ASPE
ASSISTANT SECRETARY FOR
PLANNING AND EVALUATION

OS
PCOR
TF

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

REPORT

Linking State Medicaid Data and Birth Certificates For Maternal Health Research

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

by
RAND Health Care

October 2023

OFFICE OF THE ASSISTANT SECRETARY FOR PLANNING AND EVALUATION

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

THE OFFICE OF HEALTH POLICY

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

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

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

CONTRIBUTING AUTHORS
RAND Health Care
Sara Heins, Zachary Predmore, Annetta Zhou, and Emily Hoch
ASPE
Violanda Grigorescu, Scott Smith

Suggested Citation: Heins, S., Grigorescu V., Predmore, Z., Zhou, A., Hoch, E., Smith, S. State Medicaid Linkages for Maternal Health Research. Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services. September 2023.

EXECUTIVE SUMMARY

Medicaid pays for nearly half of all live births in the United States.[1] Women and infants covered by Medicaid typically live in households with lower incomes, have more complex medical needs, and bear a disproportionate burden of maternal mortality and poor infant health outcomes compared with women and infants with private insurance. Mothers' Medicaid data, which includes enrollment and eligibility information, and claims and encounter data (hereafter "claims"), linked with infant birth certificates (hereafter "birth certificates") hold the potential to be a vital source of data for maternal and infant health research. While Medicaid claims data contain key information on diagnosis and service utilization for pregnant individuals, birth certificates often contain information on pregnancy outcomes not captured or incompletely captured in these claims.

Despite the utility of linked Medicaid and birth certificate data, performing such linkages across different states and multistate maternal health research with these data poses challenges. Each state administers its own Medicaid program, and states vary widely in the types of data that are linked to Medicaid claims and the methods for these linkages. Researchers who want to use multistate linked data must obtain them separately from each state, and the data obtained may not be comparable due to the different data linkage methodologies used by states.

Approach

Previously, the Office of the Assistant Secretary for Planning and Evaluation (ASPE), the U.S. Department of Health and Human Services (HHS) published a 2022 report on findings from a literature review of publications related to linking Medicaid claims with birth certificates and from discussions with representatives from participating states about their linkage processes.[2] To further the goal of building data capacity for patient-centered outcomes research (PCOR), the objectives of this project and report thereof were to:

- Engage selected state and national technical experts and leaders to develop a standardized methodology to link Medicaid and birth certificate data that could be used by multiple states;
- Develop a standardized procedure to make a proposed linked dataset available to outside researchers; and
- Create a research agenda of maternal health topics, including an example research protocol addressing one research question that could be answered using the proposed linked dataset.

Five states, represented in the prior report, agreed to assist with the implementation of these new objectives. In addition, representatives from four federal agencies (National Institutes of Health, the Centers for Disease Control and Prevention[CDC], the Health Resources & Services Administration, and the Centers for Medicare & Medicaid Services) were also invited to participate and provide input. A series of meetings were convened with *state and national technical experts* in data linkage methodology, data security, and data access. These experts contributed to developing guiding principles on several aspects of data linkage methodology and data security and access. Input was then sought from *state and national leaders* from state Medicaid, vital statistics, maternal health, and epidemiology departments during meetings that were convened along with technical experts to discuss and answer questions or concerns prior to finalizing the agenda and approaches described in this report.

Results

The group developed the following guiding principles that could be used to create a multistate linked database

for research using birth certificate and Medicaid claims data:

- **A standardized methodology to link these datasets that can be used by multiple states**
  The standardized methodology for linking mothers' Medicaid data with infant birth certificate data includes the following:
  - The same linkage software that allows for probabilistic matches be used by all states. Link Plus, a free, customizable software package developed by the CDC, is expected to facilitate the broadest involvement of multiple states.
  - The mother's Social Security Number be used in the first deterministic linkage step.
  - Mother's first and last names and 5-digit zip code, and mother's and infant's dates of birth be used in a subsequent probabilistic linkage step.
  - Other variables (e.g., mother's middle name, father's last name, and mother's race) only be used in a manual review of matches.
- **A standardized procedure to make a central database with core data elements available to outside researchers**
  The standardized procedures include the following:
  - States submit deidentified, linked claim-level data, which include a core set of data elements, to a central database, beginning with 2016 data and including at least one year of claims data before and after each delivery (where available).
  - Re-identification risk be assessed for both restricted and public use datasets prior to release for research.
  - A standardized application be created for researchers to obtain data from multiple states.
  - Institutional Review Board approval and data use agreement be required for accessing and using restricted data (individual and claim levels); and public use (aggregated) linked data be available online.

In addition, a **research agenda** was discussed, and the following topics were selected:
- Examining changes in maternal and infant health outcomes and Medicaid-covered service utilization following postpartum Medicaid extension (this topic was selected as the top priority by the group and developed into a sample research protocol).
- Examining changes in telehealth use during pregnancy and postpartum periods during and after the coronavirus disease 2019 (COVID-19) public health emergency for different risk groups.
- Examining how changes in access to reproductive health services impacted maternal health and birth outcomes.

Conclusion

Over the course of multiple meetings and written feedback from a diverse group of state and national experts and leaders, a set of guiding principles for linking Medicaid and birth certificate data and making these linked data available was developed. These guiding principles can be implemented to create a multistate linked database for research on maternal health, including PCOR. Through this standardized and centralized approach, researchers would be able to access these linked datasets through restricted use (claim-level) and public use (aggregated) options to answer important maternal health research questions. The study protocol described in this report (Appendix B) is an example of a study that could be done using the proposed restricted use data. In the future, these proposed data linkage activities could be expanded to include additional states or additional sources of data.

TABLE OF CONTENTS

TABLE OF EXHIBITS

CHAPTER 1. INTRODUCTION

Background

Medicaid pays for nearly half of all live births in the United States, including more than half of all births by non-Hispanic Black and Hispanic mothers.[1] Women and infants covered by Medicaid bear a disproportionate burden of maternal mortality and poor infant health outcomes, making them a key population group for the U.S. Department of Health and Human Service's (HHS's) efforts to address such outcomes. Medicaid data, which includes eligibility and enrollment data as well as claims and encounter data, (hereafter "claims") linked with live infant birth certificates (hereafter "birth certificates") are vital sources of data for maternal and infant health research. Birth certificates contain key information not captured in claims (e.g., parent education level, infant gestational age at birth) or captured inconsistently (e.g., low birth weight) in claims. Many high-quality studies use these linked datasets to investigate maternal and infant health research topics, as highlighted in a 2022 report.[2] However, most studies reviewed and included in this previous report were single- state studies.

To conduct research using multistate linked Medicaid and birth certificate data, data must be linked using consistent methodologies, and multistate data must be accessible to researchers. If Medicaid and birth certificate data are not linked using consistent methodologies across states, linked data from these states may not be comparable. For example, if one state uses a *deterministic* (i.e., exact matching) methodology based on the mother's name and Social Security Number (SSN) and another state uses a *probabilistic* approach (i.e., assigning a probability of a true match based on the degree of agreement across many variables), the matched populations of the two states may differentially exclude unmatched individuals from the study population. In the first state, mothers with commonly misspelled, non-Anglo names or without a SSN may be disproportionately excluded from the linked dataset, while the second state may have a linked dataset that is more representative of the population. These differences in population representativeness between the two states may impact the results of analyses using data from the two states.

Purpose

The overall goal of this project is to improve data capacity for patient-centered outcomes research (PCOR) and other research, as well as program outcomes measures related to maternal health, by making available a methodology for developing high-quality, multistate linked Medicaid-birth certificate data. In this project, a series of meetings with technical experts and leaders at the state and national level was held to address the following specific goals:

- Develop a standardized methodology to link Medicaid and birth certificate data that can be used by the participating states and potentially other states;
- Develop a standardized procedure to make a central database with core data elements available to outside researchers; and
- Create a research agenda on maternal health, including the development of an example research protocol as a use case.

Improving maternal health is a national priority and has been identified as a top priority by HHS,[3] as well as a focus area for the Office of the Secretary Patient-Centered Outcomes Research Trust Fund (OS-PCORTF) portfolio.[4] The OS-PCORTF portfolio is focused specifically on expanding data capacity to address maternal health through research, and it is aligned with the three goals of this project.

Organization of Report

This report is organized by chapter, starting with a description of the methods used to meet the three goals stated above (Chapter 2). Next, current state linkage efforts and a standardized linkage methodology are described (Chapter 3). Steps to make multistate linked data available to researchers (Chapter 4) and presentation of a set of maternal health research priorities that can be addressed using linked data (Chapter 5) follow. The report concludes with a summary and considerations for the future implementation of the developed approach (Chapter 6).

CHAPTER 2. METHODS

Participant Selection

Seven states that met the following criteria were invited to participate:
- Had at least 50,000 births in 2021;[5]
- Regularly linked Medicaid and birth certificate data, as defined in the previous report;[2]
- Had adequate quality Medicaid claims and encounter data (as defined by no "high concern" or "unusable" ratings in the Transformed Medicaid Statistical Information System (T-MSIS) Data Quality Atlas for 2020-2021 inpatient and other therapy fee-for-service claims and managed care encounter data, where applicable);[6] and
- Had previous collaboration within their state and with outside agencies, as defined in the previous report.[2]

Initial contacts from these seven states were identified using internet searches and references from the previous report. Representatives from five of the seven states agreed to participate: California, Colorado, Kentucky, North Carolina, and Ohio. Each of these five states was asked to recommend one or more individuals who could fulfill the following roles:
- An expert in the methodology used to link the state's Medicaid claims with birth certificate data (ideally the person responsible for conducting the linkages);
- An expert in the secure storage of linked Medicaid-birth certificate data and access procedures for internal use and external researchers (e.g., the person in charge of the data use agreement [DUA] process for researchers who want to use the linked data); and
- A leader (e.g., in Medicaid, vital statistics, or maternal health epidemiology) who can provide feedback on the approach developed by the group, as well as institutional buy-in and support to the final agreed-on guidance.

