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

James Weaver MS\textsuperscript{1,2}, Jill Hardin PhD\textsuperscript{1,2}, Patrick B. Ryan PhD\textsuperscript{1,2,3}

\textsuperscript{1}Janssen Research & Development, LLC, Raritan, NJ, \textsuperscript{2}Observational Health Data Sciences and Informatics (OHDSI), New York, NY, USA, \textsuperscript{3}Columbia University, New York, NY, USA
Background

• Current mother-offspring pairs (MOPs) in US databases lack generalizability to the commercial claims population

• Our MOP linkage algorithm measures linkage confidence and establishes an evaluation framework for assessing generalizability to the US commercial claims population

• Observational database studies of pregnancy exposures will address:
  – gap of drug safety information during pregnancy due to trial exclusions of pregnant women
  – lack of evidence on benefits and risks of drug use during pregnancy and resultant birth outcomes
  – Insufficient sample sizes, difficult and expensive new data collection, recall bias
  – includes longitudinal follow up, real-world treatment practice patterns

OHDSI community call 09/12/17
Objectives

• Develop an 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
Data Source

OHDSI community call 09/12/17
• Constructed 3 cohorts:
  1. MOPs, mothers and their associated offspring
  2. All-mothers
  3. All-offspring

• All mothers
  – Women with live birth outcome using pregnancy episode identification algorithm

• All offspring
  – People 0 years of age at observation period start
• MOP mothers
  i. Pregnancy episodes with live birth outcome
     • N mothers = 2,378,762
  ii. Have family identifier code (PAYER_PLAN_PERIOD.FAMILY_SOURCE_VALUE) and birth during insurance coverage period
     • N mothers = 2,378,234
Cohort Construction

• MOP offspring
  i. Share family identifier code (PAYER_PLAN_PERIOD.FAMILY_SOURCE_VALUE) and overlapping observation time with mother, 0 years of age at observation period start
    • Mother family_source_value = offspring family_source_value
  ii. Insurance coverage start date within mother’s observation period
    • N mothers = 1,742,530
    • N offspring = 2,212,687
    • N records = 2,221,528

Live birth event/pregnancy end date

Mother observation time
Mother/offspring insurance coverage
Offspring observation time
Cohort Construction

- MOP offspring
  iii. Exclude inaccurate linkages, duplicate records; e.g. 1 offspring associated with multiple mothers
  - N mothers = 1,729,530
  - N offspring = 2,203,687
  - N records = 2,203,687
Cohort Construction

• MOP offspring
  iv. Restrict where DOB is within 60 days of pregnancy episode end date
    • N mothers = 1,661,987
    • N offspring = 1,928,114
    • N records = 1,928,114

Mother observation time
Mother/offspring insurance coverage
Offspring observation time
Offspring DOB

Live birth event/pregnancy end date

60 days
Cohort Construction

- Correspondence between algorithm birthdate and inferred birthdate
  - 92% within 4 weeks

  92% within 4 weeks
  52% within 2 weeks
  Live birth event/pregnancy end date

  3% same day
Cohort Characterization

- Observation time, days from index

<table>
<thead>
<tr>
<th></th>
<th>MOPs</th>
<th>All</th>
<th>Std. diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mothers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Before birth</td>
<td>-1029</td>
<td>-1074</td>
<td>0.06</td>
</tr>
<tr>
<td>After birth</td>
<td>1071</td>
<td>997</td>
<td>0.07</td>
</tr>
<tr>
<td>Offspring</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>After birth</td>
<td>1054</td>
<td>1003</td>
<td>0.05</td>
</tr>
</tbody>
</table>
Cohort Characterization

- Observation time, days from index

<table>
<thead>
<tr>
<th></th>
<th>MOPs</th>
<th>All</th>
<th>Std. diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mothers</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Before birth</td>
<td>-1029</td>
<td>-1074</td>
<td>0.06</td>
</tr>
<tr>
<td>After birth</td>
<td>1071</td>
<td>997</td>
<td>0.07</td>
</tr>
<tr>
<td><strong>Offspring</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>After birth</td>
<td>1054</td>
<td>1003</td>
<td>0.05</td>
</tr>
</tbody>
</table>
Cohort Characterization

