Medicaid/CHIP
Health Care Quality Measures

TECHNICAL ASSISTANCE BRIEF

Number 4 ▪ January 2014

STRATEGIES FOR USING VITAL RECORDS TO MEASURE QUALITY OF CARE IN MEDICAID AND CHIP PROGRAMS

Background

Medicaid and CHIP provide prenatal, labor and delivery, and postpartum services for a large proportion of births in the United States, covering 48 percent of all births in 2010 (Markus et al. 2013). As the largest single payer for maternity care, Medicaid and CHIP play a key role in promoting access to care and ensuring the quality of care during the perinatal period. The Medicaid/CHIP Core Sets of health care quality measures contain nine maternity care measures to help drive quality improvement efforts at the state and national levels. The child Core Set includes six maternity measures (timeliness of prenatal care, frequency of ongoing prenatal care, behavioral health risk assessment for pregnant women, Cesarean rate, low birth weight rate, and well-child visits in the first 15 months of life) and the adult Core Set includes three maternity measures (antenatal steroid use, elective delivery, and timeliness of postpartum care).[1] Two of these measures, low birth weight rate and Cesarean rate, are specified for use with vital records, although linkage to Medicaid/CHIP administrative data is often necessary to identify women covered by Medicaid or CHIP.

Many states have encountered challenges in reporting the low birth weight and Cesarean measures because of barriers to accessing vital records and limited capacity to link vital records with Medicaid/CHIP administrative data.[2] Building states' capacity to calculate and report these measures is a high priority because of CMS's emphasis on improving birth outcomes for women covered by Medicaid/CHIP.[3]

To support states' efforts to access and use vital records for quality measurement and quality improvement, this TA brief discusses strategies for accessing vital records, factors affecting the use of Medicaid/CHIP administrative data, methods for linking vital records with Medicaid/CHIP data, issues to consider when constructing the Core Set maternity measures, and other TA resources available to states. The primary goal of this TA brief is to increase the number of states reporting and using the Core Set maternity measures.

About This Brief

This technical assistance (TA) brief discusses strategies for accessing vital records for quality measurement and improvement efforts related to maternal and infant health care in Medicaid and the Children's Health Insurance Program (CHIP). The brief also gives guidance and describes resources available to states for linking vital records and Medicaid/CHIP data to increase states' capacity to report two measures in the Core Set of children's health care quality measures: low birth weight rate and Cesarean section rate. These measures are specified using vital records data, although linkage to Medicaid/CHIP administrative data is often necessary to identify women covered by Medicaid or CHIP. Building states' capacity to use vital records is a high priority for the Centers for Medicare & Medicaid Services (CMS) as part of its efforts to improve the health of mothers and infants covered by Medicaid or CHIP.

[1] States have reported five of the child Core Set maternity measures since federal fiscal year (FFY) 2010; the behavioral health risk assessment measure will be reported for the first time in FFY 2013. States will begin reporting the three adult Core Set maternity measures in FFY 2013. The maternity Core Set is available online at http://www.medicaid.gov/Medicaid-CHIP-Program-Information/By-Topics/Quality-of-Care/Downloads/2013-Core-Set-of-Maternity-Measures.pdf.

[2] In FFY 2012, 15 states reported the low birth weight rate and 12 states reported the Cesarean rate. More states reported the other child Core Set maternity measures: 43 states reported the well-child visits in the first 15 months of life, 31 states reported the timeliness of prenatal care, and 25 states reported the frequency of prenatal care.

[3] See "Maternal and Infant Health Care Quality," available at http://www.medicaid.gov/Medicaid-CHIP-Program-Information/By-Topics/Quality-of-Care/Maternal-and-Infant-Health-Care-Quality.html.

■ This technical assistance brief is a product of the Children's Health Care Quality Measures Core Set Technical Assistance and Analytic Support Program, sponsored by the Centers for Medicare & Medicaid Services. The program team is lead by Mathematica Policy Research, in collaboration with the National Committee for Quality Assurance, Center for Health Care Strategies, and National Initiative for Children's Healthcare Quality.

Accessing Vital Records for Quality Measurement and Quality Improvement

States link vital records with Medicaid/CHIP administrative data for reporting the Core Measures and a number of additional purposes, including monitoring additional outcome variables, calculating the fraction of births in a state paid by Medicaid/CHIP, and obtaining data on maternal risk factors. More than 20 states currently link vital records and Medicaid/CHIP data.[4] The following examples demonstrate the value of linked data for improving maternal and infant health care.

- Washington State has linked birth certificates and Medicaid administrative data to support program evaluations since 1990. The First Steps Data Base has been used to evaluate, monitor, and improve programs for pregnant women and infants. Specific applications have included identifying high-prevalence risk factors in the Medicaid population (such as smoking during pregnancy) and analyzing body mass index (BMI) data on birth certificates, which revealed the need for counseling on obesity and exercise in addition to food insecurity and insufficient weight gain. Broad stakeholder support exists for using vital records for analytic purposes.
- In Illinois, the legislature mandated creation of an enterprise data warehouse (EDW) with numerous linked data sets. The Moms and Babies Data Mart includes data from Vital Records; Medicaid; the Adverse Pregnancy Outcome Reporting System; the Special Supplemental Program for Women, Infants and Children (WIC); and Family Case Management (Smith 2011). The data system has been used to develop a better understanding of factors predicting adverse pregnancy outcomes and to provide more focused case management services to women with high-risk pregnancies.
- Iowa has conducted annual linkages of vital records and Medicaid data since 1989. The state has used these data to evaluate its maternity case management programs and target women with high-risk pregnancies who are likely to benefit from these services. The following state profile highlights Iowa's experience with linking vital records and Medicaid data.

[4] No definitive source identifies states that currently link Medicaid/CHIP and vital records data. This estimate is based on recent efforts by CMS and the Centers for Disease Control and Prevention (CDC) to assess state capacity for accessing and using vital records. One publicly available list of states that link these data is available at https://mchdata.hrsa.gov/tvisreports/MeasurementData/HSCI/Hsci09Search.aspx?MeasureNum=09A&SurveyCategoryId=10. This list shows states that report on the Title V Health Systems Capacity Indicator 09A.

State Profile: Iowa's Approach to Linking Medicaid and Vital Records Data

The Iowa Medicaid-Birth Certificate Match project is conducted through an interdepartmental agreement between the Iowa Department of Health Services and the Iowa Department of Public Health. The project provides information on the characteristics of pregnant Medicaid enrollees, their use of services, and their birth outcomes. The results are used to guide the development, implementation, evaluation, and improvement of Medicaid programs and policies.

Iowa has performed an annual linkage of Medicaid and birth certificate data since 1989. The linkage is performed by matching several variables in the birth certificate and the claims attributed to the mother and infant, including mother's first, last, and maiden names; county of residence; infant's date of birth; and infant's last name. Iowa uses Link Plus software developed by the Centers for Disease Control and Prevention (CDC) and originally designed to be used by cancer registries. In recent years, its use has expanded to other applications and it has become a common linkage tool for researchers and organizations that maintain public health data. Over the years, Iowa has improved the efficiency of its linkage process such that the matching takes approximately two weeks to complete.

Iowa's linked Medicaid claims and vital records data have been available for research and program monitoring purposes for more than 20 years. One study used the data to analyze the impact of a primary care case management (PCCM) program on birth outcomes among Medicaid enrollees, finding no significant differences in prenatal care utilization, gestational age at delivery, or infant birth weight between women in PCCM counties and those in fee-for-service counties (Schulman et al. 1997). These findings have been used to better evaluate case management programs and target women with high-risk pregnancies who are more likely to benefit from these services (Alexander and Mackey 1999).

