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

James Weaver MS[1,2], Jill Hardin PhD[1,2], Patrick B. Ryan PhD[1,2,3]

[1]Janssen Research & Development, LLC, Raritan, NJ, [2]Observational Health Data Sciences and Informatics (OHDSI), New York, NY, USA, [3]Columbia 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 advances prior work by applying additional criteria to increase linkage confidence and by evaluating generalizability to the US commercial claims population.

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

Methods: The Truven Health MarketScan Research Database (1/1/2000-4/30/2016) 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 code and who have observation time overlapping with a person who is 0 years of age at database entry. These candidate offspring were then restricted to those whose birth date is within 60 days of the mother's pregnancy episode end 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 standardized difference of means to assess generalizability.

Results: The MOPs algorithm identified 1,661,987 mothers and 1,928,114 offspring. MOPs covered 70% of all-mothers (N=2,378,762) and 50% of all-offspring cohorts (N=3,853,277); 92% of observation start dates of MOP offspring were within 4 weeks of the pregnancy end 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:
  - The gap of drug safety information during pregnancy as premarketing randomized clinical trials exclude pregnant women due to ethical considerations [1]
  - The lack of evidence on benefits and risks of drug use during pregnancy and resultant birth outcomes [2]
  - Information on drug risks during pregnancy [3-12] 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 practice patterns

OBJECTIVE

- 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

METHODS

Data Source

- Medical and pharmacy claims from Truven Health MarketScan Commercial Claims and Encounters (CCAE) database between January 1, 2000 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 vocabularies for each domain (conditions, drugs, procedures, etc.)
- Algorithm developed using CCAE database transformed into the OMOP CDM

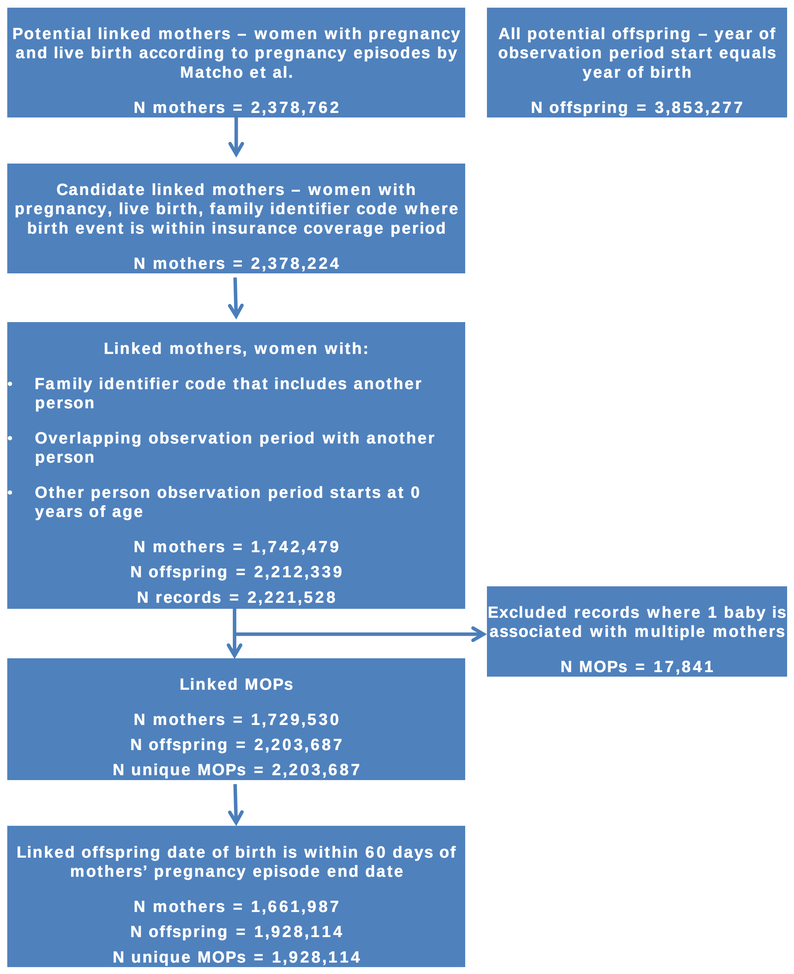
Cohort Construction

- 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**
  - I. 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

Cohort Characterization

- Before- and after-birth 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 constructed for all cohorts; gender covariate constructed for offspring cohorts

Figure 1. Maternal offspring pair algorithm attrition diagram



Cohort Characterization, cont'd.

