations for states that decided to link these data and best practices for approaching those considerations.

Inform the federal and research community of the project accomplishments and availability of the linked dataset

The final accomplishment of CCOULD was to disseminate the linked dataset to the research community, informing patient-centered outcomes researchers of the availability and utility of the dataset. The team accomplished this through publications and presentations for federal and nonfederal audiences. The project team delivered webinars through the PCORTF research network on different occasions to share progress and outcomes of the project. In 2018, the team presented early plans for the project as part of a summit on "Addressing the Opioid Epidemic: Harnessing the Power of Data for Patient-Centered Research." In 2020, the team delivered a webinar on "Building Data Capacity for Childhood PCOR: The Childhood Obesity Data Initiative and Linking State Welfare Data for Outcomes Research." In April 2023, the team presented the major outcomes of the project, as well as lessons learned, on a ASPE-sponsored webinar.

The project team presented at several research and academic conferences. In 2019, the team presented the project plans at the Child Welfare Evaluation Summit, on "State Medicaid and Child Welfare Data Linkages for Outcomes Research." The team also presented on the project's plans and goals at the 2019 American Association of Health and Human Services Attorneys on "Data Considerations for Implementing the Family First Prevention Services Act." In 2022, the team presented at several conferences. At the Annual Conference of the Association for Public Policy Analysis and Management (APPAM), the team was part of a panel on "Advances in Data Infrastructure at the Nexus of Child Welfare and Related Systems." The team presented at the 2022 Annual Conference of the Child Welfare League of America on "Considerations for Linking Child Welfare Data," based on lessons from this project. State child welfare agency leaders and staff, advocates, and researchers attend this conference. The team also presented at the American Medical Informatics Association Annual Symposium, on "Federal Initiatives to Increase Access to and Linkage of Longitudinal Data Sets to Enhance Data Capacity for Patient-Centered Outcomes Research." In 2023, the team returned to APPAM as part of a panel on "Using Medicaid Administrative Data to Provide Human Services Insights." The presentation "Child and Caregiver Outcomes Using Linked Data (CCOULD) to Explore Cross-Cutting Medicaid and Child Welfare Issues" described the process for

developing the CCOULD dataset, how to access it, and example research questions the data could address. In summer 2023, the project team presented on the CCOULD dataset for researchers through NDACAN as a part of the summer training series.

Lastly, the team prepared an [ASPE report](#) summarizing additional analyses made possible using the CCOULD dataset and guidance for completing those analyses. The report outlined sample research questions that can be explored using the CCOULD dataset, with the goal to increase understanding around the utility of the data and generate interest from researchers in the field. Although not a complete list, the research questions illustrate the types of analysis the data can address. Exhibit 6 highlights example research topics and example questions related to those topics that CCOULD data can address. These topics aim to motivate researchers to use the data for their research, but the topics are not comprehensive.

Exhibit 6. Example Research Topics and Questions that CCOULD Data Can Address

Research Topics	Example Research Questions
Prevalence of SUD among children and caregivers engaged in child welfare systems	<p>How do claims-based indicators of SUD in Medicaid data (that is, diagnosis codes) compare with indicators of SUD in the child welfare data?</p> <p>Is one system better than the other at identifying SUD?</p> <p>Among people covered by Medicaid, how does the prevalence of SUD compare across children and caregivers based on child welfare system engagement?</p>
Relationship between SUD or mental health conditions and child welfare engagement	<p>How do child welfare outcomes for children with parents with SUDs (for example, out-of-home removal rates, reunification rates, and length of time in out-of-home care) compared to those without identified SUDs? How do these outcomes differ by children’s demographic characteristics?</p> <p>To what extent are postpartum depression or other particular mental health conditions associated with child welfare involvement, care experiences, and outcomes?</p> <p>How does psychotropic medication usage differ among child welfare-engaged children and adults compared to those without child welfare engagement?</p>
Composition of SUD treatment recipients	<p>What demographic characteristics, clinical characteristics, and medical care use patterns are associated with parents or caregivers receiving Medicaid-funded SUD treatment?</p> <p>What demographic characteristics are associated with parents or caregivers receiving Title IV-E funded SUD treatment?</p>
Relationship between SUD treatment and child welfare outcomes	<p>How does receipt of SUD treatment by caregivers with children involved with child welfare systems relate to child welfare outcomes, such as length of time in foster care, repeat maltreatment incidences, and reentry into foster care?</p> <p>How does having a parent receiving SUD treatment affect placement decisions?</p> <p>How does receipt SUD treatment by parents or caregivers with children engaged in child welfare systems differ by caregivers’ demographic characteristics?</p>
Relationship between Title IV-E services on child and caregiver child welfare and health outcomes	<p>Which Title IV-E services are used most often among child welfare involved families?</p> <p>How does use of Title IV-E services differ among caregivers of various demographic groups?</p> <p>What is the relationship between receipt of Title IV-E services and caregiver health outcomes (e.g., increase in primary care visits, indicators of improved health, etc.)</p>