Representatives from the following four federal agencies were also invited to participate because of their roles in supporting maternal health research, vital statistics, and Medicaid programs: the National Institutes of Health (NIH), the Centers for Disease Control and Prevention (CDC), the Health Resources & Services Administration (HRSA), and the Centers for Medicare & Medicaid Services (CMS).

The state and national experts in linkage methods and in data security and access participated in smaller technical expert groups, which met to discuss the current linkage methodologies and processes in the five participating states and to develop the guiding principles for one consolidated approach. The technical experts also participated in larger group meetings with state and national leaders where these proposed guiding principles were reviewed prior to being finalized.

Appendix A lists the participants, their affiliations, and their roles in this project.

Process for Developing the Linkage Methodology

After an initial meeting introducing this project's goals and all participants, a series of meetings with state and national technical experts in data linkage methodology was convened to discuss state-specific processes and develop standardized linkage and data-sharing processes.

Specifically, representatives from the five participating states submitted written details on the methods used to link Medicaid and birth certificate data in their states and their processes for securely making data available

to researchers. Based on state input regarding the importance of different methodological decisions and the feasibility of implementation in their states, guiding principles were drafted on data linkage methodology. Similarly, technical experts in data security and access discussed current processes for securing linked data, making linked data available to researchers, and what data would be included in a central database consisting of linked, deidentified datasets with core data elements from five states. They discussed how these data could be shared through restricted and public use models. The larger group reviewed the linkage methodology, provided written feedback, and discussed it further during subsequent meetings.

To develop a research agenda for the use of the linked data, a list of research priorities was compiled based on suggestions from participants and based on the literature review conducted for the 2022 report.[2] These potential research priorities were discussed at a subsequent large group meeting with both technical experts and leaders, where the group also determined the top research priority and identified other important research topics. Subsequently, the smaller technical expert group discussed options for a research protocol and provided input on the final draft presented in this final report as Appendix B.

The state and national technical experts and leaders provided their input on the final report. The smaller technical expert group met a final time to make minor clarifications to the linkage methodology and security and access procedures. The larger group convened one final time to discuss participants' feedback on a draft of this report and to clarify any questions or concerns. Altogether, participants met nine times to provide input at multiple stages of the project.

Additional detail on the linkage methodology development process and a schedule of meetings for this project is provided in Appendix C.

CHAPTER 3. GUIDANCE FOR IMPLEMENTING THE DATA LINKAGE METHODOLOGY

This chapter summarizes the methods used by states to link mothers' Medicaid data to infant birth certificates (Table 3.1) and provides considerations and guiding principles for implementing a standardized data linkage methodology. Additional detail on states' current linkage activities and the specific definitions and uses of each variable in the linkage methodology are presented in Appendix D.

Table 3.1. Comparison of Current State Methodologies to Link Mothers' Medicaid Data to Infant Birth Certificates

	CA	CO	KY	NC	OH
Population and Data Definitions					
Link Out-of-State Births	No	No	Yes	Yes	Yes
Require Delivery Claim	Yes	Yes	No	Yes	No
Additional Data Linkages Informing Mother's Medicaid to Infant Birth Certificate Linkages					
Link Hospital Discharge Data to Mother's Medicaid Data	Yes	No	No	Yes	No
Link Mother and Infant Medicaid Data*	No	No	No	Yes	Yes
Mother's Medicaid and Infant Birth Certificate Variables Used in Matches					
Mother's SSN (D)	No	Yes	Yes	Yes	Yes
Mother's Name (P)	Yes	Yes	Yes	Yes	Yes
Father's Name (M)	No	Yes	Yes	No	Yes**
Mother's Date of Birth (DOB) (P)	Yes	Yes	Yes	Yes	Yes
Delivery Date (Mother's Medicaid) and Infant's DOB (infant birth certificate) (P)	Yes	Yes	Yes	Yes	Yes
Mother's Residence (e.g., address, zip code) (P)	Yes	No	Yes	Yes	Yes
Medicaid Client Number/Enrollment Number	No	Yes	Yes	No	Yes
Phone Number (M)	No	No	No	No	Yes
Mother's Race (M)	Yes	No	Yes	No	No
Delivery Method (M)	Yes	No	Yes	No	No
Linkage Software					
In-House Developed SAS Programs	Yes	Yes	Yes	Yes	Yes
Link Plus***	No	No	No	Yes	Yes
LINKS***	Yes	No	No	No	No
IBM InfoSphere***	No	No	Yes	No	No

NOTES: D = strongly suggested for use in a deterministic linkage step; P = strongly suggested for use in a probabilistic linkage step; M = strongly suggested for use in manual review only, where available.

* Mothers' and infants' Medicaid data are not directly linked; instead, they are each independently linked to birth certificates.

** Used only in manual review.

*** Probabilistic linkage software.

Populations and Datasets Linked

Considerations

All five states link mothers' Medicaid enrollment and claims data with infant birth certificate data, and all five states include data from all the ways in which Medicaid benefits are offered in a state (i.e., fee-for-service Medicaid, Medicaid managed care, and emergency Medicaid).

Some states also link additional data to the mothers' Medicaid data, such as infant Medicaid data (North Carolina and Ohio) or hospital discharge data (California and North Carolina), allowing that additional data to be used to link to birth certificate data (e.g., the infant's gender from infant Medicaid data or the delivery hospital ID number from hospital discharge data). Three states (Colorado, Kentucky, and Ohio) noted that, as

of this writing, they do not link mothers' Medicaid delivery claims to hospital discharge data, and they were unsure whether they would be able to do so. Three states (California, Colorado, and Kentucky) noted that the Medicaid family identifier was of poor quality within their states, making direct linkage of mother and infant Medicaid data difficult. Therefore, variables only available in infant Medicaid data or hospital discharge data would likely not be available to all states.

Three states (California, Colorado, and North Carolina) only linked mothers who had a Medicaid delivery claim, while the other two states (Kentucky and Ohio) did not require a Medicaid delivery claim for linkage. These two states still linked mothers that did not have a delivery claim if they had a prenatal and/or postpartum care claim and infant birth certificate data matched to a mother who was enrolled in Medicaid at the time of delivery. States noted that this approach allowed them to include Medicaid-covered mothers who may have given birth out of state or at home, mothers who were retrospectively enrolled in Medicaid to cover the delivery, and mothers who were otherwise missing a delivery claim.

All states participate in state exchanges of birth certificate data,[7] and three states (Kentucky, North Carolina, and Ohio) indicated that they currently link births that have occurred out of state (e.g., if the mother was traveling or gave birth in a neighboring state), provided that the mother was enrolled in Medicaid in the linking state at the time of delivery. States also noted that there was typically a delay of a few months up to one year in receiving these out-of-state data. There is variation by state in the number and percentage of out-of-state births: States having population centers close to other states have more out-of-state births, so the impact of excluding out-of-state births varies by state.

Guiding Principles

Based on input from the project participants, the following principles emerged:
- Due to feasibility considerations, only variables from mothers'' Medicaid and infant birth certificate data be used in this linkage. Mother-infant Medicaid data linkage and Medicaid-hospital discharge data linkage might be considered in the future.
- Enrollment, claims, and encounter data from all types of Medicaid programs (i.e., fee-for-service, managed care, emergency Medicaid) be included for each state contributing data to the multistate linked database, and that all types of Medicaid claims (e.g., inpatient, office visits, and pharmacy) be linked.
- Birth certificates be linked to mothers' Medicaid data even when there is no Medicaid delivery claim, provided that mothers had prenatal and postpartum Medicaid claims and other linkage criteria are met (described later in this chapter); prenatal and postpartum claims definitions be shared by states through a GitHub (an online site for collaborating on code development) repository.
- While out-of-state births to state residents may be included in linkages, in the interest of providing timely data, only data available at the time of linkage must be included (see the "Linkage Timeline and Updating Linkages" section below for additional details).

Variables Used in Linkages

Considerations

Four of the states directly use SSNs in linking Medicaid and birth certificate data, while the fifth (California) uses SSNs only to link Medicaid and hospital discharge data, because SSNs from the birth certificate data are not provided to the department performing the linkage. States generally noted that SSNs were a critical component of the matching process, while Ohio noted that their match rate improved by approximately

10 percentage points when the mother's SSN was added to its matching protocol. Participants agreed that a history of collaboration among agencies and data security measures may overcome challenges posed by sharing SSN data across state agencies. Also, while participants noted the importance of SSNs, they also noted that they were not sufficient for linkage on their own because SSN data are often missing or incorrect.

States also noted the importance of allowing for *fuzzy* matches (i.e., matches that are similar but not exact), particularly on such variables as name and DOB. Using a method like Soundex (an algorithm that allows for matching on similar phonetic spellings) was mentioned as being important considering that names and other information may be written down by someone relaying information verbally. Similarly, states noted that allowing a window around DOB is important because sometimes the date that the infant is enrolled in Medicaid is incorrectly entered as the infant's DOB. For example, Ohio found that allowing a two-week window addressed this issue. Overall, allowing for fuzzy matches greatly improved state match rates.

