

Best Practices for Linking Medicaid and Separate CHIP Eligibility Records

Stephen Kuncaitis, Paul M. Montebello, Matthew Hodges

For many years, policymakers and others with an interest in public health insurance have had access to person-level eligibility and claims data for the Medicaid and Medicaid expansion Children's Health Insurance Program (CHIP) through the Medicaid and CHIP Statistical Information System (MSIS) and the Medicaid Analytic eXtract (MAX) system—the only sources for uniform Medicaid claims and eligibility data for all 50 states and the District of Columbia. Comparable person-level data for separate CHIP programs have not been available. To satisfy a growing need for separate CHIP data, CMS contracted with Mathematica Policy Research to provide technical assistance (TA) to states for reporting these data to MSIS. In this issue brief, we summarize best practices for integrating Medicaid and separate CHIP data into MSIS. States may find this guidance useful in reporting separate CHIP data to MSIS and in preparing for major eligibility system changes associated with implementation of health care exchanges and streamlined eligibility determination processes.

Introduction

CMS currently encourages states to report separate Children's Health Insurance Program (CHIP) data to the Medicaid and CHIP Statistical Information System (MSIS). Doing so requires states to link and merge records for individuals present in both Medicaid and separate CHIP eligibility and claims systems. In this issue brief, we describe best practices utilized by states that have successfully integrated Medicaid and separate CHIP data into MSIS. These practices are widely applicable to other states with different eligibility and claims systems and with varying degrees of program connectivity.

There are two major benefits to having separate CHIP data included in MSIS. First, CHIP is a major coverage program for children; having separate CHIP data in MSIS makes it possible to tell a more complete story about public coverage for low-income children. Second, the inclusion of CHIP eligibility and claims data offers researchers a greater understanding of how CHIP enrollment and utilization compare to those of Medicaid. Furthermore, methods for assigning common identifiers may

About This Series

The MAX Medicaid and CHIP policy issue brief series highlights the essential roles the Medicaid and CHIP Statistical Information System (MSIS) and the Medicaid Analytic eXtract (MAX) play in program analysis. MAX is a set of annual, person-level data files on eligibility, service utilization, and payments that are derived from state reporting of eligibility and claims data to MSIS. MAX is an enhanced, research-friendly version of MSIS that includes final adjudicated claims based on dates of service, reconciled eligibility status, and data that have undergone additional quality checks, corrections, and enhancements. As part of a CMS-sponsored technical assistance (TA) effort, this issue brief provides basic guidance to states for reporting separate Children's Health Insurance Program (CHIP) data to MSIS. For more information about technical assistance for CHIP reporting, please visit: <http://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/S-CHIP.html>.

also help states prepare for implementation of the Affordable Care Act of 2010 (ACA), which calls for fully integrated and seamless eligibility determination and enrollment processes across Medicaid, CHIP, and subsidized health plans offered through health insurance exchanges.

Background

CHIP programs fall into two categories: Medicaid expansion CHIP and separate CHIP. Eligibility and claims management systems for separate CHIP programs are often distinct from Medicaid. As a result, integrating information for individuals in Medicaid and separate CHIP requires methods and procedures for linking records for individuals enrolled over time in both programs.

States are not required to submit data on separate CHIP programs to MSIS; however, since October 1, 2010, CMS has encouraged them to submit complete eligibility and claims data

for separate CHIP programs in the format used for reporting Medicaid and Medicaid expansion data.¹ Prior to this, states were only able to report a subset of separate CHIP eligibility, and no claims data. Of the 43 states with separate CHIP programs, 6 states have completed or initiated expanded MSIS reporting, 21 states report only limited separate CHIP data, and another 16 report no separate CHIP data.²

As more states opt to report separate CHIP data to MSIS, there will be a continued need to ensure consistency with MSIS reporting standards, and systems will need to assign person-level MSIS identifiers that are unique across programs and over time. For example, a child enrolled in both Medicaid and separate CHIP during a particular fiscal quarter may have enrollment in the former for one month and the latter for two. Both enrollment spells should be reported with one unique MSIS person identifier.