More recently, researchers have used the linked data set to test the reliability of the Medicaid payment indicator on the birth certificate, an element added to the standard form in 2003 (Kane and Sappenfield 2013). The study found a high level of agreement, suggesting that the payment source indicated on the birth certificate is as valid as the linked data source (although this finding might not generalize to other states). Another study examined the age at which Medicaid-enrolled children had their first dentist visits for those who obtain services at federally qualified health centers (Kuthy et al. 2013). The linked Medicaid and birth certificate data were linked to abstracted dental charts, demonstrating the broad range of analytic questions to which the linked data set can be applied.

Although nearly half the states have experience accessing and using vital records for monitoring or evaluation, other states have encountered barriers. The most common barrier concerns confidentiality of the vital records data. All states must adhere to the Health Insurance Portability and Accountability Act of 1996 (HIPAA), which addresses the security and privacy of health data, including release of personally identifiable information (such as name, address, and Social Security number) and protected health information.

States often use two vehicles to facilitate—but also govern and protect—sharing of individual-level data from vital records: data use agreements (DUAs) and institutional review board (IRB) clearance. In Arkansas, for example, the Health Statistics Branch of the Department of Health routinely links vital records and hospital discharge data for newborns, and has established a data-sharing infrastructure (Bronstein et al. 2009). State agencies might be able to negotiate standing DUAs that facilitate linking Medicaid/CHIP and vital records data for annual calculation of the quality measures. However, if analytic needs change over time (for example, because measure specifications change), states should ensure that DUAs are updated to reflect the desired data uses. In addition to DUAs, states can require clearance by an IRB or data review committee to provide legal authorization for release (even for intra-agency use). Washington State, for example, has conducted Medicaid–birth certificate linkages for two decades, with legal authorization from the Washington State IRB.

Another factor affecting states' use of vital records data for reporting on the child Core Set of measures is the schedule for data release. Many states produce an annual data release. As a result, states report that vital records data often lag behind Medicaid/CHIP data, given the time required to finalize (certify) the birth certificate data. For example, although California processes birth certificate data files 45 days after the end of the month, data reconciliation and certification can take more than a year. With greater emphasis on improving birth outcomes, some states, such as Louisiana, are updating their vital records systems to reduce the lag. The state reported reducing the time between birth and registration from 72 to 17 days, on average. [5] Because uncertified data might be available before certified data, some states could decide to use preliminary (uncertified) data until final (certified) data are available. Others might decide to report rates for an earlier period. In either case, states can update their rates when new data become available. One additional factor to consider is that several states have not yet adopted the 2003 revision to the U.S. Standard Certificate of Live Birth, which means that some fields might not be available and comparable to other states (see the technical appendix for details)

Using Medicaid/CHIP Claims Data to Supplement Vital Records Data

States may use Medicaid/CHIP eligibility and claims data to supplement vital records data in two ways: (1) to develop denominators of pregnant women and newborns covered by Medicaid/CHIP and (2) to augment the birth certificate data with diagnoses identified from claims data. [6] In developing a data linking strategy, states should consider the following factors:

1. Structure of Medicaid/CHIP data systems. In many states, Medicaid and CHIP data systems are separate. In such cases, states might have to integrate data from the two systems to link with vital records. The steps for integrating Medicaid/CHIP data files are similar to those for matching other types of records (as discussed later in this brief).
2. Denominator size. Exhibit 1 shows the denominators for the two Core Set measures that use vital records. Some states might have to combine data across years to ensure a sufficient number of Medicaid/CHIP-covered live births meeting the denominator definition. Because the specifications for the Core Set measures do not stipulate minimum denominator sizes for statistical precision, states would have to decide whether to combine data across years based on the number of births. An important consideration is the availability of historical data should a state decide to combine data across multiple years.

Exhibit 1. Denominator Definitions for the Two Medicaid/CHIP Core Set Measures that Use Vital Records

Measure	Denominator Definition
Live Births Weighing Less than 2,500 Grams	Resident live births in the reporting period with Medicaid and/or CHIP as the payer source
Cesarean Rate for Low-Risk First Birth Women	Live births at or beyond 37 (37.0) weeks gestation to women who are having their first delivery and are singleton (no twins or beyond) and are vertex presentation (no breech or transverse positions)

[5] See http://dhh.louisiana.gov/index.cfm/page/411/n/194 for more information on the Louisiana Birth Outcomes Initiative. See http://www.katc.com/news/dhh-launches-tracking-system-to-collect-pre-term-birth-data/ for more information on changes to the state's birth registration system.

[6] Although the technical specifications for the low birth weight measure do not indicate the use of Medicaid/CHIP data to calculate the denominator, insurance status often is not recorded on the birth certificate; several states reported linking birth certificate and Medicaid/CHIP data to identify the population for the measure.

3. Managed care penetration. Managed care penetration rates can affect the quality and completeness of utilization data to the extent managed care organizations' encounter data (MCOs) do not capture specific visits, services, or episodes. All states require MCOs to submit encounter data and some have standardized reporting requirements that lead to substantially complete encounter data. In addition, various resources are available to guide state agencies in validating these data to mitigate this issue.[7]
4. Reimbursement policies. State reimbursement policies can also affect the consistency with which key data elements are reported. For example, use of so-called global billing for prenatal care and delivery could affect the availability of diagnoses and procedure codes in claims data for the Cesarean section measure.
5. Timeliness of claims data submissions. Lags in provider submission of claims can affect the availability of final data in time for the Core Set reporting cycle. Many states allow up to 12 months for a health care provider to submit claims for payment. Some states note that the rates they report are provisional due to the 12-month claim submission period.
6. Data quality. State data systems vary in the types of automated quality checks or validations performed on the Medicaid/CHIP eligibility and claims data. Consequently, data anomalies and data quality vary from state to state. The prevalence of missing data or outlier values on matching variables can affect the matching approach and outcomes.

These factors can have implications for a state's approach to the data linking process, in terms of data preparation (such as integrating separate Medicaid and CHIP data systems), time frame for the data (to account for data lags or small population sizes), or selection of matching variables (due to missing data or outlier values). States should carefully assess their data before beginning the linking process.

After obtaining vital records data and Medicaid/CHIP data and taking key considerations into account, states will have to undertake a multistep process to perform the linkages. The technical appendix describes an approach to link the data in six major steps: (1) deduplicate IDs in the Medicaid/CHIP eligibility files, (2) prepare and clean the data, (3) link mother and infant Medicaid/CHIP records, (4) link Medicaid/CHIP data and vital records, (5) assess the data linkage results, and (6) prepare the final data file. The technical appendix outlines the basic objectives of each step and discusses the software packages that can be used.

For Further Information

Background information on the Medicaid/CHIP Core Sets of Child and Adult Health Care Quality Measures, guidance on collecting and reporting the measures, and technical specifications for each measure can be found on the Medicaid Quality of Care Performance Measurement Webpage. Guidance on using the measures for quality improvement is also available through webinars archived on the Medicaid Quality of Care Webpage.[8]

The Center for Medicaid and CHIP Services also supported the development of a data dictionary that provides additional guidance for conducting linkages between vital records and Medicaid/CHIP data for a Medicaid Medical Directors Network perinatal project (in partnership with the Agency for Healthcare Research and Quality and the Health Resources and Services Administration).[9]

For TA related to calculating and reporting the child and adult Core Set measures, contact the TA mailbox at MACqualityTA@cms.hhs.gov.

This technical assistance brief was prepared by Keith Kranker, So O'Neil, Vanessa Oddo, Miriam Drapkin, and Margo Rosenbach, Mathematica Policy Research.

