Linking Administrative Data to Improve Understanding of Child Maltreatment Incidence and Related Risk and Protective Factors: A Feasibility Study

November 2021

Andrea Mraz Esposito, Claire Smither Wulsin, Leah Pranschke, Andrew Burwick, and Rebekka Nickman

---

Submitted to:

Office of Planning, Research, and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services
200 Independence Avenue, S.W.
Washington, D.C. 20201
Attention: Jenessa Malin and Christine Fortunato
Project Officers
Contract number: HHSP233201500035I
OPRE report number: 2021-251

Submitted by:

Mathematica
111 East Wacker Drive, Suite 3000
Chicago, IL 60601-4303
Phone: (312) 994-1002
Fax: (312) 994-1003
Matt Stagner, Project Director

This page has been left blank for double-sided copying.

Contents

Tables

Figures

OVERVIEW

Introduction

Accurate and ongoing surveillance of the incidence of child maltreatment and related risk and protective factors can help to inform policy and programs as well as shape prevention and intervention efforts. One promising approach to capturing this information is by linking local, state, or federal administrative records, such as those from child welfare, health, social services, education, public safety, and other agencies.

The Child Maltreatment Incidence Data Linkages (CMI Data Linkages) project identified five research groups (sites) with experience using linked administrative data to examine child maltreatment incidence and related risk and protective factors and supported these sites to enhance their approaches to administrative data linkage through acquisition of new data sources, use of new methods, or replication of existing methods. The CMI Data Linkages project team conducted a study to assess the feasibility of enhancing data linkage and analysis efforts to produce new information on child maltreatment and to identify promising practices and contextual and organizational factors related to using linked administrative data.

Primary Research Questions

Three overarching research questions motivated the study:

1. What are promising methods or practices, within and across the sites, for linking administrative data to inform the incidence of child maltreatment and related risk?
2. What contextual and organizational factors promote or impede enhancement of existing administrative data linkages?
3. What novel information about child maltreatment incidence and related risk and protective factors can be gleaned through enhancement of linked administrative data?

Methods

The study team – researchers at Mathematica and Washington University in St. Louis - identified key activities involved in the data linking and analysis process that the five sites would undertake to design and implement their enhancements: (1) developing research questions and exploring data partnerships, (2) sharing and accessing data, (3) preparing data sets and completing data linkages, (4) conducting analyses, and (5) reporting results. The study team also identified factors that might influence the sites' work. Using the activities and potential factors as a guiding framework, the study team conducted a qualitative, cross-site feasibility study involving coding, collecting, and analyzing data from multiple sources to gather information on sites' experiences. The team interviewed key respondents from participating sites once in the early phase of each project (within four months of the start date) and a second time approximately one year later. Respondents included principal investigators (PIs) and co-principal investigators (co-PIs) from each site; administrators, directors, or managers at partnering government agencies or other organizations that provided administrative data; information technology or data managers; and research staff. Additional sources of data included site documents, notes and materials from cross-site learning network meetings and other technical assistance activities, and quarterly questionnaires on sites' use of resources.

Key Findings and Highlights

- The feasibility study identified promising methods or practices, within and across the sites, for linking administrative data to inform understanding of the incidence of child maltreatment and related risk.
- Contextual and organizational factors including child welfare system structures, child welfare policies and definitions, the legal and policy contexts for data use, and the existing data infrastructure influenced the feasibility of enhancing data linkages.
- The experiences of the sites offer evidence that enhancing administrative data linkages is a feasible approach to addressing high-priority questions about child maltreatment incidence and related risk and protective factors. Each site was able to accomplish its intended enhancement and yield novel information from it.

EXECUTIVE SUMMARY

Accurate and ongoing surveillance of the incidence of child maltreatment and related risk and protective factors can help to inform policy and programs as well as shape prevention and intervention efforts. In 2014, the Institutes of Medicine (IOM) and the National Research Council (NRC) issued a report underscoring the critical need to determine the true incidence and prevalence of child maltreatment in order to understand the magnitude of the problem, at both a national- and state-level (Institute of Medicine and National Research Council 2014). The report also highlighted the importance of understanding individual-, family-, and community-level risk and protective factors associated with child abuse and neglect.

The Child Abuse Prevention and Treatment Act (CAPTA), first authorized in 1974 and reauthorized regularly since then, requires the examination of a wide range of topics related to the incidence of child abuse and neglect with the aim of informing efforts to better protect children from maltreatment and improve the well-being of victims of maltreatment. These topics include, but are not limited to, trends in number and severity of cases, incidence of screened out and investigated (substantiated and unsubstantiated) cases, and incidence and prevalence of maltreatment by a wide variety of demographic characteristics.[1]

Although there have been many advances in the literature regarding the incidence of child abuse and neglect and the types of related risk, much remains unknown. Developing an accurate estimate of its incidence and achieving a full understanding of the risk and protective factors involved continue to be challenges for the field. Various national sources of child maltreatment incidence data currently exist, including administrative data and surveys. For example, the National Child Abuse and Neglect Data System (NCANDS) examines annual trends in child abuse and neglect for cases reported to child protection authorities. The National Incidence Survey (NIS), last conducted in 2006, measured the number of children reported for maltreatment, provides information on unreported cases from mandated reporters, and examines characteristics of the maltreated population. However, these data sources do not provide a complete picture of the scope and characteristics of child maltreatment and they struggle to account for unreported cases and variation in child welfare policies and practices across states and over time.

One promising approach to addressing the limitations of existing data sources is by linking local, state, or federal administrative records, such as those from child welfare, health, social services, education, public safety, and other agencies. This approach may help improve the quality, usefulness, interoperability, and availability of child maltreatment data. The Office of Planning, Research, and Evaluation (OPRE) and the Children's Bureau (CB) within the Administration for Children and Families (ACF) sponsored the Design Options for Understanding Child Maltreatment and Related Risk and Protective Factors (Design Options) project.[2] With input from key stakeholders and experts, this project identified research questions related to child maltreatment and related risk of importance for policy and practice and identified potential survey and administrative data sources that could be leveraged to examine the research questions of interest. One key finding from the project was the potential to use linked administrative data to obtain accurate information on the incidence of child maltreatment (Institute of Medicine and National Research Council

---

[1] For a full list of topics see CAPTA Sec. 105. Research and Assistance available at https://www.govinfo.gov/content/pkg/USCODE-2017-title42/html/USCODE-2017-title42-chap67.htm.

[2] For more information on the Design Options project, see https://www.acf.hhs.gov/opre/project/design-options-understanding-child-maltreatment-incidence-2015-2017.

2014; Jonson-Reid et al. 2016). However, the project determined additional work was needed to understand how innovative approaches to linked administrative data could be leveraged to advance the child maltreatment knowledge base and brought to scale.

Building on this work, OPRE in partnership with the CB funded Mathematica and its partners at Washington University in St. Louis to conduct the Child Maltreatment Incidence Data Linkages (CMI Data Linkages) project. This project aimed to examine how linking administrative data may improve the ongoing and accurate surveillance of child abuse and neglect. OPRE and the project team used several methods to develop a preliminary list of possible sites,[3] including posting a call for innovative data linking projects in OPRE's newsletter and searching for recent literature focused on integrated data. Five sites were selected using the following criteria: alignment with CMI Data Linkage project objectives; access to and quality of administrative data; experience conducting research with administrative data; proposed enhancement; organizational leadership and capacity; ability to conduct the proposed project in the specified time frame; interest in undertaking the project; willingness to assist with dissemination and communication; and variability in contextual factors.

The five sites were supported to enhance their administrative data linkages by:

- Using innovative methods to link and analyze administrative data;
- Linking novel administrative data sources; or
- Replicating an existing data linkage or analysis approach in a new geographic area or jurisdiction.

Support included funding, technical assistance (TA); participation in a cross-site learning network (CSLN); and webinars, where sites received feedback on their projects from experts in the field.

The CMI Data Linkages project team used information from the sites' experiences to assess the feasibility of enhancing data linkage and analysis efforts to produce new information on child maltreatment. This report presents findings from the CMI Data Linkages Feasibility Study (Feasibility Study).

A. Research questions

Three overarching research questions guided this study:

1. What are promising methods or practices, within and across the sites, for linking administrative data to inform understanding of the incidence of child maltreatment and related risk?
2. What contextual and organizational factors promote or impede enhancement of existing data linkages?
3. What novel information about child maltreatment incidence and related risk and protective factors can be gleaned through enhancement of linked administrative data?

B. The CMI Data Linkages sites

The CMI Data Linkages project team identified five sites that proposed projects featuring different types of enhancements to existing administrative data linkage efforts to address questions about child maltreatment incidence and the related risk and protective factors (Table ES.1). This report refers to sites by using the name of the lead research organization(s): Alaska Department of Health and Social

---

[3] The project team defined a site as a group comprising a researcher, investigator, research group, or center affiliated with a university, nonprofit organization, or public agency. The sites could include representatives of agencies or organizations that collect administrative data and make the data available (also known as data stewards).

Services/Oregon Health Sciences University (ADHHS/OHSU), Children's Data Network/California Child Welfare Indicators Project (CDN/CCWIP), CDN/Rady Children's Hospital (CDN/Rady), Center for Social Sector Analytics and Technology (CSSAT), and University of Alabama School of Social Work (UA-SSW).

Table ES.1. CMI Data Linkages project titles and descriptions

Replicating the Alaska Longitudinal Child Abuse and Neglect Linkage (ALCANLink) methodology (ADHSS/OHSU)	The ALCANLink approach used a population-based mixed-design strategy to integrate two sets of data: (1) those births that were sampled and mothers who subsequently responded to the Pregnancy Risk Assessment Monitoring System survey and (2) child welfare and other administrative data. Alaska partnered with Oregon to replicate this methodology and to estimate and compare, the cumulative incidence to first report, screen-in, substantiation, and removals by age 9.
Methods to estimate the community incidence of child maltreatment (CDN/CCWIP)	This site focused on developing a methodology that used administrative data to estimate the number of children who were victims of abuse or neglect. The site produced upper and lower bounds of estimates that reflected the number of children who were identified by the child welfare system as victims of abuse or neglect as well as those who were victims but not identified as such by the system. The site tested the methodology using data from California and explored the potential for using it in other states.
Using hospital data to predict child maltreatment risk (CDN/Rady)	This site tested the predictive value of integrating hospital data with vital birth records, statewide child protection records, and vital death records to identify children who may be at an elevated risk of maltreatment. The site focused on validating a statewide predictive risk model by determining the extent to which children identified to be "at high-risk" of maltreatment are also at elevated risk of injury, poor health outcomes, and mortality in childhood. The site used machine learning methods to train probabilistic algorithms for linking hospital-system data to other administrative data sources. These data linkages aimed to better characterize the demographics and public service trajectories of Rady Children's Hospital patients.
Understanding the effect of the opioid epidemic on child maltreatment (CSSAT)	This site contributed to the knowledge about the opioid epidemic's potential effects on child maltreatment by examining the associations among multiple indicators of child maltreatment, child welfare system involvement, and individual- and community-level risk factors from several data sources across Washington.
Examining child maltreatment reports using linked county-level data (UA-SSW)	This site examined how risk and protective factors relate to child maltreatment reports at the county level across the nation. The site linked county and state-level data from the National Child Abuse and Neglect Data System (NCANDS) to county- and state-level data from US Census, Bureau of Labor Statistics, Center for Disease Control, National Center for Health Statistics, and other sources. The site aimed to explain widely varying state- and county-level maltreatment rates and to develop valid ways to operationalize county-level child maltreatment risk.

ADHSS = Alaska Department of Health and Social Services; CCWIP = California Child Welfare Indicators Project; CDN = Children's Data Network; CSSAT = Center for Social Sector Analytics and Technology; OHSU = Oregon Health Sciences University; Rady = Rady Children's Hospital-San Diego; UA-SSW = University of Alabama School of Social Work.

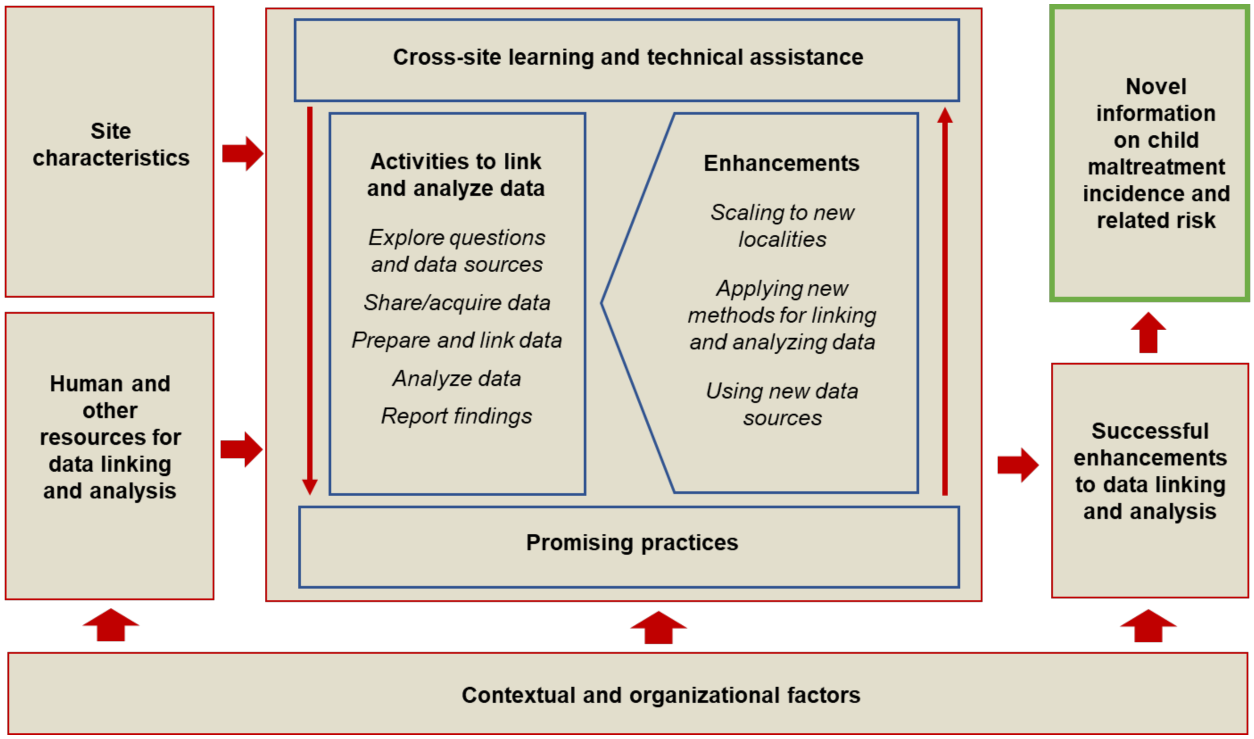
C. Study methods

To guide the Feasibility Study, the project team developed a conceptual model (Figure ES.1) based on the main activities that the sites would undertake to design and implement their data linkage and analysis enhancements, as well as the factors that might influence these activities. Sites were expected to engage in five main activities as they completed their projects:

1. *Developing research questions and exploring data partnerships*. In this phase, the Feasibility Study team anticipated that sites would define the focus of their research projects, identify potential data sources, and begin establishing relationships for accessing data, if necessary.
2. *Sharing and accessing data*. Activities during this phase might include developing data sharing agreements, establishing approaches to protecting confidentiality and privacy of data, and securing necessary approvals, such as approvals from an institutional review board (IRB).
3. *Preparing data sets and completing data linkages*. This phase might involve assessing the characteristics of data sets, processing and cleaning data, and implementing data linkage methods.
4. *Conducting analyses* to answer research questions. In this phase, sites would use a variety of methods to analyze the linked data sets and would determine the key findings from those analyses.
5. *Reporting the results of their research* through various dissemination activities. This phase might include identifying audiences for reporting and presenting results of analyses in a range of formats.

The conceptual model (Figure ES.1) depicts the hypothesized pathway by which sites successfully enhance their data linkage or analysis and ultimately produce novel information on child maltreatment incidence and related risk. This conceptual model guided the collection and analysis of data for the Feasibility Study. The chapters in this report align with the boxes in the figure, exploring site characteristics (Chapter 2); activities to link and analyze data, enhancements, and promising practices (Chapter 3); contextual and organizational factors (Chapter 4); and successful enhancements to data linking and analysis that lead to novel information on child maltreatment incidence and related risk (Chapter 5).

Figure ES.1. Conceptual model



To address the Feasibility Study research questions, the project team conducted a qualitative, cross-site study involving coding, collecting, and analyzing data from multiple sources to gather information on sites' experiences as they conducted their projects from 2019 through 2021 (Table ES.2). The Feasibility Study team analyzed the data by coding the documents, notes, and interviews using NVivo qualitative analysis software. Through multiple rounds of analysis and triangulation across data sources, the Feasibility Study team refined the themes and findings that are presented in this report. To analyze quantitative data from the questionnaire on resource use, the study team calculated descriptive statistics and conducted cross-site comparisons.

Table ES.2. Data sources and descriptions

Data source	Description
Site documents	The study team reviewed documents such as project plans, interim memos, and final memos. These documents included appendices with materials such as data use agreements (DUAs), descriptions of administrative data sets, and data dictionaries.
Notes from CSLN, expert webinars, and TA	The study team reviewed documents from the CSLN activities and webinars with experts, such as slide presentations, as well as TA meeting notes.
Interviews with key informants	The study team conducted interviews with representatives of the participating research groups, agencies, and organizations, once early in the early phase of each project and a second time approximately one year later. The team used semi-structured protocols to ensure that the interview questions aligned well with the study constructs and that a consistent set of topics was covered across all sites.
Questionnaire on resource use	To systematically gather details on the resources necessary for sites to successfully link and analyze their data, the team administered a quarterly questionnaire during the project period, for a total of five rounds.

CSLN = cross-site learning network; TA = technical assistance.

D. Key study findings

RQ1: What are promising methods or practices, within and across the sites, for linking administrative data to inform understanding of the incidence of child maltreatment and related risk?

The study team identified promising practices for each of the five project activities (Table ES.3). Site representatives and the study team identified these practices as distinct strategies or actions that helped them complete each activity.

Table ES.3. Promising methods and practices for linking administrative data

Phase	Promising practices
1. Developing research questions and exploring partnerships	Including partners, especially data partners, in early discussions may help researchers strengthen their collaborations. This process also offers an opportunity to highlight the benefits of data sharing to new data partners and develop buy-in for the project.
	Preexisting relationships with data partners and familiarity with data sets can inform the focus and scope of projects to enhance data linkages.
2. Sharing and accessing data	Researchers can build trust with data partners through familiarity with laws and regulations regarding data access and by collaborating with liaisons within public agencies.
	Organizations that are interested in enhancing data linkages may be able to modify or amend existing data use agreements (DUAs) or research permissions to conduct their work. A broadly specified IRB that covers analyses of linked administrative data may also facilitate these types of projects.
	The PI's experience and knowledge of the IRB process facilitates the approval process, as does the competency of university IRB processes in general. This experience can be leveraged to require fewer revisions to the submitted IRB package.
	Although DUAs often require stringent data security protocols, research centers working with administrative data might already have such protocols in place. Collaborations with external entities to conduct data linkages are an additional means to ensure the privacy and confidentiality of personally identifiable information.
	Using publicly available data or data that are frequently accessed by a variety of users (where there are established procedures in place to access the data) can simplify data sharing.
	Plans and timelines for projects involving administrative data linkages should anticipate delays, especially related to data acquisition, and identify opportunities to accelerate other activities or prepare for analyses.
3. Preparing data and completing linkages	Accessing specialized expertise on the content of administrative data sets supports the preparation for linkages.
	A variety of approaches may be useful for linking administrative data. Technical familiarity with the data and the flexibility to tailor approaches to the content of a specific data source may be important to successful linkages.
	When working with a third party to complete linkages, clear communication regarding the linkage approach can help the research team have confidence that linkage algorithms will operate as expected.
	Machine learning techniques and tools may be an efficient method for linking larger databases.
4. Conducting analyses	Researchers can adapt and use a wide range of analysis methods to explore child maltreatment with administrative data, including Bayesian model averaging and latent transition analyses.
	Deep familiarity with child welfare data and reflection about on-the-spot interim analyses can support exploratory approaches to estimating incidence.
	Researchers can consider mobility in sample populations and jurisdictional differences in child welfare administrative practices, definitions, and policies when exploring results and interpreting analytic findings—but clear-cut solutions to these challenges can be difficult to identify.

Phase	Promising practices
5. Reporting results	Disseminating results in a variety of formats and channels may help ensure that the results of enhanced data linkages are shared with a range of audiences and stakeholders.
	Prioritizing the reporting of findings to data partners may improve partnerships, build trust, and support policymakers in applying the results of research on child maltreatment and incidence.
	Dissemination may also support replication of methods to enhance administrative data linkages to generate knowledge on child maltreatment incidence and risk and protective factors.

IRB = institutional review board.

RQ2: What contextual and organizational factors promote or impede enhancement of existing data linkages?

Sites conducted their projects in different contexts and brought different organizational capacities to their work. The Feasibility Study investigated the influence of contextual factors, such as characteristics of the child welfare systems in the jurisdictions that were the focus of the sites' analyses, child welfare policies and definitions that affected the accessibility and content of administrative databases, state and local statutes that governed data access and use, and whether an integrated data system or other infrastructure for linking administrative data existed in a site.

- **Child welfare system structures.** The overall stability in child welfare systems appears to have facilitated sites' work to enhance administrative data linkages. The structure of child welfare systems affected how some sites articulated their research questions and approached their analyses of administrative data.
- **Child welfare policies and definitions.** Variation in child welfare policies and definitions across jurisdictions was a challenge for sites as they analyzed data. However, familiarity with the data and prior knowledge of the potential variation allowed researchers to consider this challenge as they planned and implemented their projects. In some cases, jurisdictions' policies related to child welfare facilitated efforts to use administrative data to explore child maltreatment incidence.
- **Legal and policy contexts for data use.** Local policies or regulations on data use affected the design and implementation of some projects. In some sites, laws or policies created circumstances that required adjustments to project plans; in others, the legal and policy context facilitated sites' efforts.
- **Existing infrastructure for data linkage.** Several states where sites conducted their work support clearinghouses or maintain agencies with the specific purpose of integrating administrative data from multiple state agencies.

Organizational capacities are the attributes or abilities that might enable entities participating in a CMI Data Linkages project to complete the intended enhancements and accomplish the project's goals. The Feasibility Study explored several types of capacities that research groups, organizations, or agencies participating in the process might require to be successful, including leadership, effective partnerships, technical infrastructure and expertise, and the ability to translate the results of research into policy.

- **Leadership to promote data linking.** In the sites that accessed individual-level data from states, support among policymakers and state agency leaders for data-driven policymaking and quality improvement efforts led data stewards to participate in the CMI Data Linkages project.
- **Partnership quality.** A history of data sharing between partners was important but not sufficient for successful partnerships. For new and existing partnerships, a collaborative relationship characterized by mutual trust between data stewards and researchers was critical to facilitating the data linkage

enhancements. The quality of relationships with the data stewards was as important or more important than the specifics of the data sharing agreement in promoting ongoing collaboration. Less robust partnerships with data stewards appeared to affect some sites' ability to gain access to data.

- **Technical expertise.** By and large, representatives of the sites reported that they had the technical capacity or expertise necessary to accomplish their enhancements. The sites drew on a unique body of expertise and level of existing infrastructure to complete their projects.
  - *Legal and regulatory expertise.* Team members from the sites' lead organizations had the necessary expertise to facilitate the development of data sharing agreements and research approvals, despite not always having specialized legal or regulatory infrastructure or support within their organizations. The sites' data partners had authorities designated to determine the legality of and provide support for sharing data. Team members in several sites noted that their previous involvement in IRB committees offered expertise that was useful to their projects.
  - *Data management infrastructure and expertise.* Sites generally had an existing, robust infrastructure for data management, which allowed them to accomplish the work involved in the CMI Data Linkages projects.
  - *Data linkage and analytic expertise.* All the sites had existing linkage protocols to complete their enhancements. Their deep familiarity with some of the data sets or elements within the data sets was critical to accurate linkage, analysis, and interpretation of the data.
- **Translation capacity.** The sites had a track record of communicating findings based on administrative data linkages to key audiences in the child welfare field.
- **Access to and sufficiency of resources for conducting data linkages projects.** Sites benefitted from in-kind contributions from institutions or agencies that hosted the research center or team involved in the CMI Data Linkages projects. Representatives of several sites noted that the financial resources provided by a single grant typically would not be sufficient to cover the cost of projects like this.
- **Cross-site learning and TA.** Sites reported that they valued the opportunity to learn from their peers, appreciated the stature their participation in the federal project brought to their work, and benefitted from the accountability structure that the larger CMI Data Linkages project imposed.

RQ3: What novel information about child maltreatment incidence and related risk and protective factors can be gleaned through enhancement of linked administrative data?

Sites reported novel findings from their CMI Data Linkages research projects across five outcomes that were aligned to the CAPTA research priorities:

1. **The incidence of child maltreatment.** Two sites (ADHSS/OHSU and CDN/CCWIP) produced new state-level estimates of incidence in Oregon and California, respectively. Two sites (CDN/Rady and UA-SSW) produced estimates of the incidence of child welfare system involvement in counties in San Diego, California, and nationwide, respectively.
2. **The incidence and prevalence of child maltreatment by demographic characteristics.** Findings from two sites (CDN/CCWIP and ADHSS/OHSU) add to knowledge of the characteristics of child maltreatment victims and their families. Sites also produced new information about disparities in the risk of children who experience maltreatment (CDN/CCWIP, ADHSS/OHSU, UA-SSW).

3. **Risk and protective factors related to child maltreatment.** Two sites (CSSAT and UA-SSW) offered novel information about community-level risk and protective factors related to maltreatment outcomes. Studies investigating these factors shed light on the causes, prevention, and sociocultural dimensions of child maltreatment—topics that are prioritized in the CAPTA research agenda.
4. **The trajectory of child maltreatment reports.** Longitudinal linkages of individual-level child welfare data allowed two sites (CDN/CCWIP and CSSAT) to examine transitions across outcomes among children involved in the child welfare system.
5. **Methods for estimating the incidence of child maltreatment.** Three sites employed novel methods for estimating the incidence and predicting the risk of child maltreatment. Two of the sites (CDN/CCWIP and ADHSS/OHSU) offered new means of understanding the nature and scope of child abuse and neglect and the national incidence of child abuse and neglect, both CAPTA priorities. The other site (CDN/Rady) validated a relatively new model for predicting risk of future child welfare involvement.

By drawing on the experiences of the CMI Data Linkages sites, this study identified key lessons learned about using administrative data linkages to better understand child maltreatment as well as considerations for future applications of linked administrative data (Table ES.4).

Table ES.4. Lessons learned from the CMI Data Linkages projects and considerations for future applications of linked administrative data

Topic	Lessons learned
Feasibility of enhancements to administrative data linkages	• The experiences of the CMI Data Linkages sites offer evidence that enhancing administrative data linkages—through acquisition of new data sources, use of new methods, or replication of existing methods—is a feasible approach to addressing high-priority questions about child maltreatment incidence and related risk and protective factors.
	• Sites benefitted from existing infrastructure and relationships, which took time and effort to establish and maintain prior to their involvement in the CMI Data Linkages project.
	• A distinctive component of the CMI Data Linkages project, cross-site collaboration, also supported enhancements by providing researchers a space to brainstorm and compare experiences and approaches.
Data sources and linking methods	• Hospital data may provide useful information about maltreatment injury or other health-related outcomes, though they can be challenging to obtain and interpret.
	• The sites' projects illustrated how linkages of varying levels of complexity—regarding the level of linkages and number of data sources—can yield new information for the field.
Potential of sites' approaches to linking and analysis to produce new knowledge	• Some project approaches illustrated how administrative data linkages may support efforts to improve estimates of children who experience maltreatment but are not included in data reported by child welfare agencies.
	• Sites' use of linked hospital data suggests that this approach could contribute to improving maltreatment surveillance.
	• Linkages involving data with geographic identifiers illustrated the potential for using administrative data to assess patterns of incidence and related factors at the county level—and to explore whether and how policy, practice, and demographic variation at that level might affect observed child maltreatment outcomes.
Challenges and limitations of sites' approaches	• Although sites' existing relationships, expertise, and infrastructure proved helpful in many circumstances, even existing relationships with data stewards did not guarantee smooth processes for sharing additional or new data.
	• Sites needed to adapt to changes in circumstances and address unforeseen challenges that affected their project plans.
	• While the sites' experiences underscored the value and potential of linked administrative data as a data source to better understand child maltreatment incidence, they also highlighted the assumptions and limitations inherent in these approaches.
Advantages and disadvantages of using linked administrative data to estimate maltreatment incidence	• The National Child Abuse and Neglect Data System provides data that are not currently available from other sources—specifically, standardized, national-level data on maltreatment.
	• Administrative data linkages to estimate maltreatment incidence may offer advantages relative to the National Incidence Survey. For example, linked administrative data may be able to account for unreported cases and variation in child welfare policies and practices across states and over time.

I. INTRODUCTION

Speaking in 1977 at the dedication of the headquarters building of what is now the U.S. Department of Health and Human Services (HHS), Senator Hubert H. Humphrey said, "The moral test of government is how that government treats those who are in the dawn of life, the children ..." Healthy development leads children to grow into productive adults and strong citizens. However, adverse events in childhood, such as abuse and neglect, have lifelong impacts that undermine children's progress toward a successful adulthood.

Child maltreatment remains a problem of grave concern in the United States. In fiscal year 2019, child protective services agencies nationwide received about 4.4 million referrals for allegations of abuse or neglect, which involved 7.9 million children. The referrals for about 656,000 of these children were substantiated and the children were labeled as victims of abuse or neglect (HHS 2021). Child maltreatment affects numerous individual and family outcomes related to health, employment, relationships, and self-sufficiency (Currie and Widom 2010; Danese et al. 2009; Jonson-Reid et al. 2012). For example, adults who were maltreated as children complete fewer years of education, have lower IQ scores, and are less likely to have a skilled job (Currie and Widom 2010). According to one estimate, the annual economic burden resulting from substantiated cases of child maltreatment is $428 billion, due to factors such as health care, criminal justice, and special education costs (Peterson et al. 2018).

Addressing the problem of child maltreatment requires a clear understanding of its scope and associated risk and protective factors. Accurate and ongoing surveillance of the incidence of child maltreatment and related risk and protective factors can help to inform policy and programs as well as shape prevention and intervention efforts. In 2014, the Institutes of Medicine (IOM) and the National Research Council (NRC) issued a report underscoring the critical need to determine the true incidence and prevalence of child maltreatment in order to understand the magnitude of the problem, at both a national- and state-level (Institute of Medicine and National Research Council 2014). The report also highlighted the importance of understanding individual-, family-, and community-level risk and protective factors associated with child abuse and neglect.

The Child Abuse Prevention and Treatment Act (CAPTA), first authorized in 1974 and reauthorized regularly since then, requires the examination of a wide range of topics related to the incidence of child abuse and neglect with the aim of informing efforts to better protect children from maltreatment and improve the well-being of victims of maltreatment. These topics include, but are not limited to, trends in number and severity of cases, incidence of screened out and investigated (substantiated and unsubstantiated) cases, and incidence and prevalence of maltreatment by a wide variety of demographic characteristics (Box I.1).identifies multiple research priorities related to child maltreatment, including improving surveillance of child maltreatment (Box I.1). However, challenges remain in developing an accurate estimate of the incidence of child maltreatment and achieving a full understanding of the risk and protective factors.

A. Current national data sources may not provide a complete understanding of the scope and characteristics of child maltreatment

Various national sources of child maltreatment incidence data, including administrative data and surveys, currently exist. The following are examples:

- The **National Child Abuse and Neglect Data System (NCANDS)** examines annual trends in child abuse and neglect for cases reported to child protection authorities and screened in (or investigated). All 50 states, the District of Columbia, and Puerto Rico voluntarily submit administrative child welfare data to NCANDS. States provide two files to NCANDS. The NCANDS Child File includes case-level data on reports of alleged maltreatment that were investigated or received an alternative response, including maltreatment types and dispositions, child and caregiver risk factors, services provided, staff data, and perpetrator data. The NCANDS Agency File contains agency-level aggregate statistics, including data on prevention services, screened out referrals, family preservation and reunification services, maltreatment deaths, and children referred for early intervention services through Part C of the Individuals with Disabilities Education Act.
- The **National Incidence Survey (NIS)** measures the number of children maltreated and examines characteristics of that population. It has been conducted four times between 1979 and 2006. The NIS examines reported cases plus information on unreported cases from "sentinels," who are community professionals such as educators who are mandated by law to report suspected child abuse and neglect (Sedlak et al. 2010).
- **U.S. Department of Justice (DOJ)** surveys such as the annual National Crime Victimization Survey (NCVS; Bureau of Justice Statistics 2019) and the National Survey of Children's Exposure to Violence[4] (NSCEV; Office of Juvenile Justice and Delinquency Prevention 2020) ask children (NatSCEV) and teens (NCVS and NatSCEV) and adults (NCVS) about reported and unreported crime and abuse.

---

[4] The NSCEV was conducted three times from 2008 to 2014.

Box I.1. CAPTA research agenda related to the incidence of child maltreatment

CAPTA supports research focused on the national incidence of child maltreatment, including the following (U.S. Congress 2017):

- the extent to which incidents of child abuse and neglect are increasing or decreasing in number and severity;
- the incidence of substantiated and unsubstantiated reported child abuse and neglect cases;
- the number of substantiated cases that result in a judicial finding of child abuse or neglect or related criminal court convictions;
- the extent to which the number of unsubstantiated, unfounded and false reported cases of child abuse or neglect have contributed to the inability of a State to respond effectively to serious cases of child abuse or neglect;
- the extent to which the lack of adequate resources and the lack of adequate training of individuals required by law to report suspected cases of child abuse and neglect have contributed to the inability of a State to respond effectively to serious cases of child abuse and neglect;
- the number of unsubstantiated, false, or unfounded reports that have resulted in a child being placed in substitute care, and the duration of such placement;
- the extent to which unsubstantiated reports return as more serious cases of child abuse or neglect;
- the incidence and prevalence of physical, sexual, and emotional abuse and physical and emotional neglect in substitute care;
- the incidence and prevalence of child maltreatment by a wide array of demographic characteristics such as age, sex, race, family structure, household relationship (including the living arrangement of the resident parent and family size), school enrollment and education attainment, disability, grandparents as caregivers, labor force status, work status in previous year, and income in previous year;
- the extent to which reports of suspected or known instances of child abuse or neglect involving a potential combination of jurisdictions, such as intrastate, interstate, Federal-State, and State-Tribal, are being screened out solely on the basis of the cross-jurisdictional complications; and
- the incidence and outcomes of child abuse and neglect allegations reported within the context of divorce, custody, or other family court proceedings, and the interaction between this venue and the child protective services system. ◢

For a full list of topics see CAPTA Sec. 105. Research and Assistance available at https:/www.govinfo.gov/content/pkg/USCODE-2017-title42/html/USCODE-2017-title42-chap67.htm

These data sources have supported an increased understanding of maltreatment; however, they do not provide a complete picture of the scope and characteristics of child maltreatment. A number of factors have contributed to the challenge in obtaining accurate and ongoing surveillance of child maltreatment, including the lack of consistently used data constructs and variables, potential biases in the data related to race and poverty, and the lack of longitudinal data with a large sample (Institute of Medicine and National Research Council 2014; Jonson-Reid et al. 2016). Two overarching challenges involve accounting for child welfare definitions and policies that differ across states and communities and cases that go unreported to child protection authorities.

1. Varying child welfare laws and policies

One major challenge to understanding the national incidence of child maltreatment is that state child welfare laws and policies vary on dimensions such as their definitions of child maltreatment, including whether emotional abuse, domestic violence, or prenatal drug exposure are considered maltreatment; mandatory reporting requirements; and legal standards for substantiating child maltreatment, such as differences in states' use of alternative responses (also known as differential response) to reports which offer services to families without completing an investigation. Child welfare administrative data (such as the data reported in NCANDS) are based on these diverse state laws and policies. These variations result in widely different estimates of maltreatment: for example, about 2 per 1,000 children in Pennsylvania were substantiated victims in 2019, while more than 18 per 1,000 children were substantiated victims in West Virginia (HHS 2021). Victimization rates based on varying dimensions make it difficult to understand the cause of the differences and whether the differences are real or reflect differences in definitions, laws, and policies (Institute of Medicine and National Research Council 2014; Jonson-Reid et al. 2016). The variation also makes it hard to estimate a national prevalence of and recent trends in child maltreatment and complicates efforts to develop appropriate prevention and treatment responses. The NIS and other surveys have tried to standardize policy variations, but their sampling frames have not been sensitive enough to fully account for differences across states (Institute of Medicine and National Research Council 2014; Jonson-Reid et al. 2016).

2. Unreported cases

Another challenge to understanding the full extent of child maltreatment is that child welfare administrative data (such as the data reported in NCANDS) only include cases reported to child welfare agencies. The NIS has found that many situations that qualify as maltreatment go unreported; estimates suggest that rates are up to 50 percent higher than reported, although these rates only apply to cases of neglect (Sedlak et al. 2010). The studies and surveys that have attempted to estimate unreported cases have limitations. The DOJ surveys mentioned above are designed to capture unreported incidents of abuse. However, individuals and families who could not be located or refused to respond to the surveys and other data collection efforts might differ systematically from those who did participate. For example, the nonparticipating families may have different levels of exposure to violence and abuse than the participating families (Finkelhor et al. 2015). The NIS also captures information on unreported cases. However, it surveys mandated reporters; it does not include nonmandated reporters such as family and neighbors, who make up an estimated 31 percent of the reporters in NCANDS (HHS 2021). The NIS is also limited as a tool for the ongoing surveillance of maltreatment incidence because it is conducted infrequently, is not longitudinal, is not fully sensitive to state policy variations, lacks geographic details that would allow analyses across states or communities, is subject to survey response bias, and is resource intensive to administer. The DOJ surveys and the NIS provide valuable information on unreported abuse and neglect; however, they likely still underestimate the national incidence of maltreatment.

B. New federal efforts to address the challenges in estimating child maltreatment incidence

One promising approach to addressing these limitations is by linking local, state, or federal administrative records, such as those from child welfare, health, social services, education, public safety, and other agencies. This approach may help improve the quality, usefulness, interoperability, and availability of child maltreatment data. The Office of Planning, Research, and Evaluation (OPRE) and the Children's Bureau (CB) within the Administration for Children and Families (ACF) sponsored the Design Options for Understanding Child Maltreatment and Related Risk and Protective Factors (Design Options) project.[5] With input from key stakeholders and experts, this project identified research questions related to child maltreatment and related risk of importance for policy and practice and identified potential survey and administrative data sources that could be leveraged to examine the research questions of interest. One key finding from the project was the need for a database to provide state-level information on the laws and policies related to definitions of child maltreatment and the related risk and protective factors over time. OPRE, in partnership with the Children's Bureau, funded Mathematica and its partners to conduct the Definitions and Policies Related to the Incidence of Child Abuse and Neglect and Related Risk project. The project produced the State Child Abuse and Neglect (SCAN) Policies Database,[6] which compiles data on the definitions and policies that states use in their surveillance of child maltreatment. The SCAN Policies Database is available for researchers to link with other data sources to answer key questions related to child maltreatment incidence.

Another key finding from the Design Options project was the potential to use linked administrative data to obtain accurate information on the incidence of child maltreatment (Institute of Medicine and National Research Council 2014; Jonson-Reid et al. 2016). However, additional work is needed to understand how innovative approaches to linked administrative data can be leveraged to advance the child maltreatment knowledge base and brought to scale. Building on this work, OPRE, in partnership with the CB, funded Mathematica and its partners at Washington University in St. Louis to conduct the Child Maltreatment Incidence Data Linkages (CMI Data Linkages) project, which is the focus of this report. The project aimed to examine how linking administrative data may improve the ongoing and accurate surveillance of child maltreatment. .

OPRE and the CMI Data Linkages project team recruited and enrolled five sites,[7] which proposed projects featuring different types of enhancements (Box I.2) to existing administrative data linkage efforts to address questions about child maltreatment incidence and the related risk and protective factors.

Box I.2. Administrative data linking enhancements

CMI Data Linkages sites enhanced their existing linked administrative data efforts by doing the following:

- Using new types of administrative data in analyses
- Using new methods for linking administrative data sets and analyzing linked data
- Replicating a successful approach to data linking and analysis in a new geographic area or expanding a current data linking and analysis effort to include a new jurisdiction or wider geographic area◢

OPRE and the CMI Data Linkages project team supported the sites through funding, technical assistance (TA), facilitation of a cross-site learning network, and webinars with experts that gave sites the

---

[5] For more information on the Design Options project, see https://www.acf.hhs.gov/opre/project/design-options-understanding-child-maltreatment-incidence-2015-2017.
[6] The SCAN Policies Database is available at http://www.scanpoliciesdatabase.com/.
[7] The project team defined a site as a group comprising a researcher, investigator, research group, or center affiliated with a university, nonprofit organization, or public agency. The sites could include representatives of agencies or organizations that collect administrative data and make the data available (also known as data stewards).

opportunity to receive feedback on their projects from experts in the field. In addition, the CMI Data Linkages project team used information from the sites' experiences to assess the feasibility of enhancing data linkage and analysis efforts to produce new information on child maltreatment.

