MOTHER-INFANT LINKAGE

FREQUENTLY ASKED QUESTIONS & APPENDICES

Prepared by the Sentinel Operations Center
February 2019
Version: 1.0.0

Table of Contents

Modification History

Version	Date	Modification	By
1.0.0	February 2019	• Original version	Sentinel Operations Center

I. BACKGROUND

The Mother-Infant Linkage (MIL) Table was added to the Sentinel Common Data Model (SCDM) in October 2018.

This enhancement was added to enable routine evaluation of medical product exposures during pregnancy on outcomes in infants.

Below is a list of Frequently Asked Questions and Appendices to help guide users through findings.

II. FREQUENTLY ASKED QUESTIONS

1. Why is it useful to link mothers and infants?

Mothers and infants initially appear in the Sentinel Distributed Database (SDD) at a patient level without any linkage of any set of patients to any other set of patients. In order to evaluate the effect of exposures during pregnancy on outcomes in infants, it is necessary to link mothers/deliveries with infants. Appendix A contains the structure of the Mother-Infant Linkage table that can be used for querying when links of mothers and infants is required.

2. Did the Sentinel Operations Center (SOC) conduct the linkage?

No, the SOC does not have access to patient-level data. Only the Data Partners (DPs) could perform the linkages with their data. The SOC identified mothers/deliveries and infants via a distributed package. The DPs then performed their linkages; see questions below for details on the linkage process.

3. How many Data Partners participated in mother infant linkage?

Six: Aetna, Harvard Pilgrim Health Care Institute, HealthCore, Inc., Humana, Inc., OptumInsight Life Sciences Inc., and Vanderbilt University Medical Center, Department of Health Policy, through the Tennessee Division of TennCare of the Department of Finance and Administration which provided data

4. Was this a one-time linkage?

No, this data expansion project had two rounds of linkages with the DPs on multiple extract, transform, and loads (ETLs). Going forward, there will be routinely refreshed Mother-Infant Linkage (MIL) tables after future ETL refreshes.

5. What are the processes for this entire project?

Per Data Partner, the steps are as follows:

a) The SOC distributes the mother-infant identification (MI ID) program package to DPs; see Appendix B for details on how deliveries and infants are identified.

b) DPs execute the MI ID program package and return results to SOC.

c) SOC reviews the results to ensure that the program ran properly at each DP site and that resulting aggregate data appears reasonable; the SOC then approves the MI ID program package execution.

d) DPs complete linkage using their own processes and source data; see questions below for details on the linkage process.

e) SOC distributes the MIL table quality assurance (QA) program package to DPs.

f) DPs execute the MIL QA program package and return results to SOC.

g) SOC evaluates results from the MIL QA program package at each DP site and returns reports to DPs that may contain questions for elaboration.

h) DPs respond to report.

i) SOC approves MIL table or DPs recreate their MIL table.

j) Iterate Steps d) through i) as needed.

6. Is the linkage process standardized across participating Data Partners?

Identification of mothers/deliveries and infants is standardized as it is dependent on the Sentinel Common Data Model (SCDM) databases and a SOC-created program package to identify mothers/deliveries and infants. Further, matching on family subscriber number[1] is virtually standard across DPs with potentially slight variation. However, linking methods for those not initially matched on subscriber ID may vary depending on the identifying information each DP has in their source data, such as names and addresses.

7. What type of match method (i.e., probabilistic vs. deterministic[2]) is used to link mothers to infants?

The DPs varied on their use of these two methods. In most cases, deterministic matching was used to link subscriber IDs and other variables to each other.

8. On which identifying factors were mothers and infants matched?

A family subscriber number was used for most of the linkages For those not linked by family subscriber number, names and addresses were used to link mothers to infants. One DP had access to birth certificate data for all infants and this data source was also used to link deliveries to infants. Other DPs did not have access to birth certificate data; such access would have required extensive time and processes and would have included only a very limited set of jurisdictions (e.g., states) containing birth certificate data.

9. What if a mother and infant share a member subscriber ID?

This is a common data problem with health insurers, as an infant is typically assigned the mother's subscriber ID so that the insurer can pay for health services delivered to the infant early in the infant's life, before the infant is enrolled and assigned their own subscriber ID. This can create uncertainty in linkage and DPs were advised to link only when they had confidence of the link.