[7] A primer for states on encounter data validation is available at http://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/downloads/MAX_PDQ_Task_X_EncounterDataPrimerforStates.pdf. State Medicaid data validation reports are available at http://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/MAX-Validation-Reports.html. More information on CMS efforts to strengthen state reporting of encounter data is available at http://www.medicaid.gov/Federal-Policy-Guidance/Downloads/SMD-13-004.pdf.

[8] To access these resources, see http://www.medicaid.gov/Medicaid-CHIP-Program-Information/By-Topics/Quality-of-Care/Quality-of-Care-%E2%80%93-Performance-Measurement.html and http://www.medicaid.gov/Medicaid-CHIP-Program-Information/By-Topics/Quality-of-Care/Quality-of-Care.html, respectively.

[9] More information about these resources is available at http://www.medicaid.gov/Medicaid-CHIP-Program-Information/By-Topics/Quality-of-Care/Maternal-and-Infant-Health-Care-Quality.html

References

Alexander, Judith W., and Marlene C. Mackey. "Cost Effectiveness of a High-Risk Pregnancy Program." Care Management Journals, vol. 1, no. 3, 1999, pp. 170–174.

Borck, Rosemary, Allison Hedley Dodd, Ashley Zlatinov, Shinu Verghese, and Rosalie Malsberger. "MAX 2008 Chartbook and Appendix Tables." Baltimore, MD: Centers for Medicare & Medicaid Services, 2012. Available at [https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/MAX_Chartbooks.html].

Bronstein, J., C. Lomatsch, D. Fletcher, T. Wooten, T. Lin, R. Nugent, and C. Lowery. "Issues and Biases in Matching Medicaid Pregnancy Episodes to Vital Records Data: The Arkansas Experience." Maternal and Child Health Journal, vol. 13, 2009, pp. 250–259.

Buescher, Paul. "Method of Linking Medicaid Records to Birth Certificates May Affect Infant Outcome Statistics." American Journal of Public Health, vol. 89, no. 4, 1999, pp. 564–566.

Campbell, K., D. Deck, and A. Krupski. "Record Linkage Software in the Public Domain: A Comparison of Link Plus, the Link King, and a 'Basic' Deterministic Algorithm." Health Informatics Journal, vol. 14, no. 1, 2008, pp. 5–15.

Campbell, K. "Record Linkage and Consolidation Algorithms in "Rule Your Data with the Link King." Olympia, WA: Washington State Division of Alcohol and Substance Abuse, 2013. Available at [http://www2.sas.com/proceedings/sugi30/020-30.pdf].

Czajka, John, Audra Wenzlow, and Julie Sykes. "Development of the Medicaid Analytic Extract Enrollee Master (MAXEM) File, 2005 and 2006." Washington, DC: Mathematica Policy Research, September 14, 2010. Available at [https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/Downloads/MAXEMFinalReport.pdf].

Czajka, John, and Shinu Verghese. "Social Security Numbers in Medicaid Records: Reporting and Validity, 2009." Washington, DC: Mathematica Policy Research, January 15, 2013. Available at [http://mathematica-mpr.com/publications/PDFs/health/SSN_Medicaid.pdf].

Ferrante, A., and J. Boyd. "A Transparent and Transportable Methodology for Evaluating Data Linkage Software." Journal of Biomedical Informatics, vol. 45, no. 1, 2012, pp. 165–172. doi:10.1016/j.jbi.2011.10.006.

Gyllstrom, M, J. Jensen, J. Vaughan, S. Castellano, and J. Oswald. "Linking Birth Certificates with Medicaid Data to Enhance Population Health Assessment: Methodological Issues Addressed." Journal of Public Health Management and Practice, vol. 8, no. 4, 2002, pp. 38–44.

Hamilton, Brady E., Joyce A. Martin, and Stephanie J. Ventura. "Births: Preliminary Data for 2012." National Vital Statistics Reports, vol. 62, no. 3. Atlanta, GA: National Center for Health Statistics, September 6, 2013. Available at [http://www.cdc.gov/nchs/products/nvsr.htm].

Kane, Debra J., and William M. Sappenfield. "Ascertainment of Medicaid Payment for Delivery on the Iowa Birth Certificate: Is Accuracy Sufficient for Timely Policy and Program Relevant Analysis?" Maternal and Child Health Journal, 2013, pp. 1–8.

Kuthy, Raymond A., Bhagyashree Pendharkar, Elizabeth T. Momany, Michael P. Jones, Natoshia M. Askelson, Donald L. Chi, George L. Wehby, and Peter C. Damiano. "Factors Affecting Age at First Dental Exam for Medicaid-Enrolled Children Seen at Federally Qualified Health Centers." Pediatric Dentistry, vol. 35, no. 4, 2013, pp. 100E–106E.

Markus A. R., E. Andres, K.D. West, N. Garro, and C. Pellegrini. "Medicaid Covered Births, 2008 through 2010, in the Context of the Implementation of Health Reform." Women's Health Issues, vol. 23, no. 5, 2013, pp. e273–e280.

Martin, J.A., E.C. Wilson, M.J.K. Osterman, E.W. Saadi, S.R. Sutton, and B.E. Hamilton. "Assessing the Quality of Medical and Health Data from the 2003 Birth Certificate Revision: Results from Two States." National Vital Statistics Reports, vol. 62, no. 2, July 22, 2013. Available at [http://www.cdc.gov/nchs/data/nvsr/nvsr62/nvsr62_02.pdf].

Newman, T.B., and A.N. Brown. "Use of Commercial Record Linkage Software and Vital Statistics to Identify Patient Deaths." Journal of the American Medical Informatics Association, vol. 4, no. 3, 1997, pp. 233–237.

Schulman, E.D., D.J. Sheriff, and E.T. Momany. "Primary Care Case Management and Birth Outcomes in the Iowa Medicaid Program." American Journal of Public Health, vol. 87, no. 1, 1997, pp. 80–84.

Smith, Gwen. "Administrative and Vital Records Data Housed in the Enterprise Data Warehouse (EDW)." Presented at the CMS 2011 Medicaid/CHIP Quality Conference. Available at [http://www.medicaid.gov/State-Resource-Center/Events-and-Announcements/Downloads/Quality-Conference-2011/Gwen-Smith-Administrative_Vital_Records_Data.pdf].

Winkler, W.E. "Record Linkage Software and Methods for Merging Administrative Lists." RR2001/03 Statistical Research Report Series. Washington, DC: Bureau of the Census, Statistical Research Division, 2001. Available at [http://www.census.gov/srd/papers/pdf/rr2001-03.pdf].

