UK Health Data Research Alliance

# Meeting Report

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

### Introduction

The Alliance convenes a Special Interest Group devoted to the OMOP Common Data Model[1] to bring together user communities, including those who are transforming data to OMOP and those who are using transformed data for research. The Alliance is also working 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 second meeting of the UK Health Data Research Alliance OMOP Special Interest Group was held on Zoom on Wednesday, the 8th of November 2023, from 13:00 to 14:30 (UK time).

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 the intention to make the meeting more interactive through a poll and breakout sessions.

### Presentations

#### Craig Sachson (Columbia DBMI/OHDSI): An Introduction to the OHDSI Workgroups

[Presentation 08Nov-OHDSIWorkgroups.pdf]

- Craig introduced the OHDSI workgroups ( https://www.ohdsi.org/workgroups/). There are 30 active workgroups which are developed according to community needs and interest. Each workgroup has regular meetings facilitated via Microsoft Teams, with their own specific channels to share information and chat. The OHDSI community is always seeking for new collaborators, those who want to learn or actively participate in different topics. The current working groups are listed below, in 5 main categories:
- Observational Data Standards & Management
    - Common Data Model
    - FHIR and OMOP
    - Healthcare Systems
    - Network Data Quality
    - Vocabulary (CDM Subgroup)
    - Vaccine Vocabulary
- Methodological Research

[1] The Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) is an open community data standard, designed to standardize the structure and content of observational data such as Electronic Health Records (EHR). It is developed and maintained by the Observational Health Data Sciences and Informatics (OHDSI[1]) programme.

    - Methods Research
    - Patient-Level Prediction
- Open-Source Analytics Development
    - ATLAS/WebAPI
    - HADES
    - Open-Source Community
    - Phenotype Development & Evaluation
- Clinical Applications
    - Clinical Trials
    - Dentistry
    - Eye Care & Vision Research
    - Health Equity
    - Oncology
    - Perinatal and Reproductive Health
    - Surgery and Perioperative Medicine
- Community-focussed working groups
    - Early-Stage Researchers
    - Education
    - Steering Group
    - Asia-Pacific – APAC
    - Latin America
- There are also regional chapters and national nodes to address problems specific to their location.
- New workgroups can be established by identifying an area of focus that can impact the OHDSI mission and reaching out to the Steering Group for assistance.

## Asieh Golozar (AstraZeneca): The OMOP Oncology Extension

[Presentation Oncology research in OMOP_HDRUK.pdf

https://ohdsi.github.io/OncologyWG/index.html

- OHDSI Oncology Working Group came together to develop standards and work on an extension that will become a part of OMOP later on. This is because of three problems:

    1. Cancer is a rare disease, with many mutations.
    2. Cancer needs a lot of details because of its complexity.
    3. There are no standards. We don't have good terminologies and vocabularies.

- The OHDSI Oncology working group have been working on solutions:
    - OHDSI Oncology network – data from many institutions can be analysed together.
    - OMOP Oncology Module: three main components:
        1. cancer disease model
        2. cancer treatment model
        3. cancer episode model.
    - Terminologies - identifying proper representation of some of the other pieces using existing terminologies, combining them with a set of cancer modifiers and OMOP genomics developed to represent somatic mutations in OMOP, and we use HemOnc for treatment episodes.

UK Health Data Research Alliance

    - Genomics: Genomic markers need to be turned into features of an analytical dataset. OMOP genomics is built from by multiple knowledge bases. So far, we have 120,000 variants, over 600 genes and 19,000 coding regions.
- In summary, the Oncology module enables observational cancer studies in a network setting.
- Oncology WG Structure:
    - Outreach & Research WG
    - Vocab & Development Subgroup
    - Genomic subgroup