Use of services for children with substance use-related illnesses	<p>What services funded by Medicaid and child welfare systems are used by children born with substance use related illnesses, and their parents?</p> <p>How are those services related to health and child welfare outcomes?</p> <p>What differential diagnoses are most commonly associated with parental SUD?</p> <p>How do substance use related illnesses differ among children of different demographic groups, whether engaged in child welfare systems or not?</p>
Medicaid spending for child welfare-engaged families	<p>How much does Medicaid spend on SUD treatment among parents with children in the foster care system? How much does Title IV-E spend?</p> <p>Does spending on parents' SUD treatment result in subsequent savings to Medicaid and/or Title IV-E?</p> <p>Does spending on SUD treatment and/or Title IV-E spending affect foster care spending?</p> <p>How long do children retain Medicaid enrollment following exits from foster care?</p>
Health outcomes for child welfare-engaged children	<p>What diagnoses are most common among children engaged in child welfare systems?</p> <p>How do health outcomes for children engaged in child welfare systems differ by demographic characteristics?</p> <p>What diagnoses are associated with different forms of child maltreatment?</p> <p>How are the timing of diagnosis and treatment of health problems in children related to timing of child welfare engagement?</p>

Conduct data analysis demonstrating value of the dataset for patient-centered outcome research

The team submitted two manuscripts to peer-reviewed journals. The first paper, “The Child and Caregiver Outcomes Using Linked Data (CCOULD): A New Data Resource from Two States to Identify Ways to Improve the Welfare of Children and Their Families,” introduced the dataset to researchers and provided a demonstrative analysis on the prevalence of mental health and substance use issues. The analysis also explored gaps in the receipt of mental health and substance abuse treatment. Among other key findings, the analysis demonstrated that in Florida and Kentucky, the prevalence of mental health disorders, SUDs, opioid use disorders, and alcohol use disorders was much higher among Medicaid enrollees with children involved in the child welfare system relative to the age and gender matched Medicaid enrollees without children involved in the child welfare system.

Exhibit 7 provides some of the results to the analysis. It finds that in 2020, caregivers covered by Medicaid who had children involved in child welfare systems were more likely to have any mental health or SUD than adults covered by Medicaid who did not have children involved in child welfare systems. This is consistent across mental health and SUDs, and particularly acute for opioid use disorder. Almost one-third

(30 percent) of caregivers receiving Medicaid with children involved in child welfare had a diagnosed opioid use disorder, compared to 8 percent of adults receiving Medicaid without children in child welfare systems. It is important to note that the adult population receiving Medicaid without children in the child welfare system might or might not have children. It is not possible to gain this information from the data used for analysis. These results are examples of more detailed results that will be available in the published manuscript.

Exhibit 7. Prevalence of Mental Health and SUDs Among Caregivers with Child Welfare System Involvement and Matched Group of Adults Receiving Medicaid for Florida and Kentucky, 2020 (percentages)

	Caregivers with children involved with child welfare system (%)	Matched adults without children involved in child welfare systems (%)
Any mental health or SUD	59	33
Mental health disorder	45	28
SUD	42	12
Opioid use disorder (OUD)	30	8
Alcohol use disorder	18	5

N = 58,551.