States found that some variables were less useful in the deterministic and probabilistic linkage steps. For example, states noted that the father's name and the mother's middle name generally did not provide a lot of additional matches and were not worth the additional computation time. However, these variables were still occasionally useful in *manual review*, a process in which a reviewer holistically examines a subset of matches to determine whether they are likely true or false. States' manual review processes are described in greater detail in Appendix D.

Guiding Principles

Based on input from the project participants, the following principles emerged:
- The mother's SSN be used in deterministic linkages of Medicaid and birth certificate data.
- The mother's first and last names, residential 5-digit zip code, and DOB and the infant's DOB (from birth certificate data) and delivery date (from Medicaid data) be used in probabilistic linkage steps.
- Soundex be used to allow for inexact name matches, and that a window of two weeks around the infant's DOB be allowed.
- Other variables, such as the father's name, the mother's middle name, the mother's premarital name, the mother's race, and the delivery method be used only in a manual review step.

Software, Scoring/Weighting Methods, and Validation Methods

Considerations

The type of scoring methodology used by a state is closely related to a state's matching software and matching mechanism. There are two main purposes for assigning match scores. The first is to differentiate between duplicate matches when the match is supposed to be one to one. The second is to define the threshold of whether two records should be considered a match. The ideal scoring methodology should be relatively easy to implement, provide sufficient information for de-duplicating matches, have low type I (incorrectly excluding a true match) and type II (incorrectly including a false match) error rates, and be easy to standardize across different states and settings.

Barriers to implementation include the cost of software, programming skill needed, and the burden and reliability of the manual review processes. For instance, the LINKS SAS package, IBM InfoSphere MDM software (commercial), and Link Plus (government) software all automatically generate scores for probabilistic matching. However, commercial software might allow for more customization and might be more computationally efficient than Link Plus. Some participants also noted that another available probabilistic matching software (MatchPro)[8] may have an advantage over Link Plus, although no participating states were

currently using this other option, which made feasibility of implementation unknown. Link Plus provides both the software and documentation at no cost. A challenge of Link Plus is that it can require a lot of virtual memory and processing power. California, which does not currently use Link Plus, noted concerns about the software being able to handle large numbers of observations.

Steps that can be taken to make the process more efficient include performing a deterministic match step first and using a blocking step to limit the pairs of potential records compared for matching. However, participants noted that taking the step to first match deterministically and then setting those matches aside may limit a feature of probabilistic matching software that assigns scores based on the distribution of variables within the population. For example, the probabilistic matching algorithm assigns higher scores for matching on an unusual last name than for matching on a common one, but it would be referencing an incomplete distribution of names in the population if the deterministic matches were first removed.

While there were varying perspectives on which software to use, participants agreed that standardization of method would be easier if all participants used the same software. Link Plus was the most commonly endorsed software given that: (1) a plurality of states already use it; (2) it is free and fairly straightforward; (3) it is customizable; (4) it includes built-in features to allow for fuzzy matching, scoring probabilistic matches, and conducting manual review; and (5) it was developed by the CDC, potentially making training and support more feasible. States that did not currently use Link Plus indicated a willingness to learn, particularly if external support could be provided.

Most states also include some type of manual review process that they generally agreed was time-intensive but essential for maintaining match quality.

Guiding Principles

Based on input from the project participants, the following principles emerged:

- States use Link Plus to make implementation of standardized linkage methods more straightforward.
- An expert group assists states with using Link Plus and other technical issues that may arise during linkage.
- States begin with deterministic matching on the mother's SSN to reduce the total number of matches and, thus, the computation time needed for probabilistic matching.
- States next probabilistically match on the mother's name, the mother's and infant's DOB, and the mother's zip code, using Link Plus features that allow for fuzzy matching. The probabilistic matching process using Link Plus will generate match scores. Testing this probabilistic matching step with and without first removing the deterministic matches will help determine whether the marginal improvement in match rate is worth the marginal increase in computation time.
- States iteratively test state-specific score thresholds for automatic acceptance and for manual review. These thresholds should provide a comparable level of confidence in matches across states and balance the sensitivity and specificity of the matches with the time needed to conduct manual reviews.
- A subset of manual reviews be independently conducted by two individuals to establish interrater reliability. Depending on the results of the interrater reliability, states can determine the appropriate percentage of matches that should undergo review by more than one person and whether additional reviewers are needed.

- Match rates be compared by race/ethnicity to determine if certain groups have lower match rates and could thus be underrepresented in the linked data. If this is the case, steps be taken to improve the match rate in these populations (e.g., manually reviewing at a lower threshold score).
- In the case of duplicate matches when there should be a single match (e.g., one birth record matching to two mothers' Medicaid enrollment records), the match with the higher score be retained. Special attention must be paid to multiple births (e.g., twins) and births with short inter-pregnancy intervals so that they can be differentiated from duplicate matches with similar first names and identical last names. Participants have discussed sharing SAS programs that they have developed to address these issues through a GitHub repository.

Linkage Timeline and Updating Linkages

Considerations

Key considerations in determining the linkage timeline include a balance between ensuring the receipt of high-quality input data files and producing timely and accurate matched data. States differ in the frequency and promptness with which Medicaid claims and vital statistics data are reported and validated. Having a longer runout period may help ensure more-complete and more-accurate data are used for matching, but this may delay the availability of timely matched data.

Link Plus' scoring algorithm is also affected by the complete set of records in the input files. Because of this, adding additional records and rematching with data that include previously matched records may change the match score and matching status of previously matched pairs (i.e., matching the first six months of data versus matching a whole year's worth of data might produce different match results for records generated in the first six months). This type of rematching could marginally improve accuracy. However, revising previously matched pairs makes the match process computationally burdensome and poses challenges for the comparability of research results that use matched data from different points in time.

Guiding Principles

Based on input from the project participants, the following principles emerged:
- States perform linkage of calendar year data annually to contribute to the standard core dataset; states might consider performing linkages more frequently for internal purposes.
- A runout of approximately six months following the end of the included claims be used to balance timeliness of matching with data completeness (although this could be reassessed during implementation). Rematching might be unnecessary with a sufficiently long runout period. Thus, previously unmatched records, but not necessarily previously matched records, could be included in subsequent matches.

Implementation and Next Steps

In building on the findings of this project, states that agree to implement the principles detailed in this chapter could do so to create a database of linked Medicaid and birth certificate data for research. The implementation process could be iterative, with states sharing additional details on data cleaning and standardization, testing different score cutoffs to generate comparable match quality (in terms of confidence level in matches) across the states' datasets, and developing procedures for the manual validation of matches. In the future, participating states could contribute deidentified data to a centralized, multistate linked database (described in Chapter 4) that could be used by researchers and policy analysts to generate stronger, more actionable cross-state insights into the impact of policies on maternal and child health. The methodology could also be

used in the future by additional states, and the database could be expanded to include data from more states. States also indicated that an expert group on data linkage would be valuable to assist states in implementing the guiding principles in this report.

CHAPTER 4. DATA SECURITY AND ACCESS

Models of Data Access for External Researchers

Participating state and national experts reviewed different potential models for securing and accessing linked Medicaid-birth certificate data. These models reflect existing ways that other types of multistate health data are centralized and shared with researchers. Participants noted that both restricted use access models for claim-level data (i.e., requiring a formal application and review process) and public use access models for aggregated data (i.e., data available online with no application required) could be valuable to researchers. Table 4.1 presents the potential variables to be included in restricted and public use data (described in detail later in the chapter).

Table 4.1. Potential Variables and Level of Detail Included in Restricted and Public Use Data

Variable	Restricted Use Data	Public Use Data
Mother's Medicaid Record Number (M,B)	Encrypted identifier (allows the following of mothers across multiple Medicaid-covered deliveries)	No
Geographic identifier (M,B)	3-digit zip code	State and county
Social vulnerability index score (M, B)*	Continuous score	Categorical score
Mother's Age (M,B)	Date of birth	Age category
Infant's DOB (M,B)	Birth month and year	Birth month and year
Mother's Diagnosis Codes (M)	7-digit ICD-10 code	3-digit ICD-10 code
Dates of Service (M)	Days before/after delivery	Trimester
Risk Factors During Pregnancy; e.g., diabetes, hypertension, cigarettes smoked during pregnancy (B)	Include all	Include all
Mother's Demographic Information; e.g., race/ethnicity (M,B) and education (B)	Include all	Include all
Father's Demographic Information; e.g., race/ethnicity and education (B)	Include all	Include all
Date of Last Live Birth or Other Pregnancy Outcome (M,B)	Months	Months
Maternal Morbidity; e.g., maternal transfusion (B)	Include all	Include all
Other Delivery Characteristics; e.g., multiple births, birth order (B)	Include all	Multiple births
Infant Characteristics; e.g., APGAR score, NICU admission (B)	Include all	Include all
Method of Delivery (B, M)	Include	Include
Provider Type (M)	Rendering provider taxonomy code	No
Percentage of Federal Poverty Level	Percentage (where available)	Percentage category
Medicaid Eligibility Reason	Categorical, by month	No

NOTE: APGAR = Appearance Pulse Grimace Activity Respiration; B = variable found in birth certificate data; ICD-10 = International Classification of Disease, Tenth Revision; M = variable found in Medicaid claims; NICU = neonatal intensive care unit.

*Not in dataset, but this score can be merged in at the census tract level prior to deidentification and submission.