- Prevalence of condition occurrence, procedure occurrence, and drug exposure were estimated
  - Conditions and procedures were coded using the SNOMED-CT ontology
  - Drugs were aggregated at 3rd, 4th, and 5th levels of the Anatomical Therapeutic Chemical Classification System (ATC) [27] or at the RxNorm [28] ingredient level
  - Covariates constructed based on the proportion of persons with the occurrence observed in 365 days before birth for mothers and 365 days after birth for offspring
- 54,779 and 46,130 covariates constructed for the mother and offspring cohorts

Cohort Comparison

- Comparisons 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

- 1,928,114 unique MOPs were identified covering 70% of all-mothers (N=2,378,762) and 50% of all-offspring cohorts (N=3,853,277)

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

Covariate	MOP mothers	All-mothers	Std. diff. mean
Age group: 10-14	0	0.0002	0.0179
Age group: 15-19	0.004	0.0317	0.2101
Age group: 20-24	0.0572	0.1167	0.2127
Age group: 25-29	0.2864	0.2821	0.0094
Age group: 30-34	0.4229	0.3753	0.0972
Age group: 35-39	0.2423	0.2138	0.0681
Age group: 40-44	0.0573	0.0523	0.0223
Age group: 45-49	0.0042	0.0049	0.0107
Age group: 50-54	0.0003	0.0011	0.0322
Age group: 55-59	0	0.0002	0.0189
Index year: 2000	0.0014	0.0025	0.0248
Index year: 2001	0.0097	0.011	0.0124
Index year: 2002	0.0172	0.0193	0.0158
Index year: 2003	0.0335	0.0344	0.0046
Index year: 2004	0.0504	0.0491	0.0059
Index year: 2005	0.062	0.0599	0.0089
Index year: 2006	0.0617	0.0597	0.0084
Index year: 2007	0.0732	0.0714	0.0069
Index year: 2008	0.0798	0.0783	0.0057
Index year: 2009	0.0969	0.0934	0.0119
Index year: 2010	0.0938	0.0957	0.0063
Index year: 2011	0.1163	0.1131	0.01
Index year: 2012	0.1221	0.1266	0.0136
Index year: 2013	0.1018	0.1039	0.0069
Index year: 2014	0.1054	0.108	0.0082
Index year: 2015	0.0821	0.0872	0.0183
Index year: 2016	0.0398	0.0445	0.0231
Index month: 1	0.0924	0.0877	0.0166
Index month: 2	0.0897	0.0841	0.0201
Index month: 3	0.1006	0.0941	0.0221
Index month: 4	0.0982	0.092	0.0212
Index month: 5	0.1024	0.0958	0.022
Index month: 6	0.0981	0.094	0.0136
Index month: 7	0.0969	0.0933	0.0121
Index month: 8	0.097	0.0924	0.0157
Index month: 9	0.0946	0.0914	0.011
Index month: 10	0.1086	0.1066	0.0064
Index month: 11	0.0947	0.1053	0.0352
Index month: 12	0.06	0.1083	0.1745
Mean obs. time before birth, days	-1029.2	-1073.5	0.0566
Mean obs. time after birth, days	1071.1	996.8	0.0693

Figure 2. MOP mothers vs. all-mothers cohort comparison on condition occurrence, procedure occurrence, and drug exposure

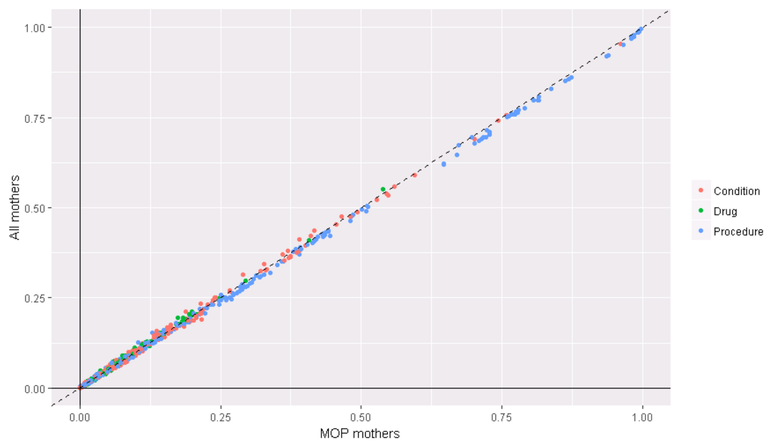
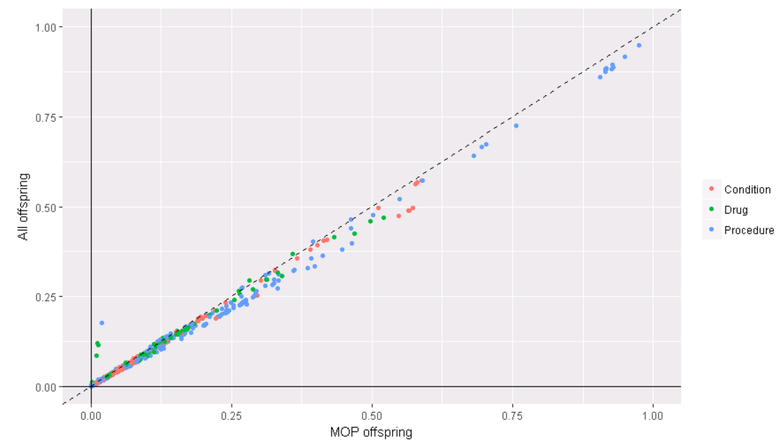


Table 2. Key demographic covariate proportions / means and standardized differences of MOP offspring and all-offspring

