Mother-Infant Linked Data: Methodology, Case Studies, and Cohort Development for Investigating Prenatal Exposure and Neonatal Outcomes

Jill Hardin[1,2], Alexis Krumme[1,2], David Kern[1,2], James Weaver[1,2], Clair Blacketer[1,2]
[1] Janssen Research & Development, LLC, [2] Observational Health Data Sciences and Informatics (OHDSI), New York, NY

Background

Linkage of maternal and infant records from routinely collected healthcare data facilitates research on prenatal exposure, prenatal comorbidity, and infant health outcomes[1]. A recent study developed a mother-infant linkage using two commercial claims databases[2]. We present two case studies demonstrating the use of this linkage and provide a step-by-step methodological guide to develop linked mother and infant cohorts.

The first case study focuses on exposure to antiepileptic seizure medications during pregnancy. The second case study identifies a cohort of infants affected by Hemolytic Disease of the Fetus and Newborn (HDFN), a condition where maternal and fetal red blood cell incompatibility can lead to maternal isoimmunization and subsequent anemia in the fetus and neonate.

The approach for utilizing the linked mother-infant cohorts differs based on the diagnoses or exposures in the mother and/or infant records. In the first case study, an index date is assigned at the first drug exposure during the pregnancy episode considering all infants as exposed based on the mother's exposure during pregnancy. The second case study highlights the challenges of identifying and indexing a condition when diagnosis can occur in pregnancy or neonatally. Specifically, the mother's index start date is set during the pregnancy episode, while the infant's index date is the birth event. The pair (mother and infant) are included in the cohort if either mother or infant has a code for isoimmunization or HDFN.

For each case study, we provide SQL code to navigate the Common Data Model (CDM) and explain the step-by-step process of extracting the mother-infant linkage data using the fact relationship table in combination with pregnancy cohorts built in ATLAS (accessible at https://github.com/OHDSI/ATLAS).

Methods

We conducted this analysis using two observational databases, both US administrative claims databases (Table 1), the Meritave MarketScan Commercial Claims and Encounters Database (CCAE) and Optum's de-identified Clinformatics® Data Mart Database (Clinformatics®), transformed to the Observational Medical Outcomes Partnership (OMOP) Common Data Model version 5.4 [3].

We used ATLAS to create cohorts by defining pregnancy episodes and identifying relevant exposures or diagnoses. In the first case study, we identified pregnancy episodes concurrent with antiepileptic seizure medications dispensing (ATC3 ingredient-level drug concepts listed in Table 2). In the second case study, we identified pregnancy episodes with concurrent diagnostic codes for maternal or infant isoimmunization (the precursor to HDFN) or HDFN due to Rhesus alloantibodies (SNOMED codes in Table 2).

Results

There were 5,252,372 pregnancy episodes ending in a live birth in CCAE and 2,830,694 in Clinformatics® with 4,304,632 linked infants in CCAE and 1,850,278 in Clinformatics® (Table 1). Note that the mother-infant relationship is not a one-to-one relationship as mothers can have multiple linked infants. Figure 1

illustrates the logic for building the exposed infant cohort for case study 1, with the first step involving an ATLAS cohort. The second step involves executing SQL code using the ATLAS cohort and the fact relationship table to link exposed mothers and infants. After executing this code, the number of pregnancy episodes exposed to antiepileptic medication in the fourth month or later of pregnancy was 4,725 in Clinformatics®and 8,700 in CCAE, linked to 3,769 linked infants in Clinformatics® and 5,256 in CCAE.

Figure 2 illustrates the logic for building the isoimmunization or HDFN pregnancy and infant cohorts for case study 2, with the first step involving two ATLAS cohorts. From the ATLAS cohort for isoimmunization or HDFN, there were 155,398 (Clinformatics®) and 264,770 (CCAE) pregnancy episodes ending in a live birth and 9,278 (Clinformatics®) and 7,773 (CCAE) infants. The third step involves executing SQL code using the ATLAS isoimmunization or HDFN pregnancy cohort and identified 141,981 (Clinformatics®) and 192,206 (CCAE) pregnancies linked to infant records. The fourth step links the isoimmunization or HDFN infants to the pregnancy episode and identifies 6,028 (Clinformatics®) and 4,828 (CCAE) infants. The final SQL step identifies the final isoimmunization or HDFN linked pregnancy and infant cohort and identifies 145,391 (Clinformatics®) and 194,979 (CCAE) episodes and infants.