Restricted Access Models

One model discussed by the group was the Federal Statistical Research Data Centers (FSRDC),[9] which house confidential data from such agencies as the U.S. Census Bureau and the Substance Abuse and Mental Health Services Administration. Researchers using data through the FSRDC must have Special Sworn Status and must analyze data on-site at one of 33 FSRDC throughout the country. Group participants noted the security of this data model, but they added that using FSRDC presents access barriers for researchers because not all research organizations are in close proximity to FSRDC or have personnel with Special Sworn Status.

Another model discussed by the group was the Research Data Assistance Center (ResDAC),[10] which manages researcher access to Medicare and Medicaid data extracted from the T-MSIS. Researchers can access all years of T-MSIS data that are available through a virtual research data center or receive a defined data extract per request. All 50 states, the District of Columbia, and two U.S. territories currently submit data to T-MSIS, so states are familiar with ResDAC's data security and access processes. Researchers interested in Medicaid policy are also familiar with ResDAC, and multiple participants endorsed using this model. Adding restricted state Medicaid data linked with birth certificates will enrich the portfolio of data made available to researchers studying Medicaid covered health services and subsequent health outcomes.

One participant noted the National Child Abuse and Neglect Data System (NCANDS) as another potential model.[11] Under this system, states voluntarily submit data on reports of child abuse and neglect to the data system each year. This system allows electronic delivery of restricted use case-level data or public use aggregated data. NCANDS is managed by the National Data Archive on Child Abuse and Neglect (NDACAN) at Cornell University through a contract with HHS. NDACAN works closely with data submitters and users and provides technical support and frequent activities for education and engagement. Currently all five participating states submit data to NCANDS.

States that make their linked Medicaid-birth certificate data available to outside researchers use different methods to do so, including allowing researchers virtual access to data or through a secure file transfer, provided that all other data security standards are met.

Public Access Data Model

For public use data, the group discussed the CDC National Center for Health Statistics' (NCHS') Wide-ranging Online Data for Epidemiologic Research (WONDER) natality data.[12] Researchers can query WONDER to get aggregated reports on customized crosstabulations and subpopulations, with suppression of cell sizes <10 or that could lead to the identification of individuals. Participants from California noted that their state required suppression of results between 1 and 10. Participants suggested a guiding principle to use the most conservative approach (i.e., suppressing cell sizes <11). Through a public access data model, no formal application process would be required, and no individual-level data would be available.

Guiding Principles

Based on input from the project participants, the following principles emerged:
- Participating states submit restricted use individual- and claim-level linked data to a central repository, such as ResDAC, that uses a similar overall model of access and security as that used for T-MSIS data.
- Managers of the central repository frequently engage with states and researchers on issues of security and access in order to identify gaps and solutions.
- Participating states make available aggregated public use data through an online query system, such as CDC WONDER, suppressing cell sizes <11, at a minimum.

Characteristics of Restricted and Public Use Data

After discussing the overall models of how data could be stored and accessed, group participants discussed specific characteristics of the restricted and public use datasets.

Participants noted the following considerations:

- Importance of the right data time frame—such as having data on the preconception period available (where possible) and including at least one year of data after delivery (where possible)— would allow longitudinal research on maternal health or study of the impact of such policies as states' recent postpartum Medicaid expansion.
- Inclusion of an encrypted Medicaid record number for the mother in the restricted use dataset to allow researchers to follow mothers longitudinally across multiple annual files and pregnancies.
- Submission of claims data begin in 2016, because this was the first full year that all states had transitioned to T-MSIS, which included some new standardization of Medicaid data. This was also the first year a new standardized birth certificate was used by all states and the first full year in which the ICD-10 diagnosis codes were used.
- Careful consideration of the variables and level of detail that should be included in restricted and public use data. Participants noted that their initial viewpoints were based on their experiences making data available for researchers but that a full review of reidentification risk for both restricted and public use files would still be necessary and that certain values of variables might need to be suppressed (e.g., public use files might include only the ten most common ICD-10 diagnoses and group the rest as "other"). They also noted that each state might want to do its own initial assessment of reidentification risk the first time data were submitted, but that if the central data repository (e.g., ResDAC) was able to meet each state's standards, it could perform this action moving forward. For public use files, at a minimum, frequencies of variable combinations <11 would be suppressed in query results. Different groupings of values would need to be explored and assessed to ensure that individuals could not be identified through combinations of aggregated data reports.
- Allowance of some variation in the level of detail provided for a specific variable. For example, in the restricted use data file, it could be possible to use more-granular 5-digit zip codes in urban areas but only provide county geographic identifiers in more-rural areas, where the risk of reidentification is higher. Participants also noted that merging the Social Vulnerability Index (SVI) score,[13] based on census tract data and including the SVI score instead of the identified census tract could be a way to prevent patient reidentification while including details about important neighborhood characteristics.

Guiding Principles

Based on input from the project participants, the following principles emerged:
- Participating states submit linked claim-level data beginning with 2016 data.
- Participating states include at least one year of claims data before and after delivery.
- An assessment of reidentification risk for both restricted and public use datasets be conducted in a timely manner agreed on by all participating states. This may involve an assessment by the central data repository (e.g., ResDAC) and/or an assessment by each participating state.

Application Process and Data Use Agreement for Restricted Data

Currently, all participating states have processes in place to make their data available to external researchers in some form, although not all states provide individual-level linked Medicaid-birth certificate data.

States that make these linked data available to researchers typically require Institutional Review Board (IRB) approval and a DUA for restricted use data, even when direct identifiers, such as names, have been removed. States also require a description of the research or a business agreement. Some states also have training

requirements for researchers, such as the Collaborative Institutional Training Initiative human subjects training.

In addition, states have requirements regarding the environment in which the data need to be stored, the signatures (i.e., required from the data custodian and an authorized representative of the requestor), the data elements shared, the length of data access, the dissemination process, and the review of research output by the state. States pointed to their requirements related to the deletion of data at the end of the research project but with consideration for the research peer-review and publication timeline.

Although there are many similarities across the five participating states, there are also differences in their processes. To consolidate these processes, participants discussed whether it would be possible to have a single IRB review process for accessing multistate data. Two states raised concerns about the possible centralization of the IRB process and suggested that state leaders would not be willing to cede their authority to approve individual research proposals to a central body. However, participants noted that if the central data repository met states' security and access standards, a centralized IRB process could be considered. Other participants emphasized the value of only requiring a single IRB approval and DUA to increase the likelihood that researchers use multistate linked data. They also noted that using a central data repository with which states were familiar, such as ResDAC, could increase confidence in the process. States emphasized that any shared data would need to be backed up with language on penalties for data release and rules on what researchers must do before submitting derivative work to journals to assure state leaders that data will be used properly. One additional common agreement among participating states was related to the standardization of and signoff on the process for data destruction by external researchers once the projects were complete.

Guiding Principles

Based on input from the project participants, the following principles emerged:
- To access and use the restricted linked data from the states hosted in the central repository, a research proposal including the variables of interest, IRB review, and completion of a DUA be required. The current ResDAC review process used for accessing T-MSIS data might be applied. This review process is focused on data security and should not compromise the independent integrity of research projects.
- Legal review of state DUAs and access procedures be conducted to help determine a single DUA template, research proposal format, and review process that will be used to access multistate data while meeting each state's legal requirements.
- The DUA include a requirement for data to be destroyed at the end of the project, although extensions may be granted to accommodate the publication process.
- States be informed about the dissemination products and acknowledged as providing the data and performing the linkages to make the data, including patient-reported outcomes, available for research.

Implementation and Next Steps

To implement sharing of the linked data underlying the restricted and public use datasets, participants emphasized that they would need buy-in and approval from top state leaders at departments of health care services or public health. Participants noted that they would also need legal review and approval on several of the topics discussed, including the processes for external researchers to access data, publish results, and destroy data once projects were completed. Current participating state processes for approving data access to external researchers varied greatly, so substantial effort would be needed to harmonize these processes across all five states. Generally, states suggested that the more layers of security to access a dataset, the likelier state

leaders were to approve the process (i.e., a model like that used for ResDAC would be more likely to get buy-in).

To help get buy-in, states proposed that emphasizing the use cases and the concrete insights that could be gained (i.e., the ability to assess the impact of specific policies) would help encourage state leaders to overcome the challenges associated with sharing data. Specific use cases of linked data proposed by the technical experts and state and national leaders are discussed in the next chapter. States also suggested limits on the amount and level of detail of data available in datasets to balance their usefulness for research with risk of reidentification.

CHAPTER 5. RESEARCH AGENDA ON MATERNAL HEALTH

Motivation

The third goal of this project was to develop a research agenda including several high-level topics and research questions for the use of the linked Medicaid-birth certificate data, as well as a research protocol for one of these research questions. This effort included identifying pressing maternal and child health research topics that could be answered using the linked data and that would benefit from comparison across several states. This research agenda and protocol could serve as a demonstration of OS-PCORTF efforts to improve maternal health through expanding data capacity and research.

Approach

Prior to the meeting discussing the national research agenda, participants in the large group proposed potential topics and research questions. In addition, representatives from ASPE and RAND proposed other topics and research questions based on a previous literature review conducted of studies using linked data.[2] Participants then decided on the highest priority research question for inclusion in the national research agenda.