## Workshop session: The minimum viable OMOP data set

### Ed Burn (NDORMS) Use Case 1: OHDSI UK study: "The use of fluoroquinolones in UK primary care and hospital settings")

[Presentation: OMOP SIG mvp.pdf]

https://www.ohdsi-europe.org/index.php/national-nodes/uk

- This presentation explored what to do and prioritize in mapping data to OMOP for organisations that are still in the early stages, and to help rapid adoption.
- Highlighting the OHDSI UK studyathon, which will be done in collaboration with MHRA, one of the two topics will be looking at fluoroquinolones.
- In the simplest approach, two tables (Person and Observation_period) will be required to generate population level denominator cohorts
- This is not quite sufficient for research, but if you map one extra table, e.g., the Drug_exposure table in this case, that will then allow you estimate prevalence, incidence, and drug outcomes.
- Adding a fourth table, Visit_occurrence, that will allow you to see when individuals were hospitalised and with that, we could fulfil first objective. So, with those four tables filled in, you could characterise incidence and also stratify by setting.
- For our second objective (Drug exposure mapping), it is important to understand the granularity of the mapping and its completeness. For duration, we need to have the drug start and end dates.
- Also need Condition_occurrence table to fill in the diagnosis.
- Adding other tables can allow additional analyses.
- **Summary**
    - Defining minimum viable OMOP data set(s) can help
        - prioritise initial mapping efforts
        - achieve early "early wins"
    - Constrain number of domains being mapped (rather than on number/ characteristics of people)
    - Start with defining research questions –> then define minimum viable OMOP data set to answer these
    - Will still need to find a way to get to a complete mapping to realise full potential of the source data

### Geoff Hall (Leeds) Use Case 2: Cancer

[Presentation – HDRUK MEDOC Nov2023.pdf]

- We are part of a consortium of 6 centres that are working across Europe on observational cancer data who came together with an aspiration to adopt OMOP to allow a common data model across six countries to support federated analysis of data (DigiONE pilot).
- In the UK, we have a highly comprehensive candidate data set with 360 data items for cancer, and 80 items in the chemotherapy data set, and we realised we needed to minimise that. So, the group have defined the minimal essential description of cancer (MEDOC), 38 concepts that are the absolute minimum to perform effective observational studies.
- Four DigiONE OMOP studies are in development: lung, breast, ovarian and pan-cancer. I will focus on the pan cancer study, as it absolutely exemplifies the idea of the absolute minimum OMOP conversion you could do to be able to contribute to a study.
- The study will look at the impact of Covid-19 on cancer care. The 38 data item MEDOC data set was minimised to 8 OMOP fields which can be mapped to three tables.
- To know about the patient, we need to know person ID, gender, date of birth, and if they have died, a date of death. Cancer diagnosis will be the condition concept and we'll use the date of diagnosis as the start date. If we capture the last date the patient came to clinic, we therefore can articulate the numbers of cancer in the six centres across Europe.
- I hope this shows that there are incredibly useful projects that you can do with very minimal data mapping. Mapping a very small number of core concepts in a small number of tables that allows you to join the Community, and that is the message that Dani, others and I involved in setting this group up are trying to get across to people across the UK.

## Breakout Sessions

The suggested questions for the breakout rooms were:

1. What are the 1 or 2 most important use cases of OMOP data sets?
2. What is the “minimum viable data” that needs to be mapped for each use case?
3. How can we agree a generic “minimum viable data set” that will work for multiple studies?

After the breakout sessions, a member of each group reported back to the group.

### Group 1

- One consideration that's probably different in different places and potentially unique to the UK is how governance can affect your strategy for mapping; you have to make strategic decisions on what you're going to map depending on when governance will arrive for different types or subsets of data.
- Gordon Milligan (Dundee) shared their experience of contributing to international study of thrombocytopenia (COVID vaccine side effect)

UK Health Data Research Alliance

## Group 2

- The group were enthusiastic about the minimal data set concept. The idea of starting small and growing was popular as many had felt overwhelmed by the size of the challenge and really liked the idea of starting small.
- A question was where these ideas for studies have come from, and what is the process for this group to start proposing these simple projects?
- The role of the big suppliers of EHR systems: TPP and EMIS in primary care, EPIC and others in secondary care, should be recognised. Can they contribute to OMOP mapping? Can this be embedded into the regional SDE programme?

