MI-CRE 2025 Annual Research Symposium and Policy Forum

Synthetic Creation of OMOP Oncology Patients (SCOOP)

Investigators: Argall, B.[1], Churches, T.[1,2], Niknam, G.[1,2,3], Kennedy, G.[1,2,3]

Author Affiliations:

[1] Faculty of Medicine & Health, UNSW Sydney, Australia

[2] Ingham Institute of Applied Medical Research, Liverpool, Sydney, Australia

[3] Maridulu Budyari Gumal (SPHERE) Cancer Clinical Academic Group, Australia

Presenter's email address: b.argall@student.unsw.edu.au

Disclosure of Interests Statement: None

Is the presenter an HDR student? Masters by coursework student

Has this research been submitted or presented elsewhere? If so where and when? No

Abstract

Research Context: Common Data Models (CDMs) facilitate the secondary research use of electronic medical records. The Observational Medical Outcomes Partnership (OMOP) CDM has been extended to support the oncology domain. This supports the complexities specific to oncology data, including its episodic nature. Synthetic data access is needed to accelerate onboarding of new users and development of further software for these types of complex and sensitive data. Existing software, such as Synthea, allow users to generate synthetic medical data; however, the existing lung cancer module is highly limited, offering only two treatment pathways. The limitations are especially apparent when used for visualisation.

This project focuses on using publicly available data to inform flexible parameterisation of patient care pathways in lung cancer, ensuring that the widely used LUCAP clinical quality indicators will be assessable in the simulated population. Modular development is emphasised to allow adaptation of the population generation methods for the needs of future users.

Engagement Process: HemOnc is a collaboration by hematology and oncology professionals which provides databases on systemic anticancer therapy agents and standard-of-care regimens. Thus, HemOnc facilitates appropriate selection of clinical protocols to simulate lung cancer patient diagnosis and full treatment. HemOnc is integrated into the OMOP oncology extension, included extended dosing information not available in the standard OMOP vocabularies. The exhaustive nature of HemOnc means that generated data can reflect changes in practice over time; allowing meaningful representation of change over time.

A subset of simulated patient journeys will be presented to oncologists to evaluate face validity. Additionally, engagement with the HemOnc editorial board and OMOP oncology work group leadership has been initiated to ensure utility for end users.

Project Outcomes: The core logic and basic treatment selection based on staging and year are in place; work is ongoing.

Reflection and Implications: Visualisation of complex, longitudinal patient journeys is ripe for innovation but necessary test data are difficult to access. Including changes such as the episode model, substantially richer selection and definition of systemic anti-cancer regimens and ability to reflect changes in practice over time will greatly enhance the utility of synthetic data produced.