Identified Topics for the Research Agenda

Prior to the above-mentioned meeting, participants noted general interest in topics related to the postpartum Medicaid extension, telehealth use during pregnancy, and substance use during pregnancy. The topic of Medicaid expansion under the Affordable Care Act (ACA) was also identified as a broad topic of interest, using linked data in the previous literature review.[2] The topic of access to reproductive health services was also brought up by participants during the meeting. Participants discussed these five broad research topics, then proposed and refined one or more research questions within each topic. The large group determined what they viewed as the top research priority of the five presented for development into a research protocol. Most participants selected *"How did postpartum Medicaid extension under the American Rescue Plan Act impact maternal postpartum service use and health outcomes?"* Table 5.1 below presents the topics, research questions, and key discussion points about the feasibility of answering these questions with the linked Medicaid-birth certificate data and the potential benefit to policymakers of answering these questions.

Table 5.1. Key Considerations for Identified Research Topics on Maternal Health Using Linked Multistate Medicaid-Birth Certificate Data

Topic	Questions Related to Identified Topics	Considerations
Postpartum Medicaid extension	**How did postpartum Medicaid extension under the American Rescue Plan Act impact maternal postpartum service use and health outcomes?***	**Participants noted that this was a topic of great interest to state Medicaid and public health leadership. While they were interested in maternal and infant health outcomes, they noted that changes in service use were more likely to be observed in the short term, while changes to maternal and infant health outcomes may take longer to observe.**
Telehealth	How often was telehealth used for prenatal and postpartum care before, during, and after the COVID-19 public health emergency? Does telehealth use differ between normal and complicated pregnancies?	Participants noted interest in understanding whether telehealth was used for normal or complicated pregnancies and how utilization might impact patient outcomes. They also noted that telehealth was rarely used before the COVID-19 public health emergency.
Substance use	How have laws prioritizing drug treatment access for pregnant individuals impacted maternal treatment for substance use, live births, and infant health outcomes?	No participants indicated interest for this research question. However, they did note an interest in substance use during pregnancy more broadly and suggested examining rates of substance use treatment as an outcome of interest in broader analyses.
ACA Medicaid expansion	How did Medicaid expansion under the ACA impact Medicaid enrollment and care utilization among pregnant individuals?	No participants indicated interest in this research question. We note that the previous report identified several published studies examining this topic, so it may not represent a large gap in the literature.
Reproductive health services	How did changes in access to reproductive health services impact maternal health and birth outcomes?	Participants noted a broad interest in understanding how changes to access to reproductive health services impacted maternal health and birth outcomes, and they mentioned such topics as access to doulas and home care, abortion, and contraception.

NOTE: * Participants selected this question as their top research priority.

Elements of Research Protocol

Once the highest priority research question had been identified, the technical expert group met to discuss specific aspects of a research protocol to address this question, such as definition of the study population, outcome and covariate measurement, and approaches to analysis. Each of these aspects of the research protocol were considered in developing a preliminary set of options and considerations that the participants reviewed (e.g., measures of service utilization to consider for outcomes, variables that could be included as control variables in the analyses, and potential analytic methods) prior to finalizing the sample research protocol (Appendix B). This protocol serves as a motivating use case for linking data across multiple states and could be used in the future for a study conducted with linked data by state and federal government researchers, as well as external researchers, as appropriate.

Considerations

Participants noted that the use of a consistent data linkage methodology and availability of variables will be key to implementing this research protocol. Medicaid data must include enrollment, eligibility, and encounter data. Additional variables include type of Medicaid coverage (i.e., traditional, pregnancy and delivery, or emergency), number of months enrolled pre- and post-delivery, and family income/size. Birth certificate data must include the mother's age, race/ethnicity, and education; presence of certain maternal complications or

infections; abnormal conditions of the newborn; infant APGAR score, and the interpregnancy interval.

Guiding Principles

With regard to the research protocol for the highest priority research topic, the following principles emerged:
- The first question addressed with the linked data set be: "How did postpartum Medicaid extension under the American Rescue Plan Act impact maternal postpartum service use and health outcomes?"
- This study to include the population and key variables of interest, specified in Appendix B.
- The protocol be refined based on feedback from other stakeholders and experts invited to participate in the proposed implementation phase.

CHAPTER 6. CONCLUSION

Over the course of nine meetings and multiple rounds of written feedback from technical experts and state and national leaders in Medicaid, vital statistics, and maternal health, a set of guiding principles for linking state Medicaid and birth certificate data and making these linked data available to researchers was collaboratively developed.

Despite the small number of participating states (five) and the short project life cycle, there was significant interest among participants in sharing knowledge and collaborating on data linkage efforts. The guiding principles outlined in this report are the product of this joint and long-standing interest to explore viable methods for making multistate linked Medicaid-birth certificate data available to maternal health researchers.

States will have the opportunity to implement the guiding principles described in this report and link mothers' Medicaid data with infant birth certificates using standardized methods. It is important to note that, while the guiding principles described in this report are detailed and cover many aspects of data linkage methodology and data access and security, implementation of these principles would require states to iteratively collaborate on additional details.

In the future, this project could be expanded to include more states and additional sources of data that are already linked by some states, such as hospital discharge data or electronic health record data, which would allow more-complex longitudinal research studies on maternal health.

Overall, this project had an important contribution to exploring methods for linking state Medicaid and birth certificate data. Implementing these guiding principles will complete the journey toward achieving the goal of creating a centralized, multistate database of state-contributed linked Medicaid-birth certificate data with common data elements from a commonly defined population. Making these multistate linked data available to researchers with expertise in PCOR and comparative effectiveness research, or generalist researchers, as well as epidemiologists and program analysts will result in more longitudinal and complex analyses on various topics, including service utilization and health outcomes, to inform policy on maternal health.

APPENDIX A. STATE AND NATIONAL EXPERT PARTICIPANTS

Table A.1. State and National Expert Participants

State or Agency	Name(s), Title(s), and Affiliation(s)	Project Role
California	Daniel Jordan, Chief of Data Fulfillment Branch, California Department of Health Care Services	Data Access and Security
California	Dr. Muree Larson-Bright, Chief, Data Science Branch, California Department of Health Care Services; Dr. Regan Foust, Executive Director and Senior Research Scientist, Children's Data Network at the University of Southern California	Data Linkage Methods
Colorado	Alexandra Denman, Data Analyst, Colorado Department of Health Care Policy and Financing; Kirk Bol, Manager of Vital Statistics Program, Colorado Department of Public Health and Environment	Data Linkage Methods, Data Access and Security
Kentucky	Dr. Matthew Walton, Researcher, University of Kentucky, Office of Health Data and Analytics, Kentucky Cabinet for Health and Family Services	Data Access and Security
Kentucky	Angela Taylor, Biomed Informatics Data Architect, University of Kentucky, Office of Data Analytics, Kentucky Cabinet for Health and Family Services; Lynn Ng, Database Analyst, Office of Data Analytics, Kentucky Cabinet for Health and Family Services	Data Linkage Methods
Kentucky	Tracey Jewell, Senior Epidemiologist and Manager, Division of Maternal and Child Health, Kentucky Department of Public Health; Dr. Henrietta Bada Director, Division of Maternal and Child Health, Kentucky Department for Public Health, Cabinet for Health and Family Services, Professor of Pediatrics, College of Medicine, University of Kentucky	State Leader
North Carolina	Dr. Chandrika Rao, Interim Director, State Center for Health Statistics; Director, Central Cancer Registry, Division of Public Health, North Carolina Department of Health and Human Services	Data Access and Security
North Carolina	Robert Lee, Statistical Services Branch Manager, State Center for Health Statistics/North Carolina Department of Health and Human Services	Data Linkage Methods
North Carolina	Kathleen Jones-Vessey, Epidemiologist, North Carolina Division of Public Health, Women's and Children's Health Section, North Carolina Title V Office/North Carolina Department of Health and Human Services	State Leader
Ohio	Lorin Ranbom, Director, Ohio Colleges of Medicine Government Resource Center	Data Access and Security
Ohio	Dr. Michael Nau, Assistant Director of Applied Research & Analysis, Ohio Colleges of Medicine Government Resource Center; Habteab Gebreab, Medicaid Health Systems Administrator, Ohio Department of Medicaid	Data Linkage Methods
Centers for Disease Control and Prevention (CDC)	Cordell Golden, Chief, Data Linkage Methodology and Analysis Branch, National Center for Health Statistics	Data Linkage Methods, Data Access and Security
Office of the Assistant Secretary for Planning and Evaluation (ASPE)	Dr. Nancy DeLew, Associate Deputy Assistant Secretary, Office of Health Policy	National Leader
Centers for Disease Control and Prevention (CDC)	Dr. Irma Arispe, Director, Division of Analysis and Epidemiology, National Center for Health Statistics	National Leader
Centers for Medicare & Medicaid Services (CMS)	Dr. Lindsey Wilde, Deputy Director for the Division of Business and Data Analysis	National Leader
Health Resources & Services Administration (HRSA)	Dr. Catherine Vladutiu, Senior Epidemiologist, Maternal and Child Health Bureau	National Leader
National Institutes of Health (NIH)	Dr. Rebecca Rosen, Director, Office of Data Science and Sharing; Dr. Alison Cernich, Deputy Director, Eunice Kennedy Shriver National Institute of Child Health and Human Development	National Leader

APPENDIX B. RESEARCH PROTOCOL

Background

As part of the 2021 American Rescue Plan Act,[14] states had a new option to extend Medicaid coverage to 12 months postpartum. Prior to this extension, pregnant individuals typically lost Medicaid coverage 60 days after delivery if they were not otherwise eligible for Medicaid based on their income or other qualifying reasons. Extending Medicaid coverage to a full year postpartum is significant because mortality and other severe complications stemming from pregnancy and delivery often occur well beyond 60 days postpartum.[15]

As of June 22, 2023, 35 states and the District of Columbia have implemented this extension, another nine states are planning to implement this extension, three states are planning more-limited extensions (i.e., less than 12 months), and three states have no plans to extend postpartum Medicaid coverage.[16]

All five of the states participating in this project have implemented the 12-month postpartum Medicaid extension, although there were some slight state-specific variations in the timing and implementation. California extended Medicaid to 12-months postpartum for Medicaid-enrolled mothers with a diagnosed mental health condition, effective August 1, 2020, and then expanded such coverage for all Medicaid-enrolled mothers, effective April 1, 2022. Kentucky, Ohio, and North Carolina all extended Medicaid coverage to 12 months postpartum, effective April 1, 2022, while Colorado's postpartum Medicaid extension had an effective date of July 1, 2022.