The second manuscript, “Linked Child Welfare and Medicaid Data in Kentucky and Florida Highlights Racial Disparities in Access to Care,” examined racial disparities in the receipt of mental health and SUD treatment among caregivers with children involved in child welfare systems. One of the key findings of this analysis was that among caregivers with a mental health diagnosis, the percentage with a mental health or SUD-related emergency department visit or hospitalization was higher among Black (25 percent) than White (22 percent) individuals in the past year.

VI. Lessons Learned for Projects Linking Child Welfare and Medicaid Data

This project generated a number of important lessons for similar efforts to link state human services records with health claims. The lessons can be useful for states as they pursue linking projects, as well as for federal and nongovernmental programs looking to support state data-linking projects. This section highlights five of the most important lessons learned. More details on lessons learned for this project appear in the report “[Linking Child Welfare and Medicaid Data: Lessons Learned from Two States.](#)”

Recruiting States Requires Time and Resources, Engaging Multiple Stakeholders, and Adapting to Changing Circumstances

A key lesson from this project was to continually identify how data-linking projects support operational goals of state agencies, which can change over the course of such projects. In isolation, state agencies consider data projects valuable, though they often fall in priority due to limited resources and competing

demands related to casework and program administration needs. This became most apparent when the COVID-19 pandemic emerged as a public health crisis, which occurred just as recruitment entered its final stages. As state leaders turned their attention to the pandemic, data linking became a low priority and the project team had to find innovative ways to demonstrate how this project could support improving operations and services to vulnerable families. Linking efforts can be successful when they generate state interest by pointing to specific information state agencies could learn from the linked data that would be of clear value to state agencies that oversee child welfare or Medicaid programs.

Recruitment efforts can also benefit from having a thorough understanding of states' data and operational readiness to undertake data linkage. Staff representing the state or agency leadership involved during the recruitment process might not fully understand their state's capacity, and the project team needs to clearly understand what technical assistance and capacity building it will need to ensure state success. When states lack the technical capacity to perform data-linking tasks, helping a state form partnership with universities or other entities can supplement research and analytic capacity.

States Need Support in Navigating Data Governance

Data-linking projects require a strong understanding of the governance structures around the datasets to link. Federal and state rules on privacy and security govern state child welfare and Medicaid data systems, and states must navigate these rules successfully to link data across systems. On their own, states might not have the capacity to navigate this complexity. The data linked in this project were subject to state and federal privacy and human subject protection regulations. From the outset, the project team had planned for support on federal rules such as HIPAA, but also needed to develop resources to help states address their own rules. Decisions about data sharing, privacy, and human subjects for CCOULD required involving multiple stakeholders from the project team and the state agencies, including staff with technical expertise in child welfare and Medicaid data, legal counsel, HIPAA privacy officers, institutional review boards, and contracts administrators. Linking efforts can take advantage of existing technical assistance resources, such as OPRE's [Confidentiality Toolkit](#) and the [Children's Bureau-CMS Toolkit for Data Sharing for Child Welfare Agencies and Medicaid](#). States could also benefit from revisiting their laws pertaining to the sharing of child welfare records for research and evaluation, given the evolution in the availability of electronic data and the greater emphasis on preventing child removals.

Aside from governance rules, the institutional structure of state agencies—including data ownership, location of agencies in the government organizational system, and historical relationships between agencies—influence collaboration. Project teams need to understand these structures and develop strategies to proactively break down agency silos and facilitate collaboration and communication across offices that might not have established processes for doing so. For many states—including the states in this study, but also other states that the project team surveyed as part of the recruitment process—child welfare and Medicaid data exist in a fragmented assortment of collection points that do not communicate with one another.

States can also benefit from revisiting their laws pertaining to sharing child welfare records, given the evolution in the availability of electronic data since the appearance of these laws, as well as the greater emphasis on family preservation and prevention of maltreatment.