This report presents findings from the CMI Data Linkages Feasibility Study (Feasibility Study). The rest of this chapter provides background on the potential for administrative data linkages to improve understanding of child maltreatment and related risk. It then describes the goals and methods of the Feasibility Study. Chapter II describes the site selection process, the characteristics of sites that participated in the CMI Data Linkages project, planned enhancements to administrative data linkage and analysis, the goals and objectives of each site's research project, the data sources, the participating organizations and project leadership, and the status of each site's project. Chapter III describes sites' experiences with completing the activities necessary to conduct their projects and highlights promising practices gleaned from these experiences, and describes sites' perceptions of the role of cross-site learning and TA activities and supports as they pursued intended enhancements. Chapter IV discusses the contextual and organizational factors that promoted or impeded sites' projects. Chapter V summarizes findings from the Feasibility Study, lessons learned from the CMI Data Linkages sites, and considerations for future applications of linked administrative data. Appendix A provides a comprehensive overview of the project each site conducted.

C. Potential of linked administrative data to support estimation of child maltreatment incidence and understanding of risk and protective factors

Linked administrative data provide a growing opportunity to accurately capture the incidence of child maltreatment and related risk and protective factors. It can be used by external researchers and stakeholders to address questions about the incidence of child maltreatment as well as by administrators and researchers within the child welfare system to inform prevention and intervention efforts.

1. Administrative data

Administrative data typically are collected by public agencies and may include client-level service records, other documentation of program implementation, records of transactions or registrations, or similar information (Connelly et al. 2016). Although administrative data are not collected specifically for research purposes, these data offer several advantages compared to primary data collection efforts such as surveys (Penner and Dodge 2019). The following are examples:

- The scale of administrative data may facilitate analysis of subgroups of individuals, which might be too small for reliable analysis in other data sources.
- Some administrative data sources collect information about services or system interactions over time, which may facilitate longitudinal analysis.
- The inclusion of all individuals who participate in a program or receive a service might also reduce sampling biases that arise in survey data.

2. Linked administrative data

Linked administrative data are data sets from different agencies or systems, such as health, social services, education, public safety, and others, whose records for individuals or families have been matched based on common identifiers. Some states and localities maintain integrated data systems or data warehouses, which are systems for systematically and continuously linking data from multiple agencies and combining them into a single database.

Linking administrative data sets may help address some of the limitations of the existing efforts to estimate maltreatment incidence by addressing policy variation and identifying unreported cases. Linked administrative data can also provide a more complete picture of the experiences of individuals and families. The following are examples:

- Linking records from various sources, such as crime reports and hospital admissions, to child welfare records may help to capture unreported incidents of maltreatment (such as Putnam-Hornstein and Needell [2011]; see Widom [1989] for an early example of sampling using administrative data).
- Linking administrative data across systems or data sets has the potential to reveal patterns in families' interactions with multiple service providers or systems and show the trajectories of families across systems, helping policymakers learn about co-occurring issues and services and laying the groundwork for comprehensive approaches to help families (see, for example, Drake, Jonson-Reid, and Sapokite [2006]; Goerge and colleagues [2010]; Penner and Dodge [2019]; Slack and colleagues [2017]; and Weigensberg and colleagues [2018]).
- Linking child welfare records with other data sources, including survey data such as census data, may help improve estimates of the incidence of child maltreatment by controlling for county demographics and economic differences. The linkage of additional demographic data can help account for variations in child welfare laws and practices (Dumas et al. 2015).

However, working with linked administrative data also presents challenges. Some child welfare agencies lack the infrastructure and expertise to successfully leverage and link their administrative data with other data systems. These challenges may be compounded by a lack of resources to support effective and efficient implementation and analysis of data. A major challenge for many external researchers is the ability to access the data (see, for example, Office of Management and Budget [2016] and Penner and Dodge [2019]). These challenges may include identifying appropriate data sources, securing approvals for data sharing across the various agencies involved, and securely transferring and storing the data. Circumstances and possibilities related to the use of administrative data differ across state and county jurisdictions. Promising practices regarding administrative data linkage may be context specific and difficult to reproduce in new localities. Due to data security policies or infrastructure, for example, it might be possible to link two data sources in one state or locality but not another. Additional challenges may involve completing accurate linkages, addressing issues related to the quality and completeness of available data, and understanding individuals' trajectories across systems (Harron et al. 2017; Penner and Dodge 2019). To expand the use of linked administrative data to inform understanding of child maltreatment at the national-level, the field needs to know more about the potential novel information that administrative data linking can provide along with the challenges involved and promising practices researchers are using to overcome them.

D. The CMI Data Linkages Feasibility Study examined promising data linkage practices and contextual and organizational factors

In addition to funding and supporting the five sites that the CMI Data Linkages project team selected, the team also conducted a study of the feasibility of using enhanced linked administrative data analyses to provide more accurate and ongoing surveillance of child maltreatment. The overall goal of the Feasibility Study was to examine promising data linkage practices that could inform the ongoing and accurate surveillance related to the incidence of child abuse and neglect as well as future prevention and treatment efforts.

1. Research questions

Three research questions motivated the study:

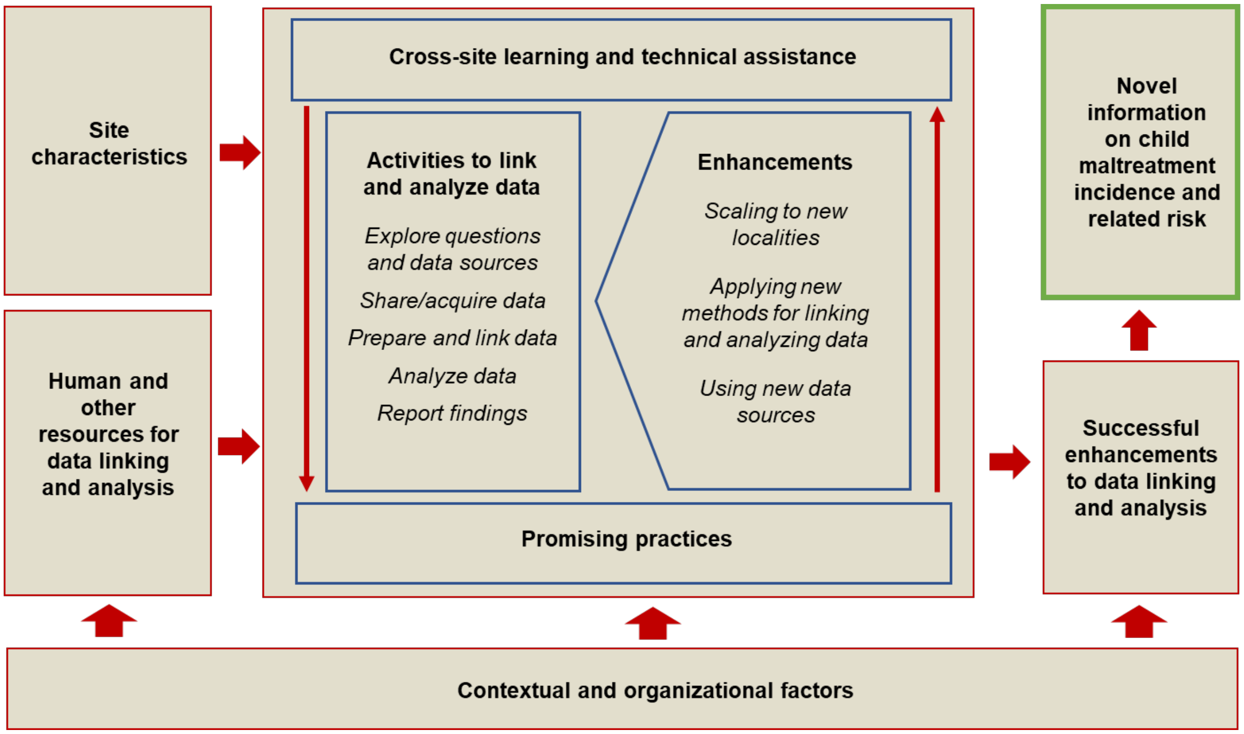
1. What are promising methods or practices, within and across the sites, for linking administrative data to inform understanding of the incidence of child maltreatment and related risk?
2. What contextual and organizational factors promote or impede enhancement of existing administrative data linkages?
3. What novel information about child maltreatment incidence and related risk and protective factors can be gleaned through enhancement of linked administrative data?

The study team tailored the Feasibility Study questions and analysis to reflect the nature of each site's research project.

2. CMI Data Linkages Feasibility Study conceptual model

To guide the Feasibility Study, the project team developed a conceptual model (Figure I.1) based on the main activities that the sites would undertake to design and implement their enhancements, as well as the factors that might influence these activities. The conceptual model also depicts the hypothesized pathway by which sites would successfully enhance their data linkage or analysis and ultimately produce novel information on child maltreatment incidence and related risk. This conceptual model and the constructs described below guided the collection and analysis of data for the Feasibility Study.

Figure I.1. Exploring the feasibility of using linked administrative data to examine the incidence and risk of child maltreatment



- **Site characteristics** were expected to shape sites' projects and provide necessary inputs to the data linking and analysis process. The characteristics that the Feasibility Study considered included the participating partners in each site, such as the agencies or organizations that provided administrative data and research groups that provided content and analytic expertise; the previous experience in data linking and analysis that these partners brought to the project; and the history of relevant partnerships among the organizations.
- Sites were required to provide different amounts and types of **human and other resources** to implement their projects based on the project's context. The Feasibility Study considered the type, amount, and provider of resources that sites used and whether stakeholders considered available resources to be sufficient for accomplishing each site's project.
- The CMI Data Linkages study team expected that sites would implement five broad **activities to link and analyze their data**: (1) explore possible research questions, data sources, and partnerships; (2) share and access the data; (3) prepare and link the data (including, assessing the quality of the data); (4) analyze the linked data; and (5) assess how findings add to the existing knowledge base and report on and disseminate the findings. The Feasibility Study explored the steps, approaches, and methods that sites undertook or applied in each activity; the challenges that they encountered; and the **promising practices** that they adopted to address the challenges and complete each activity. The team anticipated that some sites would be building on existing projects and that they would not necessarily proceed through these activities at the same pace or in the same way. For example, some sites might already have permissions for and access to the administrative data sources they planned to use.
- The **cross-site learning network** activities highlighted common analytic and logistical issues across sites and provided a forum to support sites in addressing these issues. The Feasibility Study explored the role of cross-site learning, TA, and expert consultation, as well as participants' perceptions of the benefits of these activities.

Box I.3. Cross-site learning network and webinars

The CMI Data Linkages project team facilitated a cross-site learning network for sites to interact with each other and network with experts in the field. Each site was assigned a liaison from the team. The liaisons led monthly calls with their assigned sites to identify challenges and facilitate connections with experts to help resolve them. The team facilitated a series of expert webinars where the sites presented preliminary findings from their projects and sought advice and feedback from invited experts. The team also convened quarterly meetings for the sites to discuss their projects as well as the following topics:

- Overview of the federal CMI Data Linkages project, the goals and structure of the cross-site learning network, and the sites' projects
- Objectives and methods of the CMI Data Linkages feasibility study
- Presentations by the sites to provide updates on their projects and discuss their proposed enhancements
- Discussions of data linking methods, challenges encountered, and lessons learned
- Discussions of dissemination goals, intended audiences, and best practices
- Presentations by the sites on their findings, challenges, lessons learned, and next steps

- Roundtable discussion on whether and how the sites' projects advanced the field's knowledge of child maltreatment incidence and risk

- A variety of **contextual factors** could influence the sites' approaches and results. The Feasibility Study investigated the influence of contextual factors, such as characteristics of the child welfare systems in the jurisdictions that were the focus of the sites' analyses, child welfare policies and definitions that affected the accessibility and content of administrative databases, state and local statutes that governed data access and use, and whether an integrated data system or other infrastructure for linking administrative data existed in a site.
- The **organizational capacity** of each of the participating partners—that is, the attributes or abilities that enabled these entities to accomplish their goals or mission—might also affect a site's approach and ability to complete the intended enhancement. The Feasibility Study explored several types of capacities that research groups, organizations, or agencies participating in the process might require to be successful, including leadership, effective partnerships, technical infrastructure and expertise, and the ability to translate the results of research to policy.
- Sites' expected **enhancements to administrative data linkages** might occur fully, partially, or not at all. The Feasibility Study explored the extent to which the intended enhancements were achieved and the factors that participants identified as key facilitators or inhibitors to accomplishing the objectives of each site's project.
- Successful completion of a site's project could lead to **novel information about child maltreatment**. The Feasibility Study explored the types and topics of the sites' findings and participants' assessments of whether and how their findings contributed to the knowledge base on child maltreatment.

3. Data collection efforts

To address the Feasibility Study research questions, the project team conducted a qualitative, cross-site study involving coding, collecting, and analyzing data from multiple sources to gather information on sites' experiences as they conducted their projects from 2019 through 2021 (Table I.1). Each project lasted approximately 18 months to 24 months.

Table I.1. Feasibility Study data collection activities

Data Collection activities	2019				2020				2021
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1
Site documents submitted	X				X			X	X
Site meetings with project liaison	X	X	X	X	X	X	X	X	
Cross-site learning activities and expert webinars		X	X	X	X	X	X	X	
Interviews with key respondents		X	X	X		X	X		
Questionnaire on resources			X		X	X	X		

Q = quarter

- **Site documents.** Key documents included a project plan, interim memo, and final memo authored by the research team at each site.
    - Each site completed a template outlining the site's plans for their project. The **project plan** provided an overview of the project and described each site's objectives, research team and partners, expected data sources, expected methodology for the project, factors that might help or hinder the project's progress, and the timelines and budgets for the project.
    - Each site also drafted **interim and final memos** that discussed the context for their projects; the status of key partnerships; progress toward accessing administrative data; approaches to cleaning, processing, and linking data; the quality and accuracy of data linkages; methods for analyzing linked data; preliminary finding; challenges and successes they have experienced; technical assistance received and needed; and dissemination activities.

  The Feasibility Study team also collected and reviewed a range of additional documents from each site, including data sharing agreements and memorandums of understanding (MOUs), descriptions of administrative data sets, data security plans, institutional review board (IRB) applications and/or other approval documents, and summaries of findings. Because their projects varied, not all sites provided all types of documents.

- **Notes and materials from cross-site learning network meetings, webinars with experts, and monthly consultations between the research team and CMI Data Linkages project liaisons.** Materials included notes from the meetings that sites had with their liaisons and notes of discussions from the quarterly cross-site learning network meetings (Box I.3). Materials also includes slide presentations that the site research teams prepared about their projects for the webinars with experts and notes from discussions and feedback from the experts during those webinars. The Feasibility Study team reviewed the materials to catalog the topics covered, identify the issues encountered and discussed by sites, identify the types of assistance provided to sites, and explore whether sites shared solutions to common challenges.

- **Interviews with key respondents from participating partners.** The Feasibility Study team conducted two rounds of interviews with representatives from each site to gather insights into the site's characteristics as well as important project activities and decisions—including, how sites prioritized research questions, identified useful administrative data sources, and selected their approaches to link and analyze data. Interviews ranged from 30 minutes to 90 minutes. Respondents included principal investigators and co-principal investigators from each site; administrators, directors, or managers at relevant government agencies or other organizations that provided administrative data; information technology or data managers; and research staff. The specific respondents differed among sites depending on the characteristics of participating research groups, organizations, and agencies involved in the projects. The number of respondents ranged from three to seven per site.

  The Feasibility Study team conducted one round of interviews during the early phase of each project (within four months of the start date) and a second round approximately one year later. The study team used a discussion guide for each interview that was structured according to the conceptual model established for the study. Interviewers tailored the discussion guide to the specific circumstances of each site and the role of each respondent. Team members recorded notes during each interview using a standard template.

- **Questionnaires on sites' use of resources.** The Feasibility Study team used a brief questionnaire to gather data about the time that participating partners spent on different activities (aligned with the activity categories in the conceptual model above), the dollar value of non-personnel resources used (such as expenses related to information technology or fees for data access or IRB approval), and in-kind contributions that supported the project. The questionnaire requested information on the number of hours that participating individuals spent on activities related to the site's data linkage project (whether or not this individual's time was supported by CMI Data Linkages funding). To explore variations in the use of resources over time, the study team fielded the questionnaire multiple times. For each questionnaire, project leaders were asked to provide their best estimate of resource use during a period covering three to eight months. (The first questionnaire covered a longer period because sites began their projects at different times.)

4. Analytic approach

Analysis for the Feasibility Study was primarily qualitative. The Feasibility Study team applied the framework method (Box I.4) of qualitative analysis to identify commonalities and differences among sites and specify descriptive or explanatory themes regarding the study's questions (Gale et al. 2013; Ritchie et al. 2013).

Box I.4. The framework method of qualitative analysis

The framework method is a systematic approach to organizing and analyzing qualitative data that involves five main steps:

1. Becoming familiar with the data
2. Developing an analytical framework
3. Applying the framework to the data through coding
4. Arranging the data into matrices organized by site
5. Interpreting the data ◢

Becoming familiar with the data. Members of the Feasibility Study team read notes from interviews, meetings that sites had with their liaisons, discussions from the quarterly cross-site learning network meetings, and other data sources to gain an understanding of the data. During this process, team members began making initial notes about themes or findings related to the research questions and topics.

Developing and refining the analytical framework. We identified and cataloged data relevant to each research question and topic by coding the data. To prepare for the data coding process, we created an initial high-level coding template based on the elements of the conceptual model and the topics or concepts associated with each element of the model. The template specified codes so that we applied them consistently. We coded the data using NVivo qualitative analysis software.

Coding the data. Two staff members on the Feasibility Study team had primary responsibility for coding data. To ensure consistency in coding, the two staff members conferred regularly to discuss the definition and application of codes. The Feasibility Study team coded data from documents, notes, and interviews.

Creating matrices. Next, we used NVivo to automatically create framework matrices organized by site and research question or topic. This step helped to structure the data and prepare it for analysis. In the matrix cells, we developed preliminary analytic statements for each site related to each construct.

Interpreting the data. This step of the process entail reviewing the matrices for each site and identifying key themes and findings across the sites. The matrix format enhances understanding of the data within and across sites by facilitating comparisons and highlighting patterns or contradictions across sites. Members of the Feasibility Study team discussed emerging themes and initial findings related to research questions. Through multiple conversations and triangulation across data sources, the Feasibility Study team refined these themes and findings. The study team then will document our interpretations of the data by producing a matrix articulates cross-site data interpretations related to research questions and topics, such as the organizational capacities required to complete each phase of linking and analyzing administrative data. The matrix included examples that supported each theme or finding.

To analyze quantitative data from the questionnaire on resource use, the study team calculated descriptive statistics (such as mean, median, and range of time spent on tasks) and conducted cross-site comparisons. These statistics addressed measures such as the number of hours individuals in the sites spent on project activities and the allocation of time across different types of activities. The study team also cataloged the types of non-personnel resources that the sites reported using for their projects, as well as the in-kind contributions they received.

5. Considerations for interpreting findings from the Feasibility Study

Several aspects of the Feasibility Study's design and context should be taken into account when interpreting its findings:

- All of the sites began their projects with an existing foundation for data linking and analysis—that is, their efforts built on previous projects to link and analyze data—so the sites do not represent cases in which researchers are starting from scratch. The Feasibility Study team documented the projects' starting points and their existing organizational capacity. Thus, when considering the feasibility of implementing these enhancements, the additional resources to conduct a similar study from a different starting point must be considered.
- Because sites' projects progressed at different paces, the level of detail available for each project varied. For example, project activities in some sites were disrupted by the COVID-19 pandemic. Work to address pandemic-related challenges took priority for both the project leaders and the partners providing the data, which resulted in delays in completing work on the site's CMI Data Linkages projects. Some sites experienced other challenges acquiring data, which slowed their progress. Although the Feasibility Study team intended to conduct interviews at roughly similar milestones across sites (for example, after the site had acquired data), this was not always possible due to scheduling constraints. The status of the project in each site might have affected the activities the respondent was able to discuss and respondents' perspectives on factors that facilitated or hindered project activities. To the extent possible, the Feasibility Study addressed these gaps by obtaining data from other sources.
- Because the Feasibility Study team did not have direct access to the administrative data sets that the sites used, the team relied on respondents' summaries and assessments of key characteristics of the data, such as content, quality, and completeness. For example, sites reported on the extent of missing data for the population or subpopulations of interest, on whether data sets included key variables of interest, and on quantitative indicators of the success of record linkage.

These considerations notwithstanding, the findings from the Feasibility Study presented in this report offer insights into the practices, facilitators and constraints, and potential for developing new information on child maltreatment using administrative data linkages.

II. THE CMI DATA LINKAGES SITES

To accomplish the goals of the CMI Data Linkages project, it was important to engage sites that could enhance the knowledge base on linking and analyzing administrative data as well as address key questions regarding child maltreatment. In consultation with OPRE, the CMI Data Linkages project team identified sites that had the experience and capacity to undertake innovative approaches to linking and analyzing administrative data within the project's limited time frame.

This chapter describes the selection and key characteristics of the CMI Data Linkages sites. It first describes the process and criteria used to identify and engage sites, then it summarizes the characteristics of the sites that participated. Finally, this chapter profiles the five individual sites that were selected, describing each site's planned enhancement and project focus, expected sources of data for linkages, the organizations involved and their experience with linked administrative data, and the status of each site's project at the time this report was prepared.

A. Site selection process

The CMI Data Linkages project team engaged in a site selection process from winter 2017 though summer 2018. The team defined a site as a group comprising a researcher, investigator, research group, or center affiliated with a university, nonprofit organization, or public agency. The sites could include representatives of agencies or organizations that collect administrative data and make the data available (also known as data stewards). The team expected that sites would represent existing partnerships between entities with administrative data and entities with the technical skills for linking and analyzing the data. The sites might analyze data from one or more agencies within a single jurisdiction or several jurisdictions, or they might analyze national data.

The purpose of the site selection process was to maximize the possibility that the selected sites would be able to (1) yield new information about child maltreatment incidence and related risk and protective factors through successful completion of an enhancement to existing data linkage efforts and (2) inform the feasibility of using linked administrative data to obtain novel maltreatment information, including identification of crosscutting promising practices as well as site-specific lessons related to different geographic areas and policy contexts. The CMI Data Linkages project team also sought to ensure that the study captured meaningful variation across sites.

The team used several methods to develop a preliminary list of possible sites:

- OPRE posted and circulated a public call for input on relevant data linking work in its December 2017 newsletter. OPRE invited readers to share innovative projects from academic researchers, state and local administrators, and others who used linked administrative data to examine child maltreatment.
- The CMI Data Linkages project team consulted with the team that conducted OPRE's Design Options project and other experts about potential sites identified through their work and knowledge of the field.
- The CMI Data Linkages project team conducted a targeted scan for recent literature that focused on integrated data.

- The CMI Data Linkages project team reviewed key articles from the Design Options team and the Actionable Intelligence for Social Policy (AISP) website for resources that might identify possible sites or researchers working with integrated data.
- The CMI Data Linkages project team conducted a targeted search of electronic databases and selected journals that were likely to include additional literature of interest to the team.

From the preliminary list of 59 sites identified through these steps, the CMI Data Linkages project team reduced the number of possible sites down to the top 20 sites that (1) had experience with child welfare data, (2) had used or proposed a data source unique to the analysis of child maltreatment, (3) were involved in either ongoing data integration efforts or a onetime linkage effort, and (4) appeared likely to be able to obtain the relevant data.

The CMI Data Linkages project team then collected additional information about each of the 20 sites. The team identified the number of times a site was mentioned by the sources noted above—for example, the number of commenters responding to the OPRE newsletter solicitation who mentioned the site's work. The team reviewed publicly available materials and conducted a small number of exploratory calls with representatives of some of the potential sites. The team then assessed each of the 20 sites against the project's site selection criteria (Table II.1).

Table II.1. Site selection criteria

Selection criteria	Indicators
Interest in undertaking the project	• Motivated to scale or expand existing practices • Willingness to participate in the Feasibility Study • Interest in participating in the cross-site learning network
Access to and quality of the administrative data being used (or planning to be used)	• Availability and quality of administrative data • Track record of completing key steps in obtaining, linking, and analyzing data • Accessibility of relevant databases that will facilitate rapid implementation of data linkages
Experience in conducting research with administrative data	• Experience in securing IRB approvals • Experience in addressing privacy protections when working with administrative data • Experience in negotiating data sharing agreements
Organizational leadership and capacity	• Productive working relationships between key participants in potential projects—such as the agencies that collect administrative data (data stewards) and the parties that will undertake the linkage and analysis • Data linkages identified as an organization or agency priority • Organizational capacity to add new linkages or scale existing linkages to new sites
Ability to conduct the proposed project in the specified time frame	• Staff capacity to meet project expectations and complete the project within 18 months
Extent to which the proposed project addresses priority CAPTA research questions	• Interest in pursuing research questions relevant to the Feasibility Study and aligned with CAPTA research priorities • Intent or ability to add to the knowledge base about the incidence of and risk and protective factors associated with child maltreatment • Demonstrated interest in continuing explorations and proposals for the research areas prioritized by participants in the Design Options project • Intent to communicate the results of investigations to the field and an interest in promoting new approaches to data linkage and analysis

Selection criteria	Indicators
Potential for applying innovative methods in linking and analyzing data	• Expertise in applying appropriate analytic methods—including new or advanced methods—to link databases and answer research questions; methods may include natural language processing, predictive analytics, multilevel modeling, geospatial analysis, Bayesian approaches that would allow researchers to account for uncertainty attributable to missing data as well as for sampling and parameter uncertainty, and others
Use of diverse or unexplored data sources	• Availability of administrative data sources that are infrequently used for analysis of child maltreatment incidence
Variability in contextual factors	• Geographic location • Size of local population and the child welfare system providing the administrative data • Demographics of the population that comes into contact with the child welfare system • State and local child welfare policies, such as standards and processes for substantiating reports of child maltreatment • Level of administrative data (federal, state, and/or local)

CAPTA = Child Abuse Prevention and Treatment Act; IRB = institutional review board.

The team used a detailed matrix to document its assessment of each site based on the site selection criteria. For each criterion, the team rated a site as either strong, acceptable, or needs support based on the information collected during the site selection process. The CMI Data Linkages project team narrowed the list of potential sites to nine based on this assessment – particularly interest, access to data, geographic variation, and variation with respect to the level of data to be used (national, state, city, or county). The team then contacted these sites to discuss the CMI Data Linkages project in more detail and ask them to describe the site's current work, desired work, and any possible challenges or barriers to participation in the project. After further considering these factors, especially alignment between sites' proposed research questions and priorities for the CMI Data Linkages project, the team invited five sites to participate in the project.

B. Characteristics of the CMI Data Linkages sites

The CMI Data Linkages sites all demonstrated capacity to complete innovative projects involving administrative data linkage. They varied in ways that reflected the diversity of the field of child welfare research and their proposed data sources. Table II.2 summarizes the key characteristics of the CMI Data Linkages sites. This report refers to sites by using the name of the lead research organization(s): Children's Data Network/California Child Welfare Indicators Project (CDN/CCWIP), Center for Social Sector Analytics and Technology (CSSAT), Alaska Department of Health and Social Services/Oregon Health Sciences University (DHHS/OHSU), CDN/Rady Children's Hospital (CDN/Rady), and the University of Alabama School of Social Work (UA-SSW).

The five sites shared some similarities—for example, in all sites, an academic institution or research center was the lead or co-lead organization, and all sites had previous experience with conducting research using administrative data. The sites also varied in meaningful ways, including their project enhancements and research questions; the types, sources, and level (national, state, county) of data planned for linkages; the number and types of partners involved; and the composition and size of their research teams.

Table II.2. Characteristics of CMI Data Linkages sites

Site name	Project title	PI	Lead research organization	Type of data linkage enhancement	Project goals	Geographic scope of analysis	Data stewards and providers	Planned data sources
ADHSS/OHSU	Replicating the ALCANLink methodology	Jared Parrish	Alaska Department of Health and Social Services Oregon Health Sciences University (graduate student and faculty advisor)	Replication	Test the internal and external validity of the ALCANLink methodology, which integrates survey data with child welfare and other administrative data to examine child maltreatment incidence, predictive factors, and disparities	Alaska and Oregon	Oregon Health Authority Oregon Department of Human Services	Oregon PRAMS survey data (2009–2011); vital records data; child protective services records (2009–2018); ALCANLink data
CDN/CCWIP	Methods to estimate the community incidence of child maltreatment	Emily Putnam-Hornstein	University of Southern California School of Social Work, Children's Data Network (research center) University of California, Berkeley School of Social Welfare, California Child Welfare Indicators Project (research center and child welfare data repository)	New methodology	Develop administrative data–based methodology to generate estimates of the upper and lower bounds of child maltreatment	California	California Department of Social Services California Department of Public Health	Child protective services records (1998–2018); birth records (1999); death records (1999–2017)

Site name	Project title	PI	Lead research organization	Type of data linkage enhancement	Project goals	Geographic scope of analysis	Data stewards and providers	Planned data sources
CDN/Rady	Using hospital data to predict child maltreatment risk	Emily Putnam-Hornstein	University of Southern California School of Social Work, Children's Data Network (research center)	New data sources	Characterize children with medical encounters relative to overall birth population, validate two predictive risk models (PRMs) designed to predict future system involvement and explore the use of hospital records as predictors in PRMs	California and San Diego County	Rady Children's Hospital–San Diego California Department of Social Services California Department of Public Health	Child protective services records (1998–2018); birth records (1999); death records (1999–2017); San Diego County hospitalization records (2010–2016)
CSSAT	Understanding the effect of the opioid epidemic on child maltreatment	Joseph Mienko	University of Washington School of Social Work, Center for Social Sector Analytics and Technology (research center)	New data sources	Explore how opioid use impacts maltreatment risk, child welfare system contact, and substantiation and placement decisions	Washington State	Washington Department of Children, Youth, and Families Washington Department of Health	Child protective services records (1999–2018); birth records (1999–2017); hospitalization records (1999–2017); hospital billing data; death records (1999–2017); prescription monitoring program data (2012–2018); American Community Survey data related to county-level demographics and other measures

Site name	Project title	PI	Lead research organization	Type of data linkage enhancement	Project goals	Geographic scope of analysis	Data stewards and providers	Planned data sources
UA-SSW	Examining child maltreatment reports using linked county-level data	Brenda Smith	University of Alabama School of Social Work (faculty member and graduate students)	Replication	Examine how risk and protective factors relate to child maltreatment reports at the county level across the nation, with a focus on rural and majority racial or ethnic minority counties Explain widely varying state- and county-level maltreatment rates and operationalize county-level child maltreatment risk	Nationwide	Publicly available data from the National Data Archive on Child Abuse and Neglect	Census data; NCANDS data (2012–2015)

Source: CMI Data Linkages Feasibility Study data.

ADHSS = Alaska Department of Health and Social Services; ALCANLink = Alaska Longitudinal Child Abuse and Neglect Linkage Project; CCWIP = California Child Welfare Indicators Project; CDN = Children’s Data Network; CSSAT = Center for Social Sector Analytics and Technology; NCANDS = National Child Abuse and Neglect Data System; OHSU = Oregon Health Sciences University; PI = principal investigator; PRAMS = Pregnancy Risk Assessment Monitoring System. Rady = Rady Children’s Hospital-San Diego; UA-SSW = University of Alabama School of Social Work.

PI affiliation and previous experience with integrated data. Sites' PIs were affiliated with academic institutions and a government agency. PIs in three sites (CDN/CCWIP, CSSAT, and CDN/Rady) were leaders of university-based research centers that were dedicated to analytics using administrative data and focusing on issues related to child welfare as well as children's health and well-being. These PIs previously conducted multiple studies involving the linkage methods and data that they used for their CMI Data Linkages projects. Another PI (UA-SSW) was a faculty member who worked independently in a public university's school of social work. She previously conducted research on county-level maltreatment rates using administrative data and public-use data sets. The PI in the fifth site (ADHSS/OHSU) was based in a state public health department. He had experience developing state-level administrative data linkages to examine child maltreatment incidence and predictive factors as well as conducting other epidemiological research.

Enhancements. Sites' enhancements to existing data linkage and analysis efforts involved replication of an existing approach in a new geographic area, the addition of new data sources, or the use of new methods for linkage and analysis. Two sites (UA-SSW and ADHSS/OHSU) focused primarily on replicating a linkage and analysis approach that they had undertaken previously. In one site (UA-SSW), this effort involved scaling an analysis approach from about 600 medium and large counties to include counties nationwide. A second site (ADHSS/OHSU) attempted to replicate a linkage and analysis effort completed in one state to another state. The intended enhancement in another two sites (CSSAT and CDN/Rady) was the addition of new data sources to previously completed linkages. At both sites, the new data sources included health records. The fifth site (CDN/CCWIP) incorporated new methods for linkage and analysis by applying longitudinal and cross-sectional analyses of linked administrative data to validate or adjust estimates of the incidence of maltreatment.

Goals and research questions. Sites' goals and research questions aligned with CAPTA research priorities focused on (1) understanding the nature and scope of maltreatment, generally and by demographic characteristics; (2) exploring the causal and preventative dimensions of maltreatment incidence; and (3) examining trends in and outcomes related to maltreatment reports (Table II.3).[8,9] Four sites (CDN/CCWIP, ADHSS/OHSU, CDN/Rady, and UA-SSW) addressed research questions focusing on estimating state- or county-level maltreatment incidence. Three sites (CDN/CCWIP, ADHSS/OHSU, and UA-SSW) sought to examine the characteristics of child maltreatment victims and their families. Two sites (CSSAT and UA-SSW) addressed questions related to risk and protective factors for maltreatment. Two sites (CDN/CCWIP and CSSAT) explored the trajectories of children with maltreatment reports. Three sites (CDN/CCWIP, ADHSS/OHSU, and CDN/Rady) explored novel methods for estimating maltreatment incidence, including predictive risk models.

---

[8] Some sites had research questions that spanned multiple topic areas.

[9] The CAPTA Reauthorization Act of 2010 can be found at https://www.govinfo.gov/content/pkg/USCODE-2017-title42/html/USCODE-2017-title42-chap67.htm.

Table II.3. Alignment of sites' research questions to CAPTA research priorities

	Site				
Focus of research questions	**ADHSS/ OHSU**	**CDN/ CCWIP**	**CDN/ Rady**	**CSSAT**	**UA-SSW**
Estimating state- or county-level maltreatment incidence	X	X	X		X
Characteristics of child maltreatment victims and their families	X	X			X
Risk and protective factors for maltreatment				X	X
Trajectories of children with maltreatment reports		X		X	
Novel methods for estimating maltreatment incidence	X	X	X		
CAPTA research priority					
(A) The nature and scope of and (O) the national incidence of child abuse and neglect	X	X	X		X
(ix) The incidence and prevalence of child maltreatment by demographic characteristics	X	X			X
(B) The causes, prevention, assessment, identification, treatment, cultural and socio-economic distinctions, and the consequences of child abuse and neglect				X	X
(O)(i) The extent to which incidents of child abuse and neglect are increasing or decreasing in number and severity		X		X	
(O)(vii) The extent to which unsubstantiated reports return as more serious cases of child abuse or neglect		X		X	

CAPTA = Child Abuse Prevention and Treatment Act.

Geographic scope. The geographic scope of sites' analyses varied from a single county to nationwide. Three sites (CDN/CCWIP, CSSAT, and CDN/Rady) conducted projects focused on the individual states where the lead agency was located. One of these three sites (CDN/Rady) analyzed statewide data and data from a single county. Another site (ADHSS/OHSU) analyzed data from two different states, and a fifth site (UA-SSW) analyzed nationwide county-level data.

Data sources. Data sources for the sites' analyses included restricted-use administrative data sets and publicly available data (Table II.4). All sites planned to use child welfare administrative data. Four sites (CDN/CCWIP, CSSAT, ADHSS/OHSU, and CDN/Rady) obtained individual-level data from the child welfare data stewards in their states, while a fifth site (UA-SSW) used publicly available NCANDS county-level data from the National Data Archive on Child Abuse and Neglect, a voluntary, national data collection and reporting system that is a project of the Children's Bureau within ACF. Four sites (CDN/CCWIP, CSSAT, ADHSS/OHSU, and CDN/Rady) planned to link vital records (including birth and death records) to child welfare administrative data. Two sites (CSSAT and CDN/Rady) planned to link hospitalization records and other health records to child welfare administrative data. Three sites (CSSAT, ADHSS/OHSU, and UA-SSW) planned to conduct analyses using data from population-based surveys by linking records at either the individual or county level.

Table II.4. Planned data sources, by site

Site	Planned data sources				
	Individual-level child welfare data	Publicly available county-level child welfare data	Vital records	Hospitalization records	Population-based surveys
ADHSS/OHSU	X		X		X
CDN/CCWIP	X		X		
CDN/Rady	X		X	X	
CSSAT	X		X	X	X
UA-SSW		X			X

Partnerships. The number of participating agencies and organizations involved in each site's project, aside from the lead research organization, ranged from zero to five agencies and organizations. Sites' partners included data stewards, government agencies, and other research organizations. The site accessing publicly available data (UA-SSW) did not partner with any organizations to complete its project. Another site (ADHSS/OHSU) partnered with five organizations, including a university, three data stewards, and a unit within a state agency that was responsible for linking state agency data.

Research team composition. The teams completing the work in each project varied in size from 3 to 10 members. On average, they included 6.8 members. All teams included a PI (or co-PIs) and research staff. Three project teams included representatives of the agencies that owned the data being used for the research.

C. Profiles of the CMI Data Linkages sites

This section provides an overview of each of the five sites' intended projects and the status of each project at the time this report was prepared. Site-specific summaries included in Appendix A present detailed information about the context, data sources, methods, findings, and lessons learned in each site.

Replicating the Alaska Longitudinal Child Abuse and Neglect Linkage (ALCANLink) methodology (ADHSS/OHSU)

Enhancement to administrative data linkage and analysis sites planned to undertake

This site aimed to replicate an administrative data linkage project that was first developed in Alaska in another state, Oregon.

Goals and research questions

The original project in Alaska linked child welfare and vital records data with data from the Pregnancy Risk Assessment Monitoring System (PRAMS),[10] a population-based survey administered by 47 states that explores pregnancy-related attitudes, experiences, and outcomes. This linkage previously facilitated analyses of the cumulative child maltreatment incidence, risk and protective factors, and disparities based on the population-based data in the PRAMS sample. The project team attempted to acquire similar data and conduct similar analyses in Oregon to test the external and internal validity of the ALCANLink approach and examine factors that supported or hindered its replication.

The site planned to address three main research questions:

(1) Does the cumulative incidence of the time to first maltreatment report in Oregon differ from that observed in Alaska?
(2) Are the cumulative incidence to first report, contact, and substantiation estimated through the Oregon PRAMS linkage consistent with a full Oregon birth cohort linkage to child welfare?
(3) What are the key components required for successful replication of ALCANLink methods?

Planned data sources

The site expected to link Oregon child welfare administrative data, data from the PRAMS survey, and vital records. The planned analyses also involved data from the ALCANLink project to support comparisons of incidence estimates between Oregon and Alaska.

Participating organizations

The PI was an epidemiologist with the state health and social service agency in Alaska who led the development of the original linkage approach. A graduate student and a faculty member at a public university in Oregon were responsible for coordinating the data sharing agreements, securing the research approvals, and conducting the analyses. Staff members of the Oregon public health agency participated in project planning and facilitated access to state administrative data. The Oregon child welfare agency participated as a data owner. In addition, a state-operated data warehouse working with the Oregon public health and child welfare agencies served as a repository for the administrative data and completed the data linkages to create the data set used for analysis.

Project status

The site secured data use agreements to access the planned data, completed the planned linkages, and undertook initial analyses related to the cumulative incidence of child maltreatment based on linkages to PRAMS and birth cohort data. Due to delays related to the COVID-19 pandemic, the site was unable to address its secondary research questions related to (1) whether demographic population frequency distributions confound between state comparisons and (2) whether the Alaskan cohort information inform and improve estimates of cumulative incidence in Oregon in the absence of information accounting for population loss.

---

[10] Additional information about PRAMS can be found at https://www.cdc.gov/prams/index.htm.

Methods to estimate the community incidence of child maltreatment (CDN/CCWIP)

Enhancement to administrative data linkage and analysis sites planned to undertake

The enhancement in this site focused on applying established analytic methods—longitudinal and cross-sectional analyses—in new ways to generate estimates of the upper and lower bounds of maltreatment incidence using administrative data.

Goals and research questions

This site focused on developing a methodology that used administrative data to estimate the number of children who were victims of abuse or neglect. The site attempted to produce the upper and lower bounds of estimates that reflected the number of children who were identified by the child welfare system as victims of abuse or neglect as well as those who were victims but not identified as such by the system. The site planned to test the methodology using data from California and explore the potential for using it in other states.

Two research questions motivated this site's project:

(1) How would official estimates of maltreatment victims vary if we assumed that every child substantiated as a victim in 2015 had been exposed to maltreatment (but not classified as a victim) in the year immediately before substantiation, two years before substantiation, and so forth, based on other CPS contacts that occurred?