It is important to note that these policy changes also occurred during the COVID-19 public health emergency. During this time, the Families First Coronavirus Response Act (FFCRA) required states to maintain enrollment of most Medicaid enrollees,[17] meaning that many individuals who typically would have lost Medicaid coverage 60 days postpartum did not. Congress removed the FFCRA continuous eligibility condition and, as of March 2023, required states to begin redetermining Medicaid eligibility and terminate enrollment for individuals who were no longer eligible.[18] The primary objective of this proposed research study is to determine how extending postpartum Medicaid coverage to one year impacted the access and utilization of maternal health care. Extending postpartum Medicaid coverage may have several benefits to mothers. First, the changes in postpartum Medicaid policies in states will likely result in individuals staying enrolled in Medicaid longer and retaining access to needed medical care in the postpartum period. However, changes in postpartum length of Medicaid coverage immediately following the American Rescue Plan Act will likely be modest due to the FFCRA. Postpartum Medicaid extension may also lead to modest increases in Medicaid-covered services used, such as postpartum visits, mental health visits, substance use treatment, and long-acting contraceptive (e.g., intrauterine device, injectable contraceptives) in the year following delivery. This primary objective includes two aims:

- **Aim 1 (Primary):** Examine how extending postpartum Medicaid coverage to one year impacted maternal Medicaid enrollment during the COVID-19 public health emergency.
- **Aim 2 (Primary):** Examine how extending postpartum Medicaid coverage to one year impacted Medicaid-covered service utilization among mothers with a Medicaid-covered delivery during the COVID-19 public health emergency.

The secondary objective is to examine longer-term impacts of postpartum Medicaid extension after the end of the COVID-19 public health emergency. One outcome of interest for this objective is the *interpregnancy interval* (the number of months between pregnancies) following index pregnancy. Knowing this is important because short interpregnancy intervals (less than 18 months) are associated with an increased risk of infant

mortality, low birth weight, and pre-term delivery.[19] Individuals with extended postpartum Medicaid coverage may more easily obtain highly effective forms of birth control and subsequently increase their interpregnancy intervals.

Given the policy context of FFRCA during the initial implementation of postpartum Medicaid extension, it is also of interest whether postpartum Medicaid extension policies helped prevent the loss of Medicaid coverage following the end of the COVID-19 public health emergency. Thus, the secondary objective includes two aims:

- **Aim 3 (Secondary):** Examine how extending postpartum Medicaid coverage to one year impacted interpregnancy intervals among mothers with a Medicaid-covered delivery during the COVID-19 public health emergency.
- **Aim 4 (Secondary):** Examine whether the impact of extending postpartum Medicaid coverage during the COVID-10 public health emergency on maternal Medicaid enrollment was sustained after the end of the public health emergency and how the results may differ by state.

Methods

Data and Study Population

This example research protocol uses the proposed multistate restricted-use dataset with core data elements described in Chapter 4. This dataset includes birth certificate data from all Medicaid-covered deliveries occurring between January 1, 2017, and December 31, 2023, linked with available mothers' Medicaid claims and enrollment data up to one year before and after delivery.

Outcome Measures

The following outcome measures from mothers' Medicaid claims and enrollment data in the year following delivery will be included: number of postpartum months of maternal Medicaid enrollment, number of postpartum visits, receipt of mental health services, and receipt of substance use treatment services, identified using billing codes specified by the Healthcare Effectiveness Data and Information Set (HEDIS) definitions for postpartum visits, any mental health services, and alcohol or other drug dependence treatment,[20] respectively (separately by whether or not a telehealth modifier was present),[21] and receipt of long-active contraceptives identified using billing codes.[22] These outcomes are listed in Table B.1, along with measures of other characteristics that may be used as covariates or subgroup definitions.

Control and Subpopulation Variables

Several other variables will be included in the analyses, either as controls or to help stratify by subpopulation of interest. These variables are drawn from birth certificates, Medicaid claims, or both. The technical expert group provided input to determine whether variables should be drawn from Medicaid data, birth certificate data, or both in cases where they are available in both datasets. Dates of pregnancy will be defined using the obstetric estimation of gestation from the birth certificate. All birth certificate data elements are defined using the U.S. standard certificate of live birth.[23] These variables are summarized in Table B.1.

Table B.1. Variables Used in Analyses

Variable	Data Source	Aim(s)/Variable Type
Number of postpartum months of maternal Medicaid enrollment in the year after delivery	Medicaid enrollment	Aims 1, 4/Outcome
Number of postpartum visits in the year after delivery (in person and telehealth)*	Medicaid claims	Aim 2/Outcome
Number of visits for mental health services (in person and telehealth) in the year after delivery	Medicaid claims	Aim 2/Outcome
Use of substance use treatment services (in person and telehealth)	Medicaid claims	Aim 2/Outcome
Long-acting contraceptive use in the year after delivery	Medicaid claims	Aim 2/Outcome
Interpregnancy interval (time in months from index pregnancy to subsequent pregnancy)	Birth certificate (if present), otherwise Medicaid claims	Aim 3/Outcome
Mother's state of residence at time of delivery	State provided	Covariate/subgroup
Mother's age	Birth certificate (if present), otherwise Medicaid enrollment	Covariate/subgroup
Mother's race/ethnicity	Birth certificate (if present), otherwise Medicaid enrollment	Covariate/subgroup
Mother's education level	Birth certificate	Covariate/subgroup
Neighborhood socioeconomic status (SVI decile at date of delivery)	SVI data merged with maternal residence zip code from Medicaid enrollment file (if present) or birth certificate	Covariate/subgroup
Presence of maternal complications during pregnancy (diabetes, hypertension)	Medicaid claims and birth certificate (if complication is present on either)	Covariate/subgroup
Presence of maternal infection during pregnancy (gonorrhea, syphilis, chlamydia, Hepatitis B, Hepatitis C)	Medicaid claims and birth certificate (if complication is present on either)	Covariate/subgroup
Presence of mental health diagnosis during pregnancy	Medicaid claims	Covariate/subgroup
Presence of maternal complications during delivery (third- or fourth-degree perineal laceration, ruptured uterus, unplanned hysterectomy, admission to intensive care unit, unplanned operating room procedure following delivery)	Medicaid claims and birth certificate (if complication is present on either)	Covariate/subgroup
Abnormal conditions of the newborn (e.g., NICU admission, seizure or serious neurologic disfunction)	Birth certificate	Covariate/subgroup
APGAR score	Birth certificate	Covariate/subgroup

NOTE: *The ability to study this outcome will depend on how accurately this information is recorded in the data.

Primary Analyses

First, the distribution of the outcome variables (frequency or mean and standard deviation) will be examined by state and quarter to understand trends and possible assumption violations of future analyses.

Next, bivariate analyses will be conducted to examine the relationship between each outcome variable and key characteristics of interest: race/ethnicity, mother's education level, mother's age group (<15, 15–49 by 5-year age groups, >50), SVI decile, presence of maternal complications during pregnancy, presence of a maternal mental health diagnosis during pregnancy, and presence of infant health complications during pregnancy. This will allow us to understand how Medicaid coverage length and use of postpartum Medicaid services differed by mother and infant characteristics.

Finally, multivariable analyses will examine the potential impact of postpartum extension policies. The primary analyses will be difference-in-differences analyses separately comparing each outcome in Aims 1 and 2 among individuals with Medicaid claims for a mental health condition during pregnancy and a California Medicaid-covered delivery four months before and after August 1, 2020, the effective date of the policy extending Medicaid to 12 months postpartum for individuals with a mental health diagnosis during pregnancy. These differences will be compared over the same period with the two following groups of individuals who did not experience the policy change to control for secular trends: (1) individuals without a mental health condition diagnosed during pregnancy with a delivery covered by California Medicaid and (2) individuals with a mental health condition diagnosed during pregnancy in the four other participating states. A difference-in-differences analysis will also be conducted to compare outcomes among individuals with a California, Kentucky, North Carolina, or Ohio Medicaid-covered delivery three months before and after April 1, 2022, the effective date of the policy extending Medicaid to 12 months postpartum. This difference will be compared over the same period with outcomes among individuals with a Colorado Medicaid-covered delivery, who did not experience the policy change during this time, to control for secular trends. The hypothesis is that there will be a modest increase in the number of months of Medicaid enrollment and all categories of service utilization following the postpartum Medicaid extension policy effective dates in each state.

Secondary Analyses

Two secondary analyses are suggested:

First, the same analyses above will be repeated; however, the outcome will be defined as the interpregnancy interval between the index delivery and a subsequent delivery (Table B.1). Participants noted that this analysis would require at least an additional two years of follow-up data compared with the data required for the main analysis.