TECHNICAL APPENDIX

ADDITIONAL INFORMATION ON METHODS FOR LINKING MEDICAID/CHILDREN'S HEALTH INSURANCE PROGRAM (CHIP) AND VITAL RECORDS DATA

This appendix provides additional technical information on three topics related to linking Medicaid/CHIP and vital records data:

A. Step-by-Step Approach to Data Linking

B. Software Packages for Linking Medicaid/CHIP and Vital Records Data

C. Comparability of the 1989 and 2003 Versions of the Birth Certificate

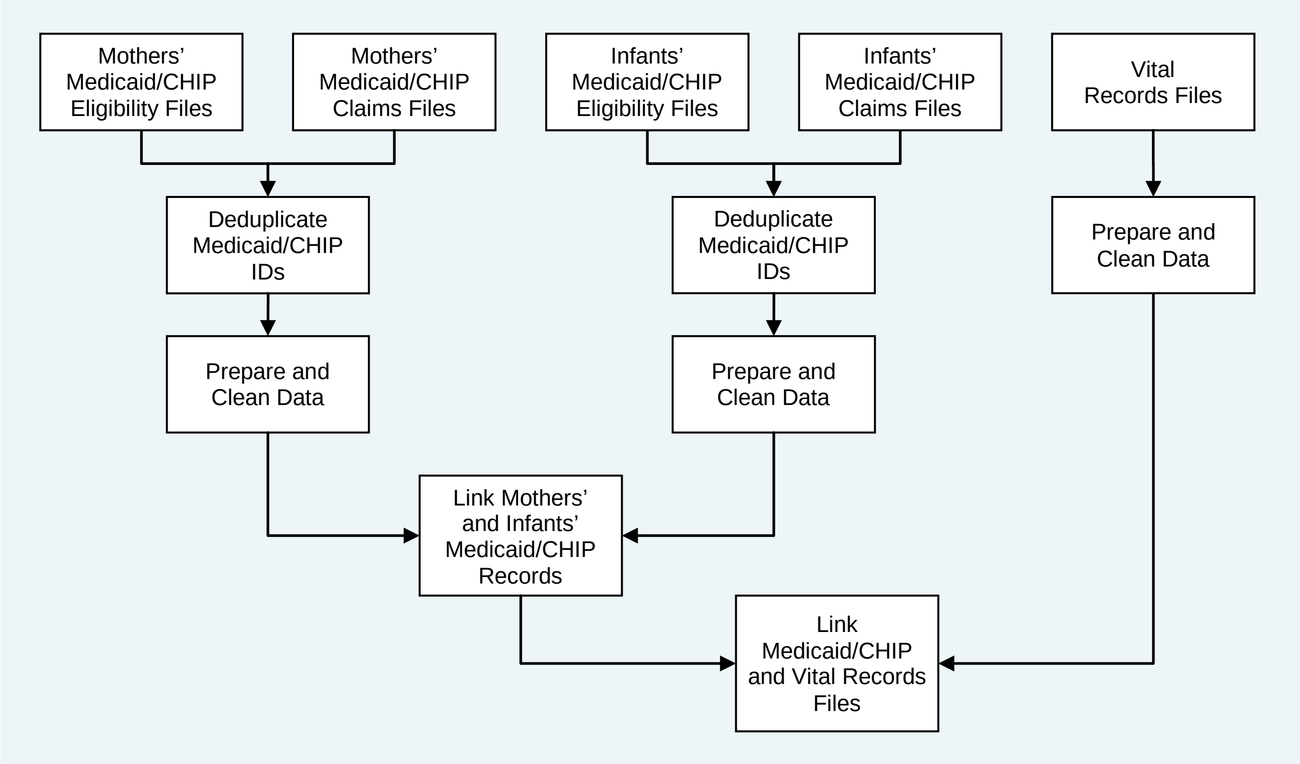
A. Step-by-Step Approach to Data Linking

Exhibit A.1 summarizes an approach to data linking. The approach uses five input data files: Medicaid/CHIP eligibility and claims data for the mother, Medicaid/CHIP eligibility and claims data for the infant, and vital records (birth certificates). This section of the technical appendix describes the following six steps in the linking process:

1. Deduplicate IDs in the Medicaid/CHIP eligibility files
2. Prepare and clean the data
3. Link mother and infant Medicaid/CHIP records[10]
4. Link Medicaid/CHIP data and vital records
5. Assess the data linkage results
6. Prepare the final data file

[10] Linking mother and infant Medicaid/CHIP records is recommended (though not essential) because it produces a higher match rate than linking vital records data to mother or infant records alone.

Exhibit A.1. Overview of the Record Linkage Process



Step 1: Deduplicate the Medicaid/CHIP IDs

The linking process requires a Medicaid/CHIP data file with reliable Medicaid/CHIP IDs and records grouped for each Medicaid/CHIP enrollee. When a Medicaid/CHIP enrollee has more than one Medicaid/CHIP ID, the records cannot be grouped for each individual and, as a result, there is a risk that more than one Medicaid/CHIP ID could match to a single birth certificate.[11] The purpose of this task, therefore, is to identify people who appear in the Medicaid/CHIP files with more than one Medicaid/CHIP ID and correct their records such that all records corresponding to the same person can be grouped together. (The women's and infants' Medicaid/CHIP IDs can be deduplicated separately.) Experience has shown that it is more efficient to deduplicate Medicaid/CHIP IDs before linking to other files to minimize the occurrence of two Medicaid/CHIP IDs being matched to the same birth certificate. This will be a larger task in states that do not routinely correct Medicaid/CHIP IDs.

Records for twins (and other multiple births) are more difficult to deduplicate and match because twins typically share many variables in common, including last name, date of birth (DOB), case ID, address, telephone number, and race/ethnicity. Researchers typically handle the deduplication and matching of multiple births separately to ensure that multiple births are not inadvertently grouped together and considered the same person.[12]

At the end of the deduplication process, states should keep multiple records (that is, rows) for each woman or infant, and assign a unique ID to indicate records that belong to the same person.[13] Medicaid/CHIP IDs should also be corrected in the claims data so that all claims associated with the same person are used for the analysis. Because the changes to the Medicaid/CHIP IDs will affect the construction of new variables, all data preparation and cleaning steps should be completed after the Medicaid/CHIP IDs have been deduplicated.

Step 2: Prepare and Clean the Data

To facilitate data management during the linking process, researchers have found it useful to prepare the input data files following three basic steps:

1. Assign unique numbers to the rows in each data set, so that output files can always be merged back to the source data.
2. Drop variables that will not be used for matching or analysis.
3. Deduplicate completely redundant (identical) rows.

The data should also be examined to ensure consistency in variable names and coding. Three coding issues frequently require attention before data linking can take place:

1. Common variables should be recoded to have the same structure in the two files. For example, if one file records DOB as three separate fields (month, day, and year) and the other file uses one field, the two files should be made to conform to the same format.
2. Variables should use common coding schemes across the two files. For example, male and female should be coded using the same values in both files.
3. In general, the same variables, such as DOB, should have identical variable names in each file.

Next, the Medicaid/CHIP data should be restricted to the population of interest.[14] That is, the mothers' file should include only women with deliveries during the measurement period and the infants' file should include only infants with DOBs during the measurement period. This will eliminate the consideration of extraneous records throughout the linkage process and reduce the probability of false matches. However, care should be taken to avoid inadvertently removing records from the file that should, in fact, be linked (for example, a typo in an infant's DOB should not cause that infant to be excluded).

[11] A Medicaid enrollee can have more than one Medicaid ID for various reasons, including when a state issues a temporary ID to infants; a name, address, or telephone number changes; a child enters foster care; a former enrollee reapplies for Medicaid under a different eligibility category; or typos prevent the data system from recognizing that two records belong to the same person.

[12] Some researchers opt to exclude multiple births from data files because they are more complicated to deduplicate and match. In the child Core Set technical specifications, multiple births are excluded from the denominator for the Cesarean section measure, but are included in the low birth weight measure.

[13] The deduplicaton process aims to identify individuals with multiple Medicaid IDs and assign a unique ID to an individual to enable grouping of all of his or her records. Some individuals will have more than one record, even after the deduplication process. For example, a person might have two addresses in the Medicaid file, but it is not possible to know which of the two addresses (if either) will appear on the birth certificate. Thus, keeping the two records increases the chance the Medicaid records would be matched successfully to the corresponding birth certificate.

[14] To ensure that all records for the same person are grouped together, this task should be conducted after the Medicaid IDs have been deduplicated.

