Developing a pregnancy algorithm in ATLAS: Applying start date offset

Rupa Makadia[1,2], Christopher Knoll[1,2], Patrick Ryan[1,2]

1. Johnson & Johnson, Raritan, NJ
2. Observational Health Data Sciences and Informatics, New York, NY

Background

Pregnancy is a complex physiological process with numerous factors that influence maternal and fetal outcomes. Understanding these factors and their impact on pregnancy outcomes requires the analysis of large-scale observational data. To facilitate such analyses, the development of pregnancy algorithms that are reproducible, sharable, and portable has become crucial. These algorithms provide a standardized approach to phenotyping and categorizing pregnancies, enabling researchers to collaborate, share methodologies, and apply them across diverse datasets. Multiple algorithms to infer pregnancy episodes have been published, such as the algorithm by Matcho et al, which uses the pregnancy outcome to begin the assessment of a viable pregnancy episode, and then estimates the start date of pregnancy based on these pregnancy markers (1). However, to date, no pregnancy algorithms have been implemented using ATLAS to enable cohort-based analyses using the OHDSI tool suite. We aimed to reproduce the logic from the Matcho et al algorithm in ATLAS. Implementation required a new feature to be developed in the ATLAS/CIRCE specification: the ability to specify a date offset to be added to a start or end date. We evaluated an ATLAS cohort definition using this new date offset feature and compared the results with the original Matcho et al. implementation.

Methods

A cohort of pregnancy episodes was created using ATLAS based on the logic in Matcho et al pregnancy algorithm. The databases used in this study included Optum© De-Identified Clinformatics® Data Mart Database – Socioecominic-(SES) (Clinformatics) dataset and IBM MarketScan® Databases Commercial Claims (CCAE). The cohort was compared to a cohort of pregnancy episodes identified via an SQL-based implementation of the pregnancy algorithm on the basis of the number of persons, episodes, and episodes that match start and end dates. Patient profiles were reviewed to assess gaps in implementation of the algorithm using the two approaches.

At a high-level, the algorithm was developed using the following logic:

1.) Creating concept sets for each start marker.
2.) Using date offset to set start markers in the index criteria (Figure 1)
3.) Hierarchy of start markers by grouping higher order start markers not to occur prior time (based on marker type).
4.) Inclusion criteria to specify at least 2 markers, age/gender criteria and criteria to remove erroneous episodes based on outcomes.
5.) Exit criteria is set to the outcome.
6.) Event persistence to 300 days or maximum length of pregnancy

Results

The results presented only include live birth outcomes for brevity. Table 1 shows the breakdown of testing for both ATLAS-based and SQL-based algorithms for both databases. The number of pregnant persons and episodes identified in the ATLAS-based algorithm was within 6% of the SQL-based algorithm in CCAE and Optum. About 70% of these episodes had agreement on start dates within 30 days. The ATLAS-

based algorithm identified more patients than the SQL based algorithm (> 100%).

Conclusions

Developing a pregnancy algorithm in ATLAS will promote consistency, transparency, and collaboration among OHDSI researchers. Additional results will be available for the other outcomes (ectopic pregnancy, still birth and abortion). Further refinement is needed for the patient profiles to ensure pregnancy episodes are correctly identified including time between subsequent pregnancies or re-try periods. The additional patients in the ATLAS-based algorithm are included due to having multiple outcomes (delivery and abortion) and need to be reclassified. The adoption of standardized frameworks and open-source tools enhances the scalability and applicability of pregnancy phenotyping with the goal of driving improvements in prenatal care and maternal-fetal health outcomes.

Figure 1. ATLAS screenshot for date offset and index criteria hierarchy.

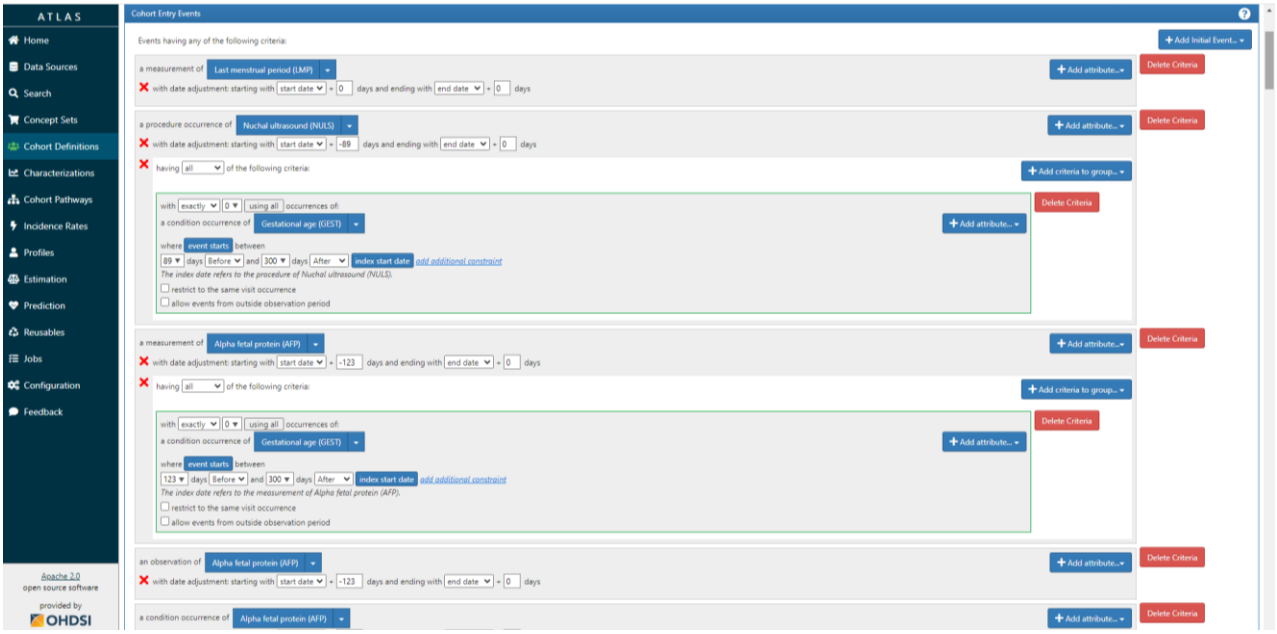

Table 1. Development metrics comparing ATLAS cohort-based episodes to episodes identified by the SQL-based pregnancy algorithm for the outcome of live birth.

CCAE		Clinformatics		Metric
N	% of SQL based episodes or persons	N	% of SQL based episodes or persons	
4,083,362	105.3%	2,147,929	104.3%	Episodes in cohort-based algorithm
3,439,608	105.0%	1,815,038	104.1%	Persons in cohort-based algorithm
852,264	22.0%	492,304	23.9%	Episodes that match start/end date exactly
2,680,732	69.1%	1,447,664	70.3%	Episodes that match within 30 days of start date
2,761,376	71.2%	1,484,988	72.1%	Episodes that match within 30 days of start OR end date

References

1. Matcho A, Ryan P, Fife D, Gifkins D, Knoll C, Friedman A. Inferring pregnancy episodes and outcomes within a network of observational databases. PloS one. 2018;13(2):e0192033.