(2) In longitudinal interactions with child welfare system from birth to age 18, are there differences in annual incidence and cumulative prevalence rates by county and demographic characteristics at birth?

Planned data sources

The site expected to use statewide, individual-level child welfare data and vital records data from California.

Participating organizations

The PI's organization for this site was the same as the one for the site focused on using hospital data to predict child maltreatment risk.

A center focused on analysis of child welfare data at a public university in California was a frequent collaborator with the PI's organization and participated as a research partner for this project. This center had a long-standing data sharing and TA partnership with the California child welfare system.

The PI's organization had a long-standing data sharing agreement with the child welfare agency. This agreement broadly authorized the use of data to address policy- and program-specific questions and epidemiological population-based research linked to other data. The PI's organization has had agreements to access vital records from the public health agency for many years.

Project status

With access to data established from the outset of the project, the site completed its planned analyses and made progress on drafting manuscripts for dissemination.

Using hospital data to predict child maltreatment risk (CDN/Rady)

Enhancement to administrative data linkage and analysis sites planned to undertake

This site planned to enhance existing linkages of statewide, individual-level administrative data by incorporating a new type of data: patient records from an individual hospital.

Goals and research questions

This site attempted to validate a predictive risk model (PRM) first used in Allegheny County, Pennsylvania. The site intended to probabilistically link children's hospital data with vital records and child welfare data and then conduct analyses to determine the extent to which children identified by the model as high risk for maltreatment were also at elevated risk of injury, poor health outcomes, and mortality in childhood.

The site planned to answer four main research questions:

(1) What are the characteristics and service trajectories of child hospital patients, compared to other children born in the county?
(2) What information does this provide about the social determinants of physical (medical), mental, and behavioral health of children served in this hospital?
(3) To what extent are children who have been identified by a statewide predictive risk model to be "at high risk" of maltreatment also at elevated risk of injury, poor health outcomes, and mortality in childhood?
(4) What is the predictive value of integrating hospital data as predictors in the predictive risk model?

Planned data sources

The site expected to use statewide child welfare data and vital records data from California and patient records of a children's hospital in San Diego County, California. The patient records included inpatient and outpatient encounters.

Participating organizations and project leadership

The PI's organization was a research center based at a private university in California that was founded to link and analyze administrative data on children and families. The project built on the center's previous work related to capturing longitudinal "snapshots" of maltreatment.

The PI's organization partnered with a research center for data analytics based at a public university in New Zealand. One of the founders of this center led development of the PRM that would be tested as part of the CMI Data Linkages project.

A third partner in the project was a children's hospital in San Diego County. The PI's organization and the hospital had not partnered before, and the hospital did not have previous experience with sharing administrative data for linkages. However, the PI's organization had previously established data sharing agreements with the state agencies that owned the child welfare and vital records data used in the analyses.

Project status

The site was able to complete new data sharing agreements for the hospital data, although doing so took longer than expected. It completed an initial round of analyses addressing all four of its research questions.

Understanding the effect of the opioid epidemic on child maltreatment (CSSAT)

Enhancement to administrative data linkage and analysis sites planned to undertake

The site's primary enhancement was intended to add new data sources—specifically, hospital billing data and county-level prescription monitoring data—to a previously completed data linkage project that involved child welfare administrative data. The site also applied a relatively new approach in the child welfare field, Bayesian model averaging, to select which variables would be the best fit for the statistical models to analyze the effects of opioid use on the risk of maltreatment.

Goals and research questions

This project attempted to contribute to the knowledge about the opioid epidemic's potential effects on child maltreatment by exploring disproportionality in the testing of mother-infant dyads for opioids at birth and in child welfare referrals related to opioid use. The team planned to examine both individual- and community-level risk factors for maltreatment and the trajectories of children with maltreatment reports.

The site planned to address three main sets of research questions:

(1) How do individual-level and county-level opioid use and misuse impact the risk of maltreatment for children and families in Washington State?

(2) How does opioid use and misuse impact child welfare system contact for children and families in Washington State? Are there are sociodemographic differences between those who are tested and those who are not tested for opioid use?

(3) How do indications of opioid use/misuse or prenatal opioid exposure at the family and county-level impact substantiation and placement decisions within households under investigation for maltreatment?

Planned data sources

In addition to statewide, individual-level child welfare records, the site expected to use multiple other types of administrative data—birth and death records, hospitalization records, hospitalization billing data, and county-level prescription monitoring data—to identify contacts with the child welfare system; risk factors and outcomes related to child maltreatment; and opioid testing, prescriptions, and overdoses. The site also planned to use county-level variables from the Census Bureau's American Community Survey (ACS) related to poverty and county-level measures of high school graduation, voter participation, and county rural status.

Participating organizations

The PI's organization is a university-based research center that undertakes a variety of projects involving administrative data from the social services system in Washington State. The research center has a long-standing relationship with the state child welfare agency. It manages the data collection and case management application for Washington State's child welfare system and receives quarterly child welfare data extracts for research activities.

Project status

The site acquired administrative data and conducted analyses to address the first of the three sets of research questions and produced a journal article on these findings.

The site acquired the necessary administrative data to address the third research question on the relationship between indications of opioid use and misuse and the substantiation and placement decisions for families under investigation. At the time of this report, the site was in the process of linking the data sets.

The site was unable to acquire hospital billing data, so it could not complete analyses addressing the second research question on how opioid use or misuse affects child welfare system contact for children, which relied on identifying mothers and infants who were tested for opioids at birth.

Examining child maltreatment reports using linked county-level data (UA-SSW)

Enhancement to administrative data linkage and analysis sites planned to undertake

This site aimed to build on its previous work that explored community factors associated with child maltreatment by scaling an existing data linkage, increasing the number of counties in the sample from 612 medium and large counties to over 3,000 counties nationwide. The site also expected to incorporate data from rural counties with small populations into the analyses.

Goals and research questions

This site attempted to (1) explain wide county- and state-level variation in child maltreatment reporting, (2) offer new insights about maltreatment reporting in rural counties, and (3) examine how risk and protective factors relate to child maltreatment reports at the county level across the nation.

The site planned to address four research questions:

(1) How well do county-level child maltreatment risk factors align with county-level rates of child maltreatment reports and victimization?
(2) What distinguishes counties with high child maltreatment risk factors but low report rates from other counties, and vice versa?
(3) Throughout the United States, how do child maltreatment risk factors and report rates in rural counties where the majority of the population are members of a racial or ethnic minority compare with risk factors and report rates in other types of counties?
(4) In counties that have higher risk factors for child maltreatment but lower child maltreatment report rates than other counties in the same region, how do county-level report rates vary by report source and race of the child subject?

Planned data sources

The site expected to link data from the NCANDS, which includes child-level data about maltreatment reports, decisions by child protective services agencies about allegations, child and caregiver characteristics, and other information.

The site also planned to use publicly available data from a variety of sources to define county demographics, socioeconomic conditions, and other characteristics. These sources included the Census Bureau, the U.S. Department of Agriculture, the Bureau of Labor Statistics, and others.

Project leadership and participating organizations

This site differed from some of the others in that it had a very small team that was not based out of a research center. The site functioned as part of a public university but did most of the work on its own. The PI hired two graduate students to serve as research assistants for the project.

The site did not establish partnerships with any agencies that were owners of administrative data. Rather, data were accessed through publicly available sources as well as through a request to the National Data Archive on Child Abuse and Neglect, which disseminates the NCANDS data.

Project status

The site was able to access the planned data sources, except for NCANDS data on counties with fewer than 1,000 reports of maltreatment. As a result, the site was able to address three of its four proposed research questions, but the data were not available to conduct analyses on counties with relatively low report rates.

III. ENHANCING ADMINISTRATIVE DATA LINKAGES: SITES' APPROACHES AND PROMISING PRACTICES

In this chapter, we describe sites' approaches to enhancing data linkages and share promising practices that emerged from their work. We structure the discussion around the following five phases of work that we expected sites to undertake to complete their planned enhancements:

1. *Developing research questions and exploring data partnerships*. In this phase, we anticipated that sites would define the focus of their research projects, identify potential data sources, and begin establishing relationships for accessing data, if necessary. (As noted, sites had a proposed project in mind as part of the site selection process but may not have defined all aspects of it.)
2. *Sharing and accessing data*. Activities during this phase might include developing data sharing agreements, establishing approaches to protecting confidentiality and privacy of data, and securing necessary approvals, such as approvals from an institutional review board.
3. *Preparing data sets and completing data linkages*. This phase might involve assessing the characteristics of data sets, processing and cleaning data, and implementing data linkage methods.
4. *Conducting analyses* to answer research questions. In this phase, sites would use a variety of methods to analyze the linked data sets and would determine the key findings from those analyses.
5. *Reporting the results of their research* through various dissemination activities. This phase might include identifying audiences for reporting and presenting results of analyses in a range of formats.

We did not expect that sites would necessarily proceed through the phases of data linking and analysis at the same pace. For example, if sites had problems accessing data, they might repeat the steps related to exploring and acquiring data. Some sites could spend more time than others on linking data if they intended to link multiple data sets. Nor did we expect progress through the phases of data linking and analysis to be linear for all sites. For instance, obstacles encountered during the process of acquiring or matching data might lead a site to refine its research questions or analytic plan.

It is also important to note that the CMI Data Linkages project's focus on enhancements meant that some sites had already accomplished work on some phases of their studies. For example, sites seeking to access new data sources to integrate into existing linkages had access to other data sources they needed for their project; often, those data sources were already cleaned and processed. Similarly, all sites had linked child welfare data on previous projects, so they were adding to or updating work they were doing already.

This chapter describes and explores the reasons for variation in sites' experiences. It also explores the level of effort required to accomplish different phases of the data linking and analysis process, based on questionnaires we fielded to the sites between March 2019 and September 2020. Data from these questionnaires can help us develop insights into how sites allocated their effort across the five phases, and how this allocation varied from one site to another. At the end of each section on an activity, we highlight promising practices that sites adopted to address challenges and move through each phase of the data linking and analysis process. Site representatives and the Feasibility Study team identified these practices as distinct strategies or actions that helped them complete each activity.

The COVID-19 pandemic affected the pace and progress of sites' studies. In describing how sites approached their work, we note conditions related to the pandemic that affected their efforts to share and acquire data and complete other steps in the process of linking and analyzing data.

A. Developing research questions and exploring partnerships

We explored several distinct activities within the first phase of the data linking and analysis process: developing research questions, creating project plans, and exploring data sets and partnerships.

1. Developing research questions

All sites' research questions built on prior work. Familiarity with relevant data and analytic approaches informed sites' research plans and helped them understand what questions might be feasible to address. For instance, the site replicating previous work in a new state (ADHSS/OHSU) understood the data sources that would be required to conduct similar analyses of cumulative incidence and explore the potential to replicate the ALCANLink methods. Similarly, another team (CSSAT) considered its study to be an extension and updating of a project involving child welfare data in Washington State, the Risk of Death and Injury Study.

To varying extents, sites included data partners in discussions as they specified research questions, or they focused questions on topics they knew to be of interest to the field as a whole. For example, the CDN/Rady team had research questions in mind because the data provider and research organization had already wanted to work together before the project's inception. Staff from the hospital had approached the site's PI to express their interest in linking hospital records to other data sources to expand their understanding of the social determinants of health. In response, the PI proposed the validation of the Predictive Risk Model (PRM). Thus, the CDN/Rady project goals were aligned with the priorities of both the research team and data partner. The ADHSS/OHSU site's initial discussions about the scope of the research project included a representative of the Oregon Health Authority (OHA) Maternal and Child Health Unit, who became a key team member. Researchers in the CSSAT site were aware of the agency's interest in a better understanding of incidence and referrals related to the opioid epidemic, although they did not consult directly with the child welfare agency to develop research questions.

Sites also worked with the CMI Data Linkages project team to refine their research questions. This process was intended to help teams focus on topics of high interest and relevance to the CAPTA research priorities. For instance, exploring methods for estimating the incidence of child maltreatment, in addition to exploring risk factors, was a high priority for the federal CMI Data Linkages project. The CMI Data Linkages team partnered with the CDN/CCWIP site to identify options and articulate research questions focused on this goal.

2. Creating project plans

In their project plans, sites specified their project teams, research questions, expected data sources, proposed methods for linking administrative data, expected analytic approaches, timelines, and the opportunities and challenges they expected to encounter during the project. The plans established clear expectations for the focus of each research project and the steps teams would take. When necessary, sites refined their plans in consultation with the federal CMI Data Linkages team.

Experience with similar projects informed sites' project plans and expected timelines. For example, project leaders in the ADHSS/OHSU site noted that the plan for replicating the ALCANLink in Oregon drew heavily on steps and methods established in Alaska.

Even with careful planning, some sites encountered delays in their work, including delays related to the COVID-19 pandemic. As described in section III.G, sites used various strategies to continue moving projects forward in the face of delays. Two examples of these strategies were preparing for analyses before receiving data (CSSAT) or relying on in-house expertise to accelerate the data linking process (CDN/Rady).

3. Exploring data sets and data partnerships

The four sites that used individual-level data (ADHSS/OHSU, CSSAT, CDN/CCWIP, and CDN/Rady) built their projects on a foundation of relationships they already had with some data partners. The one site that used publicly available data (UA-SSW) set out to integrate additional data from a data source it had used in the past. For the site replicating its approach to data linkages (ADHSS/OHSU), access to data in one state set the foundation for validating the methods in another state. Several sites had preexisting relationships that gave them access to child welfare data. The remaining site (UA-SSW) had accessed publicly available data from the National Data Archive on Child Abuse and Neglect (NDACAN). Because the site used publicly available data, the team did not develop direct relationships with public agencies that produce administrative records, but the team did draw on its experience requesting data from NDACAN.

In general, research teams' familiarity with child welfare, vital records, and health data from previous work facilitated their projects, because the teams were acquainted with the data's basic content and organization. For example, the ADHSS/OHSU site had worked with child welfare records in Alaska for an earlier project, so the team know what to expect in terms of data cleaning and record layout in Oregon. The CDN/CCWIP, CDN/Rady, and CSSAT teams already had access to and had used the child welfare data needed for their CMI linkage projects because of earlier projects. Their CMI projects benefitted from their experience using the child welfare data, because the teams did not need to spend time exploring and understanding the data.

The questions sites sought to answer, and the enhancements they pursued to answer them, drove their exploration of new data. Two of the three sites that aimed to access new data sources to add to existing linkages (CSSAT and CDN/Rady) worked to deepen their understanding of these data sources and the potential for accessing them. These projects needed outside expertise to explore their new data from hospitals (billing and diagnostic codes). For example, the CSSAT team consulted with a colleague to understand which International Classification of Diseases (ICD) codes correspond to drug tests

The third site accessing new data (ADHSS/OHSU) was focused on replicating an existing data linkage. This site team needed to confirm the replication site's suitability and explore more deeply the structure and content of the state's child welfare data. The availability of relevant administrative data, especially PRAMS and child welfare data, was a key factor related to the selection of the replication site. The team also took into consideration contextual and organizational similarities and differences between the two states (Alaska and Oregon). The co-PI worked closely with the data analyst in the Oregon Department of Human Services (ODHS) to explore the structure of child welfare data in preparation for developing data sharing agreements. For example, the co-PI confirmed that the data file would include variables indicating the timing of events like reports of maltreatment or removals.

Sites that aimed to use new methods to analyze existing data linkages (CDN/CCWIP) or scale existing linkages with publicly accessible data (UA-SSW) spent time exploring data, but the nature of their activities was distinct from prior work. In one site (CDN/CCWIP), the team already had access to child welfare administrative data through an ongoing data sharing agreement. The team's exploration focused on defining options for identifying a "focal referral"—that is, an individual child's referral event, recorded in administrative records, that would inform estimates of child maltreatment—and other issues specific to the research questions. In another site (UA-SSW), exploration focused on identifying publicly available data sources that would provide county-level information on demographics, context, and outcomes, and supplement their existing data.

Box III.1. Promising practices: developing research questions and exploring partnerships

- Including partners, especially data partners, in early discussions, such as defining project research questions, may help researchers strengthen their collaborations. This process also offers an opportunity to highlight the benefits of data sharing to new data partners to get them on board with the project.
- Preexisting relationships with data partners and familiarity with data sets can inform the focus and scope of projects that are designed to enhance data linkages. ◢

B. Sharing and accessing data

When examining sites' experiences in sharing and accessing data—the second phase of the data linking and analysis process in the project's conceptual model—we focused on four key activities: (1) developing agreements for data sharing and use; (2) protecting the data's security, confidentiality, and privacy; (3) securing IRB and other approvals; and (4) accessing the data.

1. Developing agreements for data sharing and use

Sites relied on existing and new agreements with data partners to access data necessary to complete CMI Data Linkages studies (Table III.1). Sites aiming to access new data (CDN/Rady, ADHSS/OHSU, and CSSAT) needed to identify who had the authority to grant access, and determine the appropriate processes for making requests. Understanding the structure of state agencies was important for navigating the process of completing data sharing agreements and identifying whether multiple approvals were required. Some projects (ADHSS/OHSU and CSSAT) worked with more than one state agency and had to identify the approval authority within each. In one site (ADHSS/OHSU), the core team included an advocate in one state agency (Oregon Health Authority). This person was an effective liaison with her own agency and other state agencies that were providing data or supporting the analysis. In another site (CDN/Rady), the research team had to identify the people with approval authority to share hospital admissions data: a chief administrative officer at the hospital and a transactions officer at the university medical school the hospital is affiliated with.

All sites had data use agreements (DUAs) between the PI or research organization and each separate agency (Table III.1). No sites were required to have multiparty DUAs. Multiparty DUAs can be more cumbersome to establish because they require coordination from multiple agencies.

Table III.1. Data use agreements and research approvals, by site

Site	Type of agreement/approval	Agency approving data use	Timing of agreement completion
ADHSS/OHSU	DUA: PRAMS data	Oregon Health Authority (OHA), Oregon Public Health Division, Section of Maternal and Child Health	March 2019
	DUA: vital records	OHA, Center for Health Statistics	August 2019
	DSA: child welfare data	Oregon Department of Human Services, Children, Adults and Family Division	October 2019
	DSA	Integrated Client Services	November 2019
	IRB	OHA and Oregon Health and Science University	July 2019
CDN/CCWIP	DUA: child welfare data	California Department of Social Services	Existing long-term agreement (May 2016)
	DUA: vital records	California Department of Public Health	Existing agreement renewed annually (June 2018)
	IRB	University of Southern California, Children's Data Network, State of California's Committee for the Protection of Human Subjects	June 2018
CDN/Rady	DUA: hospital data	University of California San Diego	July 2018 (before CMI Data Linkages project began)
	DUA: hospital data	Rady Children's Hospital San Diego	April 2019
	DUA: child welfare data	California Department of Social Services	Existing long-term agreement (September 2016)
	IRB	University of Southern California, Children's Data Network, State of California's Committee for the Protection of Human Subjects	June 2018
CSSAT	DUA: child welfare data	Washington State Department of Children, Youth, and Families	Existing long-term agreement (June 2018)
	IRB: vital records and hospital data	Department of Health	Existing long-term agreement (April 2015)
	IRB	Washington State IRB	July 2019, September 2019 (revision)
	IRB	University of Southern California IRB	April 2019, September 2019 (revision)

Source: Data collected for the CMI Data Linkages Feasibility Study.

Note: The UA-SSW site did not need any IRBs, DUAs, or MOUs for the data used in the CMI Linkages project.

ALCANLink = Alaska Longitudinal Child Abuse and Neglect Linkage Project; DUA = data use agreement; IRB = Institutional Review Board; PRAMS = Pregnancy Risk Assessment Monitoring System.

Provisions in sites' agreements with data partners focused on administrative, technical, and physical safeguards to protect data. For example, agreements created for CMI Data Linkages projects included requirements that data be transmitted and stored securely; that they not be moved, copied, or transmitted without safeguards; that they not be sold; that confidentiality was protected and no identifying information revealed in any research data sets or publications; that access to data would be limited to only those directly involved, and that data breaches would be reported as soon as possible. The agreements also specified that facts cited about the data must be accurate, and that the data could only be used for specified study purposes. Representatives of data partners participating in some CMI Data Linkages projects noted that researchers built trust by demonstrating familiarity with laws and regulations about data access, as well as procedures for protecting data.

Overall, there were similar provisions in all new DUAs.[11] The newly established agreements included similar provisions across agencies and agreements—including provisions about the use of data, dissemination of results, and procedures for responding to disclosures of information—but the specific details of each varied by agency. For example, several sites' agreements with the state child welfare agency (ADHSS/OHSU, CDN/Rady, and CDN/CCWIP) specified that research must support the missions of public health and child welfare agencies, and that the agency must be consulted about analysis results and dissemination products before any dissemination takes place. The ADHSS/OHSU agreements also specified a "minimum necessary information" policy: researchers must request only the data necessary to answer their research questions. Yet another agreement specified that research staff consult with the data partner (CDN/Rady) about any disclosures that might be required by law, so the data partner could consider how to respond.

The level of flexibility for sites to use acquired data for additional or alternative analysis varied by site, as allowed in their DUAs. For example, CDN/CCWIP, CDN/Rady, and CSSAT have broad DUAs with their child welfare agencies, allowing the data to be used for numerous projects and analyses. This broad license was an asset for their CMI projects because they did not have to reestablish access or permission to use the data through a new DUA. These broad DUAs still have agency review requirements for use of the data even though the sites did not have to re-establish the access or permission. For example, in the CDN/CCWIP and CDN/Rady sites, the DUA between CDN and CDSS allows for CDSS data to be used for "research purposes specifically … sanctioned in writing by CDSS." (Agreement 16-MOU-00945 between CDSS and USC, CDN, September 2016). In contrast, the ADHSS/OHSU site had permission to use child welfare data for the CMI Data Linkages project specifically, rather than broad authorization.

Using publicly available data simplified and accelerated data acquisition for some sites. No data sharing agreement was required for data sets in one (UA-SSW), but an application process was required to obtain the data from NDACAN. Because NDACAN data are de-identified data submitted by states, the data UA-SSW received were already clean. Other sites (ADHSS/OHSU and CSSAT) used some administrative data sources, such as vital records, that are frequently accessed by a variety of users. In these sites, states had established procedures for sharing these types of records, which involved a request and standardized transaction rather than a full partnership and approval process. These data sources offer advantages in terms of ease of access, but also present limitations for research projects involving

---

[11] Sites and agencies use both data use agreement (DUA) and data sharing agreement (DSA) to mean the formal agreement that clearly documents which data are being shared and how the data can be used. In this report, we refer to all of these agreements as DUAs.

administrative data linkage. Policies on these data sources are subject to change.[12] Though the process is standardized, it can still be hard to access the data in most states, especially in a way that allows individual-level linking.

2. Protecting the security, confidentiality, and privacy of data

To access and use data, sites had to meet security standards established by multiple agencies and institutions, but sites' existing protocols were stringent enough that they did not require adjustments for the CMI Data Linkages projects. Sites had standard security protocols in place, which they were able to leverage to meet CMI Data Linkages project requirements. Details on sites' standard security protocols are in Table III.2.

Sites' protocols abided by the separation principle—personally identifiable information (PII) was separated from analytic files and used only to link records. In two sites (CSSAT and ADHSS/OHSU), research staff did not have any access to PII. An external third party completed the linkages and returned a completed research file via encrypted transfer, with no identifying information. This increased data security because no team members had direct access to the individual-level records. In two other sites (CDN/CCWIP and CDN/Rady), PII was processed only on non-networked computing stations by a select group of non-research staff and used only to link records. These data are not backed up externally, only to special encrypted devices. After linkages are completed, restricted analytic data sets are stripped of all direct identifiers and created and processed on a secure data server.

In the site using only publicly accessible data from NDACAN (UA-SSW), the research team members still abided by standard security protocols. For example, they used double-password–protected computers in locked offices with encrypted cloud storage. In addition, to prevent the potential for identification of individuals, NDACAN policy does not permit the sharing of county-level data for counties with fewer than 1,000 child maltreatment reports. Because of this policy, UA-SSW was unable to obtain data on counties with fewer than 1,000 child maltreatment reports in their analysis.

3. Securing IRB and other approvals

Sites' experiences securing IRB approvals or modifications ranged from rapid approvals to prolonged delays. Three sites that aimed to add new data sources to existing data linkages (UA-SSW, CSSAT, CDN/Rady) submitted IRB amendments or modifications to existing IRB packages. Two sites (UA-SSW and CDN/Rady) completed IRB modifications that were approved relatively quickly. The modifications were approved by university IRBs. The UA-SSW analysis did not involve individual identifiable data, and thus had an easier IRB process.

One site seeking approval of an amendment through a state IRB (CSSAT) faced substantial delays. Because an initial amendment request did not include all the variables the team needed to access, they needed to submit another amendment. The team waited three months for approval of this amendment. An IRB amendment was also required due to a change in personnel for the project—specifically, the person who was linking the records. Finally, additional state and university approvals were required because the home institution of the co-PI changed. In all, IRB processing in this site lasted about six months.

---

[12] As of January 2021, a DUA is required for ordering vital records data in Washington State. The data might not be as recent as other administrative sources and are aggregated to a level that allows for analyses based on geographic- but not child-level characteristics or outcomes. See Vital Statistics Data Homepage: Washington State Department of Health.

Table III.2. Standard data security protocols, by site

Site	Security protocol
ADHSS/OHSU	Data managed according to Oregon Health and Science University (OHSU) data privacy and security policies.
	Data are stored on secure servers maintained by OHSU's Advanced Computing Center. The servers are physically located at an off-site facility with emergency power, weekly backup, multiple firewalls, and physical security. Only staff listed on the DUA and in the IRB protocol for this study have access to the data.
	OHSU staff and researchers complete annual training on research integrity, which includes data security and confidentiality.
	HIPAA compliant
CDN/CCWIP	Restricted analytic data sets are stripped of all direct identifiers and created/processed on CDN's secure data server; approved CDN researchers access data on private computing nodes requiring VPN and MFA.
	The lab is in a building with strict physical security—24-hour security requiring badge access—and lab itself is a separately keyed, locked vault containing encrypted devices with PII.
	Lab abides by CDSS Protocol for CDSS data, and public health division protocol for public health; USC Committee for Protection of Human Subjects also reviewed and approved protocol.
	HIPAA compliant
CDN/Rady	Restricted analytic data sets are stripped of all direct identifiers and created/processed on the Children's Data Network (CDN) secure data server; approved CDN researchers access data on private computing nodes requiring VPN and MFA.
	CDN's computer lab is in a building with strict physical security—24-hour security requiring badge access—and lab itself is a separately keyed, locked vault containing encrypted devices with PII.
	Lab abides by California Department of Social Services Security (CDSS) Protocol for CDSS data, and public health division protocol for public health data; University of Southern California (USC) Committee for Protection of Human Subjects also reviewed and approved protocol.
	HIPAA compliant
CSSAT	Data stored on a secure server accessible through remote desktops at the University of Washington (UW), which require UW accounts to first access servers, then access the databases stored on SQL servers.
	HIPAA compliant
UA-SSW	Data stored in double-password–protected computers in locked offices with encrypted cloud storage.

Source: Data collected for the CMI Data Linkages Feasibility Study.

DUA = data use agreement; HIPAA = Health Insurance Portability and Accountability Act; IRB = institutional review board; MFA = multifactor authentication; PII = personally identifiable information; VPN = Virtual private network.

The ADHSS/OHSU site required a new IRB for the project. Team members noted that the process for securing a new IRB approval (from a university) was relatively smooth. For example, the board granted approval within the expected time frame and did not request substantial new information about the planned approach. Team members attributed the positive experience with this process approval to an advisor's previous experience with the process.

4. Accessing data

Sites were generally successful in assessing the data needed for their CMI Data Linkages projects. Sites accessed and used 18 of the 20 planned data sources in their analyses.

Sites' experiences accessing data varied by the type of data they intended to use. Although some sites had data use agreements in place for child welfare data before the CMI Data Linkages projects began, establishing agreements to share and use these data generally required substantial time and effort. Teams also navigated challenging processes to access hospital data (CDN/Rady and CSSAT), which required negotiation with multiple parties or relatively complicated approvals. The CDN/Rady team shared their impression that the process for acquiring hospital data was more cumbersome than other types of data because private hospital data has not be used for research as much as other types of data, such as vital records. Teams using vital records data generally found these records relatively easy to access because state agencies had established procedures in place to share them.

For the two data sources that were not accessed, the types of data differed, and so did the reasons that sites did not acquire them. A site seeking statewide data from a prescription management system (CSSAT) was unable to access these data because of problems engaging the data partner, the state's department of health. The site did not have an existing relationship with this data partner and found that communicating with key contacts was difficult to sustain in the context of the COVID-19 pandemic. A site using NCANDS data (UA-SSW) was unable to access data for counties with less than 1,000 child maltreatment records. This threshold was established by ACF to lower the risk that people living in smaller counties could be identified. (See Chapter IV for additional details).

Box III.2. Promising practices: sharing and accessing data

- Researchers can build trust with data partners by making sure they know the Federal, state, and local laws and agency-specific regulations regarding data access and by collaborating with liaisons in public agencies.
- Organizations that are interested in enhancing data linkages may be able to modify or amend existing data use agreements (DUAs) or research permissions to conduct their work. A broadly specified IRB that covers analyses of linked administrative data may also facilitate these types of projects.
- The PI's experience and knowledge of the IRB process facilitates the approval process, as does the competency of university IRB processes in general. This experience can be leveraged to require fewer revisions to the submitted IRB package.
- Although DUAs often require stringent data security protocols, research centers working with administrative data might already have such protocols in place. Collaborations with external entities to conduct data linkages are an additional means to ensure the privacy and confidentiality of personally identifiable information.

- Using publicly available data or data that are frequently accessed by a variety of users (where there are established procedures in place to access the data) can simplify data sharing.
- Plans and timelines for projects involving administrative data linkages should build in room for delays, especially related to data acquisition, and identify opportunities to accelerate other activities or use the time to prepare for analyses.◢

C. Preparing data and completing linkages

In this phase of their projects, sites cleaned and processed data and linked records across data sets. In some cases, sites worked with partner organizations to execute the linkages, as described below.

1. Processing data

Sites often used protocols they established in earlier projects to access, process, and clean newly received data. They also primarily used data that had been through these procedures before. Several sites used data cleaning and diagnostics protocols that the lead organization had developed and applied before undertaking CMI Data Linkages projects. These protocols involved standardizing some fields necessary for data linkage, such as addresses and dates of birth. They also involved checking the means and ranges of key variables to find outliers or unexpected values. One team (CDN/Rady) noted that this process helped them identify variables with values they did not understand that would require clarification. As described in Chapter IV, specialized expertise with some elements of newly acquired data sets, such as diagnostic codes in hospital data, supported sites' data processing and cleaning. In at least two sites (ADHSS/OHSU and UA-SSW), the research team relied partly or fully on data sources (survey data, vital statistics data, and NCANDS data) that had undergone quality control and cleaning during data collection or preparation for public use.

In at least one site (CSSAT), initial assessments of data from one provider revealed issues related to data quality. A variable related to hospitals was determined to be unusable, and the initial extract had missing and corrupted data. As a result, the research team needed to request the re-extraction and re-transfer of the files. This process took several months, resulting in delays in the project timeline. However, the site was able to use older data to begin analyses that could be refreshed once the corrected data were received.

2. Completing linkages

To link individual-level records, sites used deterministic, probabilistic, and combined approaches. Sites selected linkage methods based on the type of data they used, their previous approaches to linkages, and the composition of their project teams. In the ADHSS/OHSU site, linkages involved a combination of deterministic and probabilistic methods, scoring, and manual matches. A state agency, Integrated Client Services (ICS), completed data linkages on behalf of the research team. After several rounds of matching, records are linked based on the highest scoring match. To integrate PRAMS and vital records data, ICS used slightly different methods for each data source. A deterministic match based on the birth certificate number was used to link PRAMS and vital records data. A probabilistic match based on names and date of birth was used to link vital records to CPS data.

One site (UA-SSW) used a direct method to link data at the county level. Data sets were merged based on a geographic identifier, the FIPS code. The site matched all counties with other data sources, with the exception of about 200 that were missing NCANDS data.

Three sites (CSSAT, CDN/CCWIP, and CDN/Rady) used machine learning techniques to complete data linkages. The CSSAT site relied on a cloud-based software product for data integration, "AWS Glue." As described below, the site adopted this method after its originally planned approach (which involved deterministic and probabilistic methods) became infeasible because of an IRB requirement that a third party complete the linkages. The software uses a machine-learning algorithm to identify and link records across data bases. The research team was able to adjust software settings to avoid false-positive matches. The team also blocked on gender to reduce the number of record-pair comparisons.

The CDN/CCWIP and CDN/Rady sites used a custom model that was developed for previous work. The model generates match probabilities based on similarities in linkage fields. Analysts conduct manual reviews of uncertain matches and use the results to "train" the model and improve its performance as new data are integrated. In the CDN/Rady site, the newly added data source of hospital records did not include Social Security numbers (SSN), a variable they typically use to link data. When the team first ran its linkages program, the match rate at the first run of the linkages program was much lower than expected. After consulting with data partners to understand the missing data pattern, the team was able to revise the linkage program to reflect the missingness of SSN in the algorithm. The match rate was higher after the revised linkages program, in the expected range.

Two sites linking individual-level data (CDN/Rady and CDN/CCWIP) reported correct match rates of 85 to 92.5 percent, respectively. The research team indicated that these rates were within the expected and acceptable range for the field (Rebbe 2019). Two sites (ADHSS/OHSU and CSSAT) were unable to report metrics related to linkage results. In one site, the research team did not yet have access to these metrics. In another, the site team was still finalizing linkages when the site had finished reports for the CMI Data Linkages project.

3. Collaborating with partners to execute linkages

Two sites (CSSAT and ADHSS/OHSU) needed to develop approaches for working effectively with an individual or agency responsible for completing data linkages independently from the research team. In the CSSAT site, the research team's agreement with the state IRB stipulated that a named individual outside the PI's organization have direct access to PII to conduct the linkages. Because of a change in personnel, this task was assigned to a staff member in a partner organization, and data linkages were not this person's primary field of expertise. To get the linkages done, the team opted to use a cloud-based software product (AWS Glue) that offers visual interfaces to control the linkage process instead of programming code. A drawback of this approach is that the linkage algorithm used in the software is not transparent to the research team, making it difficult to monitor the quality of linkages.

In the ADHSS/OHSU site, data partners required that a state agency, Integrated Client Services (ICS), complete data linkages on behalf of the research team. This agency receives and links data from multiple state programs and agencies on a monthly basis. Because of the partners' requirement, to ensure that the original ALCANLink process could be replicated and to limit unnecessary sharing of data, the site team needed to take steps to understand the linkage process and algorithm ICS would use to link new data, such as data from the PRAMS survey. Involving a separate agency in data linkage also meant that the research team was not able to monitor the quality and completeness of linkages during that process. It was therefore important to establish a high level of confidence and trust in the linkage approach from the outset. The site team held an in-person meeting with representatives from ICS to discuss the basic approach and linkage flow for each data source. The team then documented this flow in project materials and its IRB application. Ultimately, it was determined that ICS's linkage approach was close enough to the ALCANLink method.

Box III.3. Promising practices: preparing data and completing individual-level record linkages

- Consulting with staff who have specialized expertise on the content of administrative data sets supports the preparation for linkages.
- A variety of approaches may be useful for linking administrative data. Technical familiarity with the data and the flexibility to tailor linkage approaches to the content of a specific data source may be important to successful linkages.
- When working with a third party to complete linkages, clear communication regarding the linkage approach can help the research team have confidence that linkage algorithms will operate as expected.
- Machine learning techniques and tools may be an efficient method for linking larger databases.◢

D. Conducting analyses

Sites used various analysis methods, reflecting the range of research questions and data sources that their projects involved. Table III.3 summarizes the analytic methods sites used to answer individual research questions. Analysis methods included calculating descriptive statistics, cross-sectional analyses, longitudinal analyses, correlational analyses, and regression modeling, including hierarchical linear models.

Sites did not report making major adjustments to the analyses originally presented in their project plans. At least one site (ADHSS/OHSU) noted that having a clearly developed analytic plan at the start of the project was important to a successful analysis. The plan provided a clear road map to an end point for the project. An understanding of the data and experience with the planned analysis methods made it easier for sites to complete analyses as planned. For example, in one site (CDN/Rady), knowledge of ICD codes was essential, because certain codes are indicative of maltreatment in a particular age range but not another, and one code might overwrite another code indicative of maltreatment. In the site attempting to replicate a previous analysis (ADHSS/OHSU), the team aimed to use an approach that was as consistent with previous methods as possible. Another site (CSSAT) leveraged its experience with the analytic approach to move its project forward in the face of data delays. The site developed programming code for analyses based on old data and a previous research project and expected to re-run this code using new data once they arrived.

Table III.3. Analytic methods used to answer research questions, by site

Site	Research questions	Methods
ADHSS/OHSU	Does the cumulative incidence of the time to first maltreatment report in Oregon differ from that observed in Alaska?	Calculated survival function to derive cumulative incidence for Oregon PRAMS cohort; compared to previous Alaska PRAMS cohort analysis
	Are the cumulative incidence to first report, contact, and substantiation estimated through the Oregon PRAMS linkage consistent with a full Oregon birth cohort linkage to child welfare?	Calculated survival function to derive cumulative incidence for Oregon birth cohort and Oregon PRAMS cohort; compared estimates
	What are the key components required for successful replication of ALCANLink methods?	Assessed site team's own experience based on resources, technology, skill sets, and agreements required to conduct project
CDN/CCWIP	How do estimates of victimization of child maltreatment vary based on (a) number of years in an estimate window (for example, only in focal year [2015] or up to three years on either side of focal year) or (b) county-level variation in how likely victims are to be identified? [Cross-sectional strategy]	Calculated a series of revised cumulative victimization rates with both children who were substantiated victims in a base year, as well as those who were referred to CPS but were substantiated as a victim within three years on either side of the base year. Estimated county-level variation in the annual/cross-sectional substantiation rates and use that county variability to extrapolate a range of estimated maltreatment rates.
	How do annual incidence and cumulative prevalence rates differ by county and demographic characteristics at birth for children born in California in 1999? To what extent does earlier involvement with the child welfare agency predict substantiated child maltreatment? [Longitudinal strategy]	Organized CPS records longitudinally for a cohort of children born in 1999 to estimate the cumulative childhood risk of abuse and neglect in California
CDN/Rady	What are the characteristics and service trajectories of child hospital patients, compared to other children born in the county?	Calculated descriptive statistics that summarized the nature of medical encounters in hospital data and the identifying characteristics of children in linked data
	What information does this provide about the social determinants of physical (medical), mental, and behavioral health of children served in this hospital?	Calculated descriptive statistics that summarized nature of medical encounters in hospital data and the identifying characteristics of children in linked data
	To what extent are children who have been identified by a statewide predictive risk model to be "at high risk" of maltreatment also at elevated risk of injury, poor health outcomes, and mortality in childhood?	Classified risk levels of unique children in linked data using score assigned by predictive risk model; assessed alignment between ICD codes suggestive of maltreatment and model's risk scores
	What is the predictive value of integrating hospital data as predictors in the predictive risk model?	Classified risk levels of unique children in linked data using score assigned by predictive risk model; assessed alignment between ICD codes suggestive of maltreatment and model's risk scores

Site	Research questions	Methods
CSSAT	How do individual-level and county-level opioid use and misuse impact the risk of maltreatment for children and families in Washington State?	Employed Bayesian model averaging to identify variables for regression models; completed chi-square analysis on descriptive distribution table
	How does opioid use and misuse impact child welfare system contact for children and families in Washington State? Are there are sociodemographic differences between those who are tested and those who are not tested for opioid use?	Research question not answered (see section III.B.4)
	How do indications of opioid use/misuse or prenatal opioid exposure at the family and county-level impact substantiation and placement decisions within households under investigation for maltreatment?	Identified opioid-exposed infants through maternal/child ICD-9 codes; ran multistate survival model to identify placement outcomes
UA-SSW	How well do county-level child maltreatment risk factors align with county-level rates of child maltreatment reports and victimization?	Calculated bivariate correlations
	What distinguishes counties with high child maltreatment risk factors but low report rates from other counties, and vice versa?	Used hierarchical linear modeling to assess differences between rural counties with majority White population or rural counties with majority Blank population, and other counties with differing levels of rurality/demographic makeups
	Throughout the United States, how do child maltreatment risk factors and report rates in rural counties where the majority of the population are members of a racial or ethnic minority compare with risk factors and report rates in other types of counties?	Used hierarchical linear modeling to assess differences between rural majority or minority counties and other counties with differing levels of rurality and demographic characteristics
	In counties that have higher risk factors for child maltreatment but lower child maltreatment report rates than other counties in the same region, how do county-level report rates vary by report source and race of the child subject?	Research question not answered (see section III.B.4)

Source: Data collected for the CMI Data Linkages Feasibility Study.

ALCANLink = Alaska Longitudinal Child Abuse and Neglect Linkage Project; ICD = International Classification of Diseases; PRAMS = Pregnancy Risk Assessment Monitoring System.

