

Determination of Pregnancy Episodes and Outcomes within a Distributed Network of Observational Databases

Amy Matcho¹, Patrick Ryan¹, Daniel Fife¹, Dina Gifkins¹, Chris Knoll¹, Andrew Friedman¹
¹Janssen Research and Development, Raritan, NJ, USA

Abstract

Observational data can be a valuable resource for research involving pregnancy, but there are no consensus methods for inferring the period of pregnancy from administrative claims and electronic health records. We developed an algorithm for the OMOP common data model that identifies the cohort of women who are pregnant and infers the start of pregnancy, end of pregnancy, and associated pregnancy outcome. We applied this algorithm across 4 databases and conducted electronic pregnancy profile review to evaluate the performance of the phenotype. Validation results suggest it is feasible to identify pregnancy episodes with adequate inferences for pregnancy start and end to enable future epidemiologic research.

Background:

Claims databases and electronic health records (EHR) are valuable resources for evaluating prescription drug effects during pregnancy (1, 2). The use of this type of data in the study of pregnancy has several advantages including accuracy of prescription recording, longitudinal follow up, and potential linkages to infant records with large enough samples to study rare outcomes. However, direct measures of gestational age are generally not available and an algorithm must be developed and validated in order to identify the entire pregnancy episode from start to pregnancy outcome.

Methods:

An algorithm to define pregnancy episodes was developed for 4 (3 US administrative claims databases and 1 UK based EHR) data sources in the Janssen network that were converted to the OMOP CDM. One pregnancy code set and SQL program was used. Valid outcomes were determined for live births, stillbirths, spontaneous and induced abortions and ectopic pregnancies and start dates were estimated using a hierarchy of available pregnancy markers in the following order: last menstrual period, gestational age, infertility procedures, ultrasounds (U/S) of nuchal fold thickness, alpha fetoprotein tests, amenorrhea diagnoses, urine pregnancy tests and average gestational age estimates. Validation included review of 700 electronic pregnancy episode profiles and completion of a survey with 6 questions that assessed the operating characteristics of the algorithm for each. A 2nd validation was performed comparing start estimation with infertility procedures that date conception to algorithm start estimation with remaining markers.

Findings:

Figure 1 shows the proportion of pregnancies with start dates inferred by each marker by pregnancy outcome for each database. All operating characteristics of the algorithm were in agreement at least 90% of the time with reviewer chosen characteristics. 73.3% of algorithm estimated starts for live births were in agreement with infertility procedure estimated starts within 2 weeks in either direction. Pregnancy episode term lengths had expected lengths for almost all outcomes and start estimation methods.

Figure 1: Pregnancy Start Date Inferal Method by Database and Outcome



Conclusion:

An algorithm which identifies pregnancy episodes and outcomes was successfully developed with acceptable accuracy concerning all operating characteristics. This algorithm, which ascertains with reasonable accuracy starts and outcomes of pregnancies, offers a useful tool for database studies of the relationship between pregnancy outcomes and maternal exposures and illnesses during pregnancy.

References:

1. Manson, J.M., B. McFarland, and S. Weiss, *Use of an automated database to evaluate markers for early detection of pregnancy*. Am J Epidemiol, 2001. **154**(2): p. 180-7.
2. Hornbrook, M.C., et al., *Development of an algorithm to identify pregnancy episodes in an integrated health care delivery system*. Health Serv Res, 2007. **42**(2): p. 908-27.