Development and evaluation of an algorithm to link mothers and children in a US claims database

James Weaver MS,1,2, Jill Hardin PhD,1,2, Patrick B. Raymond PhD,1,2,3
1Janssen Research & Development, LLC, Raritan, NJ, 2Observational Health Data Sciences and Informatics (OHDSI), New York, NY, USA, 3Columbia University, New York, NY, USA
Contact: jweave17@its.jnj.com

ABSTRACT

Background: Maternal-offspring pairs (MOPs) exist in US databases but lack generalizability to the commercial claims population as they represent other nonrandom samples of the US. Our MOP algorithm achieved norm for applying additional criteria to increase linkage confidence and by evaluating generalizability to the US commercial claims population.

Objectives: Develop an algorithm to identify MOPs in an observational database and evaluate generalizability.

Methods: The Truven Health MarketScan Research Database (1/1/2000-4/30/2015) was transformed to the Observational Medical Outcomes Partnership Common Data Model was used. MOPs were constructed and compared to cohorts of all-mothers (identified by a pregnancy episode definition algorithm with a live birth outcome) and all-offspring (people whose birth year equals that of database entry). MOPs include all-mothers with a family identifier and who have observation time overlapping with a woman who is 0-5 years of age at database entry. These cohort calculations were then restricted to those whose birth date is within 50 days of the mother’s pregnancy episode start date. Characteristics (demographics; condition, procedure, and drug claims in the 365 days before (mothers) and after (offspring) birth) of MOPs, all-mothers, and all-offspring were compared using standard differences of mean to assess generalizability.

Results: The MOPs algorithm identified 1,261,967 mothers and 1,926,114 offspring. MOPs covered 72% of all-mothers (N=2,378,762) and 50% of all-offspring (N=3,853,277). 92% of observation start dates of MOP offspring were within 4-weeks of the pregnancy episode start date. Standardized differences of means of <0.1 were observed for 99% of mother and offspring covariates.

Conclusions: The MOP algorithm can be applied to an observational healthcare claims database and achieve generalizable results to enable further teratogenicity research.

BACKGROUND

- Use of observational databases to study pregnancy exposures would address several problems, including:
  1. The gap of drug safety information during pregnancy as premarketing randomized controlled trials exclude pregnant women due to ethical considerations.
  2. Lack of evidence on benefits of drug use during pregnancy and resultant birth outcomes [2].
  3. Information on drug risks during pregnancy [2,12] including insufficient sample size to study rare birth outcomes, new data collection that may be expensive and difficult to collect, recall bias.

- Use of large healthcare claims data could be advantageous in studying pregnancy exposure and outcomes, including longitudinal follow up, large sample sizes, and reflecting treatment patterns

OBJECTIVE

- Develop a algorithm to identify MOPs in a large, observational, US healthcare claims database
- Compare linked MOPs to all-mothers and all-offspring to evaluate generalizability to the US commercial claims population

METHODS

Data Source

- Medical and pharmacy claims from Truven Health MarketScan Commercial Claims and Enrollments (CCD) database and April 30, 2016, approximately 131.5 million patients transformed to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) [13]
- Allows database structure and content standardization; native data are mapped to standardized variables for each domain (conditions, drugs, procedures, etc.)
- Algorithm developed using CADD database transformed into the OMOP CDM

Cohort Construction

- Three cohorts were constructed and characterized:
  - MOP cohort: composed of the mother and linked-offspring pairs
  - All-mothers cohort
  - All-offspring cohort

  A mother can be linked to one or more offspring and each offspring can be linked to one mother

  Mothers can have one or more pregnancy events

- The MOP algorithm and inclusion criteria are illustrated in Figure 1

  Cohort 1: All-mothers mothers identified using the pregnancy episode algorithm
  i. Matched et al. [14] algorithm identifies pregnancy episodes; limited to mothers’ pregnancy coverage period
  ii. Restricted to those who share a family identifier code and overlapping observation time with another who is 0-5 years of age at database entry

  Cohort 2: MOPs people associated with M.O.’s L.I., are candidate offspring
  a. Restricted to those whose insurance coverage start date is within 60 days of linked mother pregnancy end date

  All-Mothers: mothers identified using the pregnancy episode algorithm [14] who have a live birth outcome

- All-offspring: people aged 0-5 years at database entry
- index date in the linked-mothers, linked-offspring, and all-offspring cohorts is the pregnancy episode end date; index date in the all-offspring cohort is date of birth

Cohort Characterization

- Description of the observation time distribution calculated for mother cohorts; after-birth observation time distribution was calculated for offspring cohorts

  Demographic (5 year age categories, index month and year), condition occurrence, procedure occurrence, and drug exposure covariates computed for all cohorts; gender covariate constructed for offspring cohorts

Table 1: Key demographic covariate proportions/means and standardized differences of MOP mothers and all-mothers

<table>
<thead>
<tr>
<th>Covariate</th>
<th>All-mothers</th>
<th>MOP mothers</th>
<th>Standardized difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age group: 15-19</td>
<td>0.05</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>Age group: 20</td>
<td>0.06</td>
<td>0.04</td>
<td>0.02</td>
</tr>
<tr>
<td>Age group: 25</td>
<td>0.10</td>
<td>0.08</td>
<td>0.02</td>
</tr>
</tbody>
</table>
| All offspring cohort
  - Linked offspring date of birth is within 60 days of mothers’ pregnancy episode end date
  - Mothers identified using the pregnancy episode algorithm [14] who have had a live birth outcome

  Linked offspring and all-offspring assessed generalizability and provided internal validation.

  Comparison between linked-mothers and all-mothers and between linked-offspring and all-offspring assessed generalizability and provided internal validation.

  Similarity between linked-mothers and all-mothers and between linked-offspring and all-offspring would suggest MOP generalizability to the commercial claims population and lend support to results from teratogenic exposure studies on this linked MOP sample.

  Proportions and standard errors calculated for covariates for cohorts

  Cohort comparisons were made by calculating the standardized difference in means for all covariates in units of the pooled standard deviation

 RESULTS

- 1,163,184 unique MOPs were identified covering 70% of all-mothers (N=2,378,762) and 50% of all-offspring (N=3,853,277)

DISCUSSION

- Illustrates the potential for pharmacoeconomics studies of maternal drug exposures to be conducted outside the context of formal pregnancy registries

- Ability to quickly and inexpensively assemble a large MOPs cohort for use in investigations of medication exposure effects during pregnancy and associated offspring outcomes

REFERENCES


CONFLICT OF INTEREST

JL, JR, and PS are full time employees of Janssen Research and Development, a unit of Johnson & Johnson. The work of this study was part of their employment. They hold pension rights from the company and own stock and stock options.