10. What if a mother gives birth to multiple infants during the same delivery?

The mother-infant linking logic at DPs addresses this issue, so long as each infant is assigned a unique member subscriber ID in the source data. The multiple infants will be linked to a single delivery. In some

---

[1] For some commercial plans when a family is all covered under a single subscriber (e.g., a parent), a family subscriber ID is assigned to each member in the family.

[2] *Deterministic matching* is where unique multiple identifiers across records are compared exactly, or compared for only a fixed number of characters, to determine whether there is a match or not. *Probabilistic matching* is "fuzzy," with each identifier being assigned a value that indicates how closely the two identifiers' values match. Sometimes differential weights are assigned to the multiple identifiers. The mathematical combination of the individual identifiers' values and weights indicates the likelihood of a match between two records.

cases, infants may have multiple member subscriber IDs in a DP's source data that get reflected with multiple corresponding patient identifiers in the DP's Sentinel database. Special direction to DPs was provided by SOC to ensure that links were made to distinct *infants*, whether as singletons or as part of multiple births. See question 20 for more information.

11. Were deliveries that occurred outside of a hospital able to be linked to infants?

If there was a medical encounter for the delivery in the DP's data, then the delivery could be identified. Most linkages were restricted to the infant's date of birth being within the interval of three days prior to delivery visit admission date through delivery visit discharge date. This three-day window is built-in to allow for out of hospital deliveries to be recorded and linked and is based on experience from the Medication Exposure in Pregnancy Risk Evaluation Program (MEPREP)[3] and Post-Licensure Rapid Immunization Safety Monitoring System (PRISM)[4] projects linking deliveries to infants.

12. How do we know that the linkages are accurate?

There were extensive QA processes in reviewing the MIL table at the DPs. The SOC QA program package provided information for review, to ensure that the MIL table was model-adherent, included only records expected, and did not contain duplicates. Aggregate data returned to SOC was intensively reviewed to be certain that linkage counts and rates appeared reasonable for the many factors of the deliveries and infants.

13. What was the rate of linking deliveries to an infant?[5]

	Total
Deliveries	5,637,969
Infants	7,849,566
Linked Deliveries	4,094,436
Linkage Rate	72.62%

14. How does this linkage rate compare to those in other similar projects?

The rate of linkage observed in this project is consistent with rates observed in the MEPREP and PRISM projects that also performed mother-infant linkage. There are a number of factors that impact linkage rates:

a) In the latest 12 months of a DP's Sentinel database, the rule used to identify infants by design identified fewer infants than in prior years. Thus, while deliveries could be identified in the final year, there were fewer infants with whom to link.

b) In some cases, infants born to a mother in one health plan may be covered by a different health plan and so would not be identified within a single DP's data. Thus, deliveries may be identified, but there would be no infants within the same plan with whom to link.

---

[3] Andrade SE, Davis RL, Cheetham TC, et al. Medication Exposure in Pregnancy Risk Evaluation Program. Matern Child Health J. 2012 Oct;16(7):1349-54. doi: 10.1007/s10995-011-0902-x.
[4] Kawai AT, Andrade SE, Zichittella L, et al. Developing the Infrastructure to Conduct Surveillance of Birth Outcomes Following Maternal Vaccination: A Project Using Influenza Vaccines and Birth Outcomes as a Use Case. https://www.sentinelinitiative.org/sites/default/files/vaccines-blood-biologics/assessments/PRISM_Influenza_Vaccines_and_Birth_Outcomes_Final_Report.pdf; Accessed October 18, 2018.
[5] All statistics in this FAQ (Questions 13 and 16) come from the MIL Quality Assurance program package for Round 2, of two rounds of linkage, with the exception that one Data Partner performed only one round of linkage.

c) The SOC had a principle that DPs should make only links for which they were absolutely confident that the links were correct. It may have been possible for some additional links to have been made, but there would have been a lower level of confidence for these links. One example includes cases where there were multiple infant identifiers, where there was uncertainty of whether these were pertaining to one infant or multiple infants. Another example was multi-generational households with members covered on the same insurance, where there was uncertainty of whether the link should be made to a mother or to a grandmother.