## Group 3

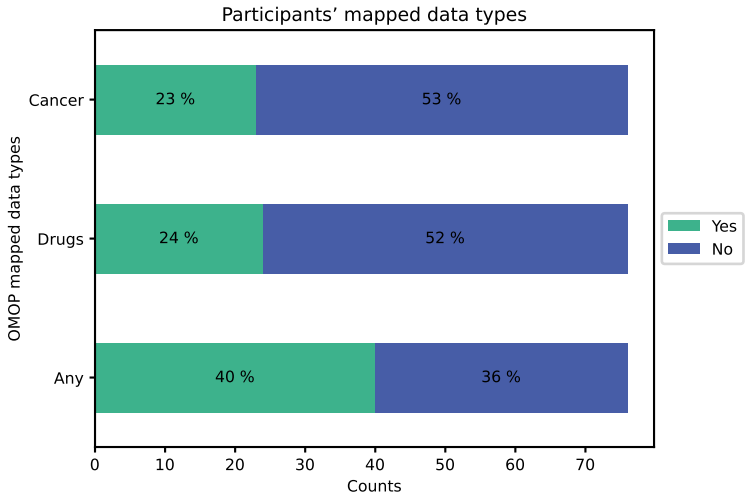
- There was a general agreement on adopting the minimum viable data sets. There were mixed experiences in the room of having mapped their data to OMOP and of those trying to come up with a minimum viable data set.
- The group discussed the importance of identifying a use case, and again the SDE'S were mentioned and to linking the acute care GP and secondary care data and see if there is a use case that can address use of data from multiple sources.
- For thematic use cases, the group mentioned mental health research and the hope that we would cover that as well as other things like diagnosis and secondary care or polypharmacy.

## Group 4

- This group had several representatives who've gone far beyond the minimal mapping of OMOP or have already done the full mapping, e.g. people from Genomics England, Great Ormond Street Hospital, BHF Data Science Centre, the Northwest SDE. So, there is a momentum towards it, but there is still a big challenge in getting to the final mapped data set.
- The discussion showed there may be a slight disconnect between researchers and the mapping, hence there need for researchers and organisations to clearly say what the priorities and reasons are.

## Closing Remarks

- The chairs closed the meeting by thanking everyone and the speakers. Prof. Prieto-Alhambra encouraged continual collaboration, and open discussion. He stated the need to be strategic nationally on the mapping of data to OMOP and bearing in mind all that is happening within the SDE's and the data at National level.



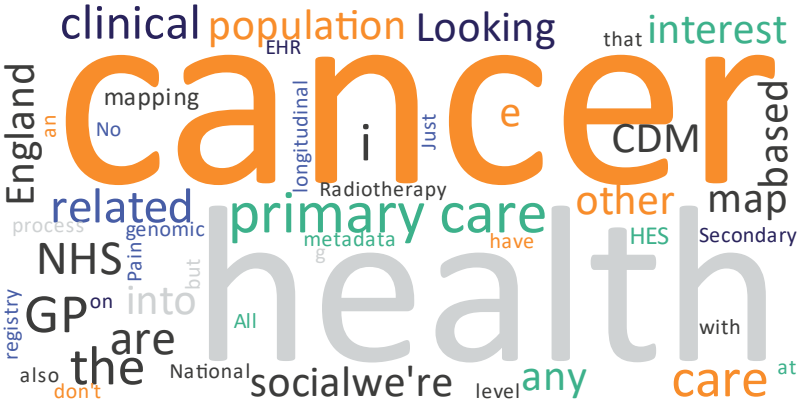

## Polling Results

For this meeting we also included a poll to get input from the attendees.