- Observation time, days from index

<table>
<thead>
<tr>
<th></th>
<th>MOPs</th>
<th>All</th>
<th>Std. diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mothers</td>
<td>Before birth</td>
<td>-1029</td>
<td>-1074</td>
</tr>
<tr>
<td></td>
<td>After birth</td>
<td>1071</td>
<td>997</td>
</tr>
<tr>
<td>Offspring</td>
<td>After birth</td>
<td>1054</td>
<td>1003</td>
</tr>
</tbody>
</table>
Cohort Characterization

- Observation time, days from index

<table>
<thead>
<tr>
<th></th>
<th>MOPs</th>
<th>All</th>
<th>Std. diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mothers</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Before birth</td>
<td>-1029</td>
<td>-1074</td>
<td>0.06</td>
</tr>
<tr>
<td>After birth</td>
<td>1071</td>
<td>997</td>
<td>0.07</td>
</tr>
<tr>
<td><strong>Offspring</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>After birth</td>
<td>1054</td>
<td>1003</td>
<td>0.05</td>
</tr>
</tbody>
</table>
Cohort Characterization

• Demographics
  – 5 year age categories
  – Index month and year

• Condition occurrence, procedure occurrence, and drug exposure covariates
  – Proportion of persons with event observed in 365 days before index for mothers and 365 days after index for offspring
  – Conditions and procedures were coded using the SNOMED-CT ontology
  – Drugs aggregated at 3rd, 4th, and 5th levels of the ATC and RxNorm ingredient level

• 54,779 and 46,130 covariates constructed for mother and offspring cohorts, respectively
Cohort Characterization

• R package and Shiny application for constructing cohorts, constructing covariates, evaluating MOP generalizability to all-mother and all-offspring, and populating MOPs information in FACT_RELATIONSHIP

• Similarity suggests MOP generalizability to the commercial claims population

• High similarity supports use of MOP sample for teratogenic exposure studies

• Demo
Results – Interactive visualization

**Figure 3. Application screen shot**

Select concept domain:

- All domains

### Mothers

<table>
<thead>
<tr>
<th>Domain</th>
<th>Name</th>
<th>Linked</th>
<th>All</th>
<th>Std diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographics</td>
<td>Age group: 20-34</td>
<td>0.0560</td>
<td>0.1177</td>
<td>0.2283</td>
</tr>
<tr>
<td>Demographics</td>
<td>Age group: 15-19</td>
<td>0.0039</td>
<td>0.0311</td>
<td>0.2085</td>
</tr>
<tr>
<td>Demographics</td>
<td>Index month: 12</td>
<td>0.0006</td>
<td>0.0105</td>
<td>0.1789</td>
</tr>
<tr>
<td>Demographics</td>
<td>Age group: 30-34</td>
<td>0.4250</td>
<td>0.2768</td>
<td>0.0988</td>
</tr>
</tbody>
</table>

### Offspring

<table>
<thead>
<tr>
<th>Domain</th>
<th>Name</th>
<th>Linked</th>
<th>All</th>
<th>Std diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Procedure</td>
<td>Procedure occurrence record for the verbatim concept observed during 3650 on or prior to cohort index: 2514094-Emergency department visit for the evaluation and management of a patient, which requires these 3 key components: A detailed history, A detailed examination, and Medical decision making of moderate complexity. Counseling and/or coordination of care with 6</td>
<td>0.1310</td>
<td>0.1608</td>
<td>0.0822</td>
</tr>
<tr>
<td>Drug</td>
<td>Drug exposure record for the verbatim concept observed during 3650 on or after cohort index: 2213474-Vaccinia virus vaccine, live, for subcutaneous use</td>
<td>0.0115</td>
<td>0.0288</td>
<td>0.4698</td>
</tr>
<tr>
<td>Drug</td>
<td>Drug exposure record for the verbatim concept observed during 3650 on or after cohort index: 2213466-Measles, mumps and rubella virus vaccine</td>
<td>0.0130</td>
<td>0.0238</td>
<td>0.4500</td>
</tr>
</tbody>
</table>
Discussion & Next Steps

- Quick and inexpensive assembly of large MOPs cohorts for investigating maternal drug exposures and associated offspring outcomes

- Potential for pharmacoepidemiology studies of maternal drug exposures to be conducted outside the context of formal pregnancy registries while being HIPAA compliant
FDA Drug Safety Communication: Selective serotonin reuptake inhibitor (SSRI) antidepressant use during pregnancy and reports of a rare heart and lung condition in newborn babies