15. Why is there a higher rate of identified infants to identified deliveries than in national natality data?

Enrollment requirements for mothers/deliveries were much more strict than requirements to identify infants; thus, more infants may have been identified than deliveries. Additionally, an infant may be covered in a one DP's health plan, while the mother is covered on another plan.

16. Linkage Statistics

Mother's Age

	10-19 Years	20-44 Years	45-54 Years	Total
Deliveries	253,183	5,342,563	42,223	**5,637,969**
Linked Deliveries	116,419	3,966,493	11,524	**4,094,436**
Linkage Rate	45.98%	74.24%	27.29%	**72.62%**

There were challenges in linkage for the older mothers' group. From discussions with DPs, the likeliest explanation is that in this age group, there is a higher rate of infants being covered in different plans from their mothers, than in the lower age groups. For example, one DP mentioned that they have a much larger percentage of Medicare membership in the older age group (45-54: 23% vs 2.4% in 20-44 age group and vs 0.28% in 10-19 age group), whose infants would be covered under another plan. It may also be due to the higher mortality rate of the newborn infants in older age women; these infants do not end up enrolling in health plans.

Encounter Type

	Inpatient Hospital Stay (IP)	Emergency Department (ED)	Non-Acute Institutional Stay (IS)	Ambulatory Visit (AV)	Other Ambulatory Visit (OV)	Total
Deliveries	5,312,558	8,215	8,215	219,646	93,093	**5,637,969**
Linked Deliveries	4,053,454	784	2,880	21,787	15,531	**4,094,436**
Linkage Rate	76.30%	9.54%	64.62%	9.92%	16.68%	**72.62%**

While all encounter types were permitted for identifying deliveries, the data show that linkage rates were highest for inpatient and institutional encounter types. This may be because, using delivery codes, these had a higher proportion of being a "true" delivery encounter as compared to those identified in other encounter types. Deliveries identified in other encounter types may be due to payment on claims that included one of the delivery codes, when a delivery did not actually occur (e.g., a prenatal care visit).

Birth type (i.e., # of infants reported to have been born in delivery)

	No indicator of # of live births	One live birth	Two live births	Three live births	Four live births	Five live births	Multiple live births; unspecified #	Conflicting codes on # of live births	Total
Deliveries	492,437	5,021,394	101,266	2,743	139	3	2,525	17,462	**5,637,969**
Linked Deliveries	152,306	3,849,340	76,441	2,035	91	1	942	13,280	**4,094,436**
Linkage Rate	30.93%	76.66%	75.49%	74.19%	65.47%	33.33%	37.31%	76.05%	**72.62%**

The highest linkage rates were observed when the Birth Type indicated one live birth (i.e., singletons).

Year

	Deliveries	Linked Deliveries	Linkage Rate
2001	40,722	34,315	84.3%
2002	40,317	33,978	84.3%
2003	39,447	34,270	86.9%
2004	40,542	35,600	87.8%
2005	41,062	36,925	89.9%
2006	42,605	38,083	89.4%
2007	210,411	163,324	77.6%
2008	230,638	180,807	78.4%
2009	574,267	466,248	81.2%
2010	552,878	451,358	81.6%
2011	561,007	449,315	80.1%
2012	563,277	428,430	76.1%
2013	570,823	431,943	75.7%
2014	569,901	439,447	77.1%
2015	572,415	439,543	76.8%
2016	570,223	412,536	72.3%
2017	417,434	18,314	4.4%
Total	**5,637,969**	**4,094,436**	**72.6%**

The drop in linkage rate from 84% in 2001 to 72% in 2016 might be due to a change in enrollment patterns, particularly a higher proportion of mothers and infants being in different plans in later years. (The lowest rate observed in 2017 is due to the issue discussed earlier regarding fewer infants being identified in the last 12 months of the DPs' Sentinel databases.)

One DP evaluated their enrollment trends among women aged 10-54 years, and infants aged 0-1 years. The enrollment of women slowly drops over the years, while the enrollment of infants drops more dramatically. The drops of the infants' enrollment plus the increase of Medicare membership leads to the slowly lowering linkage rate.