Linking Projects Need to Identify and Plan for Issues and Gaps in Data Quality and Documentation

Teams managing data-linking projects must ensure they clearly understand the data quality issues that state data systems have; particularly, on key variables needed to execute linking. These issues can include data availability, consistency, validity, and reliability. The success of linkage hinges on the quality of the linking variables. In CCOULD, for example, states had difficulty assessing how much missingness existed in proposed linkage variables, as many data systems could not answer questions about the missingness of our proposed linkage variables.

One issue was the quality and availability of caregiver information (as opposed to information for children). For CCOULD, much of the needed caregiver information, such as demographics and the caregiver's relationship to the child, was often missing or incomplete. Information on caregivers was less uniform and robust than information on children. Kentucky and Florida differed in the information collected from caregivers (such as SSNs); how they described the relationship of the caregiver to the child (for example, primary or secondary caregiver versus biological parent, foster parent, or grandparent); and the extent to which screened-in maltreatment reports collected or maintained caregivers' information. State child welfare agencies could place greater emphasis on collecting and maintaining detailed information about caregivers, including SSNs. However, this might depend on the operational necessity of collecting such information, as federal rules do not require caregivers to provide SSNs for child welfare services. The availability of caregivers' SSNs and demographic information increases the probability of linking caregivers to children, allowing for a more robust examination of child and family outcomes.

Medicaid datasets do not usually have a family identifier, so the Medicaid claims files cannot link children to parents or caregivers. Researchers first linked caregivers to the child welfare file to caregivers in the Medicaid file by first linking caregivers to children in the child welfare data. If there was a family identifier on the Medicaid data, we might have found more caregivers to have Medicaid coverage.

An additional barrier that linking projects must consider is the availability of standardized codes and definitions for key variables. Child welfare data systems generally lack standard codes or definitions in child welfare data systems. Medicaid claims data use standard nomenclature, known as the Healthcare Common Procedure Coding System, for identifying services providers use to bill for services. The American Medical Association and CMS developed and maintain this system. In contrast, ACF does not prescribe standard codes or definitions for the specific child welfare services reported to NCANDS or provided as Title IV-E prevention services, though states must report the service category (that is, mental health, substance use, or parenting support) of each Title IV-E prevention service in reporting child-specific data. States can differ in their definitions of child welfare-related interventions, such as substance abuse or mental health services, and information on the duration or intensity of service receipt is often lacking. This might be a natural consequence of Medicaid being both a federal- and state-funded system and of

deriving Medicaid data from a fee-for-service payment system. Child welfare providers are paid largely through state contracts or grants that can be reimbursed by federal funds but are not necessarily per-unit of service. As noted, including Title IV-E prevention-funded services in state data systems could, over time, increase the standardization and specificity of prevention-related child welfare services. Title IV-E prevention requirements for evidence-based services could advance the quality and completeness of certain service data collected and reported by child welfare agencies. However, Title IV-E prevention requirements will not capture all the services provided by child welfare. Using the T-MSIS, AFCARS, and NCANDS as a foundation and then refining with the states as needed to reflect their unique data variables and structures, RTI created a CDM that enabled merging the data from two state Medicaid and child welfare systems into one dataset. Future projects should build on and enhance the CDM format developed from the CCOULD project.

Aside from the quality of the data, documentation is critical for the success of a data-linking project. Child welfare agencies might not have mature data dictionaries and user guides they can share. In this project, the lack of documentation resulted in greater time demands on state and local agency personnel with institutional knowledge of the data who had to explain the data elements and infrastructure to the team charged with creating the linked dataset.

States Need Support in Strengthening and Harmonizing Data Infrastructure on Child Welfare Services

Many state child welfare systems provide services through different mechanisms and a heterogeneous network of providers. Consequently, the data systems used to collect information on families receiving these services are often not harmonized. This is particularly the case for foster care prevention services, which are relatively new in many states. When data systems are not harmonized, and particularly if they lack complete information on cases and services provided, linking efforts might not be able to provide a complete picture of the service array that families receive.