Conclusion

Our study demonstrates some benefits of the mother-infant linkage algorithm through two case studies. Access to large, linked populations enables the study of perinatal exposures, maternal and neonatal outcomes, and subgroups, which are often limited in smaller linked populations [4, 5] and registries [6-8]. This approach requires fewer study resources compared to primary data collection, and this study provides direction on how to leverage this resource in an OMOP CDM environment using OHDSI tools.

Table 1: Description of Databases used in the study

Name (Abbreviation)	Years	Country	Data Type	Clinical Visits included	Number of Persons (millions)	Numb pregn episo endin live bi
Merative MarketScan Commercial Claims and Encounters (CCAE)	2000-2021	US	Insurance Claims	Inpatient/ outpatient	157	5,252,
Optum's de-identified Clinformatics® Data Mart Date Database (Clinformatics®)	2007-2021	US	Insurance Claims	Inpatient/ outpatient	71	2,830,

Table 2: Drug exposure and diagnostic codes used in case study one and two

Case study #	Concept Names and Identifiers	Cohort Application
1	Antiepileptic seizure medications: 702661,702685,705103,711584,713192,715458,718122,734275,734354,740275,740910,742267, 744798,745466,750119,750146,751347,753860,759401,795661,797399,798874,1510417,19000921, 19004254,19005629,19006586,19018520,19020002,19021932,19023286,19023842,19054995, 19087394,19095776,19112534,19123696,35200286,35604901,36878958,37497998,40239995, 42904177,44507780	Pregnancy episode ending in a live birth
2	Isoimmunization or Hemolytic Disease of the Fetus and Newborn: 192376,195878,199891,433603,4028774,4139549,4139550,4143895,4145893,44783943	Pregnancy episode ending in a live birth and infants

Figure 1. Logic diagram for case study 1

Step 1. Generate cohort in ATLAS:

- index on livebirth pregnancy episodes
    - require prior diagnosis of epilepsy and at least one antiepileptic drug exposure starting in fourth or later month of pregnancy

Step 2. Use SQL to build linked pregnancy and infant cohort:

```
Select c.cohort_definition_id,
       c.subject_id as mother_person_id,
       c.cohort_start_date as pregnancy_start_date,
       c.cohort_end_date as pregnancy_end_date,
       f.fact_id_2 as baby_person_id
into #antiepileptic_infants
from results_truven_ccae_v2435.cohort c
join cdm_truven_ccae_v2435.fact_relationship f
  on c.subject_id = fact_id_1 and relationship_concept_Id = 40478925
where cohort_definition_id = 12870
```

Figure 2. Logic diagram for case study 2

Step 1. Generate cohorts in ATLAS:

Mother cohort:

- index on livebirth pregnancy episodes
  - require condition occurrence code HDFN during pregnancy episode

Infant cohort:

- Index on HDFN condition occurrence codes
- Age < 1 years

Step 2. Use SQL to build linked pregnancy and infant cohort:

SQL Step 1. Identify pregnancies cohort using isoimmunization and HDFN codes:

```
select count(*)
from results_optum_extended_dod_v2434.cohort c
where cohort_definition_id = 12833 S
```

SQL Step 2. Identify infants cohort using isoimmunization and HDFN codes:

```
select count(distinct subject_id)
from results_optum_extended_dod_v2434.cohort c
where cohort_definition_id = 12835
```

SQL Step 3. Identify HDFN mothers with an infant from the pregnancy episode:

```
Select c.cohort_definition_id,
       c.subject_id as mother_person_id,
       c.cohort_start_date,
       ce.condition_era_start_date as pregnancy_start_date,
       ce.condition_era_end_date as pregnancy_end_date,
       mob.baby_person_id,
       mob.baby_observation_start,
       mob.baby_observation_end
into #hdfn_linked_mothers
from results_optum_extended_dod_v2434.cohort c
join cdm_optum_extended_dod_v2434.condition_era ce
  on c.subject_id = ce.person_id
  and condition_occurrence_count = 0
  and c.cohort_start_date between ce.condition_era_start_date and ce.condition_era_end_date
join #mothers_of_babies mob
  on c.subject_id = mob.mother_person_id
where cohort_definition_id = 12833
  and datediff(day,ce.condition_era_end_date,mob.baby_observation_start) between -60 and 60
```

SQL Step 4. Identify HDFN infants linked with their pregnancy episodes:

```sql
Select c.cohort_definition_id,
       c.subject_id as baby_person_id,
       c.cohort_start_date as hdfn_start_date,
       mob.mother_person_id,
       mob.baby_observation_start,
       mob.baby_observation_end

into #hdfn_linked_babies
from results_optum_extended_dod_v2434.cohort c
join #mothers_of_babies mob
  on c.subject_id = mob.baby_person_id
where cohort_definition_id = 12835
```

