Developing a perinatal expansion for the OMOP common data model

Alicia Abellan[1], Edward Burn[2], Nhung Trinh[3], Theresa Burkard[2], Sergio Fernández-Bertolín[1], Eimir Hurley[3], Clara Rodriguez[1], Elena Segundo[1], Daniel R. Morales[4], Hedvig Nordeng[5], Talita Duarte-Salles[1]

[1] Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain.

[2] Centre for Statistics in Medicine (CSM), Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (NDORMS), University of Oxford, Oxford, UK.

[3] PharmacoEpidemiology and Drug Safety Research Group, Department of Pharmacy, and PharmaTox Strategic Research Initiative, Faculty of Mathematics and Natural Sciences, University of Oslo, Oslo, Norway.

[4] Division of Population Health and Genomics, University of Dundee, Dundee, UK.

[5] PharmacoEpidemiology and Drug Safety Research Group, Department of Pharmacy, and PharmaTox Strategic Research Initiative, Faculty of Mathematics and Natural Sciences, University of Oslo, Oslo, Norway; Department of Child Health and Development, Norwegian Institute of Public Health, Oslo, Norway.

Background

Many healthcare databases have detailed information related to the perinatal period. Algorithms exist which derive basic perinatal information from most data sources (e.g. dates, mode of delivery, pregnancy outcome) as well as mother-child linkage. Yet, more granular information especially on the infant (anomalies, weight etc.) will not be captured, but is important for perinatal research. Especially for rare outcomes, but also to compare results between countries network studies are key. Since the current Observational Medical Outcomes Partnership common data model (OMOP CDM) lacks the necessary vocabulary to accommodate detailed perinatal information, our aim was to develop and test the implementation of a perinatal expansion for the OMOP CDM to provide a structure to store detailed pregnancies, linked infants and related clinical information.

Methods

We developed a perinatal expansion for the OMOP CDM that includes two tables linked through the individual pregnancy id: a pregnancy table and an infant table. Each table includes a set of required and optional variables corresponding to each pregnancy episode. The established fields and content of new fields within the expansion was derived based on expert opinion with several rounds of review. The pregnancy table includes person and pregnancy id, pregnancy start and end dates, gestational age, mode of delivery, singleton pregnancy, and pregnancy outcomes as required fields. Optional elements include pregnancy-related data (e.g., number of fetuses, pre-pregnancy BMI), and data on previous pregnancies (e.g., parity, previous miscarriages). The infant table includes required variables to link each infant to a pregnancy (i.e., fetus and pregnancy id) and optional fields related

to birth outcomes (e.g., birth weight, APGAR score). We tested the developed expansion tables in two databases with different source data: primary care electronic health records (SIDIAP; Spain) and nationwide health registries (UiO; Norway). We developed an R package to assess the feasibility and quality of its implementation and characterised the pregnancies identified.

Results

Feasibility and quality assessment of the developed expansion tables was successful for both databases with complete data in all required variables. SIDIAP identified 672,058 pregnancies corresponding to 440,031 women during 2006-2022 and UiO 804,227 pregnancies from 491,944 women (2007-2021). Of all pregnancies, 0.4% ended in stillbirths (>20 weeks) in both databases. C-section was performed in 32% (SIDIAP) and 7% (UiO) of pregnancies. Women had a mean pre-pregnancy BMI of 24.8 (SD=4.9) in SIDIAP and 24.5 (SD=4.9) in UiO. Additionally, UiO identified 817,312 infants, with a median birth weight 3,520g (IQR=685) and APGAR score 10 (IQR=1). Data captured with the expansion was consistent with original source data.

Conclusion

We developed a perinatal expansion that will help standardise perinatal data with a high level of validity and granularity in the OMOP CDM and will facilitate the conduction of network perinatal studies such as adverse events, drug use, outcomes in the child etc. The expansion tables were successfully implemented in two large databases but need further testing and validation on a wider range of data types and populations. These tables work particularly well for source data with a high level of granularity as they provide the option to store detailed information (which information is usually not retrievable from algorithms). Yet, the required variables should be fillable also by data sources with less granular information (only including dates, mode of delivery, pregnancy outcome, and singleton pregnancy). This information is retrievable by existing algorithms for example. So, the expansion will work hand in hand with algorithm derived perinatal information for future network studies.