The final task in this step is to assess data quality, in particular the extent of missing values or coding issues on variables used for matching. The Centers for Medicare & Medicaid Services (CMS) produces anomaly tables that describe data quality issues in each state's eligibility and claims data that can help states assess data quality.[15] Identifying the number of duplicates for variables that will be used in linking is another critical consideration. In Medicaid/CHIP data, this could include records with multiple Medicaid/CHIP IDs linked to a single Social Security number (SSN) or vice versa. These duplicates should be flagged, but kept in the data file.

Step 3: Linking Mothers and Infants in Medicaid/CHIP Records

Now that the Medicaid/CHIP IDs are deduplicated and the input data files have been prepared and cleaned, the next step involves linking a mother and an infant to each other in the Medicaid/CHIP files. The linking process is more straightforward in states where mothers and infants are given a common household ID (or case ID). States that do not have a household ID (or case ID) might still be able to achieve a high match rate because mothers and infants often share identifiers in common, such as last names, addresses, and telephone numbers. Matching typically involves combinations of variables, because no single variable is sufficient to determine a match. States might want to test different combinations of variables to see how many additional matches each combination produces.

There are two general approaches to linking records: deterministic and probabilistic matching methods. States can use either of these methods to link mothers and infants in Medicaid/CHIP records, or use both methods in combination. Exhibit A.2 describes each method and discusses advantages and disadvantages of each method. Section B of the technical appendix discusses software packages that are available for implementing these methods.

Exhibit A.3 describes variables frequently used for matching mothers' and infants' Medicaid/CHIP records. Although many of the variables for matching are contained in the Medicaid/CHIP eligibility files, some are available only in the claims files. These include the hospital (or facility) name on labor and delivery claims for mothers and infants, the date of service for the labor and delivery claims, and other labor and delivery characteristics, such as plurality or method of delivery. These variables should be created from the claims data and merged to the Medicaid/CHIP eligibility files so they can be used for the record linkage process.

States that can reliably link mothers to their infants in the Medicaid/CHIP files will benefit from performing this linkage before attempting to link to vital records. For example, match rates are higher when "Jane Doe" and her son "John Doe" are linked to the birth certificate using both names than when linking the mother to the birth certificate using only the mother's name. Many mothers and infants can be linked at this stage (often more than 90 percent), but there will be some unmatched mothers and some unmatched infants. All three of these groups are passed through to the next linkage step.

If linking mothers to their infants is not feasible, states can separately link mothers' and infants' Medicaid/CHIP records to the birth certificate file. This approach is sometimes called treating the birth certificate as a hub. Because this approach involves two data linkage steps (one for the mothers and one for the infants), it might not actually save effort compared with the linkage of Medicaid/CHIP records for mothers and infants.

Step 4: Matching Between Medicaid/CHIP and Vital Records Files

This step involves linking the mother-infant pairs, unmatched mothers, and unmatched infants in the Medicaid/CHIP files–that is, claims data–to their respective birth certificates.[16] As discussed earlier, many fields are available for linking the Medicaid/CHIP and birth certificate files, particularly if mothers and infants have already been linked in the Medicaid/CHIP files (Step 3). With probabilistic matching software packages that allow only a number of limited variables, care must be used to prioritize the match variables. Generally, variables should be chosen that identify unique match pairs (such as names, DOBs, and ID numbers). However, states should also include variables that will rule out false positives, even if those variables do help form the linkages (such as the infant's gender).

[15] Claims and eligibility anomaly tables for each year of Medicaid Analytic eXtract (MAX) data are available at https://www.cms.gov/MedicaidDataSourcesGenInfo/07_MAXGeneralInformation.asp (in the data downloads section).

[16] Research has shown that adverse infant outcomes might be understated when only the newborn hospitalization record (as opposed to Medicaid/CHIP claims data) is used for matching to birth certificates. Buescher (1999) found that rates of low birth weight and infant mortality were higher when Medicaid coverage was defined based on the Medicaid delivery record (claims data) than when it was based on the newborn hospitalization record (hospital discharge data).

Exhibit A.2. Description of Deterministic and Probabilistic Matching Methods for Linking Medicaid/CHIP and Vital Records Data

Deterministic Matching Methods

For the deterministic matching method, specific data elements must match exactly to create a link between a pair of records. This approach uses multiple variables to establish a match between records, although linkage validity is higher when individuals in the data sets have unique identifiers that are shared across the data sets for linkage, such as SSNs (Czajka et al. 2010). Because unique identifiers could be subject to error, however, additional confirming variables are typically used in the matching process. For example, the following sets of variables could be used for deterministic matching:

1. Mother's SSN, DOB, last name, first name, and county for deduplicating Medicaid/CHIP IDs
2. Medicaid/CHIP household ID, DOB/date of delivery, last name, county, and hospital of birth for matching mothers and infants in Medicaid/CHIP records
3. Mother's SSN, DOB/date of delivery, infant's first name, infant's last name, mother's DOB, mother's last name, hospital of birth/delivery, and county of residence for matching mother-infant dyads to vital records
4. Mother's SSN, DOB/date of delivery, mother's DOB, mother's last name, hospital of birth/delivery, county of residence for matching unmatched mothers to vital records

Deterministic matching is relatively easy to perform. For computing efficiency, the two files should be stacked and sorted by the match variables. It is common to conduct multiple rounds of deterministic matching, starting with strict matching criteria, performing the deterministic match, setting aside the linked pairs, and then repeating the process with successively weaker matching criteria. For example, the first round might require a match on SSN, DOB, last name, first name, and zip code, whereas later rounds might relax the criteria to search for records that match on SSN, DOB, last name, first initial, and county. For each set of records in which the match variables are identical, a new deterministic match ID variable should be created to facilitate tracking of record linkages.

The quality of the match will depend on how many rounds of deterministic matching are performed (there could be dozens of rounds) and which variables are used in each round. As discussed below, developing a process to assess the quality of the linkages at each round is highly recommended.

Probabilistic Matching Methods

The probabilistic method (also known as stochastic or fuzzy matching) identifies matches between two data files based on the likelihood that a group of variables in the two files represent the same person, though variables do not have to match exactly. Probabilistic record linkage identifies matches between two data files based on a comparison of multiple data fields (match variables) in the two files, and many variables can be included in the routine in most software packages. Probabilistic matching is typically applied when there is no common unique identifier across the data sets being linked and matching must be based on data elements that can be difficult to match exactly due to discrepancies in how these elements are recorded (such as names and addresses).

In a probabilistic matching process, every record in one file is paired with every record in the other file.[a] For each pair, each variable is compared to determine whether the variables match. If the variables match, the pair receives a positive weight; if they do not match, they receive a negative weight. After the weight for each variable is calculated, the weights are summed to obtain an overall score for each pair. Pairs with higher overall scores indicate a better match than pairs with lower scores. The matched pairs are reviewed and cutoffs are determined to classify matched records into certain matches, uncertain matches, and certain nonmatches based on their overall score. Certain matches are automatically accepted and certain nonmatches are automatically rejected. The uncertain matches undergo manual review.

Weights are higher for variables with more specificity (such as DOB or family name), and lower for variables with less specificity (such as sex, race, or county of residence). The specificity of a variable is the probability that a pair is a true match in cases in which a pair matches on that particular variable. It is also possible to assign a partial weight in cases in which the variable matches closely but not exactly. For example, a partial score could be given to first names that do not match exactly but are given the same phonetic algorithms for coding name fields,[b] SSNs, or telephone numbers that are one digit off, or addresses that do not match but are geographically close.[c]

Combining Deterministic and Probabilistic Matching Methods

Deterministic and probabilistic matching methods are not mutually exclusive, and states can use a combination of the two approaches to link two files, often saving time by combining the two approaches. Typically, many observations will match exactly on many or all of the key identifying variables. These records can be matched easily with a single round of deterministic matching using strict criteria, and set aside. The remaining records can be matched using probabilistic matching methods. This approach focuses attention on the records that are most difficult to match and the probabilistic matching step will benefit from faster computation times due to the smaller files.[d] This approach avoids the resource-intensive process of developing an exhaustive list of idiosyncratic, ad hoc deterministic matching rules, although it requires the presence of one or more unique identifiers (such as SSN).

