UK Health Data Research Alliance

Meeting Report

Observational Medical Outcome Partnership (OMOP) Special Interest Group: Third Meeting

Introduction

The Health Data Research Alliance held its third OMOP Special Interest Group meeting on Zoom on Wednesday, the 9th of February 2024, from 14:00 to 16:00 (UK time). This meeting aims to bring together user communities of the OMOP common data model. Discussions in this meeting focused on the OHDSI UK study-a-thon: using OMOP data sets to answer real questions.

The Alliance also works in collaboration with related groups, such as the Observational Health Data Sciences and Informatics UK Node and the European Health Data Evidence Network (EHDEN).

The meeting was co-chaired by Geoff Hall (Professor of Digital Health, University of Leeds, and Chief Clinical Data Officer, HDR UK) and Dani Prieto Alhambra (Professor of Pharmaco- and Device Epidemiology, University of Oxford).

The chairs welcomed the attendees and mentioned their intention to make the meeting more interactive through a poll and breakout sessions.

Presentations

Prof. Dani Prieto-Alhambra: OHDSI UK Study-a-thon Introduction.

The study-a-thon started with a discussion with the MHRA (Medicines and Healthcare products Regulatory Agency), a national regulatory body in the UK, about the need for real world evidence, in the OHSDI meeting held in September 2023. A representative from the MHRA attended, presenting the need for real-world evidence for regulatory decision making, as a proof-of-concept. We agreed we would join forces to see what can be achieved with the MHRA.

We conducted studies on topics like use of medicines, although we encountered some challenges and on procedures, and different types of medical devices used in the NHS, which colleagues on the call who participated in the studyathon will shed light on through their presentations.

There is a lot we can learn from on how medical devices and implants perform in the real world. This area lacks randomised control trial and data quality evidence, and so it could benefit from real world evidence.

Annicka Jodicke: Use of Systemic Fluroquinolones in primary care and hospital settings in the UK

- While fluoroquinolones have been widely prescribed in the past to treat respiratory and urinary infections, recently prescriptions have been reduced due to the risk of severe adverse events reported by MHRA in March 2019.

- This research question was to look at overall population level drug utilization: the incidence and prevalence of use of fluoroquinolones in routinely collected data in the UK. We looked at different calendar years and different age groups in the study between 2012 and 2022.
- The preparation of the study began by designing the study protocol and received approval to run the study in our databases. We brainstormed ideas together with our colleagues from MHRA to gather relevant information.
- We looked at the population in the database that had a record of a condition occurrence during the study period and presence in the database for at least 30 days. We were more interested in new users; we included people who did not have a previous record of fluoroquinolone prescription.
- We worked with partners and NHS Trust who contributed data, Lancashire Teaching Hospitals NHS Trust, Great Ormond Street Hospital for Children (GOSH), Health Informatics Centre (Dundee), Barts Bone and Joint Health, and CPRD.
- The study was conducted in two parts: the first was a feasibility check to understand how many records of people with fluoroquinolone they would have in their database by running drug exposure diagnostics and some core diagnostics.
- We used four R-packages for the study: code list generator, incidence prevalence, patient profiles and Drug utilisation.
- The result from population level drug utilisation for primary and secondary care data from Scotland 2012-2022, reveals that the incidence of fluoroquinolones prescription slightly decreased over time, from 2012-2019 where the risk minimization measures (RMM) were issued, incidence reduced, and from 2020, incidence reduced even more.
- A further interrupted time series analyses was conducted to understand the change that happened in 2019, by monitoring any changes in the slope before and after the intervention. The slope was the same, but the incidence rate dropped as a step change. The same was done for Barts Bone and Joint Health hospital data which showed a step change around the RMM. While they implemented the RMM, they had internal recommendations on the use of ciprofloxacin.
- The characteristics of new users across data sets showed 1 million new users for CPRD whereas the paediatric data from GOSH had only 46 children. The median age of people starting treatment among children were from 6 and data from Lancashire shows a median age of 70 and in the primary care data, a median age of 60.
- We looked at whether people had antibiotics in the 30 days prior to determine if fluoroquinolone was administered as first or second line of treatment. Results suggested that many of the fluoroquinolone prescriptions that we saw were second line treatment. We also looked at comorbidities for people with new prescriptions which could make people prone to the adverse effects.
- Comparison of the fluroquinolone indication before and after the RMM suggest that the proportion of people who received fluoroquinolones with indications of urinary

tract infections and some respiratory infections decreased compared to severe infections.

- In conclusion, the RMM was effective in reducing population-level incidence of fluoroquinolone prescription, and showed slightly stronger effects of people from 60 years. New users received fluoroquinolones as a second line treatment and the proportion of prescriptions for urinary tract infection and respiratory tract infection decreased.