The second analysis will focus on determining whether state policies prevented postpartum individuals from losing Medicaid coverage at the end of the COVID-19 public health emergency. As noted above, due to the FFCRA, the observed impacts of the postpartum extension may be small or possibly not significant. However, it is possible that state postpartum extension policies had a protective effect, allowing individuals to maintain Medicaid postpartum coverage after the end of the COVID-19 public health emergency. Unlike with the implementation of the postpartum Medicaid extension, there is no state-level variation in the date that the FFCRA was no longer in effect, so a difference-in-differences design cannot be used. A regression discontinuity design can be used instead because it leverages a continuous variable (i.e., percentage of federal poverty limit) with an arbitrary cutoff that assigns beneficiaries to treatment (i.e., Medicaid eligibility). Outcomes can be compared for beneficiaries just above or below each state's Medicaid eligibility cutoff.

Limitations

A key limitation of using Medicaid claims data for research is that only care paid for by Medicaid is observed. For example, it is possible that some individuals who were no longer enrolled in Medicaid during the postpartum period obtained private insurance and received postpartum services paid for by their new insurer or received other financial assistance in paying for care from a health system or federally qualified health center. Even when services are paid for by Medicaid, the use of bundled payment codes for postpartum visits can make counting the number and timing of visits difficult.[24]

When considering potential policy impacts, the most important limitation is the passage of the FFCRA, which may decrease or even eliminate the observed magnitude of the effect of postpartum Medicaid extension in each state. Furthermore, use of health services and health outcomes were strongly impacted by the COVID-19 public health emergency. While the use of a difference-in-differences design accounts for some of these national trends, other impacts of the COVID-19 public health emergency were more localized and would not be accounted for in the study design. The proposed use of a regression discontinuity design in the secondary analysis could address this limitation if no policy effect is observed in the difference-in-differences design proposed in the primary analysis.

Policy Implications

The results of this study can be used to inform policy deliberations within states that have not yet implemented postpartum Medicaid extension. It may also inform policymakers' understanding of the impact of having Medicaid coverage on mothers' access to and utilization of postpartum care or other health care services and outcomes. In addition, the results from extending Medicaid coverage in the postpartum period may inform policy deliberations related to addressing churn in the Medicaid program more broadly.

APPENDIX C. SCHEDULE OF MEETINGS AND PROCESS DEVELOPMENT

Meetings for the current phase of this project are detailed in Table C.1. RAND and ASPE began by facilitating an initial meeting to introduce the project goals and participants (Meeting 1).

To develop guiding principles for standardized linkage and data-sharing processes, RAND convened a group of technical experts in data linkage methodology from the five participating states and the CDC's NCHS.

Each of the participating five states submitted written details on the methods they used to link Medicaid and birth certificate data and on their processes for securely making data available to researchers. RAND used this information to identify key similarities and differences and to generate questions about methodological decision points for discussion with the technical experts. Over the course of two virtual one-hour meetings on data linkage methodology (Meetings 2 and 3), the technical expert group discussed similarities and differences in these methods and processes and perceived strengths and weaknesses of different methodological decisions. RAND then drafted guiding principles based on input from the states on the importance of different methodological decisions and the feasibility of implementation in their states. ASPE and members of the technical expert group, as well as state and national leaders from different agencies, next reviewed the draft guidance on data linkage methodology. Finally, this larger group provided written feedback and contributed input through discussion and poll questions during a subsequent one-hour virtual meeting (Meeting 4). RAND revised the draft guiding principles based on this feedback.

A similar process was followed to develop guiding principles on data security and access. Technical experts in data security and access discussed current processes for securing linked data and making it available to researchers during one two-hour meeting (Meeting 5). At this meeting, the experts also discussed what data would be included in a standard core dataset consisting of linked, deidentified data from five states. In addition, they discussed how these data could be shared through restricted and public use access models. RAND drafted guiding principles that balanced states' concerns about data security and reidentification risk with their stated goal of making useful data available to researchers. The larger group reviewed these guiding principles, provided written feedback on them, and discussed them during a subsequent meeting (Meeting 6). RAND revised the draft guiding principles based on this feedback.

To develop the national research agenda, RAND and ASPE compiled a list of research priorities suggested by participants during previous discussions and based on the literature review from the prior report.[2] These potential research priorities were discussed at a subsequent large group meeting (Meeting 6) where the group voted on the top research priority and identified other important research topics. The smaller technical expert group then provided input on developing the top research priority into a draft research protocol (Meeting 8). RAND drafted a research protocol based on this input (Appendix B).

The smaller technical group met a final time to make final, minor clarifications to the guiding principles on data linkage methodology and security and access (Meeting 7). The larger group then received the full draft final report, provided written comment, and convened one final time to discuss (Meeting 9).

Table C.1. Schedule of Meetings

Meeting	Date	Participants	Meeting Topic
1	3/30/2023	All project participants (all technical experts and state and national leaders)	Project introduction
2	4/12/2023	Technical experts in data linkage methodology	Discussion of current linkage methods
3	4/20/2023	Technical experts in data linkage methodology	Development of linkage method guiding principles
4	5/10/2023	Technical experts in data linkage methodology, state and national leaders	Review draft linkage method guiding principles
5	5/22/2023	Technical experts in data security and access	Development of data security and access guiding principles
6	6/5/2023	Technical experts in data security and access, state and national leaders	Review draft guiding principles on data security and access, discuss national research agenda
7	6/15/2023	All technical experts	Finalize guiding principles
8	6/22/2023	All technical experts	Discuss protocol for researchers using linked data
9	7/17/2023	All project participants	Review final report

APPENDIX D. DETAILS OF STATE DATA LINKAGES

Populations and Datasets Linked

California begins by linking mothers' Medicaid delivery claims with the Medicaid eligibility file and hospital delivery data either deterministically if SSN is present (using SSN and delivery month) or probabilistically using the LINKS SAS package if SSN is not present (using delivery date, demographic, and diagnosis/procedure variables). The mother's linked Medicaid-hospital discharge data are then probabilistically linked to infant birth certificate data using a combination of birth event times and place variables and the mother's name and demographic variables.

Colorado begins with a mother blocking step (i.e., deterministic matching of a mother's Medicaid delivery claim to birth certificate on any of a broad set of criteria: exact Medicaid number, exact SSN, or partial name plus DOB match). Each mother's Medicaid delivery claim and each birth certificate is allowed to have multiple matches at this stage. Next, Colorado executes an infant blocking step similar to the mother blocking step in which infant Medicaid claims are linked to birth certificates. Finally, there is a selection step in which infant and mother Medicaid matches are given a score based on the number and types of variables matched, and the best score is retained for each match.

Kentucky was unable to provide detailed information on linkages, but the state noted that it begins with infant birth certificates and matches them to Medicaid claims and enrollment data using IBM Infosphere, a software that allows for probabilistic matching. Kentucky also confirmed the variables used in their linkages.

North Carolina links both infant Medicaid and maternal Medicaid data to infant birth records using probabilistic matching in Link Plus. This involves blocking steps, in which records are subset to match on blocking variables, reducing the total number of pairwise matches to be scored, and a scoring step that is used to select the most likely record pair when a birth certificate has been linked to multiple Medicaid records. Matches that meet certain more-stringent criteria are automatically retained, while matches that match on other criteria are manually reviewed by at least one reviewer.

Ohio begins by restricting the dataset to mothers who have an SSN on both Medicaid claims and birth certificates. These mothers are linked in SAS using SSN, DOB, and Soundex mother's first and/or last name. The remaining unmatched mothers are then linked probabilistically using Link Plus software with a combination of the variables mentioned above plus address, zip code, phone number, and delivery date. Mother and infant Medicaid data are then linked using a similar process, but also including the Medicaid case ID number. Finally, infants from Medicaid claims data are linked to infant birth certificates. This final step helps ensure that both the mother and the infant (already linked through Medicaid claims) are matched to the same birth certificate. Matches are selected for manual review based on their Link Plus score or if they match to multiple records (i.e., linked mother-infant Medicaid claims link to multiple birth certificates).

Variables Used in Linkages

All five states used the **mother's SSN** in some way. However, California only uses it in the Medicaid-hospital discharge linkage step, not in the Medicaid-birth certificate linkage because SSN is not in the birth certificate data they receive. SSN was used by states in both deterministic and probabilistic matching steps. States generally used SSN as an exact matching variable (i.e., no partial credit is given for transposed or missing digits), although Ohio considered inexact matches in its manual review process. States noted that SSN is

sometimes missing, and probabilistic matching on other variables is needed.

All states use the **mother's first and last names**. Some states also use the mother's middle and pre-marital names and the father's last name (sometimes, the father's last name on the birth certificate can be matched to the mother's last name on Medicaid claims if the mother's name on the birth certificate is missing or incorrect). The mother's first and last names were used in both deterministic and probabilistic matching steps. Some states allow for fuzzy matching in probabilistic matches using Soundex or the first three or four characters of a name.

All states use the **mother's and infant's DOB** for matching. The mother's DOB is in the Medicaid enrollment file, the infant's DOB can be estimated from the mother's' Medicaid delivery claims data, and both the mother's and infant's DOB are on the birth certificate. Some states allow for fuzzy matching on DOBs. For example, Colorado allows for two of three date elements to be correct in its blocking steps. Ohio allows a two-week window around the DOB, and North Carolina allows a five-day window around the DOB in probabilistic matching steps. Most states used the **mother's geographic information,** such as full residential address, street, city, zip code, or county. These variables were used in probabilistic steps and in manual review steps.