[a] Probabilistic matching can be computationally intensive because billions of pairs have to be assessed for even small data sets. To improve computation times, the data can be blocked—for example, by county—and matches and nonmatches are identified only for the pairs in the block. Blocking variables should be specific; if blocks are too large, then blocking will not substantially reduce computation times. To avoid false matches, data should be blocked multiple times using different variables (or combinations of variables), and the results should be combined.

[b] For example, the New York State Identification and Intelligence System and the phonetic SOUNDEX algorithms can be used for last names and first names, respectively.

[c] Weights will have to be obtained from a training data set or another source. A variety of approaches are available to develop and refine the weights, including the expectation-maximization algorithm. Advanced methods are available to account for relative frequencies within a variable, such as the fact that the last name Smith might appear more frequently than Vijayan.

[d] If splitting the file is inconvenient, the two approaches can be combined using a single file in two steps: (1) creating a deterministic match ID variable in a deterministic matching step and (2) including this ID as a match variable in a probabilistic matching step. To force all pairs matched at the deterministic linkage step to remain linked in the probabilistic linkage step, assign the ID variable an infinite (or very high) weight.

Exhibit A.3. Description of Variables Frequently Used to Link Medicaid/CHIP and Vital Records Data

Variable	Description
Mother's Medicaid/CHIP ID	States are required to assign a unique Medicaid/CHIP ID number to each enrollee. Some state birth certificates note a mother's Medicaid/CHIP ID number on the infant's birth certificate. Medicaid/CHIP IDs can be used to group multiple records for the same person, which is common in most eligibility files.
Mother's SSN	CMS instructs states to report available SSNs for Medicaid/CHIP enrollees and to indicate whether the state has verified the SSN with the Social Security Administration.[a] SSNs for mothers are routinely collected by vital records, but are rarely (if ever) collected for infants. However, some vital records agencies cannot release SSNs.
Mother's DOB	Reporting of an enrollee's DOB is generally complete in Medicaid/CHIP administrative data (states reported unknown DOBs for less than 1 percent of enrollees in MAX 2008). However, if a state uses CHIP funds to cover unborn children, these children might be identified in Medicaid/CHIP data under a pregnant woman's DOB or be reported as enrolled with an unknown DOB and sex. States that choose to assign enrollment to the unborn child might have relatively high numbers of enrollees with unknown DOBs.
Mother's Name	Enrollees' names are generally available in Medicaid/CHIP data, although researchers have identified challenges of using names to link records across data sets. First, names are not unique and multiple beneficiaries can have the same first and/or last name. Names also can be entered incompletely or differently in multiple databases. For example, a Medicaid/CHIP enrollee might appear in the Medicaid Management Information System under a nickname, but in vital records data that person might be reported with a full first name and middle initial. Linkage techniques can account for these discrepancies to maximize the accuracy of matches.[b]
Address	Full residence information can be available in Medicaid/CHIP data, though some information might be missing. States have used addresses, or elements of addresses such as street names or cities, to link administrative data. Similar to linkages based on names, these matches can be affected by data entry errors, abbreviations of street or city names, or by residence changes not reflected in all data sets.
Labor, Delivery, and Birth Information, Including Date of Delivery and Provider's Name	The Medicaid/CHIP claim(s) covering labor and delivery for the woman and the claims(s) covering the birth of the infant can also be used to link mother's and infant's Medicaid/CHIP records. Estimating the date of delivery to within a few days can help to link mothers to the correct infant or birth certificate. A hospital (or provider) name or ID number and type of provider (for example, hospital versus birth center) can be used to link records. Relatively rare outcomes, such as newborn intensive care unit admissions, plurality greater than one, or specific congenital anomalies, might have higher positive values, but might have low sensitivity due to measurement error.
Other Characteristics	Other characteristics that can assist with linking Medicaid/CHIP records to vital records include race, ethnicity, Medicaid/CHIP eligibility category (such as dual eligibility), family size, family income, mother's foreign-born status, and mother's education. Other characteristics that can be used to link Medicaid/CHIP records for mothers and their infants include household IDs (or case IDs) and telephone numbers.[c]

[a] An analysis of 2008 MAX data found that about 10 percent of Medicaid/CHIP enrollee records were missing SSNs (Borck et al. 2012). In most states, however, SSNs were missing or invalid for fewer than 2 percent of enrollees. A 2009 Medicaid Statistical Information System analysis found that SSNs were in the valid range for at least 99 percent of enrollees in 12 states, and for 93 to 95 percent of enrollees in 7 states (Czajka and Verghese 2013). In the remaining states, SSNs were valid for 95 to 99 percent of enrollees. However, more than one-quarter of infants enrolled in Medicaid/CHIP with full benefits were missing SSNs or had an SSN outside the valid range.

[b] Other challenges with matching by name include (1) data collection systems are built for Western naming conventions (that is, first name, middle initial, last name) rather than non-Western conventions; (2) people often change names (commonly with a marriage or divorce) and data often include a mother and baby pair with different names in the Medicaid/CHIP data and the vital statistics data; (3) affixes are often used—for example, a woman might change her maiden name to her middle name upon marriage; and (4) suffixes can appear in the last name field (for example, Smith Jr.).

[c] Telephone numbers can be used to deduplicate Medicaid/CHIP IDs and link mothers to infants. This is helpful, for example, deduplicating infants' Medicaid/CHIP ID numbers in cases in which the file is missing a large number of SSNs and/or Medicaid/CHIP ID numbers change frequently. However, telephone numbers often change, so two records could match even if they do not have the same telephone number.

Step 4: Matching Between Medicaid/CHIP and Vital Records Files

This step involves linking the mother-infant pairs, unmatched mothers, and unmatched infants in the Medicaid/CHIP files–that is, claims data–to their respective birth certificates. [17] As discussed earlier, many fields are available for linking the Medicaid/CHIP and birth certificate files, particularly if mothers and infants have already been linked in the Medicaid/CHIP files (Step 3). With probabilistic matching software packages that allow only a number of limited variables, care must be used to prioritize the match variables. Generally, variables should be chosen that identify unique match pairs (such as names, DOBs, and ID numbers). However, states should also include variables that will rule out false positives, even if those variables do help form the linkages (such as the infant's gender).

In order to link mothers and infants who remain unmatched in the Medicaid/CHIP and birth certificate files, there are two common approaches, although the basic idea is the same: unmatched mothers and unmatched infants are independently matched to the birth certificates, and mothers and infants linked to the same birth certificate are treated as a mother-infant dyad:

1. One approach is to first link the matched mother-infant Medicaid/CHIP records to the birth certificates. Then, the remaining (unlinked) birth certificates are linked to the unmatched mother's Medicaid/CHIP record, and finally these birth certificates are also linked to the unmatched infant's Medicaid/CHIP records. This approach will be required for probabilistic matching software packages that return only one matched pair for a given Medicaid/CHIP record or birth certificate (that is, one-to-one matching).
2. Another approach is to stack the Medicaid/CHIP records for mother-infant dyads, unmatched mothers, and unmatched infants. The stacked file is merged to the birth certificates, and the algorithm can produce multiple matches for each birth certificate (that is, many-to-many matching). Post-processing then assigns birth certificates to the Medicaid/CHIP records with the highest scores, allowing unmatched mothers and unmatched infants to be linked to the same birth certificate.