SQL Step 5. Identify final HDFN linked pregnancy and infant cohort:

```sql
select corrected_mother_person_id as mother_person_id,
       mother_pregnancy_start_date,
       mother_pregnancy_end_date,
       corrected_baby_person_id as baby_person_id,
       baby_hdfn_date,
       baby_observation_start,
       baby_observation_end,
       condition_era_start_date,
       condition_era_end_date
into #final_hdfn_full_cohort
from (
    select m.mother_person_id as mother_table_person_id,
           b.mother_person_id as baby_table_mother_person_id,
           case when m.mother_person_id is null then b.mother_person_id
            when b.mother_person_id is null then m.mother_person_id
            else m.mother_person_id end as corrected_mother_person_id,
           m.pregnancy_start_date as mother_pregnancy_start_date,
           m.pregnancy_end_date as mother_pregnancy_end_date,
           m.baby_person_id as mother_table_baby_person_id,
           b.baby_person_id as baby_table_baby_person_id,
           case when m.baby_person_id is null then b.baby_person_id
            when b.baby_person_id is null then m.baby_person_id
            else b.baby_person_id end as corrected_baby_person_id,
           b.hdfn_start_date as baby_hdfn_date,
           case when m.baby_observation_start is null then b.baby_observation_start
            when b.baby_observation_start is null then m.baby_observation_start
            else b.baby_observation_start end as baby_observation_start,
           case when m.baby_observation_end is null then b.baby_observation_end
            when b.baby_observation_end is null then m.baby_observation_end
            else b.baby_observation_end end as baby_observation_end,
            datediff(day,m.pregnancy_end_date,b.baby_observation_start)
    from #hdfn_linked_mothers m
    full outer join #hdfn_linked_babies b
      on m.mother_person_id = b.mother_person_id
      and m.baby_person_id = b.baby_person_id
      --145394
) a
join cdm_optum_extended_dod_v2434.condition_era ce
  on a.corrected_mother_person_id = ce.person_id
  and condition_occurrence_count = 0
  and condition_concept_id = 433260
  and datediff(day,ce.condition_era_end_date,baby_observation_start) between -60 and 60
```

Table 1. Table of pregnancy episodes and infants for case study 2

Cohort criteria	Clinformatics®		CCAE	
	Pregnancy episodes ending in live birth	Infants	Pregnancy episodes ending in live birth	Infants
Diagnosis code for isoimmunization or HDFN due to Rhesus alloantibodies	155,398	9,278	264,770	7,773
With corresponding link to pregnancy episode or infant records	141,981	6,028	192,206	4,828
[1]Final cohort	145,391		194,979	

[1]this total is less than the two above lines because the relevant diagnostic codes can occur in either mother, linked infant, or both.

References

1. Margulis, A.V. and E.B. Andrews, *The Safety of Medications in Pregnant Women: An Opportunity to Use Database Studies.* Pediatrics, 2017. **140**(1).
2. Weaver, J., et al., *Development and evaluation of an algorithm to link mothers and infants in two US commercial healthcare claims databases for pharmacoepidemiology research.* medRxiv, 2022: p. 2022.12.13.22283418.
3. Voss EA, M.R., Matcho A, Ma Q, Knoll C, Schuemie M, et al. , *Feasibility and utility of applications of the common data model to multiple, disparate observational health databases.* . J Am Med Inform Assoc. , 2015. **22**(3): p. 553-564.
4. Baldwin, E., et al., *Linking mothers and infants within electronic health records: a comparison of deterministic and probabilistic algorithms.* Pharmacoepidemiol Drug Saf, 2015. **24**(1): p. 45-51.
5. Whitmore, C.C., et al., *Building a Data Linkage Foundation for Mother-Child Pharmacoepidemiology Research.* Pharmaceut Med, 2021. **35**(1): p. 39-47.
6. Gelperin, K., et al., *A systematic review of pregnancy exposure registries: examination of protocol-specified pregnancy outcomes, target sample size, and comparator selection.* Pharmacoepidemiol Drug Saf, 2017. **26**(2): p. 208-214.
7. Huybrechts, K.F., B.T. Bateman, and S. Hernández-Díaz, *Use of real-world evidence from healthcare utilization data to evaluate drug safety during pregnancy.* Pharmacoepidemiol Drug Saf, 2019. **28**(7): p. 906-922.
8. Wyszynski, D.F., *Pregnancy exposure registries: academic opportunities and industry responsibility.* Birth Defects Res A Clin Mol Teratol, 2009. **85**(1): p. 93-101.