To achieve successful data integration when reporting to MSIS, states with separate CHIP programs could face several challenges. First, Medicaid and separate CHIP may each assign a different unique personal identifier to an individual. Second, although MSIS files are submitted to CMS by a single state agency, multiple entities—including outside contractors—are often involved in generating these data. Adequately linking and merging separate CHIP and Medicaid data can pose logistical challenges for states when the two programs are administered by separate entities, agencies, or divisions. Third, MSIS reporting is designed to maintain historical consistency by using the same unique enrollee identifier over time. For this to happen, the identifier first assigned to a person and reported to MSIS needs to remain constant for all future reporting. This issue brief provides guidance to address each of these challenges.

Methods

The recommendations reported here are based on case studies of three states (Arizona, Michigan, and Virginia) that have separate eligibility determination systems for Medicaid and separate CHIP. They also have substantial experience in matching, linking, and merging eligibility data for both Medicaid and separate CHIP. We conducted telephone interviews with state staff whose job responsibilities include data quality and systems maintenance.

Our questions were designed to assess state experience with data integration—that is, linking and merging data—for Medicaid and separate CHIP. Our questions addressed the following topics:

- History of the state’s eligibility systems for Medicaid and separate CHIP
- Methods used to match and link records, and how often they are employed

- Use of personal identifiers (how they are assigned; whether multiple identifiers are used and, if so, what is done to link them with a single, primary identifier)
- Methods used in data merging
- Obstacles, problems, or risks associated with data integration
- Advice for other states exploring options for integrating Medicaid and separate CHIP data

Findings

Interview responses revealed three steps that support successful integration of Medicaid and separate CHIP systems: (1) linking data downstream, (2) using algorithms to match records, and (3) developing a hierarchy to establish a primary unique identifier. Before a state can assign a unique identifier, it must determine where to perform the linkage, considering system and staff resources and the data elements available. Once it has completed this task, it can develop algorithms and a hierarchy to establish a unique, permanent MSIS ID.

Linking Data Downstream

Linking is a method used to create identifier keys that connect data among separate sources. Because eligibility determination is a dynamic process, maintaining unique person identifiers over time—either within a single system or across multiple systems—is a challenge. Address changes occur frequently, names are often misspelled on applications or when manually entered, and a parent may substitute his or her own SSN if a newborn does not yet have one. Consequently, reconciling the identities of enrollees with multiple identifiers is time-consuming and error prone.

In addition, it is impractical to assign and utilize unique identifiers across multiple systems that operate independently, especially when programs are administered by different agencies. Because the systems supporting eligibility determination for Medicaid and CHIP often operate independently, a person could exist in more than one system with more than one “primary” identifier.

Given these circumstances, the optimum place to link individuals should be as close as possible to the MSIS data extraction point—that is, “downstream” rather than “upstream,” where an enrollee first enters the eligibility system. Generally, this linkage would occur in the state’s Medicaid Management Information System (MMIS). Linking data downstream can (1) reduce the time and effort required to reconcile and de-duplicate multiple identifiers for the same person, (2) consolidate the effort so it occurs only once, and (3) allow for greater consistency in the linking approach employed.

Arizona, Michigan, and Virginia have all implemented downstream linking procedures. Each state links eligibility information

in their MMIS or in a downstream data warehouse, rather than in their upstream eligibility determination systems. Downstream linking gives eligibility staff more time to focus on determining eligibility. Eligibility determination systems and processes are otherwise unaffected.