One of the purposes of linking data in CCOULD was to get a complete picture of the services families received funded by child welfare agencies and Medicaid. During the period of CCOULD, states were in the process of incorporating Title IV-E prevention services information from external providers into their state data systems. Detailed information—including cost of services—was not yet available on prevention services provided to children and families. As states begin to implement the Title IV-E prevention program, information about the types of evidence-based prevention services provided, service dates, and service costs will be more readily available. Future data-linkage efforts could allow for examining the quality and utility of these data. Another lesson on data quality was that states could not readily provide detailed information on the Title IV-E prevention services provided. This was due external service provider systems storing service and cost data, child welfare services receiving payment via lump sum contracts rather than fee-for-service, and the fact that child welfare services do not currently use standardized codes, as medical services typically do (such as Current Procedure Terminology codes). To support analysis and evaluation of state child welfare data, states should

maintain high-quality data dictionaries. States might be able to use the materials that ACF has created, such as technical bulletins on data exchange standards to develop data dictionaries.³

Consider Resources Needed to Use Probabilistic Matching to Link Data Sources

To encourage states to use probabilistic matching, it is important for staff to understand the principles, strengths, and limitations of different data-linking methods. The CCOULD team proposed to use probabilistic matching to yield more linked data, but the states did not have the time or resources to pursue it. Probabilistic matching might have yielded additional linkages not captured by direct matching using SSNs or the long-identifier. However, the project timeline did not allow for an exploratory analysis of probabilistic matching methods. Future efforts should include resources to support states in incorporating these methods through, for example, technical assistance and tools to implement probabilistic matching approaches. SSNs for deterministic matching were available for the great majority of children, but they were less frequently available for caregivers. Although the project successfully linked thousands of children and caregivers in child welfare to their records in Medicaid, probabilistic methods would have likely yielded more linkages.

Based on the experience of the two states in this study, SSNs are well populated and of high quality in child welfare records. This might not be the case in other states. Kentucky pointed out that its high rate of SSNs might be a result of its decision to structure one of its public assistance divisions with the Department for Community-Based Services, the same department that contains the state’s child welfare division. Staff in the same division can help enroll people in benefits, where SSNs are critical, and investigate allegations of child abuse and neglect. These staff could have established practices that are more diligent about collecting SSNs.

VII. Directions for Future Medicaid and Child Welfare Data Infrastructure

The prior section provided lessons learned for specific projects working with states to link data. This section points out some broad directions for improving the ability of states to conduct and sustain that linking. It also points to new efforts that build directly on the success of CCOULD.

Considerations for Improving State Readiness for Sustained Linking Efforts

Data-linking efforts among health and human services programs have become more widespread in part because they help researchers and government agencies use data to understand the populations served and improve services. For child welfare services, identifiable microdata is not available at the federal level; research must rely on data from state systems. In the case of child welfare data, this project found many states are not well positioned to initiate and sustain linking efforts on their own. The states in this study faced limitations that are likely common to many states, as the project team learned through project’s

³ For example, see CCWIS Technical Bulletin #2, “Data Sharing Between CCWIS and Child Welfare Contributing Agencies,” available at: <https://www.acf.hhs.gov/cb/training-technical-assistance/ccwis-technical-bulletin-2>

state recruitment efforts and in conversations with other stakeholders through presentations and other forums.

One primary challenge states face is the lack of IT infrastructure and human capital to perform data-linking work in a continuous or repeatable way. Many states work with outdated technology not intended to facilitate this type of work. States need investment in modern technological solutions to meet privacy and security needs while also providing necessary capabilities. Often, states have a lack of subject area expertise on staff and face difficulty getting trusted external assistance. The cost of support from contractors can prohibit large-scale linking efforts.

Another limitation is the agency structure and current division of ownership and authority, which makes collaboration difficult. For many states, how they collect and store child welfare data presents problems to the sustainability of data-linking efforts. Data often reside in fragmented systems that exist at the point of collection—such as with a contracted service provider—rather than a single, centralized system. These individual systems do not always communicate with one another. Even with the best intentions and carefully crafted data use agreements, these silos stand in the way of making all necessary IT and infrastructure enhancements and cause fragmented knowledge bases to persist. Paired with siloed ownership of these data, it is difficult to implement technological solutions that can gather and store all relevant data.