The analysis process in some sites involved ongoing decision making and testing of assumptions. For example, in one site that planned a more exploratory approach to analysis (CDN/CCWIP), the team needed to weigh the pros and cons of each decision or strategy. These decisions touched on issues such as how to define maltreatment (for example, based on allegations or substantiations), how to handle multiple types of allegations within a single report, and how to define cohorts for analysis. Members of the site team noted that regular meetings including on-the-spot analyses helped the team make decisions and move the work forward. In another site (CSSAT), the analysis process included discrete steps and methods to identify variables to prioritize for modeling. The research team in this site used Bayesian model averaging to identify the best variables to include in regression models with different outcome measures. Another site highlighted that an understanding of the data includes awareness that the definition of child maltreatment incidence is complicated, subjective, and culturally informed, and that the analysis reflects these ambiguities.

Three sites (ADHSS/OHSU, UA-SSW, and CDN/CCWIP) said it was an analytic challenge to address mobility of individuals across geographic jurisdictions—both within state and out of state. In the ADHSS/OHSU site, the research team had data available in the original site (Alaska) to identify individuals who left the state in the time frame they were examining. These data were not available in Oregon, and it was not possible to censor records when sample members moved out of state. The team considered using inverse probability weighting to adjust estimates for emigration in the replication site but did not complete these analyses within the project time frame. Other sites acknowledged that cross-state or cross-county mobility was a consideration for estimating incidence rates over time. Their analyses acknowledged this limitation but did not adjust for it.

Box III.4. Promising practices: conducting analyses

- Having a clearly developed analytic plan at the start of the project can contribute to a successful analysis by specifying plans before beginning analysis.
- Researchers can adapt and use a wide range of analysis methods to explore child maltreatment with administrative data, including Bayesian model averaging (CSSAT) and latent transition analyses (CDN/CCWIP).
- Deep familiarity with child welfare data and reflection about on-the-spot interim analyses can support exploratory approaches to estimating incidence.
- Researchers can consider mobility in sample populations and jurisdictional differences in child welfare administrative practices, definitions, and policies when exploring results and interpreting analytic findings—but clear-cut solutions to these challenges can be difficult to identify.◢

E. Reporting results

Sites' dissemination of findings was delayed, partly because of the COVID-19 pandemic. (Section G has details on COVID-19 delays). Some sites were still completing analyses and preparing to share their findings as of this report's writing. Nevertheless, sites planned or had undertaken a wide range of activities to communicate the results of their projects. In this section, we focus on their planned dissemination approaches. We discuss the findings from sites' projects and their contributions to the knowledge base on child maltreatment in Chapter V.

Sites' planned formats for reporting findings included peer-reviewed journal articles, conference presentations, and briefs, as well as more informal presentations of findings through such methods

as presentations to data partners and a podcast episode. All sites were preparing manuscripts for submission to academic journals. At the time of this report's writing, several articles had been published.[13] Sites have shared their work at state, regional, and national conferences representing a variety of professional fields, including social work, pediatrics, public health, and epidemiology. One site's PI was invited to speak on a podcast from a university's school of social work. The podcast was geared toward social work students and highlighted the implications of the research for understanding the relationship between county-level characteristics and child welfare referrals.

Sites have considered a variety of audiences in their dissemination planning, including policymakers, researchers, data partners (especially child welfare agencies), and practitioners. Team members at one site (CDN/Rady) believe that health care and child welfare practitioners will be interested in the relationship between ICD codes indicative of maltreatment and actual maltreatment reports to child welfare. Team members in the ADHSS/OHSU site expected to share findings with representatives of state agencies in both Oregon and Alaska. For example, the team planned to share initial analyses of the relationship between parental stressors and child maltreatment in oral presentations to different Oregon agencies and partners and in written reports. For public agency audiences in Alaska, the team planned to share comparative and pooled analyses to describe differences, similarities, and exploration of population risks in the two states. The UA-SSW team hoped to use information from their analyses to develop county-level indices of risk and protective factors related to child maltreatment, which could be useful to policymakers, practitioners, and researchers.

As detailed in their DUAs with agencies, sites had to share results with the agencies before making them publicly available. Sites also shared preliminary and interim findings with their data partners, working to build trust in and give credence to the final results.

The ADHSS/OHSU, CSSAT, and UA-SSW sites planned to create briefs, how-to guides, and other resources for researchers interested in replicating the methods used in CMI Data Linkages. For example, the ADHSS/OHSU site planned to create a guide for stakeholders interested in replicating the ALCANLink approach, laying out key steps, data requirements, and resource needs. Teams also discussed the possibility of developing programming packages that would facilitate similar analyses by other researchers in the future.

---

[13] Rebbe, R., A.S. Bishop, J. Ahn, and J.A. Mienko, J.A. "Opioid Overdose Events and Child Maltreatment Indicators: Differential County-Level Associations." *Children and Youth Services Review*, vol. 119, 2020. https://doi.org/10.1016/j.childyouth.2020.105671.

Smith, B.D., Q. Li, K. Wang, and A.M. Smith, A.M. (2021). "A National Study of Child Maltreatment Reporting at the County Level: Interactions Among Race/Ethnicity, Rurality, and Poverty." *Children and Youth Services Review*, vol. 122, 2021. https://doi.org/10.1016/j.childyouth.2021.105925.

Prindle, J., R. Foust, and E. Putnam-Hornstein. (revise and resubmit). "Maltreatment Type Classifications and Transitions During Childhood for a California Birth Cohort." *Child Maltreatment*, forthcoming.

Putnam-Hornstein, E., E. Ahn, J.J. Prindle, J. Magruder, D. Webster, and C. Wildeman. "A Birth Cohort Study of Terminations of Parental Rights Due to Child Abuse or Neglect." *American Journal of Public Health*.

Box III.5. Promising practices: disseminating findings and methods

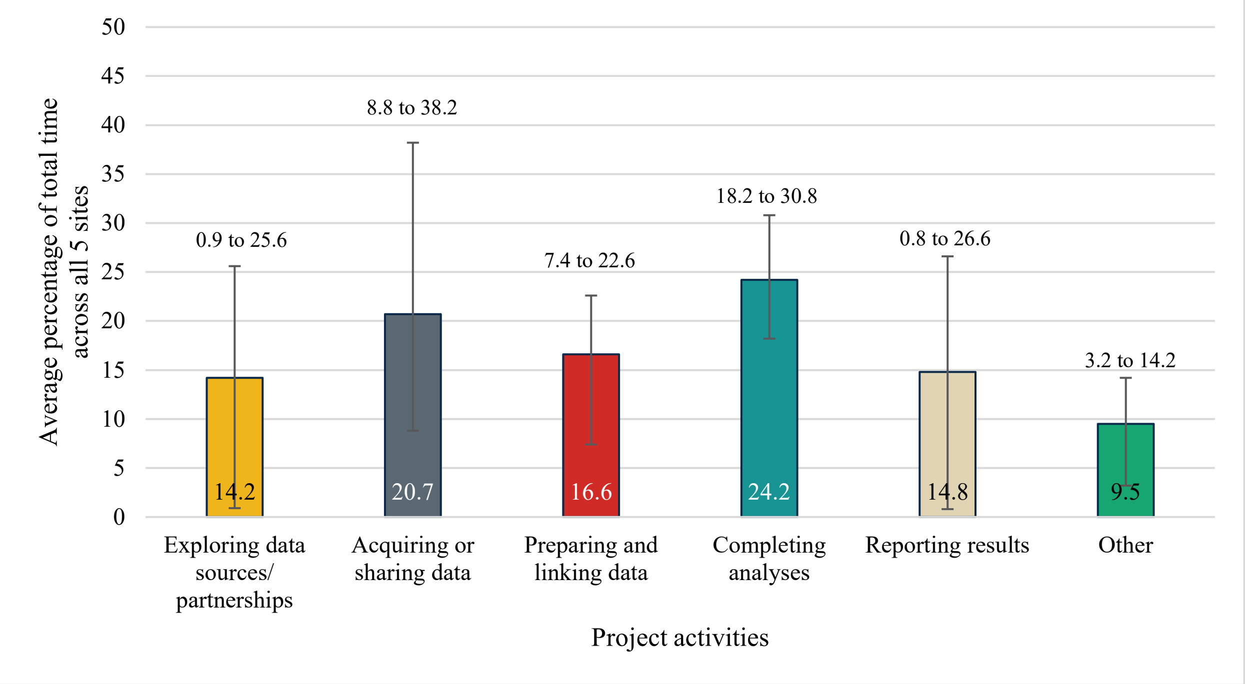
- Disseminating results in a variety of formats and channels may help ensure that the results of enhanced data linkages are shared with a range of audiences and stakeholders.
- Prioritizing the reporting of findings to data partners may improve partnerships, build trust, and support policymakers in applying the results of research on child maltreatment and incidence.
- Dissemination may also support replication of methods to enhance administrative data linkages to generate knowledge on child maltreatment incidence and risk and protective factors.◢

F. Allocation of time and effort across phases of data linking and analysis

Questionnaire data collected between March 2019 and September 2020 tell us how sites allocated their time across the five phases of data linking and analysis, when they conducted each activity during the life of their projects, and the amount of time staff allocated to the project over time.

The reported allocation of personnel hours to activities varied widely across sites. As shown in Figure III.1, on average, personnel across all sites spent the most time completing analyses (24.2 percent) and acquiring or sharing data (20.7 percent), followed by preparing and linking data (16.6 percent) and exploring data sources/partnerships (14.2 percent). The proportion of time spent exploring data sources ranged from next to none at CSSAT (0.9 percent) to 25.6 percent at CDN/CCWIP. In contrast, there was less variation in the time spent completing analysis, ranging from 18.2 percent (CDN/Rady) to 30.8 percent (UA-SSW). Sites spent an average of 10 percent of their time on other tasks, including project team meetings and general project management.

Figure III.1 Average percentage of project time by activity, across all sites



Note: "Other" project activities include project team meetings, general project management, and other activities not reflected by the other categories.

Time spent on activities aligned with the type of enhancement projects that sites pursued. For example, two sites that used new data (ADHSS/OHSU, CDN/Rady) spent 31.1 percent and 38.2 percent of their time, respectively, on acquiring data, whereas the other three sites used about 10 percent of their time acquiring data. The CSSAT site, which used some data sources it had access to at the project's outset, reported spending only 1 percent of the team's time exploring new data sources. The CDN/CCWIP team also had existing access to data but reported spending 25.6 of their time exploring new data sources. This time might have been used to investigate the potential for using the data sets in new ways.

In general, the timing of sites' activities aligned with expectations—for example, sites spent the most time exploring data sources at the beginning of the project—but the questionnaire data suggest that some activities took place throughout the sites' project periods. As expected, time spent exploring data sources and acquiring or sharing data was frontloaded between March 2019 and December 2019. Sites spent time preparing and linking data throughout the entire data collection period (an average of 16.6 percent of project time was spent during the period when questionnaires were fielded). Time spent completing analyses gradually increased to its highest (35.9 percent) from July 2020 through September 2020. A similar amount of time (34.9 percent) was spent on reporting results during that period.

G. Responding to the challenges of the COVID-19 pandemic

Even with careful planning, some sites encountered delays in their work, including delays related to the COVID-19 pandemic. Key research team members from CSSAT, UA-SSW, CDN/Rady, and ADHSS/OHSU sites had to respond to COVID-19-related data analysis and policy issues. For example, the ADHSS/OHSU PI needed to support the Alaska state public health response to the pandemic. At the UA-SSW site, the PI's administrative responsibilities at the university increased because of the pandemic, affecting the PI's availability and creating some challenges in the project's analysis phase. Research assistants were able to retrieve, clean, and link the data, but the PI maintained responsibility for reviewing their work and leading the analysis. Staff from data partners in health agencies in two sites (CSSAT and CDN/Rady) appeared to be especially affected by new demands on their time. Consequently, it became more difficult to maintain regular contact with data partners, and requests for data were not quickly fulfilled.

The CDN/Rady team was in the process of linking data when the COVID-19 pandemic began. This delayed the work, which it had to be done in person on a non-network connected computer to ensure data security. Ultimately, the person with the appropriate expertise was able to return to the research team's offices. Some sites reported that the remote work environment also made it difficult for their already distributed teams to easily coordinate with each other. This meant the analysis and data work took longer than expected.

Two sites noted that the pandemic also forced adaptations that might have been good for their work. Sites used various strategies to continue moving projects forward despite delays, such as preparing for analyses before receiving data (CSSAT) or relying on in-house expertise to accelerate the data linking process (CDN/Rady). Research organizations and state agencies also had to learn how to operate in a remote environment and have established new protocols to facilitate and streamline data sharing. These adaptations might make multi-agency data linkage projects easier to execute in the future.

IV. HOW CONTEXTUAL FACTORS AND ORGANIZATIONAL CAPACITIES INFLUENCED ENHANCEMENTS TO ADMINISTRATIVE DATA LINKAGES

This chapter discusses the contexts that the CMI Data Linkages sites conducted their research in, the organizational capacities that sites brought to their projects, the external supports provided by the broader CMI Data Linkages project team, and how these factors affected the enhancement of data linkages (Table IV.1).

With respect to contextual factors, the Feasibility Study focused on features of the analyzed jurisdictions that had the potential to affect implementation of the sites' projects. These features included the following:

- **Characteristics of the child welfare system**, such as the system's structure (for example, whether it was a state- or county-administered system) and whether major changes had recently occurred in the organization or operation of the system (for example, leadership changes, major reform efforts, or data system changes).
- **Relevant child welfare policies and definitions**, such as policies on record expungement, criteria for substantiation, and mandated reporters.
- **State and local statutes on data use**, such as whether laws or regulations exist that govern access to and use of administrative data.
- **Existing structures for data linkage**, including whether an integrated data system, central clearinghouse, or collaborative exists for linking data from various public agencies.

With respect to organizational capacity, the Feasibility Study focused on the organizational attributes or abilities that might enable the sites to complete their intended enhancements and accomplish their project's goals. Key concepts explored in the study included the following:

- **Leadership to promote data linking**, or the ability of a champion within an organization, especially a data steward, to mobilize support for using linked administrative data to build the knowledge base in child welfare.
- **Partnership quality**, or an organization's ability to establish and maintain collaborations that would support its project by facilitating consensus on research agendas and sharing data.
- **Technical infrastructure and expertise**, or the specialized infrastructure and expertise required to complete the process of linking and analyzing data—including, legal and regulatory expertise, data management expertise, analytic expertise, and project management expertise.
- **Translation capacity**, or the ability of researchers and other participants to convey findings effectively to agency leaders and other stakeholders, which might help sustain efforts to use linked administrative data.
- **Access to resources**, or the extent to which a site had sufficient staff time, funding, and materials to support its project, and how the site allocated and used these resources.

Table IV.1 Contextual and organizational factors that facilitated and hindered administrative data linkages

Facilitators
Similar contextual factors in replication site, for example: • The availability of similar data sources • The perceived similarity in terms of child welfare policies, geography (including both urban and largely rural or frontier areas), and the demographics
Prior experience linking administrative data, for example: • Staff with expertise in linking administrative data • Knowledge of federal privacy laws and laws related to administrative access in the state • Previous involvement in IRB committees • Strong collaborative relationships with state child welfare agencies. • Familiarity with data sets
Existing organizational infrastructure, for example: • Technology infrastructure and support services available • Comprehensive data management systems and processes and robust computing environments
State leadership support for data linkages
Access additional sources of funding to conduct data linkage projects
Barriers
Differing definitions of child maltreatment
Less robust partnerships with data stewards
High level of effort to complete data sharing agreements
Competing demands on staff time

As part of their participation in the broader CMI Data Linkages project, sites were offered a variety of external supports, such as cross-site learning activities and TA. The Feasibility Study explored the sites' perceptions of the role of these activities and supports in their pursuits of their intended enhancements.

How contextual factors, organizational capacity, and external supports influenced the sites' projects informs our understanding of what facilitates or hinders enhancements of administrative data linkages and the potential to implement enhancements on a broader scale.

A. Contextual factors

As noted in Chapter II, variability in contextual factors was one of the criteria used to select sites for the project. This variation provided an opportunity for the Feasibility Study team to identify considerations related to different geographic areas and policy contexts. The Feasibility Study explored factors related to the characteristics and policies of the child welfare systems in the jurisdictions that the sites were studying, the local context for data use, and existing infrastructure for ongoing linkages of public agency data.

1. Structure of the child welfare systems

The overall stability in child welfare systems appears to have facilitated the sites' work to enhance administrative data linkages. Overall, sites did not report major barriers to implementing their projects as a result of such issues as administrative changes in the child welfare system. In one site (CSSAT), the child welfare agency was involved in a larger departmental reorganization, which resulted in leadership changes. However, the child welfare contacts with whom the site had an existing relationship remained the same, and the new leaders continued to support the site's work. Sites also reported that there were no major changes to the child welfare data systems that affected their projects.

The structure of child welfare systems affected how some sites articulated their research questions and approached their analyses of administrative data. Specifically, one site (CDN/CCWIP) conducted its project within a county-administered system; thus, the site opted to explore in its analyses potential local variation in child welfare policies and administration. Other sites (CSSAT and UA-SSW) crafted research questions that acknowledged and explored potential variation in child maltreatment reporting across counties and socioeconomic contexts. As described later, the sites also attempted to control for county-level differences in their analyses or they interpreted the findings with potential differences in mind.

The governance structure of state agencies may have shaped the implementation of the projects, as noted in Chapter III. The data that most sites planned to use for their analyses were owned by different state agencies, thus requiring the sites to work with each agency to access the data. However, in the site that replicated an existing linkage (ADHSS/OHSU), child protective services and the division of public health were in the same state department where the original linkage was completed. The two data sources were governed by a single authority, which facilitated the data acquisition process.

2. Child welfare policies and definitions

Variation in child welfare policies and definitions across jurisdictions was a challenge for sites as they analyzed their data. However, familiarity with the data and prior knowledge of the potential variation allowed researchers to prepare for this challenge as they planned and implemented their projects.

- The site using national data (UA-SSW) noted that the team needed to control for or take account of state- and county-level differences in its analysis. The site used data on child maltreatment referrals that had been screened in (or investigated). However, each state has different protocols for what referrals need to be screened in. The site's analytic models included relevant variables such as whether a state had a universal reporting law as a proxy to control for these differences. The site also conducted analyses to rule out the possibility that its finding of racial and ethnic differences in reporting reflected regional reporting differences rather than differences based on the racial and ethnic composition of rural counites. However, the research team noted that the county-level data did not account for within-county variation, which may be relevant in more populous counties.
- Another site operating in a county-based child welfare system (CDN/CCWIP) also had to contend with varying policies and practices across counties, which may contribute to differing rates of investigation and substantiation of maltreatment reports. The site attempted to address this issue by using a state-level, population-based approach and by considering rates of potential maltreatment over a period of time. The site explored and discovered county-level variations in the data. Given differing practices, some counties may overreport alleged incidents of child maltreatment while others may underreport them. However, the site team attempted to smooth for this county-level variation by focusing on state-level outcomes.

  This team also conducted a cross-sectional analysis, which may help address variations in policies and practices by exploring incidence over a period of time, capturing maltreatment that may not have been substantiated in one place or time but was reflected in earlier or later substantiations. The site team intends to examine the issue of policy variability further by replicating its analysis in other jurisdictions. If the approach is validated, it suggests the approach could identify differences in maltreatment risk rather than variation in policy.

- Variation in policies and definitions across states was an issue for the site aiming to replicate a data linkage approach (ADHSS/OHSU). The state in which the replication occurred defined a category of maltreatment ("threat of harm") that was not used in the original state. When comparing cumulative incidence across the two states, the research team ran analyses that both included and excluded threat of harm as a separate category to account for any potential effect of the definitional differences.

In some cases, jurisdictions' child welfare record retention policies facilitated efforts to use administrative data to explore child maltreatment incidence. For instance, the research teams at two sites (CDN/CCWIP and CDN/Rady) highlighted the benefits of the state's approach to handling expungement of reports of child maltreatment. In this state, the central registry for child abuse operates separately from the child welfare agency's data system and is managed by a separate agency. Although reports are expunged from the central registry after 10 years, they are preserved in the child welfare data system regardless of a report's ultimate disposition. Thus, the sites had access to information on all reports made to the agency, not only those that were substantiated.

3. Legal and policy context for data use

Sites had a mixed experience as to whether data archive or state policies or regulations on data use affected the design and implementation of their projects. In some sites, laws or policies created circumstances that required adjustments to a project's plans; in other sites, the legal and policy context facilitated their efforts.

- In one site (UA-SSW), NCANDS policies that prohibited access to data from counties with fewer than 1,000 records (even when these data were aggregated) prevented the site from exploring research questions focused on rural areas.
- In two sites (CSSAT and ADHSS/OHSU), state regulations and policies regarding the use of data required sites to adjust the planned implementation of their projects. Due to state regulations, one of the sites (CSSAT) was required to seek approval from the state IRB to conduct its project. The research team noted that securing approval from the state IRB was generally more time-consuming and costly than using the site's university IRB. The site replicating its data linkage approach (ADHSS/OHSU) was prevented from conducting its own linkages due to state agency regulations regarding access to PII in administrative records. As described in Chapter III, the site was required to work with Integrated Client Services, a unit within the health department, to link the data for the project. As a result, the research team had to meet with representatives from the unit to ensure the linkage method would be comparable to the site's original data linkage approach.
- State laws and regulations in some sites facilitated the availability of certain types of data for analysis and information that supported the linkage process. One project (CSSAT) was conducted in a state that considered some types of hospital records to be vital records, which facilitated access to a data set that might otherwise be more difficult to obtain. The research team in another site (CDN/CCWIP) noted that researchers using California data can link administrative records by using Social Security numbers, which improved record match rates. Some states prohibit researchers from using this type of information in record linkages.

4. Existing infrastructure for data linkage

Several states in which sites conducted their work support clearinghouses or maintain agencies with the specific purpose of integrating administrative data from multiple state agencies. The Feasibility Study team had anticipated that existing clearinghouses, "linking hubs," or data integration initiatives, such as the ones in the Actionable Intelligence for Social Policy network, might be part of the context for sites. Thus, the Feasibility Study explored whether and how sites interacted with these clearinghouses to complete their projects.

- In Washington State, the Department of Social and Health Services (DSHS) Research and Data Analysis division maintains an integrated database with longitudinal data on individuals served by multiple state agencies. These data include records from programs overseen by DSHS and two other state agencies, the Health Care Authority and the Department of Children, Youth, and Families. The division that oversees the database also operates the state's IRB.
- In Oregon, the Integrated Client Services unit operates as a shared service between the Oregon Health Authority and the Oregon Department of Human Services (see Chapter III). The unit conducts monthly linkages of individual-level records in administrative data from a range of programs and conducts data linkages in response to specific requests from researchers.
- The California Health and Human Services Agency has partnered with the university research center that is part of the CMI Data Linkages project to link and organize administrative, individual-level records across eight agency programs and data sources. The project is known as the California Health and Human Services Record Reconciliation and Data Hub. The data integration covers such programs and records as the Supplemental Nutrition Assistance Program, the Temporary Assistance for Needy Families program, vital records, and others.

Notably, of the sites working in these states, only one directly accessed the existing administrative data linkages supported through an agency or clearinghouse. ADHSS/OHSU)partnered with the Integrated Client Services unit to complete its data linkages. As an intermediary, the unit provided a structured process for executing data use agreements, accessing data, and completing linkages. This facilitated the site's work, according to some participants in the project.

The other two sites did not use their state's existing linkage infrastructure. The site working in Washington State (CSSAT) was updating a linkage it had previously completed and that included data outside the scope of the integrated data maintained by the Research and Data Analysis division. The site did apply for approvals through the IRB that the division operates. The site working with statewide data in California (CDN/CCWIP) already had access to the relevant data sources and an established process and methods for linking data sources relevant to its research questions.

B. Organizational capacities

The sites selected for the CMI Data Linkages project all had prior experience with linking administrative data as well as existing organizational capacity from which they could draw to implement their data linkage enhancement within the project's compact timeline and budget. The Feasibility Study explored the organizational context of the sites, including leadership and support from data steward decision makers, and the quality of the relationships between key participants in the project. The study also examined the technical capacities that sites brought to their projects, including legal and regulatory expertise, data management expertise, analytic expertise, project management expertise, and research translation expertise. Finally, the study assessed sites' access to and perceptions of the sufficiency of resources to conduct their projects.

1. Leadership to promote data linking

In the four sites (CDN/CCWIP, CSSAT, ADHSS/OHSU, and CDN/Rady) that accessed individual-level data from states, support among policymakers and state agency leaders for data-driven policymaking and quality improvement efforts led data stewards to participate constructively in the CMI Data Linkages projects.

State legislators and child welfare agency leaders in three states where sites conducted research (California, Washington, and Oregon) had endorsed exploring child welfare data generally and using child welfare administrative data specifically to address policy questions. For example, in California, the child welfare agency had contracted with a university to maintain a public dashboard of child welfare metrics using data from throughout the state.

Representatives of state health and human services agencies in two sites (ADHSS/OHSU and CSSAT) were highly supportive of efforts to use administrative data to benefit policy and practice. One representative noted that using the data productively was part of being a good steward of public resources [data].

The existence of research and analytics units within state health and human services agencies also points toward administrative support for data linking in the sites. For instance, the mission of the Research and Data Analysis division within the Washington State Department of Social and Health Services is to provide decision support through analysis of administrative data, including linked data sets. Similarly, the Integrated Client Services unit in Oregon exists to facilitate ongoing linkages of administrative data from various state agencies.

2. Partnership quality

A history of data sharing between partners was important but not sufficient for successful partnerships. For new and existing partnerships, a collaborative relationship characterized by mutual trust between data stewards and researchers was critical to facilitating the data linkages enhancements.

Three sites (CSSAT, CDN/CCWIP, and CDN/Rady) had been partnering with their states' child welfare agencies for more than a decade and had built strong collaborative relationships with the agencies. These sites engaged in ongoing activities with the agencies beyond the CMI Data Linkages project and had defined goals for collaboration as well as structures for communication. In general, these sites benefitted from administrative support and engagement from child welfare agencies that were motivated to participate because they were interested in actionable findings from the research.

- One site (CSSAT) had an ongoing data partnership with the child welfare agency, which was renewed annually. The PI's research center served as a university partner that was willing to facilitate analyses of the agency's child welfare system data. The PI's research center maintained a data portal for reporting aggregate statistics on the state's child welfare population. Through this partnership, the research team also conducted a precursor study on risks for child maltreatment.
- In the other two sites (CDN/CCWIP and CDN/Rady), the state child welfare agency recognized that there were limitations to the program and policy questions that could be answered without connecting its data with other data sources. The agency recognized the potential for university-based researchers

to be valuable third parties in some data agreements. State policy included allowances for research and evaluation, which meant that university partners could receive the data more easily than state agencies. Thus, the state partnered with the PI's research center to link the agency's data with other data sources and assist with additional research and evaluation activities.

These sites also had a history of acquiring data from their states' departments of health; however, the relationships were more transactional. There was limited ongoing collaboration between the departments and the PIs' research centers.

Respondents commented that the quality of relationships with the data stewards was as important or more important than the specifics of the data sharing agreement in promoting ongoing collaboration. A data steward at one site noted that it took a long time for the individual's agency to establish a trusting partnership with the research center. At the beginning, the two entities struggled to reach an agreement that was acceptable to both sides. The data provider noted that it would be difficult for the agency to approve a project similar to the CMI Data Linkages project without previous experience working with the research team. Starting from scratch with a new organization might have been too resource intensive for the agency.

Two sites (ADHSS/OHSU and CDN/Rady) had to establish new partnerships to access data for their projects. In both cases, efforts to initiate these partnerships were facilitated by an existing relationship with individuals connected to the data stewards.

- The site replicating its approach to data linkage (ADHSS/OHSU) did not have a long history of collaborating with the data stewards in the replication state (Oregon). However, the site did have connections with some individuals at agencies and institutions within the state and these connections were instrumental for building successful partnerships with the data stewards. The data partnerships with state agencies were facilitated by a representative from the Oregon Health Agency, who championed the project. The health agency representative garnered support for the project from other members of the agency's leadership team by illustrating how the project aligned with the overall goals, priorities, and requirements of the agency's strategic plan. The health agency representative also helped the research team identify key partners and build connections within the state's child welfare agency and the unit conducting the data linkages. The site's co-PI was based at a university within the state, which offered additional connections and experience with state agencies that were data stewards.
- The CDN/Rady site had a long-standing partnership with California's child welfare agency but needed to partner with a new data provider to accomplish its CMI Data Linkages project. The site developed a strong relationship with an initial contact from the data steward (a hospital), but it took time to build support for the project among other hospital administrators. The hospital did not have a history of participating in social science research or sharing data because its primary focus was clinical care and because its administrators were unclear about how participation would benefit the hospital. Eventually, the hospital shared its data with the site, and the site is planning additional collaborative work with the data steward. For example, IRB protocols and MOUs with the data steward were amended to permit the linkage and analysis of additional hospitalization data to address questions related to the COVID-19 pandemic.

Box IV.2. Sites’ strategies to initiate and maintain partnerships

Sites relied on multiple strategies to initiate or maintain their partnerships:

- The two research centers that conducted ongoing work for child welfare agencies communicated regularly with them. The PI from one research center (CSSAT) met with the child welfare agency's data and reporting administrator at least twice per month to review the research center's activities and discuss the data and research findings. The research center affiliated with two sites (CDN/CCWIP and CDN/Rady) maintained its partnership with the child welfare agency through scheduled calls and occasional in-person meetings.
- In the site replicating its linkage approach (ADHSS/OHSU), regular communication within the project team (which included a health agency representative) supported ongoing collaboration and troubleshooting related to agency partnerships. Members of the project team noted that an in-person meeting at the outset of the project with agency partners, including the unit linking the data, established a strong foundation for communication and collaboration.
- In the site partnering with a hospital (CDN/Rady), communicating the benefits of the project to administrators-specifically, the potential advantages for managing children's care at the hospital-was a key strategy for strengthening the partnership.◢

Less robust partnerships with data stewards appeared to affect some sites’ ability to gain access to data. One site (CSSAT) had gained access to vital records data from the state health department previously but had difficulty doing so during the CMI Data Linkages project. Although the department approved a request for access to a new vital records data set, the site never received the data. It was unclear to the research team why obtaining this data set was more challenging than the other vital records data sets it had obtained. In another site (UA-SSW), the request for publicly available data was completed through a standard transaction with NDACAN, which does not engage in the same type of formal partnerships with researchers as state agencies do. Although the PI had accessed data from NDACAN previously, the site’s request for county-level data from some rural counties was not approved because HHS policy prohibits the release of aggregated data for counties with fewer than 1,000 child welfare reports to lower the risk that people living in smaller counties could be identified. The site had limited ability to pursue alternatives with the archive, partly because of the lack of a partnership mechanism.

3. Technical expertise

Representatives of the sites reported that they had the technical capacity and expertise necessary to accomplish their enhancements. The sites drew on a unique body of expertise and level of existing infrastructure to complete their projects. This section discusses the role of four types of expertise in sites’ projects: (1) legal and regulatory expertise, (2) data management infrastructure and expertise, (3) analytic expertise, and (4) project management expertise.

a. Legal and regulatory expertise

Team members from the sites’ lead organizations had the necessary expertise to facilitate the development of data sharing agreements and research approvals, despite not always having specialized legal or regulatory infrastructure or support within their organizations. A representative of the child welfare data steward in one site (CSSAT) noted that the research team’s knowledge of the strict laws related to administrative access in the state (Washington) and its understanding based on previous work of the data use that would be permitted facilitated the development of an agreement that allowed the study to proceed. Team members from two sites (CDN/CCWIP and CDN/Rady) noted that knowledge of federal privacy laws was helpful when working with data stewards to describe circumstances in which data sharing was allowable. The team members previously had created primers that they shared with data stewards to explain circumstances in which the type of research they conducted using PII was permissible, specifically with respect to the Health Insurance Portability and Accountability Act (HIPAA) and the Family Educational Rights Privacy Act (FERPA).[14]

The sites' data partners had authorities designated to determine the legality of and provide support for sharing data. These authorities had the expertise necessary to evaluate whether potential data sharing arrangements were permissible and advisable. However, a representative of a data steward in at least one site indicated that the level of effort involved in processing agreements and responding to requests was considerable. According to this person, the agency would have benefitted from having staff dedicated to processing the paperwork involved in these tasks.

Team members in several sites noted that their previous involvement in IRB committees offered expertise that was useful to their projects. This experience helped these researchers learn how other researchers approach the application process, how the committee thinks about questions being proposed, and how to best position the research organization's future applications, which facilitated their projects' application development and approval process. For example, based on her IRB committee experience, a team member in one site (ADHSS/OHSU) suggested including a graphic in the application that showed the data sources, how they were going to be linked, and who was completing the linkages. Based on the lack of questions from the review board on the data linkage process, the team member assumes the graphic facilitated the review process by clearly displaying for the review board the complicated linkage process.

b. Data management infrastructure and expertise

Sites generally had an existing, robust infrastructure for data management, which allowed them to accomplish the work involved in the CMI Data Linkages projects. All sites benefitted from technology infrastructure and support services in universities affiliated with their projects.

Sites had comprehensive data management systems and processes and experienced staff prior to initiating the CMI Data Linkages projects. The sites had robust computing environments to manage large data sets as well as system backup and recovery. The research center affiliated with two sites (CDN/Rady and CDN/CCWIP) had a dedicated information technology staff member, separate from the university information technology staff, who handled the center's information technology environment. Within their research centers, sites had experienced staff who were familiar with database management, querying, advanced statistics, and the requisite software programs.

Sites also had heavily invested in developing data security systems and processes. All sites' protocols for data security were compliant with HIPAA. Sites were able to leverage this infrastructure to facilitate data acquisition and meet the CMI Data Linkages project's security requirements (see Table III.2 for a summary of sites' security protocols).

---

[14] The HIPAA primer is available at https://www.datanetwork.org/wp-content/uploads/CDN-HIPAA-Overview_Final.pdf. The FEPRA primer is available at https://www.datanetwork.org/wp-content/uploads/CDN-FERPA-Overview_Final.pdf.

c. Data linkage and analytic expertise

Representatives of the research teams in all sites indicated that they had the requisite linkage experience, knowledge of the data sets, and analytic expertise to implement their planned enhancements.

All the sites had existing linkage protocols to complete their enhancements. Three sites (ADHSS/OHSU, CDN/Rady, and CDN/CCWIP) had previously devoted significant time and expense to establish the linkage methodologies that were applied in their CMI Data Linkages projects. The development and maintenance of these sites' linkage processes were supported by grant funding

- The PI in the ADHSS/OHSU site had documented the data linkage processes used in Alaska, anticipating that other states could replicate this linkage. These steps were intended to ensure that the linkages would be performed in the replication states in a way that allowed for cross-state comparisons.
- The research center affiliated with the other two sites (CDN/Rady and CDN/CCWIP) had developed a machine learning–based record linkage model and protocol that was used in multiple projects. In addition, the research center had a staff member whose sole responsibility was data linkages.

Sites' deep familiarity with some of the data sets or elements within the data sets was critical to accurate linkage, analysis, and interpretation of the data. Three sites (CDN/CCWIP, CDN/Rady, and CSSAT) relied on state child welfare administrative data that they had accessed regularly. These sites were aware of the information these data sources included on characteristics of child maltreatment, such as type, timing, and chronicity. Sites' familiarity with elements from other data sets also facilitated linkages.

- For example, the PI in one site (CSSAT) noted that his comprehensive understanding of the child welfare data in his jurisdiction enabled him to write the database queries for the child welfare agency to extract the data for his projects. The PI and child welfare data steward met regularly to discuss findings and confirm interpretations. Because of the PI's familiarity with the data and their collaborative relationship, each brought to the other's attention questions about the data, so they could learn from each other's experiences and ensure the data were being used and interpreted correctly.
- At a site using hospital admissions data (CDN/Rady), a staff member on the research team was familiar with the International Classification of Diseases (ICD) codes that were an important element of the data set.
- At another site (UA-SSW), the team's familiarity with the Federal Information Processing Standard (FIPS) helped identify changes in the codes for geographic areas that affected its programming for linkages.

Members of the site teams highlighted other types of analytic expertise as important to their projects, including data manipulation, coding, and knowledge of relational data systems (whether through internal project team knowledge or through consultation with the data provider). In addition, sites leveraged their expertise in statistical software packages, including SAS, R, SPSS, and modeling software packages, which were important tools to support their work.

d. Project management expertise

By and large, sites' PIs were themselves responsible for keeping their projects on track. Their project management expertise was necessary to adjust to project delays or barriers. Several sites mentioned that a project manager would have been a helpful addition to their team, especially in handling changes to plans and staff availability as a result of the COVID-19 pandemic. The two sites (CDN/CCWIP and CDN/Rady) aligned with the same research center benefitted from the involvement of a staff member whose responsibilities included overall management of the research center's multiple projects involving data linkage.

Smaller teams in some sites (UA-SSW, CSSAT, and ADHSS/OHSU) presented an impediment to the work due to the limited availability of team members and partners. While all teams faced challenges related to moving projects forward while balancing competing priorities such as academic or administrative responsibilities, these challenges were particularly acute for the smaller teams. In addition, one site's PI (UA-SSW) noted that it would have been helpful to have more access to colleagues with whom the PI could brainstorm solutions to logistical challenges and discuss the work.

4. Translation capacity

Sites had a track record of communicating findings to key audiences in the child welfare field. For example, sites had a history of publicly sharing their findings and making their research accessible through data dashboard and visualizations, conference presentations, and publishing their work in multiple formats, such as briefs, white papers, and reports. Sites had also made their findings accessible by posting them on the websites of their research organizations.

As part of existing partnerships at some sites, child welfare agencies regularly reviewed findings that researchers produced using child welfare data, which facilitated the communication of these findings to key stakeholders. A data steward in one site (CSSAT) noted that a report that the PI's research center produced for another project was well-done and useful to program staff. She commented that the findings were presented in a way that was understandable to staff who were not accustomed to reading research and evaluation findings. In another site (ADHSS/OHSU), the PI's role as an employee of a state's health and human services agency allowed him to highlight the policy implications of analyses directly to decision makers and program administrators within the agency.

5. Access to and sufficiency of resources for conducting data linkage projects

Given sites' existing capacities, the research teams noted that the personnel and financial resources available to conduct their projects were generally sufficient. However, competing demands on staff time sometimes slowed the pace of the work. Representatives of some data stewards highlighted challenges related to the demands on staff time created by processing requests for data and collaborating with researchers.

Sites benefitted from in-kind contributions from the institutions or agencies that hosted the research center or team involved in the CMI Data Linkages projects. For example, one site (UA-SSW) noted that the university the PI was affiliated with provided in-kind computing resources. At another site (ADHSS/OHSU), needed software upgrades were paid for by the PI's agency. A data steward at a third site (CDN/Rady) noted that she was not directly compensated for her time on the project. She volunteered her time because she believed in the value of the project to her institution.

Representatives of several sites noted that the financial resources provided by a single grant typically would not be sufficient to cover the cost of a study like those conducted for the CMI Data Linkages project. In addition to the in-kind resources provided, several of the sites (ADHSS/OHSU, CDN/Rady, CDN/CCWIP, and CSSAT) had additional funding from philanthropic foundations to support their research and evaluation activities. In addition, in most sites, significant amounts of funding and resources had been spent over many years to build their infrastructure and staff expertise. One respondent highlighted the role of "core infrastructure funders," who supported the site's research center. These funders invested substantial amounts annually to subsidize the secure server environment in which the team linked, processed, and analyzed administrative records. The funding was also necessary to support a full-time information technology specialist charged with overseeing this analytic environment and various staff members tasked with maintaining the relationships, agreements, and protocols required for continued access to data.

C. Cross-site learning and TA

Through their participation in the federal CMI Data Linkages project, the sites received financial support and opportunities to engage in TA and cross-site learning. These activities were intended to help sites exchange information and address challenges they faced as they implemented their projects. Sites participated in the following activities:

- Regular contact between the PIs and a liaison from the CMI Data Linkages project team for discussions of study progress, emerging challenges, and TA needs.
- Cross-site learning network sessions among sites' research teams, the CMI Data Linkages project team, and federal staff. These sessions included presentations on the sites' projects; activities to explore challenges, lessons learned, and approaches to data linking; and discussions on dissemination strategies and next steps for sites' projects (Box I.3, in Chapter I).

  Webinars with external experts to review and discuss study methods and preliminary findings. The webinars were intended to (1) inform experts, stakeholders, and federal staff of the work that was being done through the CMI Data Linkages project and (2) enable the sites to discuss their work with and receive feedback from the expert stakeholders and federal staff.[15]

Sites reported that they valued the opportunity to learn from their peers, appreciated the stature their participation in the federal project brought to their work, and benefitted from the accountability structure that the larger CMI Data Linkages project imposed. Site representatives noted that they valued the opportunities to share information through the cross-site learning network and appreciated the support that the CMI Data Linkages project liaisons provided. Sites highlighted the benefit of being able to discuss issues arising on their projects, receive feedback, and commiserate about the challenges involved in their linkage efforts.