[17] Research has shown that adverse infant outcomes might be understated when only the newborn hospitalization record (as opposed to Medicaid/CHIP claims data) is used for matching to birth certificates. Buescher (1999) found that rates of low birth weight and infant mortality were higher when Medicaid coverage was defined based on the Medicaid delivery record (claims data) than when it was based on the newborn hospitalization record (hospital discharge data).

Step 5: Assessing the Data Linkage Results

Before completing the linkage, states should assess the results to determine whether the linkage correctly identified all of the possible matches in the two data sets. States can assess the quality of their matches in several ways: First, states can begin by comparing demographics, risk factors, and other variables between matches and nonmatches. This process can identify any systematic differences between these two populations, which could provide insight into potential discrepancies between anticipated and actual rates and the reliability of the rate calculated, or if significant bias is introduced by the match method chosen. Another way to assess the method is to conduct sensitivity analyses to determine whether the set of variables included in the model is found to be robust by including or excluding other variables for matching. These analyses can help determine whether the set of chosen variables used to conduct the matching is the most optimal for identifying true matches.

The quality of a data linkage is often measured in terms of the match rate—that is, the percentage of cases that are linked between two data sets. There are two basic approaches to calculating a match rate. The first is to compute the rate using a test data set in which true linked pairs are known. The advantage of this approach is that it reveals the extent to which a matching algorithm results in false positives (that is, matching records that are not true matches). This approach, however, would require developing a test data set with known matches. Test data sets, although very useful for assessing a data linkage process, are rarely available.

Match rates are more commonly calculated based on the percentage of Medicaid-covered births linked to a birth certificate. Because this method cannot isolate false-positive matches, it could overstate the match rate. Though states will use different methods to conduct the linkage, comparisons between states can help identify significant discrepancies and potential errors in the match process.[18] Match rates typically are in the range of 90 to 99 percent, although no systematic review currently exists. For example, researchers in Minnesota achieved a 93.2 percent match rate between birth certificates and Medicaid/CHIP data using an iterative deterministic matching process. States should also review match rates by demographic characteristics or subsets of the population to identify if certain subgroups had higher or lower match rates than others. This can provide insight into the reliability and validity of the matching method used across all subgroups.

[18] The match rate can vary greatly depending on the population of interest, the outcome being studied, the availability and quality of the linking variables, and the methodology used to conduct the linkage. For example, if the vital records file does not include out-of-state births, attaining a 100 percent match rate might be impossible.

Exhibit A.4 shows the results of an iterative deterministic matching process in Minnesota, using various criteria and constraints. These results illustrate the importance of using multiple variables and multiple algorithms. The cumulative match rate increased more than 12 points across the 14 steps, from 80.6 percent to 93.1 percent (Gyllstrom et al. 2002). The authors reported that Hispanic women matched at lower rates than other races/ethnicities, and that unmatched populations were concentrated in particular geographic areas.

Exhibit A.4. Iterative Deterministic Match Procedure and Resulting Match Percentages in 1997 Minnesota Medicaid and Vital Records Linkage

Matching Criteria	Cumulative Match Rate	Additional Match Percentage
1. Exact first and last names of mother, mother's exact DOB, admission date within two months of child's DOB on birth certificate	80.62	–
2. Index function[a] on last and first names, mother's DOB, and admissions date within two months of child's DOB	83.43	2.81
3. Mother's last name (Medicaid) and mother's maiden name, along with other constraints from Step #2	84.73	1.30
4. Same as Step #3, except mother's DOB removed	85.22	0.49
5. Same as Step #3, except mother's month of birth removed	85.22	0.00
6. Index function[a] using mother's last name (Medicaid) with father's last name (birth certificate), mother's month/year of birth, and admission within 2 months of baby's DOB	90.42	5.20
7. Introduction of "Sounds Like Operator" upon matching first and last names, with mother's month/year of birth and admission date within 2 months of baby's DOB	91.65	1.23
8. Use first three months of 1998 birth data, using index of mother's name with month/year of mother's birth and baby's DOB within 3 months of admission date	91.77	0.12
9. Using 1997 birth data, expanding admission date field to 3 months before baby's DOB	91.78	0.01
10. Using 1998 births with index of father's last name on birth certificate and other constraints from Step #9	91.80	0.02
11. Using 1997 fetal death certificates, with the index of child's last name on fetal death certificate with mother's last name, month and year of mother's birth, and fetal death date within 3 months of admission	92.03	0.13
12. Birth certificates for 1997 non-Minnesota residents[b] who gave birth in Minnesota with the index of mother's name, mother's month/year of birth, and baby's DOB within 3 months of delivery	92.24	0.21
13. Addition of 1996 admission dates to search for births that occurred in 1997	92.30	0.06
14. Addition of child's last name from birth certificate to compare with mother's last name on Medicaid file	93.10	0.80

Source: Table reproduced from Gyllstrom et al. 2002.

[a] Index function is useful in cases of combined maiden name and surname, shortened first names, or imbedded numerals or titles, such as junior. The index function searches within a designated field for the nth place where the value of another data element is found.

[b] People living in Minnesota and considered residents from a Medicaid eligibility standpoint might not report Minnesota as their state of residence on the birth certificate.

Step 6: Preparing the Linked Data File

The final data file should have one record per woman and infant, which requires reconciling duplicates and discrepancies. States can reconcile values based on known data quality issues, measure specifications, or other factors. For example, the Cesarean section measure specifications identify the type of delivery through International Classification of Diseases, 9th edition (ICD-9) diagnosis codes; therefore, though the type of delivery can be captured on both Medicaid/CHIP files and vital records, the measure technical specifications would prompt a state to reconcile based on the diagnoses in the Medicaid/CHIP data. Crosswalk files are then created to link the raw Medicaid/CHIP files to the vital records data and the final paired mother/infant IDs.

B. Software Packages for Linking Medicaid/CHIP and Vital Records Data

Many software packages are available to facilitate linkage of Medicaid/CHIP and vital records data. Given scarce staff resources, these packages can reduce the burden on programming staff and ultimately save staff time. This TA brief references the software packages for informational purposes only; CMS does not endorse or recommend the software mentioned here. Three common packages include the following:

1. Link Plus is a probabilistic record linkage program developed by the Centers for Disease Control and Prevention (CDC) for use by its National Program of Cancer Registries. It is a standalone application for Microsoft Windows® that can be used with any type of data in a fixed width or delimited format to identify duplicates or link files from two different sources. The software can be obtained free of charge from the CDC website (http://www.cdc.gov/cancer/npcr/tools/registryplus/lp.htm).
2. LinkageWIZ uses probabilistic matching algorithms to match records across multiple data sources and enables the identification of duplicate records. LinkageWIZ can import data from several types of databases, including Microsoft Excel and Access. The cost of the software starts at $199 (http://www.linkagewiz.net/), depending on the size of the data set being linked.
3. Link King works with SAS and is available for free download (http://the-link-king.com). Users do not need SAS programming experience. Link King can also match data in a variety of other formats, including SPSS portable files, Excel spreadsheets, and comma-/tab-delimited files (Campbell 2013; Campbell et al. 2008).