Federal Efforts Underway that Build on CCOULD

Predating and concurrent with the CCOULD project, OPRE led the [Child Maltreatment Incidence \(CMI\) Data Linkages Study \(2017–2022\)](#) to explore “how innovative administrative data linkages can improve our understanding of the incidence of child maltreatment and related risk and protective factors”. CMI Data Linkages identified five sites using linked administrative data to examine CMI and related risk and protective factors. The project supported these sites as they enhanced their existing linked administrative data by using innovative methods to link and analyze administrative data, linked novel administrative data sources, or scaled or replicated an existing data linkage or analysis approach in a new geographic area or jurisdiction. Researchers published several papers as a result of this study, including a report titled “[Understanding the Effect of the Opioid Epidemic on Child Maltreatment](#)” (2021), which linked data from birth records, hospital admissions, death records, and Child Protective Services records in the State of Washington to learn more about the relationship between opioid use disorders and the risk of child maltreatment.

ASPE and ACF conceptualized CCOULD as a pilot project to link data in a few states. The success this pilot had in linking data in Florida and Kentucky provoked a new, larger investment to link data in more states and develop more lasting infrastructure from the HHS Office of the Secretary, outside of the PCORTF portfolio. In 2022, ASPE began a new project, [Child Welfare and Health Infrastructure for Linking and Data Analysis of Resources, Effectiveness, and Needs \(CHILDREN\)](#), funded by the Health Care Fraud and Abuse Control Program. Building directly on the lessons from CCOULD, the new project has four objectives:

1. Conduct feasibility studies to determine states' or tribes' readiness and infrastructure to develop data systems with linked records from Medicaid and public child welfare agencies, and detail steps to improve readiness.
2. Develop datasets and data systems with linked records from Medicaid and public child welfare agencies for children and parents or caregivers involved in both systems.
3. Use linked records to (a) Increase analytic opportunities to help better understand and improve child welfare operations, services, outcomes, and health services delivered to children and parents in the child welfare system; and (b) develop, pilot, and implement models to support oversight of prevention and congregate care services.
4. Document lessons learned on linking data across the agencies and the project more generally.

In addition, ASPE is taking steps to increase researchers' use of the CCOULD data. In partnership with the Institute for Research on Poverty at the University of Wisconsin-Madison, ASPE is forming a research consortium on child welfare and Medicaid use and outcomes.⁴ The consortium will fund four or five researchers to conduct research using CCOULD, and along with HHS researchers using the dataset, will host regular meetings to align research studies, share lessons about working with the data, improve technical documentation, increase exposure to the utility of the dataset, and generate findings to support patient-centered outcomes research.

VIII. Conclusion

CCOULD sought to support patient-centered outcomes research by creating a new data infrastructure to study families involved in child welfare systems and Medicaid. The project resulted in the development of a new dataset that linked Medicaid records with child welfare records for parents and children involved in the child welfare system, guidance for future linking efforts, and example analyses that can guide future research.

The original motivation behind the project was the recognition that children with parents dealing with SUD are at particular risk of child maltreatment, poor child welfare outcomes, and poor health outcomes. States can use Medicaid to fund SUD treatment interventions, as well as other wraparound services for children and their families to help mitigate the risks posed by SUD. Researchers can use the CCOULD dataset for many purposes:

- Understand the use of Medicaid services for behavioral health services, in particular SUD treatment.
- Learn how those services relate to services provided by child welfare systems.
- Gain understanding on how they relate to parent and child health and stability outcomes.
- Allow testing of the effectiveness of new interventions to provide treatment and recovery services.

⁴ For the call for proposals, see: <https://www.irp.wisc.edu/irp-extramural-large-grants-2024-2026-call-for-applications/>

- Understand how agencies can leverage the new use of Medicaid to cover institutions for mental disease for SUD treatment for parents involved in the child welfare system, and how it improves child welfare outcomes.

The lessons learned over the course of the CCOULD project serve as an equally important project outcome. Each challenge the project team faced provided key lessons that other states can use to successfully undertake data-linking work. These challenges also highlighted key barriers that states face and point to areas where the federal government or other entities can assist states attempting to undertake this work.