- For example, one respondent emphasized the usefulness of having the opportunity to present methods; explore hypotheses to explain findings (such as, differences in the incidence of reports and substantiations between Alaska and Oregon); and think through organizational challenges that came up as the project progressed. The respondent noted that the cross-site learning network activities and webinars provided an opportunity to share their work in a different way than through peer-reviewed literature, which tends to highlight novel methods and might overlook efforts that involve replication.
- A respondent from a small team with limited access to other experts in the data linking field felt that the cross-site learning network addressed a critical need for connections with other researchers.

---

[15] The project had originally planned an in-person meeting, but the meeting was changed to a series of webinars due to the COVID-19 pandemic.

- Another respondent noted that even when feedback from other sites and experts did not ultimately change the team’s approach to its work, the discussions prompted the team to rethink how to present and discuss the study’s findings.
- Respondents also appreciated exploring solutions for issues related to statistical programming and data access with the CMI Data Linkages project liaisons.

Sites appreciated the endorsement that the CMI Data Linkages project offered for their work. Multiple respondents felt that being involved in the federal CMI Data Linkages project added a beneficial sense of formality to the work, which created momentum within the team and facilitated work with state partners. Representatives of another site noted that the CMI Data Linkages activities underscored that the site’s work was part of a larger movement in the fields of child welfare and public health, was supported by a growing body of literature, and was of interest to funders and government agencies.

Sites also noted that the cross-site learning activities and TA kept the projects moving forward by creating structures for accountability. Regular check-ins with the CMI Data Linkages project liaisons helped sites monitor the progress of their work against anticipated timelines. Webinars encouraged sites to prepare papers and slides to share their initial findings and descriptions of their methods.

Sites offered several suggestions for refining or strengthening some aspects of the supports offered through the CMI Data Linkages project and the cross-site learning network:

- One respondent noted that it would have been helpful to have more opportunities to discuss issues related to data linkage methodology, such as the potential for developing a standard linkage process, various linkage applications, and a set of principles for handling linkages, because there is no single linkage paradigm.
- Several respondents noted that it was sometimes challenging to know the intended audience for the webinars and presentations—their peers, experts, or ACF. One respondent mentioned that discussions with the experts during the webinars sometimes focused on a single aspect of a study, which prevented the respondents from discussing other parts of their work that they would have liked to address. The respondent wanted to continue with the presentation but did not because it was unclear if the experts’ comments were intended to drive the focus and progression of the presentations.
- Another respondent felt that the structured activities and presence of multiple stakeholders (including a funder of the work) made it difficult to have a more informal exchange among sites. The respondent would recommend incorporating separate convenings for different types of stakeholders and participants.

Representatives of some sites indicated that connections that were initiated or deepened through the cross-site learning network might be sustained after the conclusion of the sites’ projects. For instance, some sites are exploring potential collaborations on future analyses of the linked data sets that were developed as part of their CMI Data Linkages studies.

This page has been left blank for double-sided copying.

V. FINDINGS FROM CMI DATA LINKAGES PROJECTS AND LESSONS LEARNED

Administrative data linkages can develop new knowledge about child welfare that would not be available otherwise or that would require substantial amounts of time and resources to produce using other data collection methods and data sources. This chapter discusses findings about child maltreatment that the CMI Data Linkages sites gleaned through enhancements of linked administrative data. It then explores the broad lessons learned through the CMI Data Linkages project about the potential for enhancements to administrative data linkage to inform our understanding of child maltreatment and its related risk and protective factors. Lastly, the chapter presents the advantages and disadvantages of using linked administrative data to estimate incidence compared to other methodologies (such as the NIS) and data sources (such as NCANDS).

A. Findings from the CMI Data Linkages research projects

This section presents an overview of the new information generated through the five sites' projects, as well as the areas of alignment with the CAPTA research priorities. The appendices to this report offer additional detail on the findings from each of the studies.

The sites' findings can be divided into five topics:

1. Incidence of child maltreatment
2. Incidence and prevalence of child maltreatment by demographic characteristics
3. Risk and protective factors related to child maltreatment
4. Trajectories of children with maltreatment reports
5. Methods for estimating incidence and predicting risk of child maltreatment

Each of these topics reflects research priorities outlined in the CAPTA legislation.

1. Incidence of child maltreatment

Some of the sites' findings map to two research priorities specified in CAPTA: (1) the nature and scope of child abuse and neglect and (2) the national incidence of child abuse and neglect.

a. Exploration of incidence in states

Two sites (ADHSS/OHSU and CDN/CCWIP) produced new state-level estimates of child maltreatment incidence in Oregon and California, respectively.

- **The ADHSS/OHSU team's analysis of child welfare data linked to public health and PRAMS data estimated the proportion of children born in 2009 with child welfare involvement up to age 9 in Oregon state.** This analysis found that nearly one-third of the children were reported to child welfare, and about 1 in 10 had a substantiated case of maltreatment. The team also found that a higher proportion of children were reported to child welfare before their first birthday in Oregon compared to children in Alaska.[16]

---

[16] Although it would be informative to compare these estimates based on linked data with estimates produced using child welfare administrative data only, Oregon reports incidence as the victim rate per 100,000 children annually, making comparisons with the ADHSS/OHSU team's estimates of cumulative incidence up to age 9 difficult.

- **The CDN/CCWIP team's longitudinal analysis of linked child welfare and vital records data provided new estimates of incidence by identifying the proportion of a cohort of children born in California in 1999 who were involved in child welfare during childhood (up to age 18). The analysis found that nearly one-third of children in the birth cohort were referred to child welfare at some point, with allegations of maltreatment substantiated for about one in ten children.** According to the research team, the estimates generated from this method were generally consistent with nationwide estimates of cumulative incidence from earlier studies using different methods (synthetic cohort life tables) and national data, such as NCANDS and AFCARS (for example, Kim and colleagues [2017] and Wildeman and colleagues [2014]).
- **The CDN/CCWIP team's longitudinal analysis also produced new estimates of the proportions of children born in 1999 with allegations of different types of maltreatment in different age ranges.** The team found that, among children reported to child welfare, an allegation was likeliest to occur from age 5 through 12. Smaller percentages of children with referrals experienced an allegation from infancy through age 4 and from age 13 through 17. Allegations of neglect were the most common type of maltreatment in all age groups.
- **Using a second method, the CDN/CCWIP team created new estimates of victimization based on analyses of annual cross-sections of children with referrals for alleged maltreatment between 2010 and 2018 in California.** The new estimates are designed to capture those who are substantiated as victims of abuse as well as those who are victims but do not come to the attention of the child welfare system in a given year. The team selected a base year (2013 to 2015) and examined how substantiation rates would change if they were adjusted to include children substantiated during the base year as well as children with unsubstantiated allegations in the base year but substantiated allegations in prior or later years. Overall, expanding the window for substantiation to three years results in a greater number of children considered victims of maltreatment by 72.5 percent. When this percentage increase was applied to 2015 NCANDS data, the estimated number of child victims nationwide rose from 683,487 to over 1.2 million. The findings suggest that in many cases incidence is not an isolated event of maltreatment but rather a set of conditions under which children live.

b. *Exploration of incidence in counties*

Two other sites (CDN/Rady and UA-SSW) explored estimates of the incidence of child welfare system involvement in counties.

- **The CDN/Rady team identified patterns of alleged maltreatment and child welfare involvement among a cohort of young patients of Rady Children's Hospital in San Diego County.** The cohort included children born between 2010 and 2014. In this group of hospital patients, nearly one-fifth had one or more reports of alleged maltreatment up to age 5. A small proportion of patients in the cohort had diagnosis codes specific to maltreatment (fewer than 1 percent); the vast majority of this group had referrals to child welfare. The analysis also indicated that a majority of children with child welfare involvement in San Diego County received medical care at Rady Children's Hospital—suggesting that the population served by these systems overlaps substantially.
- **The UA-SSW team produced new information on patterns of maltreatment reports in counties nationwide, with a focus on comparisons between rural counties with majority populations of color and those with majority white populations.** The analysis revealed that majority Black counties had the lowest mean child maltreatment report rate in 2015, followed by rural majority Latinx counties, and rural majority white counties. These patterns were consistent across all years in the data (2012 to 2015).

2. Incidence and prevalence of child maltreatment by demographic characteristics

The sites' research explored outcomes related to child maltreatment by various demographic characteristics at the individual and geographic levels. Findings from several sites address a CAPTA research priority related to increasing understanding of the characteristics of child maltreatment victims and their families. Sites also produced new information about disparities in the risk of children who experience maltreatment.

a. Demographics of children and families involved in the child welfare system

In addition to conducting statewide estimates of incidence, the CDN/CCWIP and ADHSS/OHSU sites explored the demographics of children and families involved in the child welfare system.

- **The longitudinal analytic strategy used by the CDN/CCWIP site produced new information about family characteristics associated with increased risk of various outcomes related to child maltreatment, such as termination of parental rights over time.** The analysis found, for example, that children born to teen mothers and children whose mothers had less than a high school degree were more likely to experience termination of parental rights compared to children born to mothers who were older or had more education. Rates of terminations of parental rights were much higher for families covered by public health insurance compared to those covered by private insurance and for children for whom paternity had not been established at birth compared to children with established paternity.
- **The ADHSS/OHSU team concluded that the Oregon and Alaska samples were similar in terms of incidence by demographic group.** The team's analysis compared maltreatment incidence based on such characteristics as race, ethnicity, maternal marital status and education, gestational age, and others. However, the team had not yet completed statistical tests for differences across subgroups at the time of this report.

b. Disparities and disproportionality in child welfare involvement

Findings from the CDN/CCWIP, UA-SSW, and ADHSS/OHSU sites add to the knowledge base on disparities in child maltreatment referral rates and involvement in the child welfare system across racial and ethnic groups.

- **The CDN/CCWIP team's longitudinal findings confirm disproportionate involvement among Black and Native American children in child welfare.** The team found that cumulative proportions of Black and Native American children who had child protection encounters, including investigations, substantiations, removals, and terminations of parental rights, were significantly higher than those of other children.
- **The UA-SSW team found that the relationship between reported child maltreatment and child poverty was different in rural, majority minority counties (primarily Black) compared to other counties.** This points to an exception to the pattern of disproportionately high system involvement in communities of color. The site team noted that this finding raises questions about service gaps related to inequities (UA-SSW et al. 2021).
- **The ADHSS/OHSU team's exploration of child maltreatment incidence among children up to age 9 in Oregon and Alaska indicated disproportionate involvement among some population groups in both states.** In both states, the proportions of Black and American Indian/Alaska Native

children with child welfare allegations were substantially higher than the proportion of white children. Estimates based on PRAMS data revealed similar disproportionality.

3. Risk and protective factors related to child maltreatment

Findings from two sites (CSSAT and UA-SSW) offered novel information about community-level risk and protective factors related to maltreatment outcomes. Studies investigating these factors shed light on the causes, prevention, and sociocultural dimensions of child maltreatment—topics that are prioritized in the CAPTA research agenda.

- **The CSSAT team's research on county-level opioid overdose rates as a risk factor for outcomes related to child maltreatment added to the literature on the relationship between opioid use and child maltreatment.** Previous studies have found positive relationships between indicators of opioid use at the county or state level, such as prescription rates, overdoses, hospitalizations, and deaths, and child welfare indicators such as caseloads (Rebbe et al. 2020). In contrast to these findings, the CSSAT team found no statistically significant relationship between opioid overdose rates and maltreatment outcomes at the county level when other demographic and socioeconomic variables were incorporated into the statistical models. However, the team noted that Washington State does not define parental substance abuse as maltreatment, although it can be taken into consideration when assessing neglect. This context gives professionals substantial discretion when assessing neglect, possibly leading to fewer referrals to child welfare. In addition, services available to families have declined due to state budget cutbacks. The team hypothesized that child welfare workers might respond to this reduction in services by becoming less likely to initiate substantiations or out-of-home placements.
- **The UA-SSW team's exploration of factors associated with patterns of child maltreatment reporting, especially in rural counties, found that rural counties with majority populations of color do not follow the typical pattern of positive association between rates of child maltreatment and child poverty.** Although child poverty rates were higher, on average, in counties with majority populations of color, the association between child poverty and child maltreatment report rates was negative. Rural counties with majority Black populations (n = 59) have the highest rate of child poverty, on average, but the lowest rate of child maltreatment reports. The team found the same pattern in rural counties with majority Latinx populations (n = 23), with a higher mean rate of child poverty and lower rate of child maltreatment reports compared with rural, majority White counties. The team noted that these findings align with studies from other countries that have found lower levels of child maltreatment reporting among highly marginalized populations (Bywaters et al. 2016; Sulimani-Aidan and Benbenishty 2013).

  The UA-SSW team also found that in rural counties, the social association rate (the number of membership organizations per 10,000 people) had a negative relationship with reports of child maltreatment. In non-rural counties, the relationship was the opposite; the social association rate had a positive association with maltreatment reporting. The team suggested that social association may function differently in different contexts. It may have a protective effect in rural places as a reflection of social capital and trust; in a non-rural context, it might lead to more opportunities for observation and reporting.

4. Trajectories of children with maltreatment reports

Longitudinal linkages of individual-level child welfare data allowed two sites (CSSAT and CDN/CCWIP) to examine transitions across outcomes among children involved in the child welfare system. These sites' studies offer information relevant to CAPTA research priorities focusing on "the nature and scope of child abuse and neglect," "the extent to which unsubstantiated reports return as more serious cases of child abuse or neglect," and "the extent to which incidents of child abuse and neglect are increasing or decreasing in number and severity" (U.S. Congress 2017).

- **The CSSAT team used a survival modeling approach to generate information on substantiation and placement decisions and transitions for infants diagnosed with prenatal exposure to opioids in Washington.** They found that opioid-exposed infants born in counties with high overdose rates were less likely to go home from the hospital after birth and more likely to be placed into foster care compared to infants born in counties with lower overdose rates.
- **Both of the CDN/CCWIP team's analytic strategies—the longitudinal method and the cross-sectional method—yielded information about the types of maltreatment children experience over time and the system responses to alleged maltreatment over time.** The team found that those children most likely to have additional child welfare involvement as they grew older were children with multiple and varied allegations of maltreatment when they were younger. The team's cross-sectional analysis found that a substantial proportion of children who are referred to child welfare but not substantiated in a given year may have allegations of maltreatment substantiated in an earlier or later year.

5. Methods for estimating incidence and predicting risk of child maltreatment

Three sites employed novel methods of estimating incidence and predicting risk of child maltreatment. These methods offer new means of understanding the nature and scope of child abuse and neglect and the national incidence of child abuse and neglect, which are both CAPTA priorities.

a. Estimating incidence of child maltreatment

Two sites (ADHSS/OHSU and CDN/CCWIP) conducted proof-of-concept projects that involved novel methods of estimating incidence.

- **The ADHSS/OHSU team found that data based on linkages between child welfare and PRAMS data may slightly underestimate cumulative incidence compared to a full birth cohort linkage—although, the differences in estimates were not statistically significant.** The team validated estimates of cumulative incidence up to age 9 in Oregon based on a linkage of PRAMS data to child welfare data, comparing estimates of incidence based on the PRAMS sample to the estimates a full birth cohort linkage would provide. Although it may underestimate cumulative incidence, the PRAMS linkage is unique in that it offers the opportunity eventually to explore risk and protective factors for child maltreatment that are captured in the PRAMS survey.

  The team also compared the results of this analysis between Oregon and Alaska, where the team had conducted a similar analysis previously. Relative to the birth cohort linkages, incidence proportions based on the PRAMS linkage in Oregon underestimated incidence to a greater degree than those in Alaska. Although the research team deemed the effort to replicate the PRAMS linkage in Oregon to be successful overall, the team noted that jurisdictions considering additional replications should also consider conducting comparisons to full birth cohort linkages to better understand potential differences in estimates.

- **The CDN/CCWIP team used new analytic strategies that offered information not only about estimated incidence at a point in time but also about changes in incidence rates and types over different periods in children's lives.** As described above, the team investigated the use of two new analytic approaches—a cross-sectional approach and a longitudinal approach—to explore incidence of child maltreatment in California.

b. Predicting risk of child maltreatment

Similar to these two sites' exploration of new methods to estimate incidence, the CDN/Rady site validated a relatively new model for predicting risk of future child welfare involvement.

- **Analyses of linked hospital and child welfare data indicated that there was alignment between risk scores for a child generated by a predictive risk model and ICD codes in the child's medical record that relate to child maltreatment.** The risk of being placed in foster care was higher, on average, among children with an ICD code that indicated or suggested child maltreatment compared to children without such codes. These findings underscore the value of medical records as a resource for validating predictive risk models in child welfare and exploring maltreatment incidence and outcomes more generally.

B. Lessons from the CMI Data Linkages sites and considerations for future applications of linked administrative data

1. Key lessons on using administrative data linkages to better understand child maltreatment

The experiences and findings of the CMI Data Linkages sites offer important lessons about the process of using administrative data linkages to study the incidence of child maltreatment and related risk and protective factors and the potential for these approaches to inform understanding of child maltreatment. Key lessons that emerged from the study include the overall feasibility of enhancing administrative data linkages, the use of different types of data and analysis methods, the potential of sites' approaches to produce new knowledge, the challenges and limitations of these approaches, and the relative advantages and disadvantages of using linked administrative data to estimate child maltreatment incidence compared to other data sources.

a. Lessons on the feasibility of enhancements to administrative data linkages

- **The experiences of the CMI Data Linkages sites offer evidence that enhancing administrative data linkages—through acquisition of new data sources, use of new methods, or replication or scaling of existing methods—is a feasible approach to addressing high-priority questions about child maltreatment incidence and related risk and protective factors.** Each site was able to accomplish its intended enhancement and yield novel information from it. The work of the ADHSS/OHSU project, specifically, supports the idea that replication of linkage approaches across jurisdictions is feasible under the right conditions. This is an important finding for researchers who may be interested in replicating the approaches taken by these five sites. Key factors in the sites' ability to move their projects forward were support for administrative data linkage among leaders in the agencies that owned the data as well as the quality of the sites' partnerships and their technical expertise and infrastructure.

- **Sites benefitted from existing infrastructure and relationships, which took time and effort to establish and maintain prior to their involvement in the CMI Data Linkages project.** The sites drew on existing relationships with data providers, existing technical expertise, and existing infrastructure to accomplish their projects. The sites nurtured relationships with data stewards through regular meetings and consideration of the child welfare agency's priorities when conducting research—for example, considering and communicating how the research could help the data providers as well as the site. PIs and co-PIs were seasoned researchers with expertise in administrative data linkage and analysis. Nearly all sites that needed agreements to access child welfare administrative data already had them in place. Respondents noted that data linkage projects require substantial effort and resources, particularly if researchers do not have existing infrastructure and experience from which to draw to conduct linkage projects. (See Chapter III for promising practices for linking administrative data and Chapter IV for a summary of contextual and organizational factors that facilitate data linkage projects.)
- **A distinctive component of the CMI Data Linkages project, cross-site collaboration, also supported enhancements by providing researchers a space to brainstorm and compare experiences and approaches.** Representatives from all sites indicated that they found the cross-site learning network meetings, webinars, and TA to be helpful. Sites used webinars to get colleagues' and other experts' opinions on their findings and strategies for interpreting results. Site representatives might have benefitted from additional opportunities for open-ended discussion and brainstorming.

b. Lessons on data sources and linking methods

- **Hospital data may provide useful information about maltreatment injury or other health-related outcomes, though they can be challenging to obtain and interpret.** Two sites (CSSAT and CDN/Rady) incorporated hospital data as an additional source of maltreatment outcomes or maltreatment injury. Representatives of both sites noted the importance of having a team member with knowledge of hospital data and ICD codes, especially ICD codes pertaining to child welfare. They mentioned that, for example, certain ICD codes are suggestive of child maltreatment at some ages but not others. Analyses involve more than mapping codes directly to types of maltreatment. In addition, collaborations with hospitals may pose distinctive challenges in terms of data sharing. Researchers may need to spend more time communicating the opportunity that data sharing offers to the hospital, especially for hospitals that are less focused on research. It may be especially important to communicate this perspective to high-level hospital administrators. Identifying individuals with authority to approve data sharing, securing the approvals, and acquiring data extracts may all take time.
- **The sites' projects illustrate how linkages of varying levels of complexity—regarding the level of linkages and number of data sources—can yield new information for the field.** Linkages need not involve individual-level data from numerous sources to yield useful insights. For example, the CDN/CCWIP project relied on linkages of just two types of data: vital records and child welfare data. Although the cleaning, processing, and linking of these data involved complex methods, the project relied on a small number of data sources. Similarly, the UA-SSW project used relatively straightforward geographic-level linkages, rather than individual-level linkages, and publicly available data.

c. Lessons on the potential of sites' approaches to linking and analysis to produce new knowledge

- **Some project approaches illustrated how administrative data linkages may support efforts to improve estimates of incidence to include children who experience maltreatment but are not represented in data reported by child welfare agencies.** Several sites used longitudinal strategies to refine estimates of maltreatment incidence. The approach used by the CDN/CCWIP site—adjusting annual estimates of maltreatment incidence through analyses of linked child welfare and birth records—produced estimates that may address underreporting on an annual basis by including children for whom maltreatment was not substantiated in a single year but who were likely to have experienced it. The ADHSS/OHSU and CDN/CCWIP sites leveraged linked child welfare and birth records to develop estimates of cumulative incidence. These approaches expand our understanding of maltreatment incidence by providing insights into the proportions of children who experience maltreatment at some point during their childhoods.
- **Sites' use of linked hospital data suggests that this approach could contribute to improving maltreatment surveillance.** The CDN/Rady site's analysis indicated substantial overlap between children with diagnoses suggesting maltreatment and children reported to the child welfare system. In general, this approach offers opportunities to understand the extent and types of maltreatment identified by hospital staff and whether children with identified maltreatment come to the attention of the child welfare system (Rebbe et al. 2020). As part of a study conducted separately from the CMI Data Linkages project, the team from the CSSAT site used linked hospital data to develop estimates of maltreatment incidence in Washington (Rebbe et al. 2021).
- **Linkages involving data with geographic identifiers illustrate the potential for using administrative data to assess patterns of incidence and related factors at the county level—and to explore whether and how policy, practice, and demographic variation at that level might affect observed child maltreatment outcomes.** Although the researchers could not identify factors conclusively with the available data, the variation they observed pointed some teams toward hypotheses that could be investigated further. For example, as described above, the CSSAT team offered hypotheses to explain findings about the association between opioid use and child maltreatment outcomes.

d. Lessons on the challenges and limitations of sites' approaches

- **Although sites' existing relationships, expertise, and infrastructure proved helpful in many circumstances, even existing relationships with data stewards did not guarantee smooth processes for sharing additional or new data.** For example, in the CDN/Rady site, CDN had begun developing a relationship with Rady Children's Hospital, both partners were interested in collaboration, and the CMI Data Linkages project offered the support necessary to move ahead. But signing data sharing agreements, completing the data extraction, and receiving the data still took the site much longer than anticipated. Similarly, the CSSAT team's relationship with one contact at the state department of health did not translate into smoothly connecting with another contact within the agency, and the site was ultimately unable to acquire some data the team intended to link and analyze.
- **The sites needed to adapt to changes in circumstances and address unforeseen challenges that affected their project plans.** All sites had to adapt to changes in working conditions, priorities, and partner availability due to the COVID-19 pandemic. For example, proven linkage procedures at Children's Data Network became challenging to execute in the context of the COVID-19 pandemic, because they required a staff member to be physically present at the research center. At least one site

(CSSAT) had to address data that were unexpectedly corrupted, and the project team needed to work with the data steward to repair these issues.

- **While the sites' experiences underscored the value and potential of linked administrative data as a data source to better understand child maltreatment incidence, they also highlighted the assumptions and limitations inherent in these approaches.**
  - First, reports and substantiations of child maltreatment in child welfare administrative data may reflect culturally and state-defined definitions of maltreatment. The sites' approaches did not directly address these potential differences across individuals and communities. Similarly, hospital data may be affected by variation in the way that medical practitioners use ICD codes and the presence of physicians with child abuse pediatric certification or child abuse pediatric teams.
  - Second, the approaches to estimate incidence are necessarily imprecise because they rely on assumptions that may or may not be correct. For example, the approach of adjusting annual rates using longitudinal data relies on assumptions about whether later substantiations represent actual maltreatment in the base year.
  - Third, successful implementation of enhancements to administrative data linkages depends upon successful collaboration with data stewards. Despite sites' deep expertise and existing relationships, they were not always able to access the data they needed to answer their planned research questions.
  - Finally, to the extent that analyses focus on a single jurisdiction, such as a state, the findings cannot necessarily be generalized to other jurisdictions. Replications of approaches across jurisdictions may help shed light on their usefulness across contexts.

e. *Lessons on the advantages and disadvantages of using linked administrative data to estimate maltreatment incidence, relative to ACF's other mechanisms including the NCANDS and the NIS*

- **The NCANDS provides data that are not currently available from other sources—specifically, standardized, national-level data on maltreatment.** While annual estimates of maltreatment incidence based on NCANDS data might be subject to underreporting, this issue might be alleviated by linking NCANDS data longitudinally, assuming children who experience maltreatment are likely to be reported over time. However, without individual identifiers, the NCANDS is not as suitable as state-level administrative data for exploring child maltreatment trajectories or risk and protective factors through individual-level linkages with other data sources, including vital records. In addition, the NCANDS provides a less complete picture of child welfare system outcomes than state administrative data because its records do not include detailed information on unsubstantiated cases and those that are screened out, or other information that may be available with state-specific administrative data. In addition, NCANDS includes information only for children who are reported to child welfare agencies. Linking state-level administrative data with other datasets may help to provide a more complete picture of maltreatment by identifying unreported case of maltreatment. Finally, NCANDS does not address cross-state differences in child welfare policies and definitions and lacks a consistent poverty indicator which could enhance interpretability of national estimates. .
- **Using administrative data linkages to estimate maltreatment incidence offers advantages relative to the NIS.** Analyses based on administrative data linkages can be implemented at a lower cost and with greater frequency than the NIS and they can potentially provide greater insight into changes in incidence over time, sampling and non-response bias is less of an issue for population

data, and administrative data linkages may be able to capture reporting from non-mandated sources and unreported maltreatment. Moreover, the NIS sampling strategy is unlikely to fully address variation across jurisdictions in policy and child welfare system response. Expanding the use of administrative data linkages to estimate maltreatment incidence in a larger number of jurisdictions may be an efficient way to address this issue.

Overall, the experiences of the CMI Data Linkages sites endorse the continued development and validation of approaches for using linked administrative data to better understand child maltreatment incidence and the related risk and protective factors. Ultimately, the information produced through these approaches might not only support stakeholders in estimating the extent of child maltreatment but also inform efforts to prevent maltreatment through appropriately targeted supports for communities, families, and children.

REFERENCES

Bureau of Justice Statistics. “National Crime Victimization Survey (NCVS).” Washington, DC: U.S. Department of Justice, 2019. Available at https://bjs.ojp.gov/data-collection/ncvs#surveys-0. Accessed September 27, 2021.

Bywaters, P., G. Brady, T. Sparks, and E. Bos. “Child Welfare Inequalities: New Evidence, Further Questions.” *Child & Family Social Work*, vol. 21, no. 3, 2016, pp. 369–380. doi: org/10.1111/cfs.12154

Connelly, R., C. J. Playford, V. Gayle, and C. Dibben. “The Role of Administrative Data in the Big Data Revolution in Social Science Research.” *Social Science Research*, vol. 59, 2016, pp. 1–12.

Currie, J., and C.S. Widom. “Long-Term Consequences of Child Abuse and Neglect on Adult Economic Well-Being.” *Child Maltreatment,* vol. 15, no. 2, May 2010, pp. 111–120.

Danese, A., T.E. Moffitt, H. Harrington, B.J. Milne, G. Polanczyk, C.M. Pariante, R. Poulton, and A. Caspi. “Adverse Childhood Experiences and Adult Risk Factors for Age-Related Disease: Depression, Inflammation, and Clustering of Metabolic Risk Markers.” *Archives of Pediatric Adolescent Medicine,* vol. 163, no. 12, December 2009, pp. 1135–1143.

Drake, B., M. Jonson-Reid, and L. Sapokaite. “Re-Reporting of Child Maltreatment: Does Participation in Other Public Sector Services Moderate the Likelihood of a Second Maltreatment Report?” *Child Abuse & Neglect,* vol. 30, 2006, pp. 1201–1226.

Dumas, A., G. Elzinga-Marshall, B. Monahan, M. van Buren, and M. Will. “Child Welfare Screening in Wisconsin: An Analysis of Families Screened Out of Child Protective Services and Subsequently Screened in.” University of Wisconsin-Madison: Robert M. La Follette School of Public Affairs, 2015.

Finkelhor, D., H. Turner, A. Shattuck, S. Hamby, and K. Kracke. “Children’s Exposure to Violence, Crime, and Abuse: An Update.” Washington, DC: US Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention, 2015.

Gale, N. K., G. Heath, E. Cameron, S. Rashid, and S. Redwood. “Using the Framework Method for the Analysis of Qualitative Data in Multi-Disciplinary Health Research.” *BMC Medical Research Methodology,* vol. 13, 2013, no. 117. https://doi.org/10.1186/1471-2288-13-117.

Goerge, R.M., C. Smithgall, R. Seshadri, and P. Ballard. “Illinois Families and Their Use of Multiple Service Systems.” Chicago: Chapin Hall at the University of Chicago, 2010.

Harron, K., C. Dibben, J. Boyd, A. Hjern, M. Azimee, M. Barreto, and H. Goldstein. “Challenges in Administrative Data Linkage for Research.” *Big Data & Society,* December 2017. Doi:10.1177/2053951717745678.

Institute of Medicine and National Research Council. “New Directions in Child Abuse and Neglect Research.” Washington, DC: The National Academies Press, 2014. Doi.org/10.17226/18331.

Jonson-Reid, M., P. Kohl, and B. Drake. “Child and Adult Outcomes of Chronic Child Maltreatment.” *Pediatrics,* vol. 129, no. 5, 2012, pp. 839–845.

Jonson-Reid, M., B. Drake, E. Weigensberg, M. Stagner, S. Collier, S. LeBarron, and L. Read. “Improving Understanding of Child Maltreatment Incidence: A Review of Data, Measures, and Methods.” Unpublished report submitted to the U.S. Department of Health and Human Services Administration for Children and Families. Chicago: Mathematica Policy Research, 2016.

Kim, H., C. Wildeman, M. Jonson-Reid, and B. Drake. “Lifetime Prevalence of Investigating Child Maltreatment Among U.S. Children.” *American Journal of Public Health,* vol. 107, no. 2, 2017, pp. 274–281. Doi:10.2105/AJPH.2016.303545.

Office of Juvenile Justice and Delinquency Prevention. “National Survey of Children’s Exposure to Violence.” Washington, DC: U.S. Department of Justice, 2020. Available at https://ojjdp.ojp.gov/research-and-statistics/research-projects/national-survey-childrens-exposure-violence/overview#:~:text=The%20National%20Survey%20of%20Children%27s%20Exposure%20to%20Violence,the%20Centers%20for%20Disease%20Control%20and%20Prevention%20%28CDC%29. Accessed September 27, 2021.

Office of Management and Budget. “Barriers to Using Administrative Data for Evidence-Building.” White paper. Available at https://obamawhitehouse.archives.gov/sites/default/files/omb/mgmt-gpra/barriers_to_using_administrative_data_for_evidence_building.pdf. Accessed June 14, 2021.

Penner, A. M., and K. A. Dodge. “Using Administrative Data for Social Science and Policy.” *The Russell Sage Foundation Journal of the Social Sciences: RSF,* vol. 5, no. 3, 2019, pp. 1–18. Doi:10.7758/RSF.2019.5.3.01.

Peterson, C., C. Florence, and J. Klevens. “The Economic Burden of Child Maltreatment in the United States, 2015.” *Child Abuse & Neglect*, vol. 86, 2018, pp. 178–183.

Prindle J., R. Foust, and E. Putnam-Hornstein. “Maltreatment Type Classifications and Transitions During Childhood for a California Birth Cohort.” *Child Maltreatment*, May 2021. Doi:10.1177/10775595211006784

Putnam-Hornstein, E., E. Ahn, J. Prindle, J. Magruder, D. Webster, and C. Wildeman. “Cumulative Rates of Child Protection Involvement and Terminations of Parental Rights in a California Birth Cohort, 1999–2017.” *American Journal of Public Health,* vol, 111, 2021, pp. 1157–1163. doi.org/10.2105/AJPH.2021.306214.

Putnam-Hornstein, E., and B. Needell. “Predictors of Child Welfare Contact Between Birth and Age Five: An Examination of California’s 2002 Birth Cohort.” *Children and Youth Services Review*, vol. 33, no. 11, 2011, pp. 2400–2407.

Rebbe, R. A Population-Based Study of Child Maltreatment-Related Hospitalizations and Child Protection Responses. Doctoral dissertation, University of Washington, 2019. http://hdl.handle.net/1773/44448

Rebbe, R., A.S. Bishop, J. Ahn, and J.A. Mienko. “Opioid Overdose Events and Child Maltreatment Indicators: Differential County-Level Associations.” *Children and Youth Services Review*, vol. 119, 2020, pp. 1-7. https://doi.org/10.1016/j.childyouth.2020.105671

Rebbe, R., M.L. Martinson, and J.A. Mienko. “The Incidence of Child Maltreatment Resulting in Hospitalizations for Children Under Age 3 Years.” *The Journal of Pediatrics*, vol. 228, 2021, pp. 228–234.

Rebbe, R., J.A. Mienko, and M.L. Martinson. “Reports and Removals of Child Maltreatment-Related Hospitalizations: A Population-Based Study.” *Child Maltreatment,* 2020. Doi:1077559520984549.

Ritchie, J., J. Lewis, C. McNaughton Nicholls, and R. Ormston (eds.). *Qualitative Research Practice: A Guide for Social Science Students and Researchers.* Thousand Oaks, CA: Sage, 2013.

Sedlak, A. J., J. Mettenburg, M. Basena, I. Peta, K. McPherson, and A. Greene. “Fourth National Incidence Study of Child Abuse and Neglect (NIS-4).” Washington, DC: U.S. Department of Health and Human Services (HHS), 2010.

Slack, K. S., S. Font, K. Maguire-Jack, and L.M. Berger. "Predicting Child Protective Services (CPS) Involvement Among Low-Income US Families with Young Children Receiving Nutritional Assistance." *International Journal of Environmental Research and Public Health*, vol. 14, no. 10, 2017, pp. 1197.

Smith, B.D., Q. Lee, K. Wang, and A. Smith. "A National Study of Child Maltreatment Reporting at the County Level: Interactions Among Race/Ethnicity, Rurality and Poverty." *Children and Youth Services Review,* vol. 122, March 2021.

Sulimani-Aidan, Y., and R. Benbenishty. "Child Maltreatment Reports in Israel: The Intersection Between Community Socioeconomic Characteristics and Ethnicity." *American Journal of Orthopsychiatry*, vol. 83, no. 1, 2013, pp. 29–36. doi:10.1111/ajop.12009

U.S. Congress. "Chapter 67—Child Abuse Prevention and Treatment and Adoption Reform." Pub. Law 93-246, Stat. 5105. 2017. Available at https://www.govinfo.gov/content/pkg/USCODE-2017-title42/html/USCODE-2017-title42-chap67.htm.

U.S. Department of Health & Human Services, Administration for Children and Families, Administration on Children, Youth and Families, Children's Bureau. "Child Maltreatment 2019." Washington, DC: HHS, 2021. Available at https://www.acf.hhs.gov/cb/research-data-technology/statistics-research/child-maltreatment.

Weigensberg, E., D. Cornwell, L. Leininger, M. Stagner, S. LeBarron, J. Gellar, S. MacIntyre, R. Chapman, E.J. Maher, P.J. Pecora, and K. O'Brien. "Superutilization of Child Welfare, Medicaid, and Other Services." Final report submitted to Casey Family Programs. Chicago: Mathematica Policy Research, March 2018.

Widom, C.S. "Child Abuse, Neglect, and Adult Behavior: Research Design and Findings on Criminality, Violence, and Child Abuse." *American Journal of Orthopsychiatry*, vol. 59, no. 3, July 1989, pp. 355–367.

Wildeman C., N. Emanuel, J.M. Leventhal, E. Putnam-Hornstein, J. Waldfogel, and H. Lee. "The Prevalence of Confirmed Maltreatment Among U.S. Children, 2004 to 2011." *JAMA Pediatrics,* vol. 168, no. 8, 2014, pp. 706–713. Doi:10.1001/jamapediatrics.2014.410.

APPENDIX A:

ADHSS/OHSU FINAL REPORT

Introduction

This appendix is an overview of the ADHSS/OHSU project, conducted as part of the CMI Data Linkages work. It was written by the site team, with the Mathematica team working with the site to ensure consistency in information, level of detail, and presentation across sites.

Overview

The Alaska Longitudinal Child Abuse and Neglect Linkage project (ALCANLink) was developed to examine over time the incidence of maltreatment, predictive and etiologic factors, and disparities related to maltreatment. ALCANLink is a population-based mixed-design strategy that integrates those births that were sampled and mothers who subsequently responded to the Pregnancy Risk Assessment Monitoring System (PRAMS) survey with child welfare and other administrative data. ALCANLink partnered with the Oregon Health Authority and Oregon Health Sciences University to replicate the ALCANLink methods. This project estimated the cumulative incidence to first report, screen-in, substantiation, and removals in Oregon by age 9 and compared the cumulative risk distributions over time with those observed in Alaska.

Partnership history

The ALCANLink replication project was conducted through a joint partnership between the Alaska Division of Public Health (AKDPH), Oregon Health Authority (OHA), and Oregon Health Sciences University (OHSU).

The principal investigator and primary contact for this project was Dr. Jared Parrish. Dr. Parrish developed the ALCANLink methodology and implemented it in Alaska. Dr. Parrish oversaw this project to ensure model adherence, timely completion, and statistical consistency. Dr. Parrish first initiated the partnership with Oregon in early 2018 by forming connections with the Oregon Health Authority and state child welfare director. Although other states expressed interest in replicating ALCANLink, the partnership between Dr. Parrish and Oregon was ultimately selected. This was due in part to the fact that Dr. Parrish's former colleague, Abigail Newby-Kew, left AKDPH to pursue her doctorate in epidemiology at OHSU-PSU.

The second major partnership, established during preliminary work conducted before the CMI Data Linkages work, was with the OHSU-PSU School of Public Health (SPH). Abigail Newby-Kew, a doctoral student in epidemiology at OHSU-PSU SPH and a former co-worker of Dr. Parrish, was hired as an independent contractor for this project. This partnership between AK and SPH was formalized through that contract. In this capacity, she was tasked with coordinating activities, facilitating data sharing and partnerships between OHA and OHSU, and conducting analyses. Ms. Newby-Kew served as the primary Oregon contact and worked closely with Dr. Parrish. Dr. Lynn Marshall, Abigail Newby-Kew's academic advisor was also part of the OHSU- PSU partnership. She advised on project analysis and reporting and has served as the Principal Investigator for the project IRB application.

The third partner in this project was the Oregon Health Authority (OHA). This partnership was also established during preliminary work conducted before the CMI Data Linkages work, though it has expanded in scope over the course of the grant. Although the partnership was not formalized as such, two of the data sources used in the project were housed within the OHA, and data use agreements between OHSU-PSU and OHA are in place. The project also went through review and received official approval from the OHA Science and Epidemiology Council. Beyond OHA's role as a primary data steward for PRAMS and Vital Records, project partners at OHA contributed to project design and reporting and facilitated communication with Child Welfare and Integrated Client Services. Suzanne Zane, the senior MCH epidemiologist with OHA, and John Putz, the principal executive manager of the MCH Surveillance Unit, were actively engaged with these aspects of the project.

Background

Although annual estimates indicate that about 10 percent of children whose ages range from newborn to 7 are reported to child welfare for maltreatment, cumulative incidence estimates that account for out-of-state emigration and competing cause mortality indicate that 32 percent of children born in Alaska have such reports before age 8—three times that of the annual estimate (Parrish et al 2020; Parrish et al 2017; Kim et al 2017). There has been internal validation of these estimates in Alaska, but external validation is needed. By investigating the proposed research questions and clearly documenting the process, external validation will establish the feasibility of these methods for a multi-state, regional, or even national model. Multi-state data based on administrative data linkages that are standardized between states will enable estimates of the cumulative incidence of maltreatment and comprehensive research projects that would deepen understanding of the factors associated with and predicting maltreatment across jurisdictional and structural boundaries.

Research Questions

1. Testing external validity: Does the cumulative incidence of the time to first maltreatment report in Oregon differ from that observed in Alaska?
    a. Do demographic population frequency distributions confound the between-state comparisons?
    b. Accounting for population loss when estimating the cumulative incidence over time in Oregon, can the Alaskan cohort data inform and improve estimates in the absence of population censoring information (i.e., administrative censorship: data including whether a child left the state and is thus no longer part of the analytic population)?
2. Testing internal validity: Is the cumulative incidence to first report, contact, and substantiation estimated through the Oregon 2009 PRAMS linkage consistent with a full Oregon 2009 birth cohort linkage to child welfare?
    a. What are the key components required for successful replication of ALCANLink methods?
    b. What partners are required?
    c. What minimal resources are necessary?
    d. What technology/skill sets are required?
    e. What challenges impact fidelity?