Other variables used by only one or two states include the mother's' Medicaid record number, the mother's' phone number, the mother's' race/ethnicity, and delivery method (i.e., vaginal or C-section). Other variables, such as infant name, infant SSN, infant DOB from infant Medicaid enrollment or claims data and hospital ID number were used by some states, but they are not discussed further here because they can only be used when the mother's Medicaid data are merged with the infant's Medicaid data or hospital discharge data from the delivery.

Software, Scoring/Weighting Methods, and Validation Methods

Ohio and North Carolina use Link Plus,[25] free software developed and supported by the CDC, to probabilistically match records. The software generates match scores based on the strength of the match. Two cutoffs are selected: a higher cutoff, above which matches are accepted as is, and a lower cutoff, for which matches should be manually reviewed between the lower and the higher cutoff. Kentucky uses IBM InfoSphere MDM to perform probabilistic matching.[26] This process assigns match scores for each variable used for matching and aggregates such scores into a total match score. California uses LINKS, a SAS package developed by the University of Manitoba that performs a staged matching process that generates and uses scores for probabilistic matching.[27] The highest scores at each stage are used to remove duplicates from ties, and staff manually review to ensure the best matches were kept. Colorado uses an in-house-developed SAS program to perform deterministic linkages only and generates scores based on exactly matched variables. States also use in-house-developed SAS programs before and after using the linkage software or programs for data preparation and cleaning.

All states do manual review to some extent, although the criteria and process for manual review differ by state. Ohio and North Carolina manually review a subset of matches based on its Link Plus match score; North Carolina also has two reviewers review some matches and a third reviewer address disagreements between the two reviewers. California manually reviews records that have multiple records and also verifies frequencies of births and demographic variables in matched data against national vital registry sources. Colorado sometimes manually validates all matches for a special purpose analysis of a smaller population and periodically validates a random sample of matches to assess accuracy.

Linkage Timeline and Updating Linkages

Currently, states vary in the frequency with which linkages are performed. California and North Carolina perform linkages annually, and Colorado and Ohio perform linkages quarterly. As Medicaid and birth certificate data might become available at different times due to enrollment or reporting delays, states typically use a runout period (i.e., only match births that happened after a certain amount of time). Ohio and North Carolina also include previously unmatched records in new rounds of matching.

ABBREVIATIONS

ACA	Affordable Care Act
APGAR	Appearance Pulse Grimace Activity Respiration
ASPE	Assistant Secretary for Planning and Evaluation
CDC	Center for Disease Control and Prevention
CMS	Centers for Medicare & Medicaid Services
COVID-19	coronavirus disease 2019
DUA	data use agreement
DOB	date of birth
FFCRA	Families First Coronavirus Response Act
HEDIS	Healthcare Effectiveness Data and Information Set
HHS	U.S. Department of Health and Human Services
HRSA	Health Resources & Services Administration
FSRDC	Federal Statistical Research Data Centers
ICD-10	International Classification of Disease, Tenth Revision
IRB	Institutional Review Board
NCANDS	National Child Abuse and Neglect Data System
NDACAN	National Data Archive on Child Abuse and Neglect
NCHS	National Center for Health Statistics
NICU	neonatal intensive care unit
NIH	National Institutes of Health
OS-PCORTF	Office of the Secretary – Patient-Centered Outcomes Research Trust Fund
PCOR	patient-centered outcomes research
ResDAC	Research Data Assistance Center
SSN	Social Security Number
SVI	Social Vulnerability Index
T-MSIS	Transformed Medicaid Statistical Information System
WONDER	Wide-ranging Online Data for Epidemiologic Research

REFERENCES

1. Martin, Joyce A., Brady E. Hamilton, and Michelle J.K. Osterman, "Births in the United States, 2019," *NCHS Data Brief*, Vol. 387, 2020. As of July 25, 2023: https://www.cdc.gov/nchs/data/databriefs/db387-H.pdf
2. Heins, Sara, Zachary Predmore, Emily Hoch, and Sangita Baxi, *Linking Medicaid Claims, Birth Certificates, and Other Sources to Advance Maternal and Infant Health*, Office of the Assistant Secretary for Planning and Evaluation, Department of Health and Human Services, 2022. As of July 25, 2023: https://aspe.hhs.gov/reports/linking-medicaid-other-data-pcor
3. U.S. Department of Health and Human Services, "Strengthening Maternal Health," webpage, June 21, 2023. As of July 25, 2023: https://www.hhs.gov/healthcare/maternal-health/index.html
4. Office of the Assistant Secretary for Planning and Evaluation, "OS-PCORTF Themes and Focus Areas," webpage, undated. As of July 25, 2023: https://aspe.hhs.gov/collaborations-committees-advisory-groups/os-pcortf/os-pcortf-themes-focus-areas
5. Osterman, Michelle J.K., Brady E. Hamilton, Joyce A. Martin, Anne K. Driscoll, and Claudia P. Valenzuela, "Births: Final Data for 2021," *National Vital Statistics Reports*, Vol. 72, No. 1, 2023.
6. Medicaid, "DQAtlas," webpage, undated. As of July 25, 2023: https://www.medicaid.gov/dq-atlas/welcome
7. National Association for Public Health Statistics and Information Systems, "State and Territorial Exchange of Vital Events (STEVE)," webpage, undated. As of July 25, 2023: https://www.naphsis.org/steve
8. National Cancer Institute, "Download Match*Pro Software," webpage, undated. As of July 25, 2023: https://seer.cancer.gov/tools/matchpro/download
9. U.S. Census Bureau, "Federal Statistical Research Data Centers," July 5, 2023. As of July 25, 2023: https://www.census.gov/about/adrm/fsrdc.html
10. Research Data Assistance Center, "Find, Request and Use CMS Data," webpage, undated. As of July 25, 2023: https://resdac.org/
11. National Data Archive on Child Abuse and Neglect, homepage, undated. As of July 25, 2023: https://www.ndacan.acf.hhs.gov/index.cfm
12. Centers for Disease Control and Prevention, "Natality Information," webpage, December 2, 2022. As of July 25, 2023: https://wonder.cdc.gov/natality.html
13. Agency for Toxic Substances and Disease Registry, CDC/ATSDR Social Vulnerability Index, July 12, 2023. As of July 25, 2023: https://www.atsdr.cdc.gov/placeandhealth/svi/index.html
14. Public Law 117-2, American Rescue Plan Act of 2021, March 11, 2021.
15. Ukah, U. Vivian, Natalie Dayan, Brian J. Potter, Gilles Paradis, Aimina Ayoub, and Nathalie Auger, "Severe Maternal Morbidity and Long-Term Risk of Cardiovascular Hospitalization," *Circulation: Cardiovascular Quality and Outcomes*, Vol. 15, No. 2, 2022.
16. KFF, "Medicaid Postpartum Coverage Extension Tracker," webpage, September 7, 2023. As of July 25, 2023: https://www.kff.org/medicaid/issue-brief/medicaid-postpartum-coverage-extension-tracker/#note-0-5
17. Public Law 116-127, Families First Coronavirus Response Act, March 18, 2020.
18. Medicaid, "Unwinding and Returning to Regular Operations After COVID-19," webpage, undated. As of July 25, 2023: https://www.medicaid.gov/resources-for-states/coronavirus-disease-2019-covid-19/unwinding-and-returning-regular-operations-after-covid-19/index.html
19. World Health Organization, *Report of a WHO Technical Consultation on Birth Spacing*, WHO/RHR/07.1, 2005.
20. National Committee for Quality Assurance, "HEDIS Measures and Technical Resources," webpage, undated. As of July 25, 2023: https://www.ncqa.org/hedis/measures/
21. Health Resources & Services Administration, "Billing and Coding Medicare Fee-for-Service Claims," webpage, August 31, 2023. As of July 25, 2023: https://telehealth.hhs.gov/providers/billing-and-reimbursement/billing-and-coding-medicare-fee-for-service-claims
22. Crissman, Halley P., Caleb Haley, Daphna Stroumsa, Anca Tilea, Molly B. Moravek, Lisa A. Harris, and

Vanessa K. Dalton, "Leveraging Administrative Claims to Understand Disparities in Gender Minority Health: Contraceptive Use Patterns Among Transgender and Nonbinary People," *LGBT Health*, Vol. 9, No. 3, 2022.

23. Centers for Disease Control and Prevention, "U.S. Standard Certificate of Live Birth," form, November 2003.

24. DeSisto, Carla L., Angela Rohan, Arden Handler, Saria S. Awadalla, Timothy Johnson, and Kristin Rankin, "Comparing Postpartum Care Utilization from Medicaid Claims and the Pregnancy Risk Assessment Monitoring System in Wisconsin, 2011–2015," *Maternal and Child Health Journal*, Vol. 25, No. 3, 2021.

25. Centers for Disease Control and Prevention, "Link Plus," webpage, August 16, 2023. As of July 25, 2023: https://www.cdc.gov/cancer/npcr/tools/registryplus/lp.htm

26. IBM, "The InfoSphere MDM Probabilistic Matching Engine Data Model," webpage, April 12, 2021. As of July 25, 2023: https://www.ibm.com/docs/en/imdm/12.0?topic=engine-infosphere-mdm-probabilistic-matching-data-model

27. University of Manitoba, Max Rady College of Medicine, "Concept: LINKS: A Record Linkage Package," webpage, March 26, 2002. As of July 25, 2023: http://mchp-appserv.cpe.umanitoba.ca/viewConcept.php?conceptID=1029