Safety Announcement
Additional Information for Patients
Additional Information for Healthcare Professionals
Data Summary

Safety Announcement

[12-14-2011] The U.S. Food and Drug Administration (FDA) is updating the public on the use of selective serotonin reuptake inhibitor (SSRI) antidepressants by women during pregnancy and the potential risk of a rare heart and lung condition known as persistent pulmonary hypertension of the newborn (PPHN). The initial Public Health Advisory in July 2006 on this potential risk was based on a single published study. Since then, there have been conflicting findings from new studies evaluating this potential risk, making it unclear whether use of SSRIs during pregnancy can cause PPHN.
Discussion & Next Steps

Paediatric and Perinatal Epidemiology
Affiliated to the Society for Pediatric and Perinatal Epidemiologic Research

578
doi: 10.1111/ppe.12004


Conclusions: AED use during pregnancy increased between 2001 and 2007, driven by a fivefold increase in the use of newer AEDs. Nearly one in eight AED-exposed deliveries involved the concomitant use of more than one AED. Additional investigations of the reproductive safety of newer AEDs may be needed.

Statins and congenital malformations: cohort study

Conclusions: Our analysis did not find a significant teratogenic effect from maternal use of statins in the first trimester. However, these findings need to be replicated in other large studies, and the long term effects of in utero exposure to statins needs to be assessed, before use of statins in pregnancy can be considered safe.

Pregabalin use early in pregnancy and the risk of major congenital malformations

CONCLUSIONS: Findings did not confirm the suggested teratogenic effects of pregabalin, although they cannot rule out the possibility of a small effect.

OHDSI community call 09/12/17
Discussion & Next Steps

June 2, 2015

Antidepressant Use Late in Pregnancy and Risk of Persistent Pulmonary Hypertension of the Newborn

Krista F. Huybrechts, MS, PhD\textsuperscript{1,2}; Brian T. Bateman, MD, MSc\textsuperscript{1,2,3}; Kristin Palmsten, ScD\textsuperscript{4}; et al

\textit{Author Affiliations} | Article Information


\textbf{Conclusions and Relevance} Evidence from this large study of publicly insured pregnant women may be consistent with a potential increased risk of PPHN associated with maternal use of SSRIs in late pregnancy. However, the absolute risk was small, and the risk increase appears more modest than suggested in previous studies.
Discussion & Next Steps

• R package deployment to OHDSI GitHub
Contact

- jweave17@its.jnj.com
- jhardi10@its.jnj.com
Potential linked mothers – women with pregnancy and live birth according to pregnancy episodes by Matcho et al.
N mothers = 2,378,762

Candidate linked mothers – women with pregnancy, live birth, family identifier code where birth event is within insurance coverage period
N mothers = 2,378,224

Linked mothers, women with:
- Family identifier code that includes another person
- Overlapping observation period with another person
- Other person observation period starts at 0 years of age
N mothers = 1,742,479
N offspring = 2,212,339
N records = 2,221,528

Excluded records where 1 baby is associated with multiple mothers
N MOPs = 17,841

Linked MOPs
N mothers = 1,729,530
N offspring = 2,203,687
N unique MOPs = 2,203,687

Linked offspring date of birth is within 60 days of mothers' pregnancy episode end date
N mothers = 1,661,987
N offspring = 1,928,114
N unique MOPs = 1,928,114

All potential offspring – year of observation period start equals year of birth
N offspring = 3,853,277
Supplementary

- 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
- MOP mothers
  i. Matcho et al. [14] algorithm identifies pregnancy episodes; limited to live births
  ii. Initial pool of women limited to those whose birth event was during insurance coverage period
  iii. Restricted to those who share a family identifier code and overlapping observation time with another person who is 0 years of age at database entry

II. MOP offspring
  a. People associated with mothers in I.iii. are candidate offspring
  b. Restricted to those whose insurance coverage start date is within 60 days of linked mothers pregnancy end date

III. All-mothers: mothers identified using the pregnancy episode identification algorithm [14] who have had a live birth outcome

IV. All-offspring: people 0 years of age at database entry

• index date in the linked-mothers, linked-offspring, and all-mothers cohorts is the pregnancy episode end date; index date in the all-offspring cohort is date of birth