17. Are there any differences or possible bias of those who are linked vs. those who are not linked?

This is possible. An additional project would be required to make this assessment, as it requires querying the SDD and the MIL table.

18. How do we obtain characteristics of mothers, such as their enrollment intervals, exposures, etc.

See answer to question 17 above.

19. How do we obtain characteristics of infants, such their enrollment intervals, outcomes, etc.

See answer to question 17 above.

20. When there are multiple infant numbers (CPatIDs) that reflect the same infant in source data, how did DPs prioritize which infant to select for linkage?

SOC advised DPs to prioritize infants who are enrolled within 30 days of the date of birth and have the longest overall enrollment period. The rule was operationalized as follows:
   a) Among multiple CPatIDs that represent a single infant, select only those where Enrollment-Start-Date minus Birth-Date <= 30 days.
       i. Then from those filtered by A above select the one CPatID that had the longest enrollment period; i.e., Enrollment-End-Date minus Enrollment-Start-Date.
   b) If no CPatID has Enrollment-Start-Date minus Birth-Date <= 30 days, then select the single CPatID with Enrollment-Start-Date closest to Birth-Date, regardless of the length of enrollment.
       i. If multiple CPatIDs have identical Birth-Date and Enrollment-Start-Date, then select the CPatID with the longest enrollment span.
       ii. If multiple CPatIDs have identical Birth-Date and Enrollment-Start-Date and identical enrollment span lengths, then do not link any of these CPatIDs.

21. Were there circumstances of multiple CPatIDs representing the same infant for which links were not made?

One DP raised the possibility of randomly selecting a single CPatID to link. This existed for ~100 CPatIDs (< 0.1%), where there were the same values of infant birth date, enrollment start date, and enrollment lengths. In this circumstance, SOC recommended no links be made as it could not be known which of the multiple CPatIDs would represent the "correct" infant or have the most utilization data.

22. What if an infant could be matched to more than one *mother's delivery*?

This can happen if a mother has more than one patient identifier in a DP's Sentinel database. Linking between infant information and mother/delivery data may then result in potential linkage to multiple mothers' deliveries. SOC advised Data Partners not to match these infants, because the accuracy of the links could not be confirmed. This was a very infrequent occurrence with 0.04% of infants that could be matched to more than one mother's delivery.

III. APPENDIX A: SENTINEL COMMON DATA MODEL SPECIFICATIONS FOR THE MOTHER-INFANT LINKAGE TABLE

Description: The SCDM Mother-Infant Linkage Table contains one record per MPatID, CPatID, and ADate. The file may include:

1. Live birth deliveries (with MPatID and ADate) that were linked to a child (CPatID);
2. Live birth deliveries (with MPatID and ADate) that were not linked to a child (CPatID, CBirth_Date, Sex, and CEnr_Start will have missing values); and
3. Children (with CPatID) who were not linked to a mother (MPatID, MBirth_Date, Age, EncounterID, EncType, ADate, DDate, Birth_Type, and Birth_Type_Primes will have missing values).

Variable Name	Type & Length	Format	Label	Values	Definition / Comments / Guideline
MPatID	Char(Site specific length)	$##.	Mother's patient ID	Unique member identifier. Text string; left justified.	Arbitrary person-level identifier. Used to link across tables. Length is DP specific. Must match mother's "PatID" value in all other SCDM tables. Blank for child-only records.
MBirth_Date	Num(4)	mmddyy10.	Mother's birth date	SAS Date.	Mother Birth_Date value from SCDM Demographic table. Blank for child-only records.
Age	Num(3)	2.	Mother's age as integer	19-54 inclusive	Mother's age as of ADate

Variable Name	Type & Length	Format	Label	Values	Definition / Comments / Guideline
EncounterID	Char(Site specific length)	$##.	Mother's delivery EncounterID	Unique encounter identifier.	EncounterID value from SCDM Encounter table, for mother's delivery encounter.[6] Blank for child-only records.
EncType	Char(2)	$2.	Encounter type	IP = Inpatient Hospital Stay ED = Emergency Department AV = Ambulatory Visit IS = Non-Acute Institutional Stay OA = Other Ambulatory Visit	EncType value from SCDM Encounter table, for mother's delivery encounter.** Blank for child-only records.
ADate	Num(4)	mmddyy10.	Admission date for mother's delivery encounter	SAS Date.	ADate value from SCDM Encounter table, for mother's delivery encounter.** Blank for child-only records.
DDate	Num(4)	mmddyy10.	Discharge date for mother's delivery encounter	SAS Date; may be null	DDate value from SCDM Encounter table, for mother's delivery encounter.** Blank for child-only records.