Sub-question 1.b was not addressed due to timing constraints.

Data

Sources

The data sources used for this project are in Table A.1.

Table A.1. Data sources

Data source	Description of data
Oregon Health Authority, Oregon Public Health Division, Section of Maternal and Child Health	2009–2011 Oregon PRAMS data set and corresponding two-year follow-up (PRAMS-2); 4,867 surveys; includes text and numeric variables for PRAMS Phase 6 Core survey questions and demographic information listed on the birth record
Oregon Health Authority, Center for Health Statistics	All Oregon resident birth and death records for 2009, and records for PRAMS respondents during 2009–2011; 47,188 records; Includes text and numeric variables.
Oregon Department of Human Services, Children, Adults and Family Division	All child welfare administrative records of allegations, including type of allegation and investigative findings for the years 2009–2018 for children born in 2009 and for children born in 2009–2011 whose mothers responded to the PRAMS survey
ALCANLink	The ALCANLink data has linked the 2009–2011 PRAMS cohort to vital records and child welfare records (allegations, type of allegation, and investigative findings) ; 3,549 survey respondents representing over 200 data elements.

Source: Project documents.

Linking process

As part of AKDPH agreements with the State of Oregon data stewards, data linkages were completed by Oregon Integrated Client Services (ICS). ICS used separate but similar processes to the ones used in Alaska (ALCAN) to link data; the latter are described in Parrish et al. (2017). The use of the exact same algorithm to link records in different states would be unwise because the algorithm should reflect the nuances of the location. ICS brought in data from multiple state programs and agencies on a monthly basis and made or maintained individual-level links using available identifiable data (that is, name, date of birth [DOB], Social Security number [SSN], race/ethnicity). These sources included both birth certificate records and child welfare records. For the Oregon Longitudinal Child Abuse and Neglect (OLCAN) request, ICS leveraged the individual-level link that already existed to create unique project IDs.

ICS used a combination of probabilistic, deterministic, and manual matching each month to make/maintain the individual-level links (Table A.2). Each month, each "class" of probabilistic matching components (one class might be Names-DOB, another might be Names-SSN, etc.) went through iterations in which the matching criteria gradually loosened. With some exceptions, most of the matching components were a mix of deterministic matching on some fields and probabilistic matching on others. Records went through multiple matching components, and the highest-scoring match was chosen at the end. Data cleaning and standardization also took place before matching, and there was a manual process of cleanup after. Much of the linking and processing of data were programmed in a software called RedPoint.

Table A.2. Data linkage methods

Data source	Variables used to link (and linking destination)	Linkage approach
PRAMS and PRAMS-2	Birth certificate number (to vital records)	Deterministic
Vital records (birth and death)	First, last, and middle names; DOB (to CPS)	Deterministic followed by probabilistic
Child protective services	First, last, and middle names; DOB (to vital records)	Deterministic followed by probabilistic

Source: Project documents.

CPS = child protective services; DOB = date of birth.

The ALCANLink program maintains and updates (annually) a master name list for the PRAMS sampled child.

For the PRAMS cohort, we first processed annual data through an Extract Transformation and Load (ETL) tool called Pentaho®. This tool systematically identifies and merges in new information for cases that have already been linked (using ID Keys and Foreign Keys).

The data were first deterministically linked based on the child’s first and last names, sex, and date of birth. Duplicates are assessed and manually resolved. Using the procedures above, second twins are identified in the base birth records and isolated and linked to ensure correct matches. Finally, probabilistic matching is performed on the remaining unmatched PRAMS records to the incoming files to specify those that need manual review. Using a weighted Jaro-Winkler edit distance scoring method, probable matches with a score of 0.89-0.99 are manually reviewed and classified. These data are compared against a standing name change repository and other resources in the state for verification.

All known variations of the name were included in a long file with a common ID and linked against incoming records. Names with the highest probability match for the common ID were included, with others removed.

Middle name and maternal and paternal information were used for manual review qualifiers. City of residence was also considered if needed. Manual review consisted of first assessing whether there were any duplicates of the name and date of birth in the birth record or incoming record sets. If true, measures to codify were taken to distinguish the record; if false, all possible administrative sources were reviewed, and once these were exhausted, if the record was not confirmed, it was rejected.

The R statistical environment was used with the RecordLinkage package. The review thresholds were established using a Petro Distribution and subsample single-layer neural network for confirmation of established thresholds. These thresholds were set to limit the amount of manual review conducted but still establish a reasonable probability of capturing all possible cases. Given that these methods attempt to replicate a longitudinal prospective cohort, losses to follow-up are mitigated if reasonably possible.

Analytic Methods

The researchers calculated the cumulative incidence using the same methods (using R) in both Alaska and Oregon.

Operationally, the researchers measured the incidence proportion of first allegations, investigations, and substantiated allegations using a Kaplan-Meier method for the full birth cohort analysis and an Aalen-based hazard method for the complex sample data. Using the survivorship function S(t), the incidence proportion was calculated as 1- S(t). Estimates use administrative data to censor observations due to death (competing cause mortality) and emigration before age 9. Due to the known and documented nonlinear incidence in the occurrence of maltreatment (that is, higher incidence in younger children) and population loss that can occur in birth populations as a consequence of death, out-of-state emigration, and missed linkages in the study design, simple proportion at the end of a the period or x/n will consistently underestimate the incidence, and these methods mitigate this limitation.

Analysis 1: Validation analysis using the 2009 birth cohort

Full birth cohort analysis. Two congruent analyses were conducted for both Alaska and Oregon among the full 2009 linked birth cohort and the PRAMS linked birth cohort. In both Alaska and Oregon, the 2009 in-state resident births were linked with child welfare and death records through 2018. We calculated age in decimal years to first event, death, or administrative censorship. Events that we considered were first report, first screen-in, and first substantiation, resulting in three age calculations for each child. We also calculated corresponding indicator variables (0 = No event, 1 = Event) for each event type per individual. Working with the survival package in R, we used a Kaplan-Meier method to calculate survival. Based on the relationship described briefly above between survival S(t) and the Incidence Proportion F(t), we were able to derive the cumulative incidence as 1-S(t). Consistent with the concept of "instantaneous risk," F(t) estimates the risk of an event at any time (t), where time (t) is calculated at an event of interest. Thus F(t) is equivalent to the summation of calculated risk over n intervals of length k. Formula A.1:

$$F_{(t)} = \int_0^t f_{(t)} dt = f_{(t)} \frac{dF(t)}{dt} = 1 - e^{-H_t} = 1 - e^{-\sum_{K=1}^{K} \lambda_k t_k} \approx \frac{I * \overline{D}}{(I * \overline{D} + 1)} = R$$

PRAMS birth cohort. In both Alaska and Oregon, the 2009 records for the child in the PRAMS sample at birth child were linked with child welfare and death records through 2018. We used the same methodology described above for the full birth cohort but with the survey package in R. To calculate standard errors, the Aalen (hazard-based) estimator was used.

We plotted 1-survival curve to visualize the cumulative incidence for both the full and PRAMS cohorts by each state. Using the PRAMS cohort, we completed this for all events considered and by subset demographics to compare the consistency in the estimation. We also tabulated F(t) within age intervals and compared these tabulated data between the PRAMS estimates and observed full birth cohort estimates.

Analysis 2: Analysis of 2009–2011 PRAMS cohort

Using the Oregon PRAMS linked cohort from 2009–2011, we calculated the cumulative incidence to first report, screen-in, and substantiation using the same methods described above. Using a three-year cohort increased our sample size, thereby reducing our standard errors. We expanded the subgroup estimates using multiple indicators available on PRAMS, but focused on maternal stressors, mental health, and substance use reported during the pre-birth period. We completed comparative analyses by group using cox-proportional hazard models.

Because the Alaska cohort can account for out-of-state emigration using a unique administrative data set, we originally intended to investigate the probability of censorship due to emigration in the Alaska cohort by demographics and investigate the utility and need for adjusting the Oregon cohort estimate with methods such as inverse probability weighting (IPW). This analysis was delayed because it took so long to obtain these data and because of the impact of COVID-19 on available staff time. These analyses will still be conducted at a later date.

Findings

Question 1, Testing external validity: Does the cumulative incidence of the time to first report in Oregon differ from that observed in Alaska?

The purpose of this descriptive research question was to fully describe the cumulative incidence of child welfare reports, investigations, and substantiations in the two populations of interest. Based on differential populations, reporting laws, and CPS investigation/substantiation policies, we anticipated differences between the two states. Describing the similarities and differences will facilitate future pooled analyses by guiding hypothesis development and identifying potential confounding factors.

We compared the incidence proportion (IP) in Alaska and Oregon before age 9 for both the 2009 resident birth cohorts and 2009 PRAMS cohorts. We considered reports, investigations, and substantiations (Table A.3). Age 9 was chosen based on data availability, as our Oregon linkage contained data through 2018. We found that before age 9, the cumulative incidence of both states' measures of child welfare involvement were similar in the full birth cohort. However, the Oregon PRAMS cohort underestimates involvement with child welfare to a greater degree than the Alaska PRAMS cohort for each measure.

Table A.3. Incidence proportion of child maltreatment allegations before age 9 for reports, investigations, and substantiations in Oregon and Alaska, 2009 full birth and PRAMS cohorts

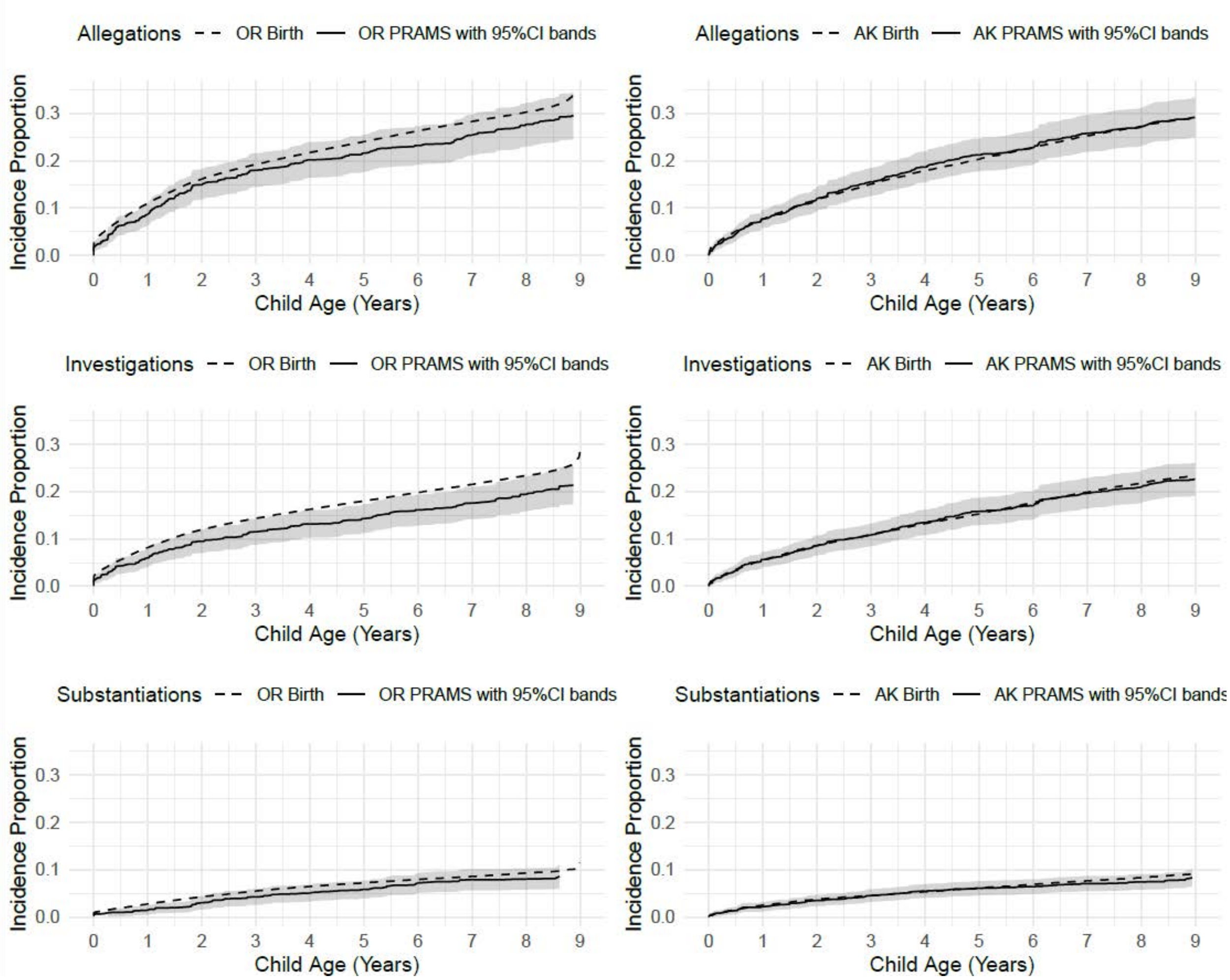
	Oregon 2009				Alaska 2009			
	Birth cohort N = 46,338		PRAMS cohort N = 1,652		Birth cohort N = 11,187		PRAMS cohort N = 1,235	
	N (%)	IP	n (%w)	IP (95%CI)	N (%)	IP	n (%w)	IP (95%CI)
Reports	15,135 (32.7)	32.0	585 (29.8)	28.7 (24.0, 33.4)	3,247 (29.0)	29.1	386 (29.0)	29.1 (25.0, 33.3)
Investigations	11,836 (25.5)	25.0	457 (21.8)	20.9 (17.1, 24.7)	2,613 (23.4)	23.5	302 (22.4)	22.6 (19.1, 26.1)
Substantiations	4,654 (10.2)	9.9	174 (8.6)	8.3 (6.0, 10.5)	1,014 (9.1)	9.1	126 (8.2)	8.3 (6.5, 10.1)

Source: Project documents.

Note: IP = incidence proportion of child maltreatment allegations; PRAMS = Pregnancy Risk Assessment Monitoring System.

Compared with children in Alaska, a higher proportion of Oregon children were involved with child welfare throughout the study period (Figure A.1). This may be particularly meaningful in the first year of life, as 11.0 percent of Oregon children are reported to child welfare before their first birthday, although only 7.6 percent of Alaska children are.

Figure A.1. Incidence proportion child maltreatment allegations, investigations, and substantiations among Oregon and Alaska 2009 Full Birth and PRAMS cohorts



Question 1.a. Do demographic population frequency distributions confound the between-state comparisons?

Although no direct between-state comparisons were attempted due to timing constraints, there was a descriptive analysis that compared the incidence of child maltreatment between states for an array of variables available on the birth certificate (Table A.4). Qualitatively, the relative incidence of maltreatment was similar in Oregon and Alaska.

Table A.4. Incidence proportion of child maltreatment allegations before age 9 for reports, investigations, and substantiations in Oregon and Alaska, 2009 full birth and PRAMS cohorts, by demographic factors

	Oregon 2009				Alaska 2009			
	Birth cohort		PRAMS cohort		Birth cohort		PRAMS cohort	
	N	IP	n	IP (95%CI)	N	IP	n	IP (95%CI)
Sex								
Female	23,201	31.6	822	29.1 (22.3, 35.9)	5,289	29.5	608	32.1 (25.5, 38.7)
Male	24,483	32.4	830	28.2 (21.8, 34.6)	5,897	28.8	626	26.2 (20.9, 31.5)
Race								
American Indian/Alaska Native	1,381	58.6	347	55.3 (43.7, 66.9)	2,949	53.5	437	52.0 (42.1, 61.9)
Black	1,303	58.8	244	59.1 (44.9, 73.2)	460	32.7	44	63.4 (8.3, 1.0)
Asian	2,157	13.3	235	12.3 (7.7, 16.9)	573	24.3	34	12.4 (0.0, 27.9)
NHOPI	835	30.3	91	28.5 (15.3, 41.6)	363	20.2	39	23.2 (1.8, 44.6)
White	40,742	32.0	909	28.6 (23.2, 33.9)	6,710	19.2	614	18.7 (14.9, 23.3)
Ethnicity								
Hispanic	9,690	32.6	386	29.1 (23.4, 34.8)	1,041	25.2	115	33.3 (14.7, 5.2)
Not Hispanic	37,995	31.9	1,266	26.9 (21.0, 32.8)	9,499	29.9	1,004	29.0 (24.6, 33.3)
Marital status								
Married	30,882	19.9	671	16.7 (12.7, 20.7)	6,911	17.1	720	17.7 (13.7, 21.7)
Unmarried	16,793	54.2	981	49.3 (37.3, 61.2)	4,263	48.7	511	48.1 (37.8, 58.4)
Maternal education								
< 12 years	9,324	49.6	368	42.2 (29.7, 54.7)	1,469	55.5	177	56.6 (36.9, 76.9)
≥12 years	38,131	27.7	1,277	25.6 (20.6, 30.5)	9,284	24.8	973	25.2 (21.0, 29.4)
Maternal age								
< 20 years	4,131	60.2	168	49.0 (25.5, 72.5)	1,113	49.2	138	51.7 (30.9, 72.5)
≥ 20 years	43,551	29.3	1,484	26.8 (22.1, 31.5)	10,072	26.9	1,096	26.2 (22.2, 30.3)
Gestational age								
Preterm	3,787	38.4	121	0.26 (0.10, 0.42)	1,183	37.5	291	35.4 (22.7, 48.0)
Term	43,858	31.5	1,531	0.29 (0.24, 0.34)	9,418	28.5	881	29.1 (24.4, 33.7)
Previous life births								
0	19,434	29.8	689	24.3 (17.7, 30.8)	4,230	26.0	494	26.1 (19.7, 32.6)
≥ 1	28,163	33.5	961	31.9 (25.3, 38.4)	6,789	31.3	714	31.1 (25.5, 36.7)

Source: Project documents.

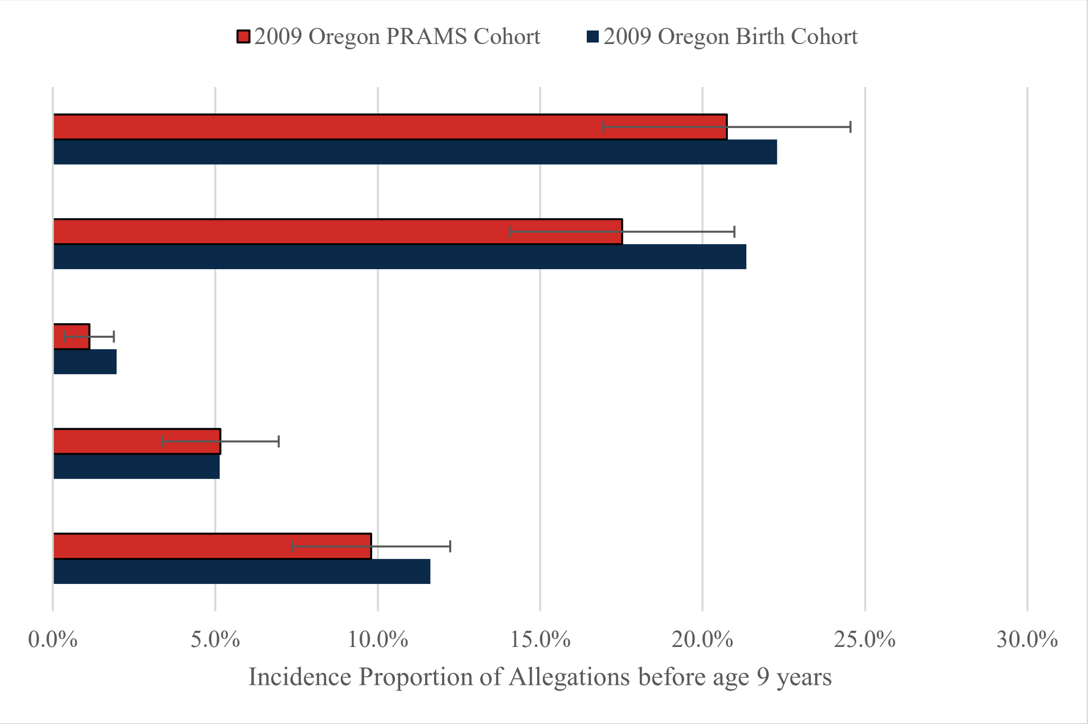
Note: NHOPI = Native Hawaiian or Other Pacific Islander; PRAMS = Pregnancy Risk Assessment Monitoring System. IP = incidence proportion of child maltreatment allegations.

Question 2, Internal Validation: Is the cumulative incidence to first report, contact, and substantiation estimated through the Oregon PRAMS linkage consistent with a full Oregon birth cohort linkage to child welfare?

This question is key to determining whether the ALCANLink method can be generalized to another region that has different demographics, different PRAMS sampling strata, and a different jurisdictional structure than Alaska. The Tables (A.3 and A.4) presented above for Question 1 illustrate the comparability in cumulative incidence at age 9 between the 2009 Oregon PRAMS and 2009 Oregon full birth cohorts. The estimated cumulative incidence of child maltreatment among the Oregon PRAMS cohort is consistently (but not significantly) an underestimate of that observed in the full birth cohort. (Full birth cohort estimate is captured within the 95 percent confidence intervals of the PRAMS estimates). We observed large confidence intervals in some demographic strata that were not oversampled by PRAMS. These include teen mothers and preterm infants. The consistent underestimate of cumulative incidence of reports is also observed for each unique type of maltreatment allegation with the exception of sexual abuse (Figure B.2). The difference is not significant for any maltreatment type with the exception of neglect.

We determined that ALCANLink can be successfully used in Oregon, but that any new jurisdiction may want to consider a comparative full birth cohort linkage to understand how PRAMS distributions may differ from those observed in the full population.

Figure A.2. Incidence proportion of child maltreatment allegations before age 9 for reports by maltreatment type: Oregon 2009 PRAMS and 2009 full birth cohorts



Question 2a: What are the key components required for successful replication of ALCANLink?

The ALCANLink method is a relatively simple linkage that requires few resources. However, we identified a number of requirements that are key for successful replication.

A strong partnership with at least one stakeholder within the health department is necessary for building relationships, gaining support from additional stakeholders, and ensuring support for navigating and understanding state systems. Although buy-in from all data stewards is necessary, they do not need to be primary members of the project team. In addition, it is important to have an advisor within the child welfare office because of state-by-state variations in the way allegations are categorized and screened. Due to the standardized nature of PRAMS and birth certificate data, a lower level of support for data interpretation is needed from these agencies. In addition to a strong team, the primary resource needed for this project is a researcher with the time and capacity to navigate the state system and identify and implement all necessary replication steps.

We found that the primary technology skill set required for this project replication is competency with R statistical analysis software. Although we approached the project with the idea that our team would also need to include an expert SAS programmer, our final replication protocol involved only R. In addition, it is necessary for someone with advanced knowledge of data linkage protocols to be involved with the project. Depending on state requirements, this individual may be a member of the team or a state agency. The ALCANLink data sets are not large, and advanced computing servers are not necessary to handle analysis.

The primary challenge that affected fidelity for this replication process was our inability to conduct data linkage ourselves. However, we worked closed with the Oregon data linkage agency to ensure that their protocol was similar to the one initially used for ALCANLink. However, without the right partners in place, replication can be challenging.

Next steps

Given the delays brought on by COVID-19, the project is behind the anticipated schedule for disseminating findings. We’ve presented initial findings to the CMI network and the Oregon Child Welfare leadership. We are also planning to present initial findings to the Oregon Public Health Research Group in February 2021. We intend to present these results at multiple conferences (for example, the Council of State and Territorial Epidemiologist and City Maternal and Child Health Epidemiology conferences). To date, presentations on the project and/or initial findings have been given at the Western Regional Epidemiology Network Annual Conference, OHUS-PSU Doctoral Seminar and other courses, and the OHSU School of Public Health Conference.

We are currently writing up our findings from the replication work and will be submitting to *Public Health Reports Journal* as a methods paper. The principal investigator is developing a short two-page replication brief that will be sent to CDC PRAMS, which is aware of ALCANLink but not the replication work.

The replication work findings that are validation-focused are less useful to policymakers but provide the foundation for future work. Initial analyses focusing on parental stressors such as IPV, economics, and mental health will be described in oral presentations to multiple Oregon agencies and partners, written reports, and peer-reviewed publications. As the development of this cohort will serve as the basis for a dissertation project, multiple manuscripts will be derived from this established cohort.

Finally, the replication findings will be presented through a webinar to Alaskan partners. This will focus on comparative and pooled analyses to describe differences, similarities, and exploration of population risks and the factors contributing to these risks.

Lessons Learned About Administrative Data Linkage Practices Related to Examining the Incidence and Risk of Child Maltreatment

Overall, ALCANLink was relatively easy to replicate. As with any data linkage project, it would have been helpful to know all the review processes required in Oregon and who ultimately had authority to approve each step. For future expansions Dr. Parrish plans to provide the principal investigator from the state a worksheet to complete to help them identify all the processes, reviews, protocols, and contacts before they begin. This would enable them to initiate all necessary processes in a timely manner and determine which steps could be contracted out or delegated to a graduate student.

We also learned the power of in-person meetings. The project principal investigator (Dr. Parrish) was able to travel to Oregon and meet with multiple partners. These in-person meetings help establish relationships and generate excitement. Based on how successful these were, we would have planned and scheduled more of them upfront. As we’ve learned with all the online meetings due to COVID-19, we often try to multitask, and can come to a meeting less prepared because of stacked meetings throughout the day. In future replication projects, in-person meetings would be preferred.

We also confirmed that although the PRAMS data were weighted to the birth population, some populations are under- or overrepresented, which can impact analyses. The weighted population tends to underestimate the observed cumulative incidence (though not to a statistically significant degree). If the likelihood of responding to PRAMS is associated with the outcome of interest, confounding due to selection bias may be introduced and should be considered when constructing etiologic/comparative assessments. It is critical that researchers conducting etiologic analyses with these data establish causal diagrams that consider the sampling design of PRAMS and potential impact of linkages.

Finally, we learned that with a limited amount of resources and consistent effort, these methods can be replicated in other jurisdictions resulting in data that can be used for comparative and inter-jurisdictional pooled analyses. This is critical, as it could enable investigations among underrepresented populations by expanding sample sizes, and also identify universal impacts. We are excited about these methods and the potential for expanded usage.

REFERENCES

Kim, H., C. Wildeman, M. Jonson-Reid, et al. “Lifetime Prevalence of Investigating Child Maltreatment Among U.S. Children” *American Journal of Public Health*, vol. 107, no. 2, 2017, p. 274 – 280.

Parrish, J.W., J.M. Fleckman, J.J. Prindle, et al. “Measuring the Incidence of Child Maltreatment Using Linked Data: A Two-State Comparison.” *American Journal of Preventive Medicine*, vol. 58, no. 4, 2020, e133 – e140.

Parrish, J. W., M.E. Shanahan, P. G. Schnitzer, P. Lanier, J. L. Daniels, and S. W. Marshall. “Quantifying Sources of Bias in Longitudinal Data Linkage Studies of Child Abuse and Neglect: Measuring Impact of Outcome Specification, Linkage Error, and Partial Cohort Follow-Up.” *Injury Epidemiology,* vol. 4, no. 1, 2017, p. 23.

APPENDIX B:

CDN/CCWIP FINAL REPORT

Introduction

This appendix is an overview of the CDN/CCWIP project, which was conducted as part of the CMI Data Linkages work. The site team authored the appendix, although the Mathematica team worked with the site to ensure consistency in information, level of detail, and presentation across sites.

Overview

The site team used California administrative data to understand the extent to which methods and methodological decisions affected estimates of the incidence of child maltreatment. The goal was to develop a methodology that could be generalized to other states and then used in conjunction with or in place of the National Incidence Study (NIS).

NIS estimates include estimates of the number of abused and neglected children who do and do not come to the attention of child protective services (CPS). The NIS is, however, expensive to conduct due to the complexities of a nationally representative sampling approach and the nature of primary data collection; consequently, its estimates are outdated. It has been shown to suffer from problems in terms of lacking the precision and statistical power needed to assess critical group differences in maltreatment (for example, by race or ethnicity). (Drake and Jonson-Reid 2011)

Improving the collection, management, accessibility, and integration of administrative records positions these data as an increasingly important source of information for research, evaluation, and policy analysis (Putnam-Hornstein, Needell, & Rhodes 2013). In the context of a population-level understanding of maltreatment exposure, linked administrative data have the potential to be prospectively leveraged as an alternative or complement to the NIS, generating more cost-effective, timely, local, and (potentially) more accurate estimates of the victims of child abuse and neglect.

Partnership history

The site team included staff affiliated with the Children’s Data Network (CDN) and California Child Welfare Indicators Project (CCWIP). See Table B.1. Data were already in hand based on the long-standing relationships between CDN, CCWIP, and California state agencies.

CDN is a data and research collaborative focused on the linkage and analysis of administrative records. In partnership with public agencies, philanthropic funders such as First 5 LA, affiliated researchers, and community stakeholders, CDN seeks to generate knowledge and advance evidence-rich policies that will improve the health, safety, and well-being of the children of California. CDN maintains data use agreements with numerous agencies that give permission to link cross-sector data and configure them longitudinally. The population-based, cross-sector data can be leveraged to develop applied and actionable research, support cost-effective program evaluations, and address policy-relevant questions.

CCWIP is a long-standing university/agency data partnership between CDSS and the University of California at Berkeley (UCB), supported through funding from CDSS and the Conrad N. Hilton Foundation. The project, housed in the UCB’s School of Social Welfare, gives policymakers, child welfare workers, researchers, and the public direct access to customizable information on California’s entire child welfare system.

CCWIP and the CDN have been data and research partners since the CDN’s inception in 2013 and are well positioned to collaborate on this project. CDN maintains a formal data and research collaboration with CCWIP, working closely with CCWIP researchers to provide technical support to state and county child welfare agencies. CCWIP is included in CDN's memorandum of understanding/contract with the California Department of Social Services (CDSS); likewise, CDN is named in CCWIP's agreement with CDSS.

Table B.1. Staff involved in the project

Name	Title	Affiliation	Role
Emily Putnam-Hornstein, Ph.D.	CDN principal investigator	USC	Oversee all aspects of the project, including design, analysis, and reporting. Serve as primary point of contact with state agency and CMI Data Linkages project.
Regan Foust, Ph.D.	CDN research scientist, co-investigator	USC	Project management, results translation/dissemination/communication, and manuscript preparation.
John Prindle, Ph.D.	CDN co-investigator	USC	Lead cross-sectional and longitudinal analytic strategies.
Daniel Webster, Ph.D.	CCWIP principal investigator	UC-Berkeley	Facilitate extract. Provide technical assistance for CWS/CMS records. Help with deliverables and manuscript preparation (co-authorship).
Stephanie Cuccaro-Alamin, Ph.D.	CCWIP analyst	UC-Berkeley	Facilitate extract. Provide technical assistance and analytic support. Help with deliverables and manuscript preparation (co-authorship).
Wendy Weigmann, Ph.D	CCWIP analyst	UC-Berkeley	Facilitate extract. Provide technical assistance and analytic support. Help with deliverables and manuscript preparation (co-authorship).
Joe Magruder, Ph.D.	CCWIP analyst	UC-Berkeley	Facilitate extract. Provide technical assistance and analytic support. Help with deliverables and manuscript preparation (co-authorship).

Source: Project plan.

Note: CDN = Children's Data Network; CCWIP = California Child Welfare Indicators Project; USC = University of Southern California; UC-Berkeley = University of California at Berkeley.

Background

The project developed a framework for an administrative-record–based methodology to estimate the number of children who are victims of abuse or neglect. The project, designed and tested using data from California, had the goal of developing a methodology that could be generalized to other states, and could produce estimates inclusive of both children who are identified as victims of abuse or neglect in a given year and those who are victims, but do not come to the attention of CPS.

It is important to note that the objective of this project was not to produce a single set of estimates held out as the "true" rate of abuse and neglect in a given community. Instead, the goal was to develop a number of upper- and lower-bound maltreatment estimates based on a range of assumptions and different methodological approaches. To document the potential to use administrative records to develop such estimates, we drew on information available through California vital birth, child protection, and death records. Thanks to the existing research infrastructure available at CDN and CCWIP, each of these data sources had already been cleaned, standardized, geocoded, and probabilistically linked using an algorithm customized to California's data.

Using records from 1998 through 2018, along with two different analytic approaches to structuring records (that is, cross-sectional and longitudinal), we exploited variability in the likelihood of CPS involvement and substantiation across:

- Counties with different policy and practice environments
- Child and family demographics, as measured for the population overall using vital birth records
- Time (that is, annual and lifetime involvement with CPS)
- Maltreatment type (that is, emotional, physical, and sexual abuse; neglect)
- Referral patterns before and after substantiation events

The observed variation allows the development of reasonable population-based approximations of child maltreatment victimization, which were compared with official substantiation rates.

Research Questions

This project has two main strategies, described below.

Cross-sectional strategy

How do estimates of victimization of child maltreatment vary based on (a) number of years in an estimate window (for example, only in focal year [2015] or up to three years on either side of focal year) or (b) county-level variation in how likely victims are to be identified?

We used California Child Protective Service records from 2012 through 2018 to identify a cross-section of children, across counties and demographics, identified as substantiated victims of maltreatment in 2015. We identified all other contacts (for example, no referrals, referrals to the CPS hotline that were screened out without investigation, referrals that were investigated but unfounded, and referrals that were substantiated) during the three years before the focal substantiation for a given child victim and the three years following substantiation. We estimated county-level variation in the annual/cross-sectional substantiation rates (conditioned on child-level characteristics gleaned from birth and CPS data) and use that county variability to extrapolate a range of estimated maltreatment rates.

Longitudinal strategy

How do annual incidence and cumulative prevalence rates differ by county and demographic characteristics at birth for children born in California in 1999? To what extent does earlier involvement with the child welfare agency predict substantiated child maltreatment?

We organized CPS records longitudinally for a cohort of children born in 1999 to estimate the cumulative childhood risk of abuse and neglect in California. With these data, we examined differences in annual incidence and cumulative prevalence rates by county and by demographic characteristics at birth.

Data

Sources

Three data sets were linked for this project (Table B.2).

Table B.2. Data sources

Name	Years	Geography covered	Source	Measures
Child protection (CWS/CMS)	1999–2018	California (statewide)	California Department of Social Services	Child protection encounters (referrals, substantiations) across time and counties
Vital birth records	1999	California (statewide)	California Department of Public Health	Birth characteristics
Vital death records	1999–2017	California (statewide)	California Department of Public Health	Deaths

Source: Project plan, interim, and final memo.

Note: CDN and CCWIP had access to all three data sets before the project began.

Linking process

CDN links and analyzes birth, child welfare, and death records (among other data sources) under approved state and university IRB protocols. The CDN uses unique identifiers created from linked birth, child welfare, and death records using previously published machine learning methods (see E. Putnam-Hornstein et al., 2020 for a detailed description). Briefly, linkages are developed using probabilistic matching methodologies that incorporate identifying information including names and dates of birth of both children and parents. ChoiceMaker, the probabilistic record linkage software that CDN uses, is based on a machine learning technique called Maximum Entropy. ChoiceMaker is based on Clues, which are Boolean tests of similarly between fields in a record pair. Each clue is assigned a weight, which is learned through machine learning on manually reviewed training record pairs. The probability of a match is output by ChoiceMaker, and all pairs above a certain threshold are "matches" and included in the de-identified analytic data set. Those pairs with a value less than a second threshold are "differs" and not include in the de-identified analytic data set. Pairs with values between the two thresholds, called "holds," are held, reviewed, and may be added to the de-identified analytic data set. The de-identified data set with unique identifiers was used for analysis by only identified CDN and CCWIP researchers.

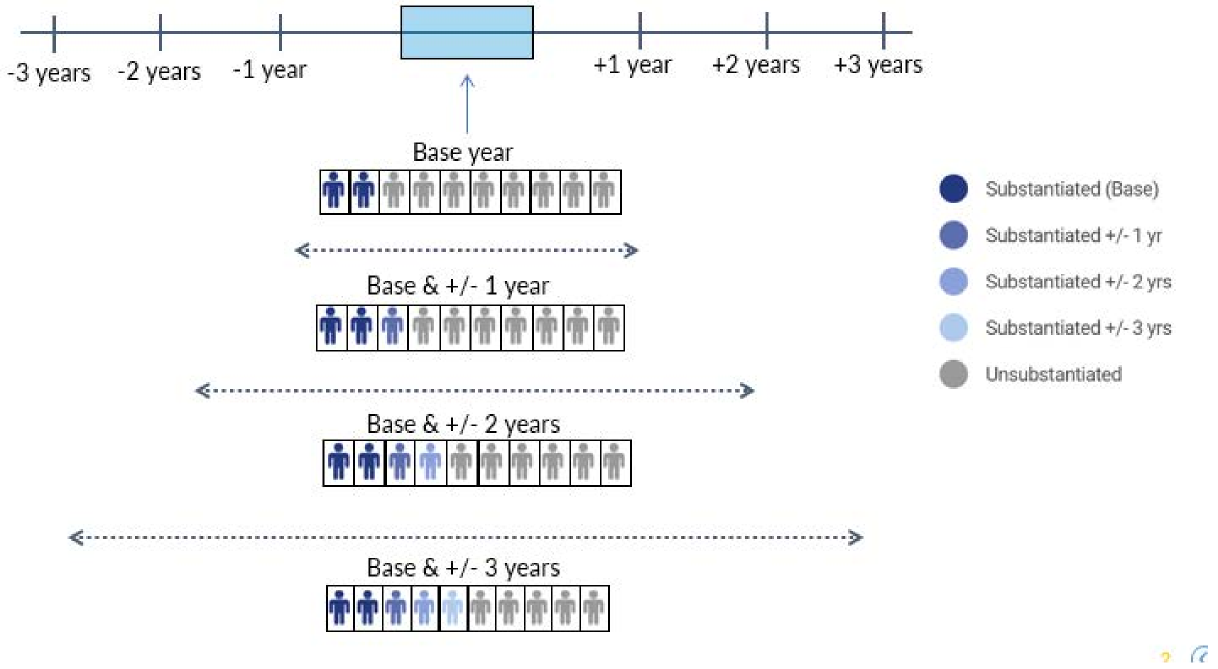
The team subjected 1,000 record pairs that represented complex clue patterns (that is, they were more difficult to categorize as a match or differ) to clerical review. We found that our model achieved a match recall rate of 92.5 percent, correctly classifying 92.5 percent of all "true" matches in the sample as matches. Meanwhile, false positive rates (pairs incorrectly identified as matches of out of all pairs) and false negative rates (pairs that were true matches but were incorrectly identified as holds or differs) were very low at 1.84 percent and 1.20 percent, respectively. This evaluation, training, and refinement process is ongoing; we are continually integrating new records, assessing match quality, and feeding that information back into our model to optimize accuracy.

Methods

Cross-sectional strategy

For our initial exploratory analysis, we used California CPS records from 2010 through 2018 to identify, for each year, a cross-section of children across counties and demographics who were referred because of alleged maltreatment. For each annual referral cohort, we then identified all other contacts (that is, no referrals; referrals to the CPS hotline that were screened out without investigation; referrals that were investigated but unfounded; referrals that were substantiated) during the three years before the focal substantiation for a given child victim and the three years after substantiation (see Figure C.1).

Figure B.1. Methodology for cross-sectional estimation of cumulative substantiation rates



We calculated annual base estimates of maltreatment victimization by examining the proportion of referred children with a substantiation during the year. We then documented how these base estimates changed if we *included children who had been referred but were not substantiated as a victim* in the specific year, but had been a substantiated victim in the year immediately prior, two years prior, and so forth. Similarly, we examined this for the years going forward with substantiations at one, two and three years following the base year. Using these data, we calculated a series of revised cumulative victimization rates that takes into account both children who were substantiated victims in a base year, as well as those who were referred to CPS but were substantiated as a victim within three years on either side of the base year. In addition, we explored maltreatment victimization rate estimates by allegation type (emotional, physical, and sexual abuse, neglect) and by geography, as well as interactions between allegation type and geography.

Longitudinal strategy

We also organized CPS records longitudinally for a cohort of children born in 1999 to estimate the cumulative risk of childhood abuse and neglect in California, providing a longitudinal record of system interactions from birth (1999) to age 18 (through 2017). With these data, we examined differences in annual incidence and cumulative prevalence rates by county and demographic characteristics at birth. Further work to model probability of exposure of abuse or neglect symptoms to mandated reporters was an opportunity to assess a range of incidence rate estimates. In addition, the longitudinal approach allowed us to examine "persistence," which would identify children whose risk for long-term negative outcomes is heightened because of their repeated involvement with the child welfare system. We regard this longitudinal strategy as a method for exploring victimization estimates based on assumptions that will serve as an extension of the cross-sectional strategy, but with additional opportunities to exploit annual vs. lifetime data.

Findings

Cross-sectional findings

In 2014, 82,388 children were substantiated as victims in California. If we add to that number the children who were referred to CPS but were substantiated as a victim within three years on either side of 2014, there was a 72.5 percent increase in the number of children classified as victims over the base estimate (142,157; Figure B.2). Applying the percent changes from the three-year cumulative analysis to the 2015 National Child Abuse and Neglect Data System (NCANDS) child victim total (n = 683,487), we found that an additional 524,614 children would be classified as victims, for a total of 1.2 million child victims in 2015 (Figure B.3).