Jennifer Lane: Rectopexy- finding surgery and devices in data

- This research explored the surgical and device data we have in the UK OHDSI Node. We worked with the MHRA to identity the need for device surveillance in an improved manner for us to identify patient safety signals sooner rather than later. We explored rectal prolapse and rectopexy.
- We started with the epidemiology of rectal prolapse and rectopexy and the surgical repair. We explored the different types of rectopexy, potential surgical complications, and if we can identify devices in our data and if there is, is it useful?
- Barts Health joined the studyathon as a data partner and has a secondary health care electronic record that covers specialist services predominantly towards the centre of London, with cardiac, cancer, renal, and paediatric services. We are a level 1 trauma centre and part of the university is the Barts London Medical School and Queen Mary University of London. We have a data set with over 2.6 million patients with a 15-year horizon.
- In this study our first two objectives were based upon phenotyping, defining the cohorts of patients that we think were presenting with rectal prolapse, patients with rectal prolapse and then to rectopexy and further went into the subtypes. This was defined from a clinical approach and from a data-driven approach using code list generator.
- We focused on OPSC coding for the surgeries, thinking about how that converts to SNOMED and how that fits in with the OMOP common data model. We then wanted to look at general and specific outcome measures; complications at 30 days, 90 days and 1-2 years.
- The R packages patient profiles, incidence prevalence and drug utilisation were also used. However going beyond that, our plan was to identify how to find our devices to bolt on to some of the studies to know how data comes into our source, how to identify the data and the unique device identifiers.
- To identify the unique device identifier was through the European Union or the US (FDA) search engines. Using the US system we were able to find the unique device identifier for a hip replacement derived from a serial number of a femur fixation, and going back to our research question on rectopexy, we wanted to identify the devices that potentially would have been used in the pelvic reconstruction that had occurred, and we found some codes in our source data, which took us to a mesh product.
- We looked at this question from an abstract sense, whether we could use the data to identify our devices and how this could translate to the whole of the OHDSI UK node. We could identify some specific implants and shared this experience with our colleagues.
- We also identified and generated device-specific cohorts we aim to use in the future. Overall, what we found was over 1.2 million device serial numbers for implanted devices that have

been used in east London, which we would take up with our partners such as Lancashire Teaching Hospitals to identity how we can use this meaningfully in the future.

- The learning point for this study was thinking about the consistency of data entry, knowing that there are different formats of unique device identifier however with a slight element of predictability, understanding UDI format in our data source and the need to develop infrastructure to analyse device data.

Vishnu Chandrabalan: OHDSI OMOP/-The Hard way is the Easy Way.

[Link to Presentation slides]

- The presentation focused on the challenges and solutions related to handling large-scale healthcare data within a complex data landscape. The speaker highlighted the difficulties of managing diverse data sources, including multiple systems operating on different platforms, such as Oracle, SQL Server, and Sybase, with some databases even using German-language schemas.
- A key aspect of the talk was the approach taken to extract, transform, and load (ETL) data into the Observational Medical Outcomes Partnership (OMOP) Common Data Model. The team adopted a structured and collaborative method, leveraging tools like DBT (Data Build Tool) to manage SQL code efficiently, ensure version control, and maintain data lineage. This approach facilitated rapid onboarding of new data engineers and streamlined data transformations.
- The presentation also emphasized the importance of near real-time data processing to support operational needs, research, and direct patient care. The team developed an automated pipeline that refreshes data daily, ensuring that updates are promptly available. The architecture follows a structured medallion model, with Bronze, Silver, and Gold data layers to manage data ingestion, quality checks, and de-identification for research use.
- Finally, the practical impact of these efforts was discussed, demonstrating how efficient data mapping and automation have enhanced data accessibility, accuracy, and usability for various stakeholders. The implementation of an Azure-based ingestion pipeline ensures scalability and reliability, supporting both immediate clinical applications and long-term research initiatives.

Useful link: https://omop-lsc.surge.sh/

Discussion

Several key topics and challenges were discussed. Participants emphasized the importance of integrating OMOP into hospital organizations beyond research funding to ensure sustainability. Various departments, including Microbiology and Paediatrics, have shown significant interest, particularly highlighted by the study-a-thon's focus on fluoroquinolone use for multi-drug-resistant TB in East London.

The discussion also covered the choice of Databricks for scalability, despite the cost, based on positive experiences with workload management and the transformation processes in cloud environments. The lack of data on medical devices was acknowledged, with efforts underway to address this by focusing on unique device identifiers and exploring how to integrate this data into the OMOP CDM.

The meeting highlighted the value of reusable code within the OMOP community, the need for standardized analytics, and the potential to create a community of collaborators to contribute and refine codes across different databases. There was a consensus on the importance of making ETL processes transparent and shareable, particularly for multi-site centres.

Participants stressed the need for tools to simplify the transformation of data into OMOP, with ongoing collaborations to reduce the learning curve and enhance reproducibility. The successful and rapid completion of studies, such as the fluoroquinolone study, was attributed to the availability of tested packages and prepared data partners.

The meeting concluded with an emphasis on balancing the need for high granularity in data with the practicalities of using pre-OMOP data for research. Standardizing data was seen as a valuable process that can improve data quality and facilitate meaningful research.

AOB

There was a call for participation in a survey aimed to identify the gaps of training needs for OMOP for researchers.