[6] If more than one delivery encounter occurs on the same ADate, then the values are based on the encounter selected from the following hierarchy: IP > ED > AV > IS > OA

Variable Name	Type & Length	Format	Label	Values	Definition / Comments / Guideline
CPatID	Char(Site specific length)	$##.	Child patient ID	Unique member identifier. Text string; left justified.	Arbitrary person-level identifier. Used to link across tables. Length is DP specific. Must match child "PatID" value in all other SCDM tables. Blank for mother/delivery-only records.
CBirth_Date	Num(4)	mmddyy10.	Child birth date	SAS Date	Child Birth_Date value from SCDM Demographic table. Blank for mother/delivery-only records.
Sex	Char(1)	$1.	Child sex	F = Female M = Male O = Other (Ambiguous or Unknown)	Child Sex value from SCDM Demographic table Blank for mother/delivery-only records.
CEnr_Start	Num(4)	mmddyy10.	Earliest infant enrollment start date	SAS Date	Earliest Enr_Start from Enrollment table

Variable Name	Type & Length	Format	Label	Values	Definition / Comments / Guideline
Birth_Type	Num(3)	1.	Singleton vs. multiple	0 = Unspecified # of live births 1 = One live birth 2 = Two live births 3 = Three live births 4 = Four live births 5 = Five live births 8 = Multiple live births, unspecified number 9 = Conflicting code(s) for number of live births	Based upon ICD-9-CM/ICD-10-CM diagnosis and procedure codes, as well as CPT procedure codes, in the health plan data for the delivery admission. Blank for child-only records
Birth_Type_Primes	Num(8)	comma15.	All birth types	2+	Multiplication of all prime numbers assigned to all Birth_Types found in delivery codes within the selected encounter: Birth_Type Prime Number 0 2 1 3 2 5 3 7 4 11 5 13 8 17 Missing/null for child-only records

Variable Name	Type & Length	Format	Label	Values	Definition / Comments / Guideline
MatchMethod	Char(2)	$2.	Prioritized method of linkage for mom-baby match, *or* reason for unlinked record	BC = Birth Certificate RE = DP maintained birth registry SI = health plan subscriber or family number LA = exact or probabilistic last name and address match based upon health plan administrative data OT = other	For <u>linked records</u>, prioritize so that only one method is listed: RE > SI > LA > BC > OT
				N1 = No subscriber/family IDs available for linkage N2 = No name/address available for linkage N3 = Neither subscriber/family IDs nor name/address available for linkage NA = no linkage made; any other reason	For cases where a mother/delivery is <u>not linked</u> to a child OR a child enrolled is not linked to a mother/delivery, the value of this variable should be one of N1, N2, N3, or NA only

IV.   APPENDIX B: IDENTIFICATION OF MOTHERS/DELIVERIES AND INFANTS

Mothers/Deliveries

Deliveries were identified using the following criteria:

1. An encounter from one year later than the start date of the Data Partner's data availability through the end date of the DP's data.
2. An encounter contained a delivery code of interest (over 700 diagnosis and procedure codes)
3. Women were between 10 years and 54 years of age inclusive on the admission date of the delivery encounter
4. No evidence of delivery for 180 days prior to any identified delivery.
5. Mothers must have had medical coverage from 180 days prior to date of delivery through the delivery date. A single enrollment gap of up to 45 days within the enrollment period is allowed.

Note that more than one delivery per mother may be identified within the DP's Sentinel Distributed Database.

Infants

1. Identified with a date of birth from one year later than the start date of the DP's data availability through the end date of the DP's data.
2. An infant must have at least one day of enrollment with medical coverage during their first year of life.