Covariate	MOP offspring	All-offspring	Std. diff. mean
Gender: Female	0.4861	0.4864	0.0005
Index year: 2000	0.0012	0.0074	0.094
Index year: 2001	0.0085	0.0109	0.0245
Index year: 2002	0.015	0.0234	0.061
Index year: 2003	0.0292	0.0392	0.0548
Index year: 2004	0.0439	0.0487	0.0228
Index year: 2005	0.0541	0.0545	0.002
Index year: 2006	0.0538	0.0576	0.0169
Index year: 2007	0.0637	0.0615	0.0091
Index year: 2008	0.0696	0.0734	0.0147
Index year: 2009	0.0845	0.0829	0.0058
Index year: 2010	0.0819	0.0874	0.0197
Index year: 2011	0.1013	0.0961	0.0174
Index year: 2012	0.1064	0.0981	0.0275
Index year: 2013	0.0888	0.0835	0.0188
Index year: 2014	0.0919	0.0851	0.024
Index year: 2015	0.0716	0.0616	0.04
Index year: 2016	0.0346	0.0288	0.0335
Index month: 1	0.0924	0.0877	0.0166
Index month: 2	0.0897	0.0841	0.0201
Index month: 3	0.1006	0.0941	0.0221
Index month: 4	0.0982	0.092	0.0212
Index month: 5	0.1024	0.0958	0.022
Index month: 6	0.0981	0.094	0.0136
Index month: 7	0.0969	0.0933	0.0121
Index month: 8	0.097	0.0924	0.0157
Index month: 9	0.0946	0.0914	0.011
Index month: 10	0.1086	0.1066	0.0064
Index month: 11	0.0947	0.1053	0.0352
Index month: 12	0.06	0.1083	0.1745
Mean obs. time after birth, days	1054.6	1003.8	0.0467

Figure 3. MOP offspring vs. all-offspring cohort comparison on condition occurrence, procedure occurrence, and drug exposure



- All-mothers younger than MOP mothers; greater proportion of 15-19 years, 20-24 years groups observed in all-mothers (standardized difference of mean >0.2)
- All-offspring greater vaccine (varicella, MMR, hepatitis A) exposure than MOP offspring (standardized difference of mean >0.35)

DISCUSSION

- Illustrates the potential for pharmacoepidemiology 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

1. Allesee, L. and C.M. Gallagher, *Pregnancy and Protection: The Ethics of Limiting a Pregnant Woman's Participation in Clinical Trials.* J Clin Res Bioeth, 2011. **2**(108).
2. Blehar, M.C., et al., *Enrolling pregnant women: issues in clinical research.* Womens Health Issues, 2013. **23**(1): p. e39-45.
3. Schaefer, C., *Drug safety in pregnancy: utopia or achievable prospect? Risk information, risk research and advocacy in Teratology Information Services.* Congenit Anom (Kyoto), 2011. **51**(1): p. 6-11.
4. Taguchi, N., et al., *Prenatal exposure to HMG-CoA reductase inhibitors: effects on fetal and neonatal outcomes.* Reprod Toxicol, 2008. **26**(2): p. 175-7.
5. Chambers, C.D., K.A. Johnson, and L.M. Dick, *Birth Outcomes in Pregnant Women Taking Fluoxetine.* New England Journal of Medicine, 1996. **335**(14): p. 1010-1015.
6. Laura Park-Wyllie, P.M., Anne Pastuszak et al, *Birth Defects After Maternal Exposure to Corticosteroids: Prospective Cohort Study and Meta-Analysis of Epidemiological Studies.* Teratology, 2000. **62**: p. 385-392.
7. Thorpe, P.G., et al., *Medications in the first trimester of pregnancy: most common exposures and critical gaps in understanding fetal risk.* Pharmacoepidemiol Drug Saf, 2013. **22**(9): p. 1013-8.
8. Brender, J.D., et al., *Nitrosatable drug exposure during early pregnancy and neural tube defects in offspring: National Birth Defects Prevention Study.* Am J Epidemiol, 2011. **174**(11): p. 1286-95.
9.Tomson, T., et al., *EURAP: An International Registry of Antiepileptic Drugs and Pregnancy.* Epilepsia, 2004. **45**(11): p. 1463-1464.
10. Tomson, T., et al., *Pregnancy registries: differences, similarities, and possible harmonization.* Epilepsia, 2010. **51**(5): p. 909-15.
11. Alice D. White, E.B.A., *The Pregnancy Registry Program at Glaxo Wellcome Company.* J Allergy Clin Immunol, 1999. **103**(2): p. S362-S363.
Kennedy D.L., K. Uhl, and K. S.L., *Pregnancy Exposure Registries.* Drug Safety, 2004. **27**(4): p. 215-228.
12. Kennedy D.L., K. Uhl, and K. S.L., *Pregnancy Exposure Registries.* Drug Safety, 2004. **27**(4): p. 215-228.
13. *Informatics, OMOP Common Data Model.* 2016 [cited 2016 10/5/2016].
14. Matcho, A., et al., *Inferring Pregnancy Episodes and Outcomes within a Distributed Network of Observational Databases.* in press, 2016.

CONFLICT OF INTEREST

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