Figure B.2. 2013 - 2015 Children with referrals: cumulative substantiation rate at -/+ 1, 2, and 3 years

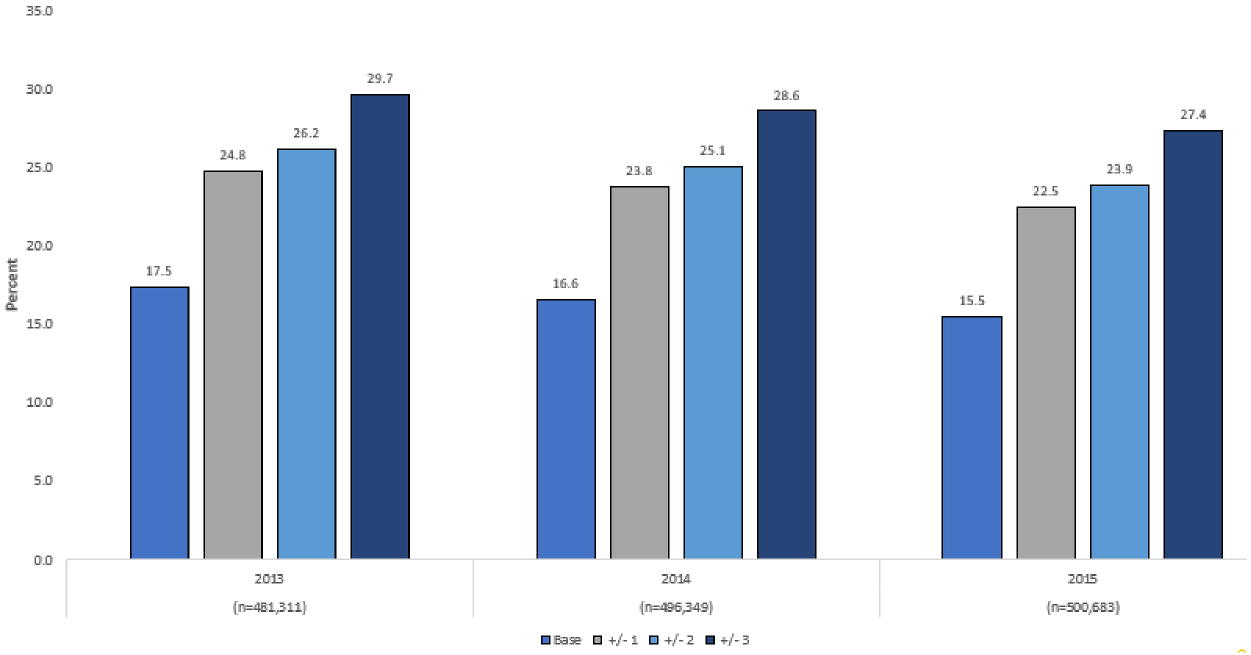
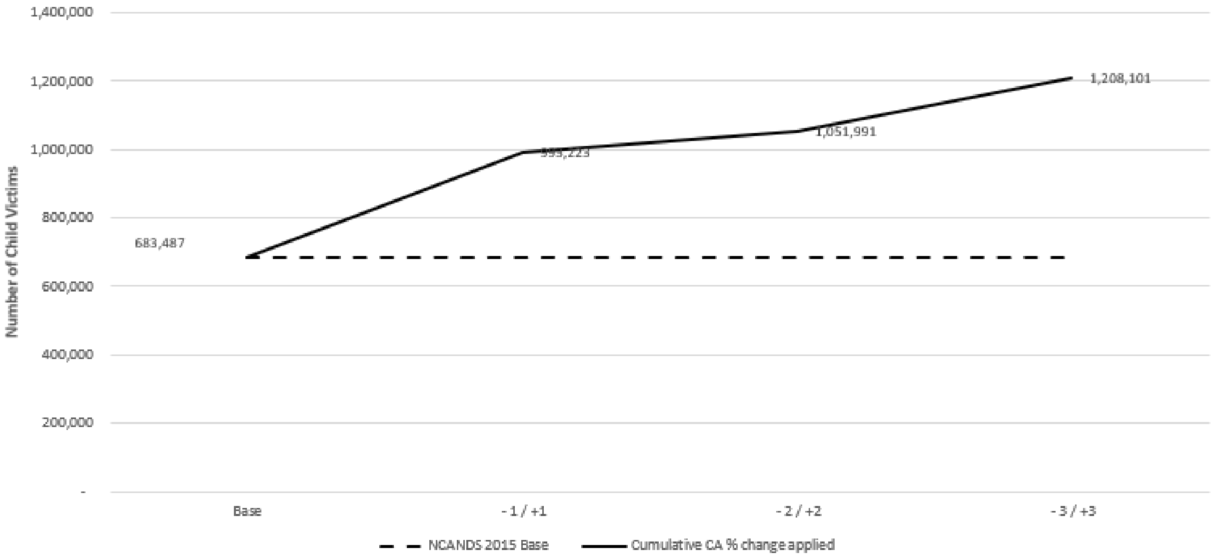


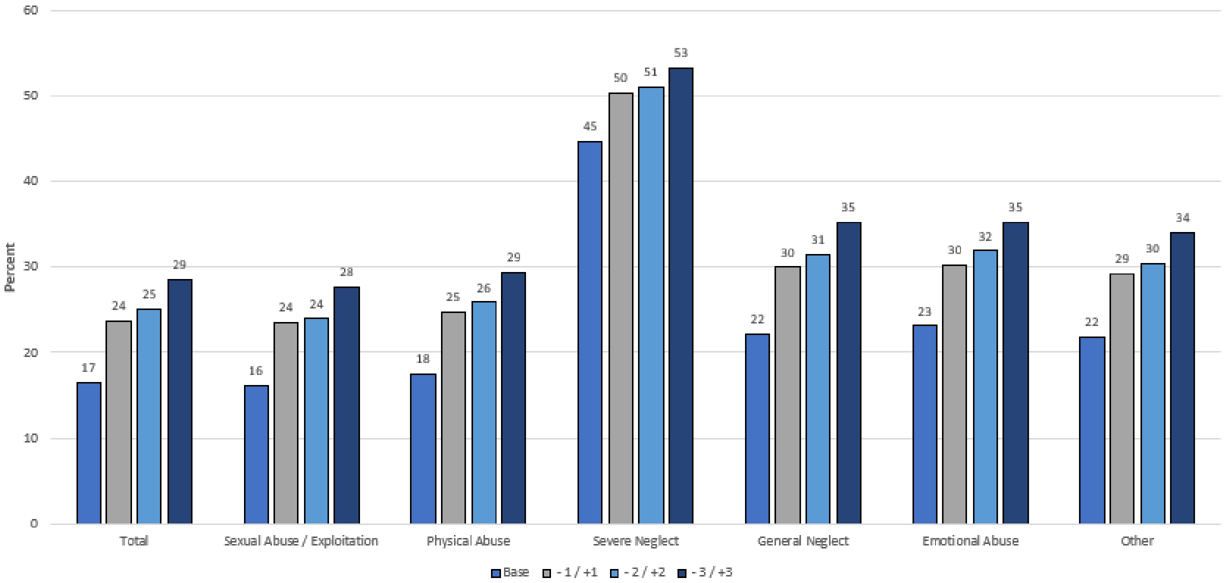
Figure B.3. 2015 U.S. child victim estimates derived using California's cumulative % change at -/+ 1, 2, and 3 years



In addition, if the 3-year cumulative window were applied on top of the 683,487 child victims in 2015 (i.e., if we included both children who were substantiated victims in a base year and those referred to CPS but substantiated as a victim within three years on either side of that base year) an additional 524,614 children would have also been counted as victims, totaling 1.2 million child victims.

We also found variation in base and cumulative substantiation rates (+/ - 1, 2, 3 years) by county and by allegation type (Figure B.4).

Figure B.4. 2014 Children with referrals cumulative substantiation rate at -/+ 1, 2, & 3 years by allegation type



Longitudinal findings

We organized CPS records longitudinally for a cohort of children born in 1999 to estimate the cumulative childhood risk of abuse and neglect in California through age 18 (through 2017). Cumulative prevalence rates by county and demographic characteristics at birth are in Table B.3 (see Putnam-Hornstein et al. 2021). In California, 519,248 children were born in 1999. Mirroring the secondary sex ratio nationally, the cohort was defined by a slight male majority. Consistent with California demographics, a plurality was born to Hispanic mothers. Slightly more than 1 in 10 children were born to teen mothers and approximately 30 percent of children were born to mothers with less than a high school degree. Overall, 92.8 percent of children had paternity established at birth.

Cumulatively, 29.9 percent of children in the cohort were investigated for alleged maltreatment and 12.1 percent were substantiated as victims of abuse or neglect. Between birth and age 18, 4.4 percent of children in the cohort experienced a removal and placement in foster care at least once. 1.1 percent experienced a legal termination of parental rights. Although the magnitude of the relative group differences varied across levels of CPS involvement, demographic patterns were directionally consistent. The cumulative proportions of Black and Native American children who had child protection encounters were significantly higher than those of other children. In the cohort overall, approximately half (51.3%) of Black and Native American (55.4%) children were investigated for alleged maltreatment before age 18. Both groups experienced termination of parental rights at twice the rate of White children in the cohort.

The likelihood of child protection involvement exhibited an inverse relationship to both maternal age at birth and maternal education levels. The rate of termination of parental rights was twice as great for children born to teen mothers versus children born to mothers 25 and older (IRR: 2.52, 95% CI: 2.36, 2.69). Likewise, children born to mothers with less than a high school degree experienced a termination of parental right at 2.6 times the rate of those who had completed high school (IRR: 2.60; 95% CI: 2.47, 2.74). Receipt of public health insurance and missing paternity were also strongly related to all levels of CPS involvement. Among children whose births were covered by public insurance, the rate of terminations of parental rights as 6 times that of children in the cohort covered by private insurance (IRR: 6.14; 95% CI 5.74, 6.57). Although only 1 in 14 children in California was born without paternity established (n = 37,513), parental rights were terminated for nearly 6 percent (n = 2,000) of those children. Children with missing paternity experienced termination of parental rights at seven times the rate of those with established paternity (IRR: 7.77; 95% CI: 7.37, 8.19).

Table B.3. Demographic characteristics of 1999 birth cohort in California (incidence rate per 1,000 children)

Characteristic	1999 Birth Cohort		Referred for Alleged Maltreatment		Substantiated as Victim of Abuse or Neglect		Removed and Placed in Foster Care		Termination of Parental Rights	
	N	%	Cumul. %	RR (99%CI)	Cumul. %	RR (95% CI)	Cumul.%	RR (95% CI)	Cumul. %	RR (95% CI)
Total	519,248	100.0	29.9	-	12.1	- -	4.4	- -	1.1	- -
Child sex										
Female	253,734	48.9	30.7	1.05 (1.05, 1.06)	12.5	1.07 (1.05, 1.09)	4.4	1.01 (0.99, 1.04)	1.1	1.06 (1.01, 1.12)
Male (ref.)	265,511	51.1	29.1	1.00	11.8	1.00	4.3	1.00	1.1	1.00
Maternal race/ethnicity										
Black	34,156	6.6	51.3	1.97 (1.95, 2.00)	24.0	2.22 (2.13, 2.23)	12.3	2.97 (2.86, 3.08)	3.2	2.47 (2.30, 2.65)
Native American	2,532	0.5	55.4	2.13 (2.06, 2.21)	30.0	2.83 (2.65, 3.02)	14.5	3.49 (3.17, 3.85)	3.9	2.97 (2.44, 3.63)
Latinx / Hispanic	252,691	48.7	32.7	1.26 (1.25, 1.27)	12.7	1.13 (1.14, 1.18)	4.0	0.96 (0.93, 0.99)	0.8	0.68 (0.61, 0.69)
Asian / PI	55,422	10.7	15.1	0.58 (0.57, 0.59)	4.9	0.46 (0.42, 0.46)	1.2	0.29 (0.27, 0.31)	0.3	0.20 (0.16, 0.23)
White (ref.)	172,188	33.2	26.0	1.00	11.0	1.0	4.2	1.00	1.3	1.00
Maternal age at birth										
< 20 years	57,693	11.1	50.4	2.14 (2.12, 2.16)	22.8	2.66 (2.60, 2.71)	9.2	2.99 (2.98, 3.08)	2.1	2.52 (2.36, 2.69)
20 – 24 years	120,519	23.2	38.0	1.61 (1.60, 1.63)	15.8	1.81 (1.78, 1.84)	5.7	1.85 (1.80, 1.91)	1.3	1.57 (1.48, 1.67)
25+ years (ref.)	340,974	65.7	23.6	1.00	9.0	1.00	3.1	1.00	0.8	1.00
Maternal education										
Less than high school	155,364	29.9	40.0	1.57 (1.56, 1.98)	17.9	1.93 (1.90, 1.96)	7.1	2.30 (2.25, 2.36)	1.9	2.60 (2.47, 2.74)
High school degree (ref.)	356,358	68.6	25.4	1.00	0.00	1.00	3.1	1.00	0.7	1.00
Birth payment method										
Public	218,643	42.1	41.7	1.96 (1.95, 1.98)	18.9	2.79 (2.75, 2.84)	7.7	4.11 (3.99, 4.24)	2.1	6.14 (5.74, 6.57)
Private (ref.)	298,178	57.4	21.2	1.00	7.1	1.00	1.9	1.00	0.3	1.00
Paternity established										
Missing	37,513	7.2	52.8	1.88 (1.86, 1.90)	28.4	2.79 (2.74, 2.85)	15.7	4.53 (4.40, 4.65)	5.8	7.77(7.37, 8.19)
Established (ref.)	481,735	92.8	28.1	1.00	10.9	1.00	3.5	1.00	0.7	1.00

Source: Project documents.

Note: Cumul. = cumulative. CI = confidence interval. RR = risk ratio. Rates of missing sociodemographic variables were low for all groups: child sex (0.03%), maternal race/ethnicity (0.05%), maternal age (0.01%), birth payment method (0.47%), and maternal education (1.45%).

Further details on the longitudinal analyses and findings can be found in:

1. Prindle, J., R. Foust, and E. Putnam-Hornstein. "Maltreatment Type Classifications and Transitions During Childhood for a California Birth Cohort." *Child Maltreatment,* 2021. Available at https://doi.org/10.1177/10775595211006784.
2. Putnam-Hornstein, E., E. Ahn, J.J. Prindle, J. Magruder, D. Webster, and C. Wildeman. "Cumulative Rates of Child Protection Involvement and Terminations of Parental Rights in a California Birth Cohort, 1999 – 2017." *American Journal of Public Health*, 2021, pp. e1-e7. Available at https://doi.org/10.2105/AJPH.2021.306214.

Next steps

The methodology developed in the cross-sectional work holds the promise of helping other jurisdictions develop alternative estimates of child maltreatment incidence through simple extensions of the window in which substantiation is counted. Findings from California suggest that many children who are reported in any given year may be experiencing conditions that at another proximate point in time the system classifies as official maltreatment. Future work should explore the extent to which these same adjustments yield similar increases in maltreatment incident estimates in other jurisdictions.

Lessons learned about administrative data linkage practices related to examining the incidence and risk of child maltreatment

We learned much from the process, and from the people we were able to collaborate with in framing these findings. Thanks in large part to the grant support, we had the time and space to apply new methodologies to and explore new conceptualizations of maltreatment. The protected time to collaborate, discuss, and present analyses that were in progress to an expanded group of researchers and stakeholders broadened our horizons and informed the development of our methodology to produce a range of community estimates of the incidence of abuse or neglect. In addition, our approach to using a base cohort of children with reports of maltreatment in any given year and then making adjustments to the substantiated victimization rate for that year by looking at how many of those children were classified as victims in the recent past, or will be in the near future, increases the number of child maltreatment victims in the United States by approximately 75 percent. Although academic researchers and others are correct to question the value of substantiation given the uncertainties about what contributes to that classification, we still believe this approach is a useful way to think about an alternative approach to estimating maltreatment incidence. It is also important to note that state comparisons and other demographic statistics reported by the Children's Bureau (i.e., annual Child Maltreatment report) continue to be strongly oriented around substantiation rather than children who are reported, meaning that this designation continues to drive policy decisions. In terms of recommendations or suggestions for other jurisdictions, gaining access to vital birth records would be useful for replication and characterization of maltreated children.

REFERENCES

Drake, B. and M. Jonson-Reid. “NIS interpretations: Race and the National Incidence Studies of Child Abuse and Neglect.” *Children and Youth Services Review*, vol. 33, no. 1, 2011. pp. 16 – 20.

Putnam-Hornstein, E., E. Ahn, J. Prindle, J. Magruder, D. Webster, and C. Wildeman. “Cumulative Rates of Child Protection Involvement and Terminations of Parental Rights in a California Birth Cohort, 1999 – 2017.” *American Journal of Public Health*, 2021, pp. e1 – e7. Available at https://doi.org/10.2105./AJPH.2021.306214.

Putnam-Hornstein, E., M. Ghaly, and M. Wilkening. “Integrating Data to Advance Research: Operations, and Client-Centered Services in California.” *Health Affairs*, vol. 39, 2020, pp. 655 – 661. Available at https://doi.org/10.1377/hlthaff.2019.01752.

Putnam-Hornstein, E., B. Needell, and A. E. Rhodes. “Understanding risk and protective factors for child maltreatment: the value of integrated, population-based data.” *Child Abuse & Neglect*, vol. 37, no. 2-3, 2013. pp. 116 – 119.

APPENDIX C:

CDN/RADY FINAL REPORT

This page has been left blank for double-sided copying.

Introduction

This appendix is an overview of the CDN/Rady project, conducted as part of the CMI Data Linkages work. It was written by the site team, with the Mathematica team working with the site to ensure consistency in information, level of detail, and presentation across sites.

Overview

The project incorporates hospital records into existing linked data to determine whether adding hospital records increased the predictive power of an existing predictive risk model (PRM) in identifying children and youth who experienced child maltreatment. The project team linked pediatric hospital records to child protection service (CPS) records and vital birth and death records to characterize children with medical encounters for accidental and non-accidental trauma relative to children with other medical encounters. The data were used to validate statewide predictive risk models built to predict future child protection involvement and explore the use of hospital records as predictors in predictive risk models (PRMs) focused on child maltreatment.

Partnership history

The Children's Data Network (CDN) is a data and research collaborative focused on linking and analyzing administrative records. In partnership with public agencies, philanthropic funders such as First 5 LA, the Conrad N. Hilton Foundation, the Heising-Simons Foundation, affiliated researchers, and community stakeholders, CDN works to generate knowledge and advance evidence-rich policies that will improve the health, safety, and well-being of the children of California.

The CDN maintains data use agreements with numerous agencies that give us permission to link cross-sector data together and configure them longitudinally. The population-based cross-sector data can be leveraged to develop applied and actionable research, support cost-effective program evaluations, and address policy-relevant questions. Emily Putnam-Hornstein, Ph.D., who served as principal investigator, oversaw all aspects of the project and was the main point of contact with state partners.

RCHSD is a nonprofit, 551-bed pediatric-care facility dedicated to excellence in care, research, and teaching. RCHSD's mission is to restore, sustain, and enhance the health and developmental potential of children through excellence in care, education, research, and advocacy. Subaward Principal Investigator Jeannie Huang, M.D., and Co-Investigator Cynthia Kuelbs, M.D., were responsible for extracting of RCHSD data.

The Centre for Social Data Analytics is in the School of Social Sciences and Public Policy at Auckland University of Technology. The Centre applies strong data science to linked data, undertaking research that tells us new things about the human experience. These valuable insights can help us measure the impact of social interventions, identify the need for new policy, and predict the likely outcomes for individuals and groups in society. PRM Validation Subaward Principal Investigator Rhema Vaithianathan, Ph.D., and team were responsible for providing technical assistance (TA) and guidance on PRM validation and coding.

Background

First, this project will allow the CDN to use machine learning methods to train probabilistic algorithms for linking hospital-system collected health data to other administrative data sources. This will ensure a high quality record match and lay a foundation for the ongoing linkage with subsequent years of hospital-based health data at both the local and state level.

This unprecedented analysis filled in the gaps for RCHSD patients by identifying key public service interactions before, during, and after hospital encounters. Ultimately, this analysis helped to form a more complete picture of the characteristics and public service trajectories of RCHSD patients compared to the universe of all children born in San Diego County. In addition, it enhanced our understanding of the social determinants of the physical and behavioral health of children served at RCHSD, including child maltreatment).

This project facilitated work to validate (and potentially improve) a model that predicts a child’s risk of future involvement with the child protection system once a referral for child abuse or neglect has been made.[17] This preliminary work suggested a relationship between identification as “high risk” and near-fatal injury and death during childhood. Applying this model to all children born in San Diego County (Birth Model) and to all children referred for maltreatment in San Diego County (Child Protective Services [CPS] Hotline Model) helped us assess the extent to which children who are identified by this model to be “at high risk” of maltreatment also are at elevated risk of injury and mortality in childhood (that is, to validate the model). It also increased understanding of the RCHSD patient population and the pediatric population of San Diego at large. In addition, it provided a mechanism for exploring any potential value that integrating hospital records into the PRM could have on increasing the model’s predictive value.

Research Questions

1. To what extent are children identified by a statewide predictive risk model (PRM) "at high risk" of maltreatment also at elevated risk of injury, poor health outcomes, and mortality in childhood? Two different models will be used to validate the PRM: (1) "birth model" predicting a child's risk of a CPS referral at birth and (2) "CPS hotline model" exploring injury and mortality risk.
2. What is the predictive value of integrating hospital data as predictors in the PRM?

Data

Sources

The data used in this work are presented in Table C.1.

Table C.1. Data sources

Data source	Description of records and sample
RCHSD patient records (RCHSD)	Records for all in- and out-patient encounters between 2010 and 2016
Vital Birth/Death records (CDPH)	Records for all births and deaths in California between 2010 and 2016
Child protection records (CDSS)	Statewide CWS/CMS records for the years 2010–2016

Source: Project documents.

---

[17] Please see http://www.datanetwork.org/research/assessing-childrens-risk-using-administrative-records-a-proof-of-concept-predictive-risk-modeling-prm-project for more information about the model that has been developed.

Linking process

RCHSD securely transmitted records for all clients seen between 2010 and 2016 to the CDN for linkage. The CDN linked RCHSD client records to birth, child protection, and death records using previously published machine learning methods,[18] and created a unique identifier (Table C.2). Briefly, linkages are developed using probabilistic matching methodologies that incorporated identifying information including names and dates of birth of both children and parents. The names and dates of birth of both children and parents and other unique-and non-unique identifiers were used to develop the proposed linkages. ChoiceMaker, the probabilistic record linkage software that CDN uses, is based on a machine learning technique called Maximum Entropy. ChoiceMaker is based on Clues, which are Boolean tests of similarity between fields in a record pair. Each clue is assigned a weight, which is learned through machine learning on manually reviewed training pair records. ChoiceMaker outputs a probability of a match, all pairs above a certain threshold are matches, pairs less than a certain threshold are differs and those pairs in between are held out for review. The dataset was then stripped of all direct identifiers, leaving only the encrypted unique identifier and an identifier that links back to RCHSD records and then placed on a secure server. Once analytic data were appended to the deidentified file, analysts were able to carry out the analyses. In this way, the team was able to maintain the separation between PII and analytic information (i.e., the separation principle, which is considered best practice for ensuring confidentiality and security in administrative record linkage.

Table C.2. Methods for linking data

Data source	Variables used to link data	Linkage approach
RCHSD patient records (RCHSD)	Patient First Name Patient Last Name DOB [date of birth] Address Zip Code SSN [Social Security number] Sex	Probabilistic
Vital Birth/Death records (CDPH)	Birth surname of decedent’s mother Date of decedent’s birth Decedent’s place of birth First name of decedent Initials of decedent’s spouse, if married Last name of decedent Middle name of decedent SSN of decedent Surname of decedent’s father Date of birth/delivery Date of child’s death Date of decedent’s death Father’s date of birth Father’s SSN First name of child	Probabilistic

---

[18] (Putnam-Hornstein et al., 2020)

Data source	Variables used to link data	Linkage approach
	Last name of child Last name of father Mother’s date of birth Mother’s first name Mother’s maiden name (birth surname) Mother’s SSN	
Child Protection records (CDSS)	Agency name Badge number Birth date Birth place Birth year – Caretaker 1 Birth year – Caretaker 2 City name Common first name Common last name Common middle name Court case number Death date Death place Department division name Driver license number Driver license state code type Gender code Mailing city name Mailing zip number Name Partner birth date Receiver safe surrender child name SSN	Probabilistic

Source: Project documents.

Analytic methods

To form a complete picture of the characteristics and public service trajectories of RCHSD patients, we matched administrative CPS records from San Diego County with RCHSD health encounter records for the cohort of all children born between 2010 and 2014. CPS and health care encounter data spanned the period from 2010–2016 and included information about the timing and type of referrals alleging maltreatment. After CDN linked records, staff added ICD-9 and 10 codes to document health care encounters. We used Schnitzer’s (2011) condensed list of ICD codes that are suggestive or confirmatory of a medical encounter for child abuse or neglect. Using these linked records, we produced descriptive statistics about our population and examined the nature of medical encounters that led to official reports of maltreatment, identifying characteristics that affected reporting likelihood; and used medical data to validate assessed differences in risk among children reported to CPS for abuse or neglect. Four analyses were conducted:

1. **Identified children with the highest risk score and an injury encounter**. We looked at all unique children in our data, classified their risk based on the highest risk score assigned for any referral

during the analytic window, and coded all associated injury encounters, regardless of when the injury occurred relative to the child abuse and neglect referral.

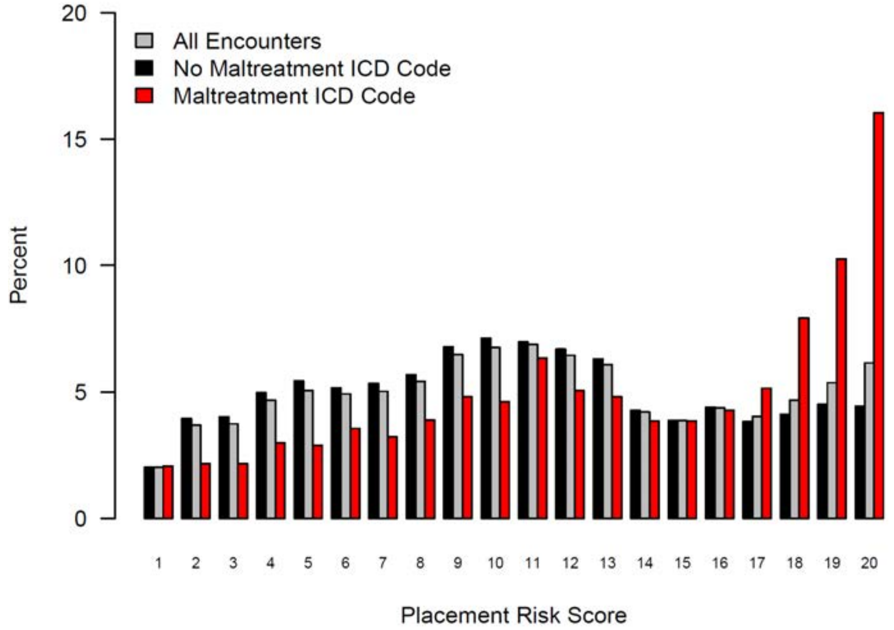
2. **Randomly selected risk score and an injury encounter.** We looked at all unique children in the data, randomly selected a CPS referral and associated risk score for each child during the analytic window, and coded their associated injury encounters, regardless of when the injury occurred relative to the selected child abuse and neglect referral.
3. **Identified the highest risk score before an injury encounter.** We looked at all unique children in the data and coded the child's risk level based on the highest risk score assigned during the analytic window, but *before* a specific injury encounter.
4. **Randomly selected risk score before an injury encounter.** We looked at all unique children in the data, randomly selected a referral and associated risk score for each child during the analytic window, and coded a medical encounter as having occurred only if the selected referral date was *before* the injury encounter.

Findings

Key findings include the following:

- Nearly one in five patients (19.7 percent) born between 2010 and 2014 who were followed by RCHSD until about age 5 had one or more report of alleged maltreatment made to CPS. This aligns with other research the team completed using statewide CPS, birth, and death records.
- Among children with a documented encounter coded as maltreatment (suggested/confirmed) in RCHSD data, most (84.9 percent) were also reported to CPS. The CPS reporting rate is lower for injuries and events with codes "suggestive," as opposed to "indicative," of maltreatment. These results reinforce the value of the list of those medical encounters for identifying the larger universe of children who may have been victims even if it was not officially documented in the medical records.
- From the perspective of the local child protection agency, 61 percent of the children who were born in San Diego between 2010 and 2014 and reported for maltreatment had a medical encounter with RCHSD.
- We found a strong association between the risk level (generated by the CPS risk model) and maltreatment-related injuries in hospital records. In addition, we found no relationship between the risk level (generated by the CPS risk model) and medical encounters that were not maltreatment-related (Figure C.1). As such, our findings not only validate a model trained to predict future CPS involvement, but also clearly indicate that maltreatment injuries are a credible indicator of child harm when viewed from the lens of the child protection system. Figure C.1. Relationship between risk for placement and hospital encounters.

Figure C.1. Relationship between highest risk score for placement and medical encounters



Next steps

We are in the process of developing multiple papers, presentations, and applications for future funding based on this work. Specifically:

1. Initial analyses investigating CPS system responses formed the basis for a COVID-focused NIH submission, "The Impact of COVID-19 on Child Maltreatment-Related Medical Encounters and System Responses Using Linked Administrative Data," submitted on October 5, 2020.
2. The team was invited to present "Use of RCHSD Healthcare Encounter Data to Validate a Child Protection Risk Model" as a poster session at the Rady Children's Hospital Interprofessional Innovations Transforming Healthcare Symposium on October 20, 2020,
3. The team has been invited to present to the Helfer Society's 2021 Pre-Conference Institute: Abusive Head Trauma & Medical Evaluation of Child Physical Abuse.
4. The team submitted an abstract, "A Longitudinal Study of Healthcare Utilization Among Infants Reported for Maltreatment", to the Pediatric Academic Societies Meeting 2020.
5. Additional papers that explore the relationship between Adverse Childhood Experiences health outcomes and the variation in identifying race/ethnicity in administrative data sources are in the works as well.

Lessons Learned About Administrative Data Linkage Practices Related to Examining the Incidence and Risk of Child Maltreatment

We learned a number of lessons based on our data linkage project—all of which would inform the considerations and recommendations we could provide to other states or localities working on data linkage initiatives related to child maltreatment incidence. Including medical records in data linkage projects is a strategic move because it makes more funding sources and grant mechanisms available for health care projects.

SSNs are infrequently available in ER /hospital records for children. We discovered this through our linkage work with Rady Children's Hospital, but have confirmed this reality with the Office Statewide Health Planning which manages state hospitalization data. Although the inclusion of SSN numbers in medical records is higher for adults, it is clear that many pediatric medical settings do not use SSNs for medical billing. Specifically, we found that less than one third of records had a SSN. This has implications for the accuracy of the linkages that can be completed and underscores the necessity of having other non-unique personal identifiers available for linkage.

Although most (~80 percent) of maltreatment documented through official child abuse and neglect ICD codes is associated with a report to child protective services by medical reporters, the CPS reporting rate is lower for injuries and events with codes "suggestive" of maltreatment. That said, data from this study reinforce the value of the Schnitzer classification list of ICD-9/10 codes for identifying the larger universe of children who may have been victims even if it was not officially documented in the medical records. (Translations between ICD-9 and ICD-10 codes has been well documented and does not pose hurdles to coding medical records.)

It is possible to set up data sharing agreements to support the use of medical and CPS records for research, but it is a different process than simply signing an agreement with a public agency. Hospitals / ERs operate within a complex health care system. Clarity is needed about who has the authority to grant access to sensitive data and which parties are needed as signatories. These roles, responsibilities, and governance structures should be documented at the outset of a project.

Maltreatment injury base rates are, as expected, low. Thinking strategically about the numbers of years that need to be accessed and the structure of longitudinal data is critical to ensuring a sample that is of sufficient size.

Consistent with what was observed in Allegheny County, Pennsylvania, findings from San Diego, California, suggest that medical records collected in other jurisdictions are also of high enough quality to serve as useful sources for linkages and validating predictive risk models.

REFERENCES

Putnam-Hornstein, E., M. Ghaly, and M. Wilkening. (2020). Integrating Data to Advance Research, Operations, and Client-Centered Services in California. *Health Affairs*, vol. 39, 2020, pp. 655–661. Available at https://doi.org/10.1377/hlthaff.2019.01752.

Schnitzer, P.G., P.L. Slusher, R.L. Kruse, and M.M. Tarleton, Identification of ICD Codes Suggestive of Child Maltreatment. *Child Abuse & Neglect*, vol. 35(1), 2011, pp. 3–17. Available at https://doi.org/10.1016/j.chiabu.2010.06.008

APPENDIX D:

CSSAT FINAL REPORT

Introduction

This appendix is an overview of the CSSAT project, conducted as part of the CMI Data Linkages work. The site team authored the appendix, and the Mathematica team worked with the site to ensure consistency in information, level of detail, and presentation across sites.

Overview

The public has raised concerns that the current opioid epidemic is increasing the risk of child maltreatment and contributing to higher caseloads in child welfare systems (Brundage and Levine 2019). This project expanded knowledge about the validity of these concerns. Based on several sources of data on a state population, the CSSAT project used multiple indicators of child maltreatment and involvement with the child welfare system. Individual and community-level risk factors were included in the analysis, which can guide the use of prevention and intervention services.

Partnership history

Overall, the state of Washington has been supportive of linking data from different entities and sharing the linked data with research partners. The state-administered child welfare agency has had a data sharing agreement with the University of Washington (UW) School of Social Work since July 1, 2007.

The partnership includes research staff (researchers and data linkers) and data partners (state agency staff). The research staff included three people: Joseph A. Mienko, principal investigator; Rebecca Rebbe, co-principal investigator, and Karen Segar. Dr. Mienko oversaw all aspects of the project and is the main contact with the state partners. Dr. Rebbe was the primary data analyst, even after she moved to the University of Southern California. Ms. Segar, the data manager at the University of Washington Harborview Injury and Prevention Research Center, completed the data linkages, which the institutional review board (IRB) requires someone outside the research analytic team to conduct.

The data partners were Tammy Cordova of the Department of Children, Youth & Families (DCYF) Office of Innovation, Alignment, and Accountability (OIAA), where she serves as the data and reporting administrator; Sarah Veele of OIAA, who is the research and analysis administrator; Katie Hutchinson of the Washington Department of Health, where she serves as the data manager and epidemiology supervisor (she facilitated access to vital records and hospital admission records); and Mariana Rosenthal, manager and epidemiologist at the Washington Department of Health Prescription Management System, who facilitated access to data from the prescription management system.

The project expanded on work using linked administrative data in the context of a preexisting ongoing relationship. Although there is an ongoing agreement for DCYF to share child protective services (CPS) data with the UW School of Social Work, the study needed new approvals from the Department of Health (DOH) and the IRB to get updated data (births, hospitalizations, and deaths). According to state law, the Washington State IRB had to approve the work because this project used Washington State administrative data concerning clients of a state agency.

The cornerstone of this partnership is an annual work plan and an associated data sharing agreement with OIAA. The data sharing agreement facilitated the quarterly transfer of more than 100 tables from the data warehouse for Washington’s transitional Comprehensive Child Welfare Information System (CCWIS), which is known as FamLink.

The relationship with DCYF facilitated access to DCYF data. Given the topical focus of the project, DCYF was the main point of contact for disseminating findings related to this project and the parent study, the Risk of Death and Injury Study, as a whole (see Rebbe et al. 2019, 2020). The study team's relationships with DOH only involved the acquisition of data needed for this study.

Background

The purpose of this study was to examine the association between county-level rates of opioid overdose events and child maltreatment indicators. Specifically, the study examined associations between county-level rates of opioid overdose events in Washington State, as measured by hospitalizations and deaths, with rates of CPS (1) reports, (2) substantiations, (3) removals, and (4) maltreatment-related hospitalizations. The study built on and extended previous research in two ways. First, it focused on opioid overdose events specifically. By disaggregating opioids from general parental substance (ab)use, the opioid epidemic's specific impact on the child welfare system can be clarified. Second, it used repeated and nuanced measures of opioid overdoses events, CPS involvement, and child maltreatment over an extended period of time to capture a more precise and accurate picture of temporal associations (Rebbe et al. 2020).

Research Questions

1. How does individual-level and county-level opioid use/misuse impact the risk of maltreatment for children and families in Washington State?
   - To what extent do counties exhibiting pronounced indications of an opioid public health emergency, as indicated by opioid overdose hospitalizations and deaths, also exhibit increased rates of child maltreatment?
   - To what extent do individual-level maternal opioid prescriptions affect in utero opioid exposure diagnoses?

2. How does opioid use/misuse impact contact with the child welfare system for children and families in Washington State?
   - Is there disproportionate testing of children/mothers for prenatal substance exposure in certain populations (racial/ethnic groups, people with different insurance status, maternal age) compared to appropriate referent groups in the entire population of children born in the state during the study time period?
   - If disproportionate testing exists, does the disproportionate representation result in disproportionate reporting rates for families experiencing opioid use/misuse or prenatal opioid exposure?

3. How do indications of opioid use/misuse or prenatal opioid exposure (POE) at the family and county level impact substantiation and placement decisions within households under investigation for maltreatment?

The project could not access the data required to answer Research Question 2. Although DOH saw value in the question, the data's sensitive nature was a persistent barrier. DOH did not have the resources available to link the data on its own, and was not able to share personally identifiable information (PII) with the team members so they could make the linkages.

Data

Sources

The state of Washington provided the data for the study—specifically, CPS records, birth records, hospital admissions, and death records.

- Birth records (approximately 1.5M rows): Records of all live births in Washington State from 1999 through 2017, a 19-year birth cohort maintained by DOH.
- Hospital admissions records (approximately 13M rows): Records from the Comprehensive Hospital Abstract Reporting System (CHARS) from 1999 through 2017, a DOH system collecting patient-level information on inpatient and community hospitals for all nonmilitary hospitals in Washington State.
- Death records (approximately 1.1M rows): Mortality records for all fatalities in Washington State from 1999 through 2017, maintained by DOH.
- CPS records
    - Person records (approximately 3.2M rows): Records for all persons with data in FamLink
    - Intake records (approximately 2M rows): Records from the Washington State child welfare system for all reports of child maltreatment from 1999 through 2018
    - Screening records (approximately 1M rows): Records from the Washington State child welfare system for all reports of child maltreatment that resulted in an investigation or assessment from 1999 through 2018
    - Removal records (approximately 100K rows): Records from the Washington State child welfare system for all removal episodes from 1999 through 2018

Linking process

The project used a combined (probabilistic and deterministic) approach to link records. The goal was to match family members across data sources, so birth and CPS records were used to identify parents of children.

The linkage process made use of the following data elements in all of the aforementioned data:

- Full names for children listed in the birth cohort, and the parents of the children
- Dates of birth for children listed in the birth cohort, and the parents of the children
- Records of race, ethnicity, and tribal membership at birth for children listed in the birth cohort and the parents of the children
- The geographic residence of the child or the child's parents at the time of a given event in the data
- The full or partial Social Security number for a child or the child's parents

The project used a combination of deterministic (blocking) and probabilistic linkages as outlined by Enamorado et al. (2019). Project staff decided to use a relatively new product from Amazon Web Services (AWS) which better accommodated the unique arrangement they had with the state with respect to linkage personnel (that is, a single person has direct access to PII).

Specifically, the project team used AWS Glue, a component of the HIPAA-compliant AWS Lake Formation tools now offered by AWS.[19] The key feature of the AWS Glue for his ongoing linkage work was FindMatches, a machine learning transformation algorithm that is a tool to identify and remove duplicate records across data sets in a manner that requires little action on the part of the linking staff. In this case, the same algorithm is used to identify and link duplicate records across data sets. The AWS Glue Developer Guide outlines the entire process.[20]

The one limitation of the approach is that FindMatches is truly a black-box algorithm. It was clear from reviews of the system logs that FindMatches used Apache Spark and a variation on k-means clustering.[21] Accuracy can be tested through more traditional post hoc approaches (for example, examination of the proportion of infant CPS referrals present in state birth records). Amazon also provides accuracy and precision metrics typical of all machine learning models. Overall, the team believed the benefits of the ease of use and good fit for the small team outweighed the opaqueness of the underlying algorithm.

Given the black-box nature of the algorithm, the AWS Glue interface limited the tuning criteria to those available – namely, balancing the risk of false positives (vs false negatives) in terms of recall versus precision, and cost versus accuracy. The chosen approach was to follow the logic of Zech et al. (2016) and err on the side of avoiding false-positive matches. The project staff maximized for accuracy (that is, conformity with true positives) and precision (that is, repeatability); and thus trained its algorithm with accuracy (versus cost) and precision (versus recall) parameters set to 0.9.[22]

Washington has a highly mobile population. Project staff were concerned that linking individuals across data sets would result in insufficient power to conduct the planned analyses. The AWS Lake Formation tool addressed this concern by giving the ability to block based on gender, given its use in deterministic linking due to the number of years of data. Additionally, some portion of the unmatched cases may represent migration to a locality outside of Washington. The project staff compared they unmatched case rates to those reported by other CMI Data Linkages sites, which were similar. They also examined U.S. Census migration rates within the same localities to determine if the rate of unmatched cases was unreasonable.