Using Algorithms to Match Records

Matching is a method of assigning a probability that data residing in separate sources represent the same person. If there is no common unique identifier across the data systems, the matching process must use other identifying information. Algorithms are formulas that utilize a predetermined set of rules to match records for the same person across different systems. Automated algorithms use advanced code and require minimal manual intervention. They match individuals on the basis of personal and demographic data elements—for example, last name, date of birth, gender, and Social Security number (SSN)—using various methods. The methods may be refined by adding numeric weights and probabilities for specific elements, and can accommodate subtle differences in the spelling of names, as well as partial address matches. Certain data elements or combinations of data elements may provide better match probabilities and may weigh more heavily in determining whether two records represent the same person. A match based on full name and SSN, for example, is better than one based on a partial last name match with a partial date of birth (for example, year and month of birth). Automated algorithmic matching processes usually produce a relatively small number of questionable matches that staff can reconcile and link manually.

Arizona uses a sophisticated matching process to automatically link enrollment records for each person as they are loaded into its MMIS. In Michigan, Medicaid and separate CHIP enrollment information is passed through the MMIS in parallel and then linked in a data warehouse, where a matching algorithm identifies duplicate records and links them. Virginia uses manual data queries during eligibility determination to identify

matching records across systems. They look forward to eliminating manual matching processes as they update their systems to integrate with a health care exchange. In addition, Arizona, Michigan, and Virginia employ data quality processes to further ensure that duplicates do not exist after all linking has been completed.

Table 1 offers examples of algorithms that can be used to identify multiple records that represent the same person by matching various identifiers. This process can be executed in a stepwise progression from the most likely matches to less reliable matches. For example, Match Key 1 uses full last name, full first name, full date of birth, and the exact SSN to establish an exact match between two sets of specific identifying variables. By comparison, Match Key 6 uses partial last name, partial first name, and full date of birth to establish a likely match between two sets of identifying variables. Less reliable matches may need further manual review and validation.

Developing a Hierarchy to Assign a Unique MSIS Identifier

Individuals often are eligible over time for both Medicaid and separate CHIP; if this occurs, they may be assigned a different person identifier in each system. In the future, health insurance exchanges will add a further level of complexity, because over time a person may be eligible for Medicaid, separate CHIP, *and* subsidized health care via an exchange. As a result, a single individual could have multiple records across all three programs. Integrating these data will require establishment of one primary unique identifier for each person across these systems.

The approach used to establish a person’s unique MSIS identifier will vary depending on whether the person is new to MSIS or was included in previous submissions. If an individual was ever enrolled in Medicaid and assigned an MSIS identifier, the state should use that as the permanent identifier for reporting

Table 1. Examples of Match Key Algorithms

| | Last Name | First Name | Date of Birth | Gender | SSN ^a | Result |
|-------------|----------------------|----------------------|------------------|--------|------------------|-----------|
| Match Key 1 | Exact (Full) | Exact (Full) | Exact (CCYYMMDD) | N/A | Exact | Auto-link |
| Match Key 2 | Exact (Full) | Exact (Full) | Partial (CCYYMM) | Exact | N/A | Auto-link |
| Match Key 3 | Soundex ^b | Soundex ^b | Exact (CCYYMMDD) | Exact | Exact | Auto-link |
| Match Key 4 | Partial ^c | Partial ^c | Exact (CCYYMMDD) | N/A | Exact | Auto-link |
| Match Key 5 | Soundex ^b | Soundex ^b | Partial (CCYYMM) | N/A | Exact | Review |
| Match Key 6 | Partial ^c | Partial ^c | Exact (CCYYMMDD) | N/A | N/A | Review |

^a States should use SSN when available.

^b A soundex categorizes values based on their phonetics. It can be used to find similar names differentiated only by variation in preferred spelling.

^c A partial name match could be limited to a certain number of characters from either or both name values. For example, a partial name match may only match on the first three characters of a first name and the first eight characters of the last name.

to MSIS. In some states, the MSIS identifier is the same as the identifier that appears on a beneficiary's identification card; however, other states may assign a different internal number to be used in MSIS. Alternatively, if an individual was enrolled first in separate CHIP, his or her separate CHIP identifier may be used as the primary unique identifier in MSIS, or a new number may be assigned. The important point is that once an MSIS identifier is assigned to an individual, that identifier should be used for all reports in the future, regardless of the program in which the person is enrolled.