In conclusion, the experience in linking data from two states through the CCOULD project provides a solid foundation for future data linking efforts. As the value of integrating data across Medicaid and child welfare program becomes more apparent, more and more states can build from this foundation.

References

- Ahrens, K. R., Garrison, M. M., & Courtney, M. E. (2014). Health Outcomes in Young Adults From Foster Care and Economically Diverse Backgrounds. *Pediatrics*, *134*(6), 1067–1074.
- Arria, A. M., Mericle, A. A., Meyers, K., & Winters, K. C. (2012). Parental substance use impairment, parenting and substance use disorder risk. *Journal of Substance Abuse Treatment*, *43*(1), 114–122. <https://doi.org/10.1016/j.jsat.2011.10.001>
- Australian Government, Australian Institute of Health and Welfare. (n.d.). SLK-581 guide for use. <https://www.aihw.gov.au/getmedia/e1d4d462-8efa-4efa-8831-fa84d6f5d8d9/aodts-nmds-2016-17-SLK-581-guide.pdf.aspx>
- Clary, E., Ribar, C., Weigensberg, E., Radel, L., & Madden, E. (2020). Challenges in providing substance use disorder treatment to child welfare clients in rural communities. US Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation.
- Coulson, T. G., Bailey, M., Reid, C., Shardey, G., Williams-Spence, J., Huckson, S., Chavan, S., & Pilcher, D. (2021). Linkage of Australian national registry data using a statistical linkage key. *BMC Medical Informatics and Decision Making*, *21*(1), 37. <https://doi.org/10.1186/s12911-021-01393-1>
- Dusetzina, S. B., Tyree, S., Meyer, A.-M., Meyer, A., Green, L., & Carpenter, W. R. (2014). *Linking Data for Health Services Research: A Framework and Instructional Guide*. Agency for Healthcare Research and Quality (US). <http://www.ncbi.nlm.nih.gov/books/NBK253313/>
- Ghertner, R., Madden, E., & Crouse, G. (2020). Availability for treatment for opioid use disorder in areas of high foster care increases. Office of the Assistant Secretary of Planning and Evaluation, Washington, DC: US Department of Health and Human Services.
- Leventhal, A. M., Pettit, J. W., & Lewinsohn, P. M. (2011). Familial Influence of Substance Use Disorder on Emotional Disorder across Three Generations. *Psychiatry Research*, *185*(3), 402–407. <https://doi.org/10.1016/j.psychres.2010.08.013>
- Radel, L., Baldwin, M., Crouse, G., Ghertner, R., & Waters, A. (2018). Substance use, the opioid epidemic, and the child welfare system: Key findings from a mixed methods study. *Office of the Assistant Secretary for Planning and Evaluation*, 1-9.
- Raghavan, R., Brown, D. S., Allaire, B. T., Ross, R. E., & Landsverk, J. (2016). Associations Between Magnitude of Child Maltreatment and Medicaid Expenditures for Psychotropic Medications. *Psychiatric Services*, *67*(8), 916–919. <https://doi.org/10.1176/appi.ps.201500211>
- Rebbe, R., Nurius, P. S., Courtney, M. E., & Ahrens, K. R. (2018). Adverse Childhood Experiences and Young Adult Health Outcomes Among Youth Aging Out of Foster Care. *Academic Pediatrics*, *18*(5), 502–509.
- Sieger, M. L., Becker, J., Philips, J., Lee, J. W., & Moore, T. E. (2023). Latent classes among substance-involved families in child welfare: Associations with treatment completion and reunification. *Children and*

Youth Services Review, 150, 107010. Weigensberg, E., et al. Superutilization of Child Welfare, Medicaid, and Other Services. (2018). Mathematica Policy Research. [Superutilization of Child Welfare, Medicaid, and Other Services \(mathematica.org\)](#).

Zlotnick, C., Tam, T. W., & Soman, L. A. (2012). Life Course Outcomes on Mental and Physical Health: The Impact of Foster Care on Adulthood. *American Journal of Public Health*, 102(3), 534–540.