Analytic Methods

The team conducted all analyses in R, after using AWS to link the data.

Research Question 1:

1. How does individual-level and county-level opioid use/misuse impact the risk of maltreatment for children and families in Washington State?

---

[19] See “AWS Lake Formation” at https://aws.amazon.com/lake-formation/?whats-new-cards.sort-by=item.additionalFields.postDateTime&whats-new-cards.sort-order=desc.

[20] See “Tuning Machine Learning Transforms in Amazon Glue” at https://docs.aws.amazon.com/glue/latest/dg/add-job-machine-learning-transform-tuning.html.

[21] See “K-Means Clustering with Apache Spark” at https://www.bmc.com/blogs/k-means-clustering-apache-spark/.

[22] See Zech et al. (2016) for a fuller explanation of the approach.

The project staff used Bayesian Model Averaging (BMA) to find the variables that were the best fit for each model with the four outcome measures (CPS report rates (regardless of screening decision), CPS substantiation rates, CPS removal rates, and maltreatment-related hospitalization rate). This method considered multiple models simultaneously and used the Bayesian Information Criterion (BIC) to identify the optimal model. After identifying the variables to include in each model, the team ran an ordinary least squares model, followed by a fixed effects panel data model with year and county as the fixed effects. Project staff tested to see if the panel data model was a better fit than the ordinary least squares model, then ran a random effects panel data model and used a Hausman test to identify if the random or fixed effects model was a better fit. For both tests, p-values < 0.05 indicated that the fixed effects model was better than the ordinary least squares or random effects models for the respective tests.

Research Question 3:

3. How do indications of opioid use/misuse or prenatal opioid exposure (POE) at the family and county level impact substantiation and placement decisions within households under investigation for maltreatment?

The project identified children diagnosed at birth with opioid exposure through either the maternal or child ICD-9 diagnostic codes. Project staff ran a multistate survival model identifying three placement outcomes: (1) birth home, (2) out-of-home placement, and (3) hospitalization or death. They included a county rate of opioid-related hospitalization or death, concentrated disadvantage, presence of a CPS report, child sex, birth payment, maternal race, maternal age at birth, and parity as covariates. They completed chi-square analysis on a descriptive distribution table.

Findings

Research Question 1:

1. How does individual-level and county-level opioid use/misuse impact the risk of maltreatment for children and families in Washington State?

Although numerous news reports have stated there is a relationship between the opioid epidemic and child maltreatment, the project found that controlling for other factors in Washington State resulted in no relationship between opioid overdose events and child maltreatment indicators. Initially, the analysis revealed positive, statistically significant relationships between opioid overdose event rates and CPS report rates (regardless of screening decision), CPS removal rates, and maltreatment-related hospitalizations. However, in full panel data models that took a number of factors into consideration, there were no statistically significant relationships identified (Table D.1). Instead, CPS reports and maltreatment-related hospitalizations increased over time, whereas CPS substantiations and CPS removals decreased in the study time period.

The findings may relate to maltreatment policies in Washington State, and should be interpreted within the context of broader systems-level policies and related practices. Although the definition of neglect in Washington State gives weight to parental substance abuse as a factor in child maltreatment, the state does not consider substance abuse maltreatment in and of itself. This ambiguity may lead to fewer reports to CPS than there are in states with more stringent definitions of neglect. Another possible explanation is the lack of resources available in Washington State during the study time period. Following the Great Recession in 2008, Washington State had to make a number of budget cutbacks to address the reduction in state revenue. This included cuts to CPS staff and social worker positions. According to the state's historical spending trends, salaries and wages for children and family services were more than $32 million lower in the 2011–2013 biennium than in the 2007–2009 biennium (Washington State Fiscal Information 2020). The decrease in substantiations and removals by year found in this study may also reflect CPS workers raising the threshold that warrants substantiation and removal in response to a lack of services available to families (Barnett et al. 1993, Giovannoni 1991).

Table D.1. Panel data model results examining relationship between opioid overdose events and child maltreatment-related outcomes

	Child maltreatment-related outcomes											
	Reports			Substantiations			Removals			Maltreatment-related hospitalizations		
Coefficient	Est.	S.E.	*p*	Est.	S.E.	*p*	Est.	S.E.	*p*	Est.	S.E.	*p*
Opioid overdose event rate	-7.79	4.19		0.07	0.91		-0.04	0.79		0.04	0.09	
Public assistance	0.27	1.25		-0.11	0.63		-0.62	0.32		-0.06	0.03	*
Population younger than 18	-9.52	5.15		-2.39	2.46		1.23	1.20		-0.15	0.14	
Unemployment	1.25	2.16		0.14	0.75		-0.75	0.49		0.04	0.04	
Year 2006	25.45	1.85	***	-1.57	0.65	*	0.001	0.22		0.01	0.02	
Year 2007	27.63	1.84	***	-2.24	1.04	*	-0.25	0.34		0.09	0.05	*
Year 2008	24.98	1.90	***	-2.68	0.84	**	-1.21	0.36	***	0.07	0.03	*
Year 2009	22.91	1.95	***	-3.65	0.87	***	-1.39	0.39	***	0.15	0.04	***
Year 2010	24.43	2.01	***	-3.22	0.92	***	-1.60	0.34	***	0.11	0.03	***
Year 2011	20.35	1.54	***	-3.61	1.10	**	-2.18	0.41	***	0.22	0.03	***
Year 2012	24.91	1.80	***	-3.90	1.07	***	-2.22	0.38	***	0.14	0.04	**
Year 2013	29.31	2.11	***	-2.26	0.96	*	-1.78	0.48	***	0.15	0.03	***
Year 2014	29.24	2.01	***	-3.03	1.07	**	-1.81	0.51	***	--	--	
Year 2015	30.89	2.14	***	-4.08	1.00	***	-2.03	0.43	***	--	--	
Year 2016	32.41	2.24	***	-5.44	1.24	***	-1.78	0.45	***	--	--	
Year 2017	36.97	2.38	***	-4.86	1.09	***	-1.49	0.38	***	--	--	
R^2	0.55			0.19			0.24			0.18		
BIC	3267.518			2367.033			1733.149			-178.7682		

Source: Author analysis.

Note: Rates are per 1,000 in the population. Public assistance, population under 18, and unemployment are *z*-scores. Each model is a fixed effects model. Data on hospitalization rates for child maltreatment were not available after 2013. Est. = estimate. S.E. = standard error. *p* = *p*-value. BIC = Bayesian Information Criterion.

* Significantly different from zero at the .05 level, two-tailed test.

** Significantly different from zero at the .01 level, two-tailed test.

*** Significantly different from zero at the .001 level, two-tailed test.

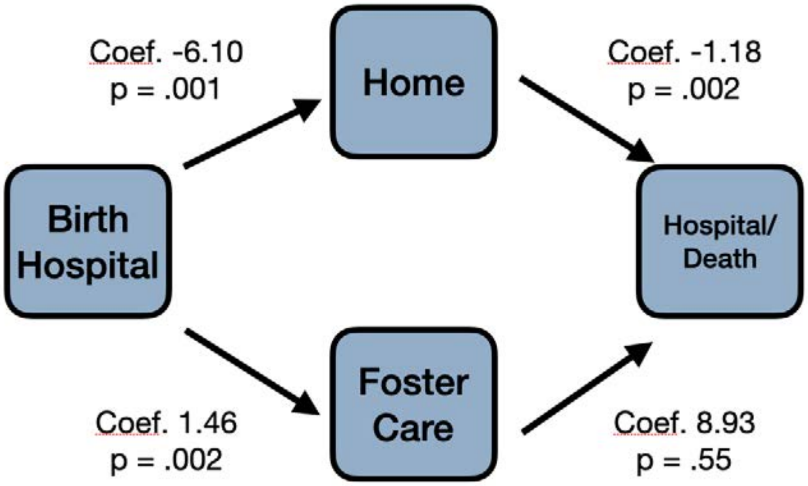
Research Question 3:

3. How do indications of opioid use/misuse or prenatal opioid exposure (POE) at the family and county level impact substantiation and placement decisions within households under investigation for maltreatment?

This analysis examined specific transitions for infants diagnosed with prenatal opioid exposure. The transitions were from (1) birth hospitalization to birth home, (2) birth hospitalization to foster home, (3) birth home (after birth hospitalization) to either a hospitalization or death, and (4) foster home (after birth hospitalization) to either a hospitalization or death. The analysis revealed that as the rates of county opioid overdose events increased (as measured by hospitalizations and deaths), children were less likely to go to their birth home from their birth hospitalization ($p < .001$). In other words, opioid-exposed infants born in counties with high rates of opioid overdose are less likely to go home after birth. Similarly, opioid-exposed infants were more likely to go to out-of-home care (foster care) in counties with higher rates of opioid overdose events ($p = .02$). For opioid-exposed children who do go home after birth, there appears to be a protective effect against experiencing hospitalization or death as the rate of opioid overdose events in the counties they are born in increases ($p = .002$). This remains true even when controlling for whether the child had a subsequent referral to the child welfare system after going home. The opioid overdose rate of birth counties was not associated with the transition from out-of-home placement to hospitalization or death.

A county's rate of opioid overdose events impacts the transitions from birth hospitalization to birth homes, birth hospitalization to foster care, and from birth homes to a hospitalization or death for children diagnosed with prenatal opioid exposure in Washington State. However, it does not impact transitions uniformly. Instead, it appears that the system engages in triaging activities and reserves placements in care for the most severe cases.

Figure D.1. County opioid overdose rate associations with child welfare outcomes



Note: Coef. = coefficient. *p* = *p*-value for coefficient. Home = Return to home. Foster Care = placed in a foster home.

Next steps

The project intends to update the analysis for Research Question 3 using the fully data through 2017. There are additional analyses that will be pursued using the data linked through 2017.

Lessons Learned About Administrative Data Linkage Practices Related to Examining the Incidence and Risk of Child Maltreatment

The 14-year relationship between Partners for Our Children and the State greatly enhanced the project staff's ability to conduct this study. The time and attention needed to sustain the relationship are vital to the ability to address new research questions as they emerge.

The major lesson the team learned during this engagement is the need for a linkage solution that meets the unique needs of a small team with limited access to fully identified data. Within the last month of this project, they were able to identify a new solution,, AWS Glue, that perfectly met the need. Glue is an extract, transform, and load (ETL) service that allows a single person to train a machine learning model on a small extract of records from two or more data sets requiring linkages. On the basis of this training data, Glue programmatically generates Python code that a researcher can run to replicate the patterns observed in the training data to the data as a whole. The only point in the linkage process at which individuals observe PII is during the training procedure, 30- to 60-minute process that the designated data linker for the IRB can manage. This approach to linking data works well for this small team and has allowed them to unblock their linkage process after months of delay.

References

Barnett, D., J. T. Manly, and D. Cicchetti. “Defining Child Maltreatment: The Interface Between Policy and Research.” *Child Abuse, Child Development, and Social Policy,* vol. 8, pp. 7–73, 1993. Norwood, NJ: Ablex.

Brundage, S.C., and C. Levine. “The Ripple Effect: The Impact of the Opioid Epidemic on Children and Families: Recommendations for an action agenda based on a multidisciplinary meeting.” United Hospital Fund. New York, 2019. Available at https://uhfnyc.org/media/filer_public/17/2c/172ca968-43aa-45f9-a290-50018e85a9d8/uhf-opioids-20190315.pdf.

Enamorado, T., B. Fifield, and K. Imai. “Using a Probabilistic Model to Assist Merging of Large-Scale Administrative Records.” *American Political Science Review,* vol. 113, no. 2, 2019, pp. 353–371.

Giovannoni, J. “Social Policy Considerations in Defining Psychological Maltreatment.” *Development Psychopathology*, vol. 3, no. 1, 1991, pp. 51–59.

Rebbe, R., A.S. Bishop, J. Ahn, and J.A. Mienko. “Opioid Overdose Events and Child Maltreatment Indicators: Differential County-Level Associations.” *Children and Youth Services Review*, 119, 2020, pp. 1–7. https://doi.org/10.1016/j.childyouth.2020.105671.

Rebbe, R., J.A. Mienko, E. Brown, and A. Rowhani-Rahbar. “Hospital Variation in Child Protection Reports of Substance Exposed Infants.” *The Journal of Pediatrics*, vol. 208, 2019, pp.141–147.e2. https://doi.org/10.1016/j.jpeds.2018.12.065

Rebbe, R., Mienko, J. A., & Martinson, M. L. (2020). “Reports and Removals of Child Maltreatment-Related Hospitalizations: A Population-Based Study.” *Child Maltreatment*, 1077559520984549–1077559520984549. https://doi.org/10.1177/1077559520984549

Washington State Fiscal Information (2020). Historical spending trends. Retrieved from http://fiscal.wa.gov/SpendHistObj.aspx.

Zech, J., G. Husk, T. Moore, and J.S. Shapiro. “Measuring the Degree of Unmatched Patient Records in a Health Information Exchange Using Exact Matching.” *Applied Clinical Informatics,* vol. 11, no. 7(2):330-40. doi: 10.4338/ACI-2015-11-RA-0158. PMID: 27437044; PMCID: PMC4941843.

APPENDIX E:

UA-SSW FINAL REPORT

This page has been left blank for double-sided copying.

Introduction

This appendix presents an overview of the UA-SSW project, conducted as part of the CMI Data Linkages work. The site team wrote the appendix, although the Mathematica team worked with the site to ensure consistency in information, level of detail, and presentation across sites.

Overview

This project focused on understanding how risk and protective factors relate to child maltreatment reports at the county level across the nation. In previous studies, researchers have assessed county-level racial disparities in child welfare involvement—for example, Maguire-Jack et al. (2015) and Putnam-Hornstein et al. (2013)—but additional research is needed to explain widely varying state- and county-level maltreatment rates.

This study builds on earlier work (Smith et al. 2018), expanding it to a national level, and incorporates counties with fewer than 1,000 child maltreatment reports, which could offer new insights about rural counties and reveal new regional or state patterns. The site team continues to work on developing valid ways to measure county-level risk of child maltreatment.

Partnership history

No formal partnerships were needed to access the data used in the project. A university professor and graduate students from the University of Alabama’s School of Social Work formed the research team. Data were publicly accessible (for example, from the U. S. Census Bureau or U.S. Departments of Labor or Agriculture) or were available upon request from the National Data Archives on Child Abuse and Neglect (NDACAN). No formal partnership existed between the research team and NDACAN, although the professor had previously used data housed at NDACAN.

Background

Previous research (Smith et al. 2017) focused on larger counties in the U.S. South, examining the association between (1) county-level child maltreatment risk and protective factors and (2) county-level child maltreatment reports and victimization rates. This project extended that work and addressed the previous study’s research questions nationally with all counties, which offers new insights about rural counties and reveals new regional or state patterns. The site team also worked to develop new ways to operationalize county-level child maltreatment risk.

New measures included factors shown to be associated with child maltreatment at the community level, including rates of child poverty, single parenthood, unemployment, food insecurity, access to health care, and other county-level characteristics. Finally, the project provided an opportunity to assess the characteristics of counties in which the risk factors for child maltreatment do not coincide with rates of child maltreatment reports. For example, some counties with high rates of child poverty and solo parenthood – two critical child maltreatment risk factors – have very low rates of child maltreatment reports and victimization. Likewise, some counties with comparatively low child maltreatment risk factors have high maltreatment and victimization rates. In addition, the site team recognized that the community causal mechanisms that influence parental behavior (that is, actual maltreatment) might differ from those that influence the reporting of maltreatment. The reports are likely affected by community members' definitions, observations, and decisions to report, which may be distinct from maltreatment behaviors (Coulton et al. 2007).

Research Questions

The site team posed four research questions.

1. How closely are county-level risk factors for child maltreatment associated with county-level rates of child maltreatment reports and victimization? Do counties having comparatively high risk factors for child maltreatment also have high maltreatment report rates?
2. What distinguishes counties with high risk factors for child maltreatment but low report rates, and vice versa? (Demographic characteristics? Protective factors, such as comparatively high rates of social association, primary health care providers, or mental health care providers)?
3. Throughout the U.S., how do child maltreatment risk factors and report rates in rural counties with majority populations of color compare with the risk factors and report rates in (a) nonrural counties with majority populations of color, (b) rural counties that are majority White, and (c) nonrural counties that are majority White?
4. In counties that have higher risk factors for child maltreatment but lower child maltreatment report rates than other counties in the same regions:
   a. How do county-level report rates vary by report source? In counties with higher risk factors but lower report rates than in comparison counties, are the report rates lower among professionals, nonprofessionals, or both?
   b. How do county-level report rates vary by the race of the child subject? That is, what is the county-level racial composition of child maltreatment report and victimization rates for all counties in the relevant regions, including rural counties?
   c. How do county-level report rates vary by report type? That is, what is the county-level distribution of the most severe allegation type for all counties in relevant regions, including rural counties?

Owing to issues with data access, the team was unable to address Question 4.

Data

Sources

Most of the data used for this project are publicly available, such as data from the U.S. census (Table E.1). The county-level data on child maltreatment comes from the National Child Abuse and Neglect Data System (NCANDS), which is housed at NDACAN. Researchers may request data from NCANDS data at the child level and aggregate to the county level using county-level identifiers. However, the county-level identifiers are not available on child level data for counties with fewer than 1,000 reports (about 78 percent of U.S. counties). The site team requested a number of variables aggregated to the county level for all counties. The site team received number of reports, number of children in reports, number of substantiated reports, and number of child victims from 2021 through 2015 aggregated at the county level. NDACAN would not release even county-level aggregates of report source or maltreatment types for all counties. NDACAN declined our request because HHS policies prohibit sharing any data (even aggregated data) from counties with fewer than 1,000 reports.

Table E.1. Data sources

Data source	Description of records and sample
NCANDS	(1) number of reports; (2) number of substantiated reports; (3) number of children in the reports; and (4) number of child victims for all U.S. counties, including rural counties (2012 – 2015)
U.S. Census	Percentage rural: Percentage of the county defined as rural by the U.S. census
U.S. Census, Small Area Income and Poverty Estimates	Child poverty rate: County-level child poverty rate for 2012 through 2015
U.S. Census, American Community Survey	Single-Parent Household Rate: Percentage of children living in a household headed by a single parent, by county (2008 to 2012) and aggregate estimates (2011 to 2015)
U.S. Bureau of Labor Statistics	Unemployment Rate: Percentage of population ages 16 and older unemployed but seeking work, by county
U.S. Department of Agriculture survey distributed by Map the Meal Gap, a program of Feeding America	Food Insecurity Rate: Percentage of the population estimated to be food insecure based on responses to a survey distributed by the U.S. Department of Agriculture
Small Area Health Insurance Estimates	Adult Uninsured Rate: Percentage of adults under age 65 without health insurance
Centers for Disease Control and Prevention’s WONDER mortality data	Injury Death Rate: Number of deaths as a result of injury per 100,000 people
National Center for Health Statistics natality files	Low Birth Weight Rate: Percentage of births under 2,500 grams Teen Birth Rate: The birth rate for 1,000 women ages 15 to 19
County Business Patterns	Social Association Rate: Number of membership organizations per 10,000 people; membership organizations include labor, professional, recreational, religious, and civic organizations
Area Health Resources File from the American Medical Association and the Health Resources and Services Administration	Primary Care Physician Rate: Number of primary care physicians per 100,000 people
Centers for Medicare & Medicaid Services Provider Identification File	Mental Health Provider Rate: Number of mental health providers per 100,000 people
Behavioral Risk Factor Surveillance System	Adult Current Smokers Rate: Percentage of adult smokers Excessive Drinking Rate: Percentage of adults who report binge or heavy drinking
Child Welfare Information Gateway	State Universal Reporting Law: State has a universal child maltreatment reporting law

Source: Project documents.

A primary concern for this project was identifying counties for which the population is too small to reliably assess county-level child maltreatment report and victimization rates. A small number of U.S. counties have populations so low that a rate per 1,000 children is subject to vast fluctuation with a change of only one or two reports. In each of the four years, we excluded about 150 to 200 such counties nationwide. We also assessed report rates and victimization rates and excluded a small number of counties with outlier rates (i.e., report rates of greater than 150 per 1,000 or victimization rates greater than 70 per 1,000). The total number of counties excluded per year because of small populations or outlier rates was 207 in 2012, 202 in 2013, 151 in 2014, and 167 in 2015. Nearly all of the counties excluded were the very small population counties. The project's findings do not apply to counties with population less than 1,000.

Another data concern involved assessing the extent of missing data. Although most data sources have data for every county, a small number are missing data for some counties. These tend to be data elements (such as the primary physician rate or mental health provider rate) that are calculated per 100,000 people and cannot be reliably calculated for counties with very small populations. Such data elements were less helpful than others for understanding risks in rural counties.

We assessed the extent of missing data by conducting a descriptive analysis of all variables. Because variables with larger percentages of missing data (such as the primary physician rate) were not critical to our main research questions or analysis, the missing data were not a problem for the analysis. We were able to address our study aims and answer Research Questions 1 through 3 with data from all counties.

Linking process

We linked data using the Federal Information Processing Standard (FIPS) code assigned to each U.S. county or county equivalent. We linked state-level data using a state code only (Table E.2).

Unlike some probabilistic linking processes, the FIPS code provided a high quality and accurate method for linking data. Any errors, such as a mis-specified FIPS code, were easily to identify by an error code or warning message in the linking process.

Table E.2. Methods for linking data

Data source	Variables used to link data	Linkage approach
NCANDS	FIPS code	Simple merge on key variable
Census and all other county-level variables	FIPS code	Simple merge on key variable
Census and all other state-level variables	State code	Simple merge on key variable

Source: Project documents.

Analytic Methods

The site team used descriptive statistics and multilevel regression models to address all research questions. We used the multilevel models to account for the fact that counties are nested within states, and therefore there is dependence between counties in the same state. (Dependency means that counties in the same state are more alike one another than counties in other states.)

We conducted the regression models to assess associations of a range of independent variables with the county-level maltreatment report rate. For most of our analyses, the dependent variable was the county-level child maltreatment investigated report rate, or the number of investigated reports per 1,000 children in a county. In some analyses, the dependent variable was the county-level child victimization rate, or the number of child victims per 1,000 children in a county. Key independent variables at the county level included continuous measures of the child poverty rate; percentage rural; and percentages Black, White, and Latinx. State-level variables included the child welfare worker/child ratio and an indicator of Medicaid accessibility. The independent variable social association rate was defined as the number of membership organizations (such as religious, labor, and political organizations) per 10,000 county residents. Because counties are nested in states, we conducted multilevel regression models with random intercepts.

Findings

Research Question 1

1. How closely are county-level risk factors for child maltreatment associated with county-level child rates of maltreatment report and victimization rates? Do counties having comparatively high risk factors for child maltreatment also have high maltreatment report rates?

To answer Question 1, we investigated the role of social association as a protective factor for child maltreatment at the county level. We examined whether social association was linked to child maltreatment reports and whether this relationship differed in rural versus nonrural counties. Additional work related to county-level risk and protective indices is ongoing.

At a bivariate level, there was a weak negative association between the maltreatment report rate and the social association rate ($r = -.07$, $p < .01$). In multilevel models accounting for child poverty, demographics, and Medicaid expansion status, the social association rate retained a negative association with child maltreatment reports. When we included county rurality status in the model, however, the association was suppressed, pointing to an interaction effect. Models run separately for rural and nonrural counties illustrated that the social association rate retained a strong negative association with child maltreatment reporting in rural counties but had a positive association in nonrural counties.

General patterns differed in rural and nonrural counties, suggesting that the role of social association could differ in different community contexts. In rural counties, as opportunities for social association went up, maltreatment report rates went down, suggesting a protective effect from membership organizations and perhaps reflecting enhanced bonds of social trust (Putnam 2000). But in nonrural counties, as opportunities for social association went up, so did maltreatment report rates. In more densely populated areas, opportunities for social association might promote more observation and formal support for struggling families (Klinenberg 2002), resulting in more maltreatment reports. Hence, social association could have different protective functions in different types of communities. (Additional findings for this research question are forthcoming.)

Research Questions 2 and 3

2. What distinguishes counties with high risk factors for child maltreatment but low report rates, and vice versa? (Demographic characteristics? Protective factors, such as comparatively high rates of social association, primary health care providers, or mental health care providers)?
3. Throughout the U.S., how do risk factors and report rates for child maltreatment in rural counties with majority populations of color compare with the risk factors and report rates in (a) nonrural counties with majority populations of color, (b) rural counties that are majority White, (c) nonrural counties that are majority White?

Research Questions 2 and 3 address the relationship between county-level child maltreatment reports and demographic and other county-level characteristics. The county-level child maltreatment report rate ranged from less than 1/1,000 children to over 119/1,000 children. Of all counties, 9.6 percent have majority populations of color, and 58.8 percent are majority rural. In 2015, of county types based on rurality and race/ethnicity, rural counties that were majority Black (n = 59) had the highest mean rate of child poverty (43 percent in 2015) but the lowest mean child maltreatment report rate (26/1,000 in 2015).

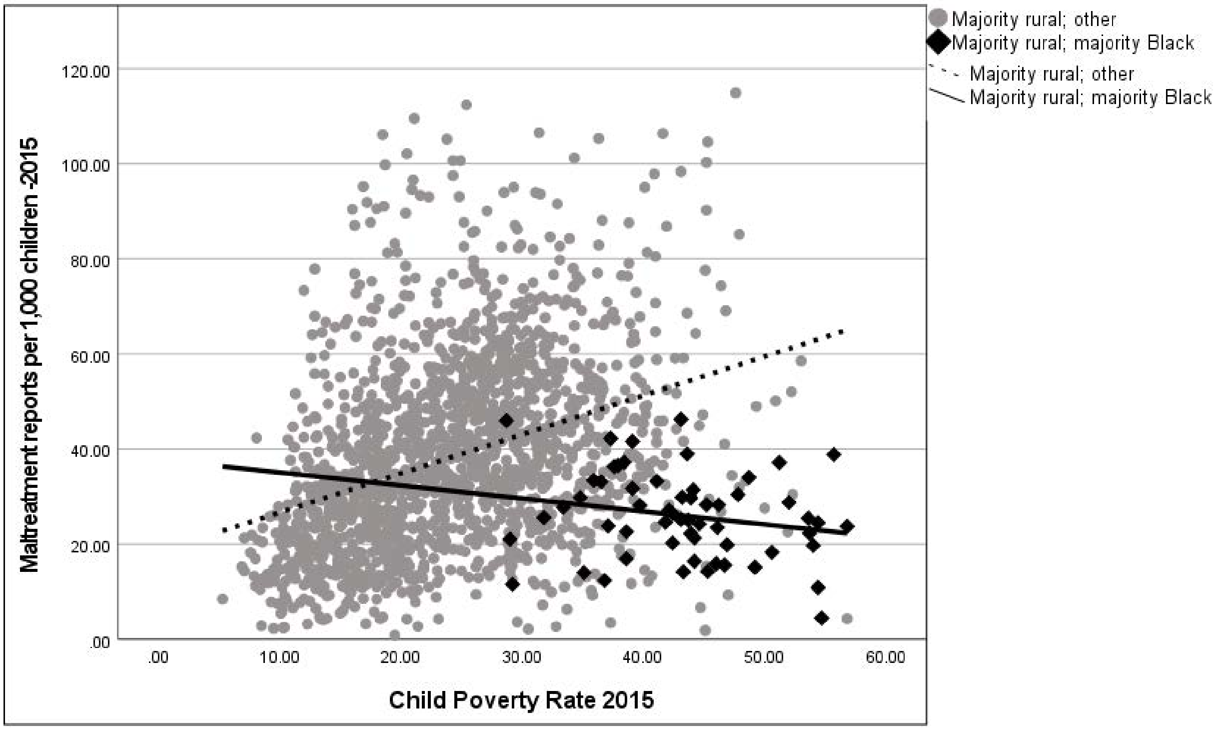
Post-hoc analysis showed that the maltreatment report rate in rural, majority Black counties differed from the maltreatment report rate in rural, majority White counties at a statistically significant level. Rural majority Latinx counties (n = 23) also had a higher mean rate of child poverty (31 percent in 2015) and lower rate of child maltreatment reports (32/1,000 in 2015) compared with rural, majority White counties (24 percent and 39/1,000 in 2015). The same patterns were evident in all four years, 2012 through 2015.

Consistent with previous research, in all rural counties, child maltreatment report rates were positively associated with child poverty. But this pattern did not apply to the small number of rural counties with majority populations of color, where child maltreatment report rates were negatively associated with child poverty. In multilevel models accounting for child poverty, majority rural counties and counties with majority populations of color generally had lower maltreatment report rates than other counties. An interaction term showed that maltreatment report rates were even lower in rural counties with majority populations of color.

We conducted a subsequent analysis to rule out the possibility that the results reflected (1) regional reporting differences rather than differences based on the racial/ethnic composition of rural counties and (2) an effect primarily driven by one racial/ethnic group rather than by populations of color generally. To test whether the results reflected lower report rates in states that also happened to have the most rural counties with majority populations of color, we repeated the analysis among southern states only, as southern states contain most rural counties with majority populations of color. Among six southern states (Alabama, Georgia, Louisiana, Mississippi, North Carolina, and South Carolina), maltreatment report rates were lower among rural counties with majority populations of color than among rural counties with majority White populations (27.9 vs. 38.5, $t = -9.1$, $p < .01$). Hence—although southern states contain most rural counties with majority populations of color, and maltreatment reporting tends to be lower in the South than in other U.S. regions—the lower report rates we found in U.S. rural counties with majority populations of color were not simply a result of lower report rates in the region.

We then tested whether a particular racial/ethnic group was driving the negative association between child maltreatment reports and child poverty in rural counties with majority populations of color. To do this, we conducted separate regression plots in rural counties to compare the race/ethnic-specific relationship between the child maltreatment report rate and child poverty. The negative relationship seen for all rural counties with majority populations of color was mainly driven by rural counties with majority Black populations, where the association between child poverty and maltreatment report rates was negative, albeit not statistically significant ($r = -.21$, $p = .12$) (Figure E.1). Among rural counties with majority Latinx populations, the relationship between the child maltreatment report rate and child poverty was positive and strong ($r = .47$, $p = .03$), as in rural counties with majority White populations. In counties with majority populations of color in which no single race/ethnic group constitutes a majority, child poverty was unrelated to the maltreatment report rate ($r = .009$, $p = .97$). The research team obtained similar findings from analyses conducted for years 2012 through 2015.

Figure E.1. Association between child poverty and child maltreatment report rates in counties that are majority rural and majority Black vs. other counties, 2015



Consistent with our previous findings focusing only on the U.S. South, the national data we analyzed revealed surprising patterns of child maltreatment reporting in rural counties with majority populations of color. We were fortunate to have access to data on child maltreatment reporting from all U.S. counties, including rural counties. To our knowledge, our work for the project was the first national child maltreatment study to disentangle county rurality from racial/ethnic composition by specifically investigating rural counties with majority populations of color.

We found that, compared with rural, majority White counties, rural counties with majority populations of color tended to have higher rates of child poverty but lower child maltreatment report rates. In addition, these counties did not have a positive relationship between child poverty and child maltreatment report rates, as seen in most counties and commonly found in previous studies. In the comparatively small number of rural counties with majority populations of color, as poverty rates went up, child maltreatment report rates went down. This negative relationship appeared to be primarily driven by low report rates in rural counties with majority Black populations.

As with most U.S. counties, in rural counties with majority Latinx populations, we found a strong positive association between child poverty and child maltreatment reporting. But child poverty rates were higher and child maltreatment reporting was lower in rural, majority Latinx counties compared with rural, majority White counties.

Our project adds to the literature on the complex and sometimes paradoxical relationships between child poverty, race/ethnicity, and child maltreatment at the community level. As with other recent U.S. studies (Wulczyn et al. 2013; Maguire-Jack et al. 2020), the project's findings indicate that simple conclusions about the relationship between community poverty, racial/ethnic composition, and child maltreatment may not apply for some child welfare indicators, or in some areas where higher rates of child poverty do not coincide with higher rates of child maltreatment or other child welfare involvement. The findings are also consistent with those of studies from other countries, which have shown surprisingly low levels of child welfare intervention in high-poverty, marginalized, or racial/ethnic minority communities (Bywaters et al. 2016; Sulimani-Aidan and Benbenishty 2013). Our findings make a case for avoiding assumptions about the level of child welfare involvement based on a community's demographic characteristics.

Besides revealing lower rates of child maltreatment reports in rural counties with majority populations of color, our project points to the distinction between child maltreatment investigated report rates and the incidence of child maltreatment. Although it is well-known that official maltreatment report rates do not reflect all child maltreatment, more research is needed on how child maltreatment comes to the attention of state authorities. Many studies have addressed racial/ethnic and urban/rural differences in trust in the police and crime reporting (for example, Burgason 2017; Desmond et al. 2016; Hamm et al. 2017; Kochel 2019), but few researchers have investigated how racial/ethnic composition or other community characteristics relate to confidence in child welfare authorities or hotlines for reporting child maltreatment. Future research should help distinguish the community characteristics that increase or reduce risks for child maltreatment from the characteristics that increase or reduce the reporting and investigation of maltreatment.

Next Steps

This project raised new questions about the child welfare response to communities of color. Black children are disproportionately represented in maltreatment reports and have greater exposure to maltreatment risk factors (Drake and Jonson-Reid 2011; Kim et al. 2017; Putnam-Hornstein et al. 2013; Wildeman et al. 2014). And many communities of color, including those with majority Black populations, have high levels of child welfare intervention (Fong 2019; Roberts 2008), prompting justified concern about invasive and authoritative overinvolvement of state agents. But what does it mean that this disproportionate level of intervention in communities of color is not evident in the smaller number of rural counties with majority populations of color, most notably in rural counties that are majority Black?

We hope these findings will not be interpreted as making a case for unnecessary intervention or more formal investigative oversight in high-poverty rural counties, but as raising questions about overlooked inequities and associated service gaps.

Lessons Learned About Administrative Data-Linkage Practices Related to Examining the Incidence and Risk of Child Maltreatment

Our project differed from the other CMI Data Linkage projects in its use of county-level rather than case-level linkages. Perhaps an important lesson is that there is still much to learn from comparing child maltreatment responses at the county and state levels. We began the project after examining the state Kids Count data map and noticing surprising patterns, such as apparently low maltreatment report rates in counties with high levels of need. Although probabilistic data linkages at the child level are complex and exceed the resources and capacity of many child welfare researchers and administrators, much can be learned from fairly simple county-level data linkages that many could conduct. It takes time to track down county-level data from various public sources, but the linkages are straightforward. Child welfare administrative data linked to publicly available data can be used more extensively to illuminate and inform child welfare practice. Analyses involving comparisons between counties and states can inform and improve child welfare services and policy.

Many of the project's accomplishments are described in the "Findings" section. We are excited about those contributions from this project. Most important, the analyses conducted for the project revealed a previously overlooked pattern in the formal child welfare response to rural counties with majority populations of color.

We encountered two primary challenges. First, NDACAN was not able to release even county-level aggregates of report source or maltreatment types for all counties. The request was declined due to policies set by HHS that prohibit sharing any data (even aggregated data) from counties with fewer than 1,000 reports. Therefore, we were able to replicate our findings over four years, however we were unable to confirm that the patterns identified from 2012 through 2015 continued. Furthermore, without county-level data on report sources, we could not address Research Question 4, which pertains to explaining the results identified in response to Question 3. One consideration stemming from this project is potentially having NDACAN allow for aggregated, county-level data on select variables from all U.S. counties available to more researchers.

The second challenge related to our organizational capacity. PI time was limited because of university administrative responsibilities, which increased as a result of the COVID-19 pandemic. The PI was therefore not able to devote time to data analysis during much of 2020, and work on Research Question 1 was delayed. The site team will conduct additional data analyses to answer Research Question 1 in summer 2021.

REFERENCES

Burgason, K.A. “Trust Is Hard to Come by: Disentangling the Influence of Structural Characteristics and Race on Perceptions of Trust in the Police in a Major Southern County.” *Journal of Ethnicity in Criminal Justice*, vol. 15, no. 4, 2017, pp. 349–371. doi:10.1080/15377938.2017.1385558

Bywaters, P., G. Brady, T. Sparks, and E. Bos. “Child Welfare Inequalities: New Evidence, Further Questions.” *Child & Family Social Work*, vol. 21, no. 3, 2016, pp. 369–380. doi: org/10.1111/cfs.12154

Coulton, C.J., D.S. Crampton, M. Irwin, J.C. Spilsbury, and J.E. Korbin. “How Neighborhoods Influence Child Maltreatment: A Review of the Literature and Alternative Pathways.” *Child Abuse & Neglect*, vol. 31, 2007, pp. 1117–1142. doi:10.1016/j.chiabu.2007.03.023

Desmond, M., A.V. Papachristos, and D.S. Kirk. “Police Violence and Citizen Crime Reporting in the Black Community.” *American Sociological Review*, vol. 81, no. 5, 2016, pp. 857–876. doi:10.1177/0003122416663494

Drake, B., and M. Jonson-Reid. “NIS Interpretations: Race and the National Incidence Studies of Child Abuse and Neglect.” *Children & Youth Services Review*, vol. 33, no. 1, 2011, pp. 16–20. doi:10.1016/j.childyouth.2010.08.006

Fong, K. “Neighborhood Inequality in the Prevalence of Reported and Substantiated Child Maltreatment.” *Child Abuse & Neglect*, vol. 90, 2019, pp. 13–21. doi:10.1016/j.chiabu.2019.01.014

Hamm, J.A., R. Trinkner, and J.D. Carr. “Fair Process, Trust, and Cooperation: Moving Toward an Integrated Framework of Police Legitimacy.” *Criminal Justice and Behavior*, vol. 44, no. 9, 2017, pp. 1183–1212. doi:10.1177/0093854817710058

Kim, H., Wildeman, C., Jonson-Reid, M., and Drake, B. “Lifetime Prevalence of Investigating Child Maltreatment Among U.S. Children.” *American Journal of Public Health*, vol. 107, no. 2, 2017, pp. 274–280. doi:10.2105/AJPH.2016.303545

Klinenberg, E. *Heat Wave: A Social Autopsy of Disaster in Chicago*. Chicago: University of Chicago Press, 2002.

Kochel, T.R. “Explaining Racial Differences in Ferguson’s Impact on Local Residents’ Trust and Perceived Legitimacy: Policy Implications for Police.” *Criminal Justice Policy Review*, vol. 30, no. 3, 2019, pp. 374–405. doi:10.1177/0887403416684923

Maguire-Jack, K., S.A. Font, and R. Dillard. “Child Protective Services Decision-Making: The Role of Children’s Race and County Factors.” *American Journal of Orthopsychiatry*, vol. 90, no. 1, 2020, pp. 48–62. doi:10.1037/ort0000388

Maguire-Jack, K., P. Lanier, M. Johnson-Motoyama, H. Welch, and M. Dineen. “Geographic Variation in Racial Disparities in Child Maltreatment: The Influence of County Poverty and Population Density.” *Child Abuse & Neglect*, vol. 47, 2015, pp. 1–13. doi:10.1016/j.chiabu.2015.05.020

Putnam, R.D. *Bowling Alone: The Collapse and Revival of American Community.* New York: Simon & Schuster, 2000.

Putnam-Hornstein, E., B. Needell, B. King, and M. Johnson-Motoyama. “Racial and Ethnic Disparities: A Population-Based Examination of Risk Factors for Involvement with Child Protective Services.” *Child Abuse & Neglect*, vol. 37, no. 1, 2013, pp. 33–46. doi:10.1016/j.chiabu.2012.08.005

Roberts, D.E. “The Racial Geography of Child Welfare: Toward a New Research Paradigm.” *Child Welfare*, vol. 87, no. 2, 2008, pp. 125–150.

Smith, B.D., E.S. Kay, and B.G. Womack. “How Can County-Level Maltreatment Report Rates Better Inform Child Welfare Practice?” *Children and Youth Services Review,* vol. 79, 2017, pp. 341–347.

Sulimani-Aidan, Y., and R. Benbenishty. “Child Maltreatment Reports in Israel: The Intersection Between Community Socioeconomic Characteristics and Ethnicity.” *American Journal of Orthopsychiatry*, vol. 83, no. 1, 2013, pp. 29–36. doi:10.1111/ajop.12009

Wildeman, C., N. Emanuel, J.M. Leventhal, E. Putnam-Hornstein, J. Waldfogel, and H. Hee. “The Prevalence of Confirmed Maltreatment Among U.S. Children, 2004–2011.” *JAMA Pediatrics*, vol. 168, no. 8, 2014, pp. 706–713.

Wulczyn, F., R. Gibbons, L. Snowden, and B. Lery. “Poverty, Social Disadvantage, and the Black/White Placement Gap.” *Children and Youth Services Review*, vol. 35, 2013, pp. 65–74.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

Mathematica
Progress Together

mathematica.org