In Michigan, the Medicaid Member ID is assigned by the Medicaid eligibility determination system and has always been the unique identifier used for MSIS reporting. Children enrolled in separate CHIP are assigned an identifier—known as the Client Identification Number, or CIN—that is unique within the separate CHIP eligibility determination system. Individuals who have been enrolled only in separate CHIP and never entered into the Medicaid eligibility determination system do not have a Medicaid Member ID for MSIS reporting. When separate CHIP-only enrollees are combined with Medicaid enrollees, the state includes a “C” as the first position of the CIN to differentiate it from the all-numeric Medicaid Member ID.

Michigan uses a hierarchy to determine whether the Medicaid Member ID or separate CHIP CIN is used as the unique and permanent identifier in MSIS. The state executed a one-time automated matching and linking of all records to establish a permanent identifier for individuals who have ever existed in both systems. During this process, individuals who had ever been present in the Medicaid eligibility determination system were assigned a Medicaid Member ID as their permanent, unique MSIS identifier. This ensured that any individual could be linked to records previously submitted to MSIS. After this initial link had been established and Michigan began reporting both Medicaid and separate CHIP enrollment and claims to MSIS, individuals who were first eligible for separate CHIP were then assigned the CIN as their primary unique identifier for MSIS reporting, even if they subsequently were eligible for Medicaid.

The hierarchy used by Michigan allows the state to maintain the continuity of its permanent MSIS identifiers and assign new permanent IDs to separate CHIP enrollees without making costly modifications to source systems. This is particularly important given that changes to eligibility determination systems will eventually occur in response to ACA, and any changes prior to that would only serve as temporary solutions.

Conclusions

Reporting separate CHIP data to MSIS requires planning, time, and organizational focus. Our discussions with staff in Arizona, Michigan, and Virginia identified three best practices for integrating Medicaid and separate CHIP data. First, link data from separate eligibility determination systems downstream. Second, establish automated processes using algorithms to match individuals with multiple identifiers. Third, develop a logical hierarchy to help select a primary unique identifier for MSIS reporting. By embracing these practices, states that opt to report separate CHIP data to MSIS can avoid unnecessary or duplicative efforts.

The ACA calls for major system changes to accomplish its goal of integrated, seamless eligibility and enrollment for Medicaid, separate CHIP, and subsidized coverage within health insurance exchanges. Assigning MSIS identifiers to individuals enrolling for the first time should be relatively straightforward, since the MSIS ID is indifferent to the source of the identifier or its associated program. Prior to any system conversion, each state is advised to develop a crosswalk that links all of the primary identifiers for individuals in different systems to a single permanent MSIS ID. The development of such a crosswalk in conjunction with the practices cited above will proactively position states to accurately report any individual over time without regard to the program in which he or she is enrolled at a specific point in time.

Mathematica is currently available to provide technical assistance to aid states with data integration efforts for separate CHIP reporting to MSIS. Each technical assistance arrangement considers a state's unique circumstances and needs. We have also developed resources that states may find useful. To contact Mathematica for technical assistance and other resources, email Paul Montebello at pmontebello@mathematica-mpr.com or Stephen Kuncaitis at skuncaitis@mathematica-mpr.com.

Endnotes

¹ In addition to this issue brief, Mathematica has developed a reporting guide for states that provides a more detailed overview of the MSIS data structure and submission process as well as guidance for reporting complete separate CHIP data in MSIS. This report can be found at <http://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/S-CHIP>.

² Alaska, the District of Columbia, Hawaii, Maryland, Nebraska, New Mexico, Ohio, and South Carolina do not operate separate CHIP programs.

For further information on this issue brief series, visit our website at www.mathematica-mpr.com

Princeton, NJ • Ann Arbor, MI • Cambridge, MA • Chicago, IL • Oakland, CA • Washington, DC

Mathematica® is a registered trademark of Mathematica Policy Research, Inc.