Other commercial record linkage software programs include AutoMatch, Bigmatch, dfPowerStudio, FRIL, GRLS, HDI, LINKS, and QualityStage, as well as a suite of software available from the U.S. Bureau of the Census (Newman and Brown 1997; Winkler 2001; Ferrante and Boyd 2012). In addition, probabilistic matching software packages are available for several commonly used computer languages, including Stata (the reclink package), the R language (the RecordLinkage package), and Python (the Freely Extensible Biomedical Record Linkage project). These packages can be customized for the linkage of Medicaid/CHIP and vital records data, but they require experience with these programming languages. States that prefer to write their own software should consider using computing language that stores the data in memory (not on the hard disk), such as Stata, Matlab, or the R programming language.

Federal matching payments may be available to support the purchase of record linkage software and professional services for data processing and analysis. States could expand their existing contracts to conduct ongoing utilization review and data analysis to support Medicaid/CHIP and vital records linkage through an administrative contract under Section 1903(a)(7) of the Act and receive a 50 percent federal matching payment. These costs, along with all other allowable costs for Medicaid administrative activities performed by the state, must be "necessary for the proper and efficient administration of the state plan."[19]

C. Comparability of the 1989 and 2003 Versions of the Birth Certificate

Two factors affecting the use of vital records for quality measurement and quality improvement include the version of the birth certificate used in the state and the reliability of necessary fields. The 2003 version of the birth certificate includes many key indicators not found in the 1989 version that can enable more robust measurement of key maternal and infant health quality indicators, including vaginal birth after Cesarean section, whether the pregnancy resulted from infertility treatment, and the presence of certain procedures during and after delivery (such as epidural or spinal anesthesia during labor and administration of steroids for fetal lung maturation). Not all states, however, have adopted the revised birth certificate. [20]

[19] Additional information on federal matching payments to support Medicaid/CHIP and vital records linkage is available at http://medicaid.gov/Federal-Policy-Guidance/Downloads/CIB-07-24-2013.pdf.

[20] As of January 1, 2012, 38 states (California, Colorado, Delaware, Florida, Georgia, Idaho, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maryland, Massachusetts, Michigan, Minnesota, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, South Carolina, South Dakota, Tennessee, Texas, Utah, Vermont, Washington, Wisconsin, and Wyoming), the District of Columbia, Guam, Puerto Rico, and Northern Marianas had implemented the revised birth certificate (Hamilton et al. 2013).

Exhibit A.5 lists the items included in the 2003 birth certificate, indicating which items were included in the 1989 version (and whether the 2003 version is comparable), as well as which items were introduced with the 2003 version. Although most states have now adopted the 2003 version of the birth certificate, the comparability and reliability of these data points across hospitals and states is still being explored. A recent report comparing the frequency with which fields on the birth certificate matched medical records found substantial variation in accuracy among eight hospitals in two states on many key items. Overall, low levels of agreement were found on the total number of prenatal care visits, history of previous preterm birth, and the presence of gestational diabetes or hypertension (Martin et al. 2013). One state had consistently higher agreement among hospitals on certain fields (such as estimated gestational age at birth, outcomes of previous pregnancies, and whether labor was induced or augmented). These findings, as well as findings from other studies, suggest that further standardization is necessary before vital records can reliably supplement current data sources for monitoring maternal and child health care quality.

Exhibit A.5. Comparability of Variables on the 1989 and 2003 Versions of the Standard Birth Certificate

Item on 2003 U.S. Standard Certificate of Live Birth	2003 Comparable to 1989?		New Item in 2003
	Yes	No	
Race – mother/father	✓	.	.
Hispanic origin – mother/father	✓	.	.
Education – mother/father	.	✓	.
Maternal marital status	✓	.	.
Cigarette smoking during pregnancy	.	✓	.
Month prenatal care began	.	✓	.
Primary source of payment for delivery	.	.	✓
Type of birth facility	✓	.	.
Number of prenatal care visits	✓	.	.
Type of birth attendant	✓	.	.
Dates of first and last prenatal care visits	.	.	✓
Risk Factors in this Pregnancy	.	.	.
Diabetes (prepregnancy and gestational)[a]	✓	.	.
Hypertension (prepregnancy, gestational, eclampsia)	✓	.	.
Previous preterm birth	.	✓	.
Other previous poor pregnancy outcome	.	✓	.
Mother had previous Cesarean delivery	.	✓	.
Maternal weight gain during pregnancy	.	.	✓
Pregnancy resulted from infertility treatment	.	.	✓

Item on 2003 U.S. Standard Certificate of Live Birth	2003 Comparable to 1989?		New Item in 2003
	Yes	No	
Obstetric Procedures			
Cervical cerclage		✓	
Tocolysis	✓		
External cephalic version			✓
Onset of Labor			
Premature rupture		✓	
Precipitous labor	✓		
Prolonged labor		✓	
Characteristics of Labor/Delivery			
Induction of labor	✓		
Augmentation of labor		✓	
Nonvertex presentation			✓
Steroids for fetal lung maturation			✓
Antibiotics received by the mother during labor			✓
Clinical chorioamnionitis diagnosed during labor		✓	
Moderate/heavy meconium staining of the amniotic fluid	✓		
Fetal intolerance of labor		✓	
Epidural or spinal anesthesia during labor			✓
Method of Delivery			
Forceps attempted		✓	
Vacuum extraction attempted		✓	
Cephalic presentation		✓	
Breech/other presentation[b]	✓		
Final route and method of delivery[c]	✓		
If Cesarean, was trial of labor attempted			✓
Newborn Information			
Birthweight	✓		
Apgar score – 5 minutes	✓		
Plurality	✓		
Gestational age	✓		
Breastfeeding at discharge			✓

Item on 2003 U.S. Standard Certificate of Live Birth	2003 Comparable to 1989?		New Item in 2003
	Yes	No	
Abnormal Conditions of Newborn	.	.	.
Assisted ventilation required immediately following delivery	✓	.	.
Assisted ventilation > 6 hours	.	✓	.
Neonatal intensive care unit admission	.	.	✓
Newborn given surfactant replacement therapy	.	.	✓
Antibiotics received by the newborn for suspected neonatal sepsis	.	.	✓
Seizure or serious neurologic dysfunction	.	.	✓
Significant birth injury	.	.	✓
Congenital Anomalies	.	.	.
Anencephaly	✓	.	.
Meningolcele/spina bifida	✓	.	.
Cyanotic congenital heart disease	.	.	✓
Congenital diaphragmatic hernia	✓	.	.
Omphalocele[d]	✓	.	.
Gastroschisis[d]	✓	.	.
Limb reduction defect	.	.	✓
Cleft lip with or without cleft palate[e]	✓	.	.
Cleft palate alone[f]	✓	.	.
Down Syndrome	✓	.	.
Down Syndrome – karyotype pending or confirmed	.	.	✓
Suspected chromosomal disorder	✓	.	.
Suspected chromosomal disorder - karyotype pending or confirmed	.	.	✓
Hypospadias	.	.	✓

Source: Detailed Technical Notes – U.S. 2010 Natality. ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/DVS/natality/UserGuide2010.pdf.

[a] Prepregnancy diabetes and gestational diabetes can be combined to be consistent with the diabetes item reported on the 1989 U.S. Standard Certificate of Live Birth.

[b] Breech and other fetal presentations at birth can be combined to be consistent with the breech/ malpresentation item on the 1989 U.S. Standard Certificate of Live Birth.

[c] Information on whether the vaginal delivery following a previous Cesarean delivery (vaginal birth after Cesarean) or whether the delivery was a primary or repeat Cesarean is not comparable.

[d] Omphalocele and gastroschisis can be combined to be consistent with the omphalocele/gastroschisis item on the 1989 U.S. Standard Certificate of Live Birth.

[e] Cleft lip with or without palate can be combined with cleft lip alone to be consistent with the cleft lip/palate item on the 1989 U.S. Standard Certificate of Live Birth.