The questions asked were:

1. Has your organisation mapped any data sets to OMOP?
2. Has your organisation mapped any drug/medication data to OMOP?
3. Has your organisation mapped any cancer data to OMOP (specifically diagnosis, surgery, chemotherapy, or radiotherapy)?
4. What data sets are you interested in?

There were 76 responses to questions 1-3. 40% of organisations had done some OMOP mapping; 24% for drug data and 23% for cancer data. For question 4, which was free text, there were 47 responses which are here shown as a "word cloud[2]" which gives an impression of the dominant themes.

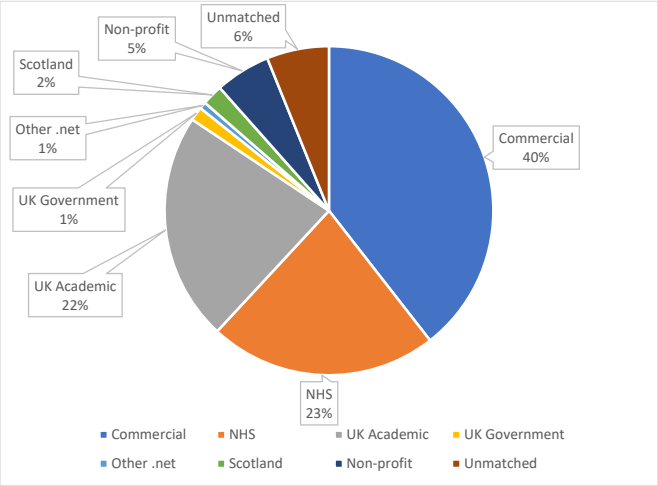


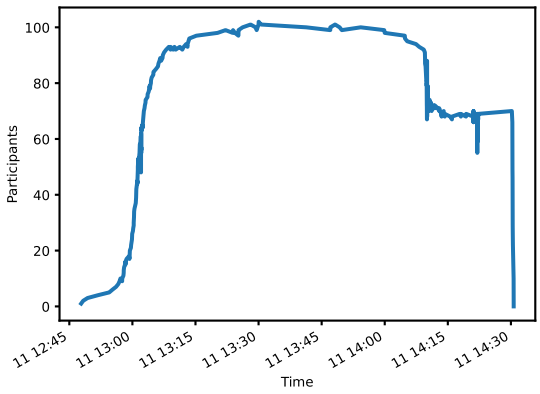

[2] Common words which were excluded were: 'data', 'set', 'sets', 'datasets', 'dataset', 'a', 'OMOP', 'and', 'of', 'to', 'in', 'we', 'our', 'for', 'can', 'as'.

UK Health Data Research Alliance

## Participation

The meeting was well attended. There were 159 people registered on Eventbrite in advance of the meeting. Affiliations inferred from email addresses included commercial organisations (40%), NHS (23%) and academia (22%). Actually attending were 117 unique users (according to Zoom), with a maximum of 102 simultaneous participants. There was a drop in participation (to about 70) when the breakout sessions commenced.




## Useful Links

- OHDSI UK National Node https://ohdsi-europe.org/index.php/national-nodes/uk
- DARWIN EU Project: https://darwin-eu.org/
- OHDSI Forums: https://forums.ohdsi.org/
- OHDSI workgroups: https://www.ohdsi.org/workgroups/
- OMOP Oncology extension: https://ohdsi.github.io/OncologyWG/index.html
- Alliance:
    - OMOP Page: https://ukhealthdata.org/projects/adoption-of-the-omop-common-data-model/
    - Data Standards page: https://ukhealthdata.org/projects/data-standards-and-quality/
- Preliminary results from OMOP adoption survey:
    - Poster: https://doi.org/10.5281/zenodo.8309722
    - Report: https://doi.org/10.5281/zenodo.8309536
- 1st OMOP SIG:
    - meeting report and presentations: https://doi.org/10.5281/zenodo.8191779
    - recording: https://youtu.be/QG_3iBw8Hqs?si=4pkSOupQ_hV_y1ew