Creating a general harmonization process flow using a multi-stage transformation and tooling that is technology agnostic could promote further adoption of OMOP across Belgium and lead to opportunities for Belgian data owners to participate in cutting-edge, observational health data studies both within Europe and beyond.

**Title: Creation of a reusable OMOP transformation workflow for Belgian electronic health record systems**

**BACKGROUND:** AZ Damaia is a mid-size regional hospital located in Oostende in Belgium with 513 hospital beds [1]. The hospital joined the European Health Data and Evidence Network (EHDEN) in early 2022, and has been working together with edenceHealth to transform electronic health record (EHR), registry, and laboratory information system (LIS) data to the Observational Medical Outcomes Partnership (OMOP) common data model (CDM). The structure of the transformations that have been implemented thus far lend themselves toward generalizable tooling for hospitals with comparable medical record structures. More specifically, the team has opted for creating pre-OMOP staging tables that link source tables captured in different vendor systems and serving various purposes within the local data warehouse. In this work, we aim to present the approach taken to transform this rich datasource into OMOP CDM by (1) describing the processes in detail, (2) detailing the inherent challenges and advantages when working with this sort of data, and (3) outlining the next steps in generalizing this approach to apply at other hospitals around Belgium with comparable EHR systems.

**METHODOLOGY:** We performed the majority of the semantic mapping for this harmonization effort using edenceReviewer, a web-application alternative to Usagi that enables collaborative review of mapping suggestions. We first deployed an automated mapping – suggestions generated by a set of algorithms curated by edenceHealth – and uploaded it to the web portal, after which members of the team at AZ Damaia manually approved or modified a total of 2140 source-to-standard mappings. In instances where community mappings were available – for ICD10CM, ICD10PCS, ATC, GRR and LOINC codes – we created lookup tables using the OMOP standard vocabulary relationships. We did not need to create custom concepts to map the observational data in this project. Note that the semantic mapping is still ongoing; we are prioritizing drugs initially because that domain has the lowest mapping coverage in our dataset, and the 20 most frequent un-mapped drugs represent nearly 1M additional records. The cancer registry also requires additional manual mapping – coverage using automated ICD03 lookups was rather poor, and we have begun looking toward regional and national cancer registries for collaborations in this effort. We intend to share the semantic mappings we have produced thus far to expedite other Belgian harmonization processes and to help validate the mappings themselves. The main structure of the Extract-Transform-Load (ETL) process is written in python; transformations are written in embedded SQL using an f-string method and the tables are constrained using SQL Alchemy models. The ETL connects to the AZ Damaia data warehouse, specifically referencing a set of views generated for the purposes of this EHDEN project. The ETL captures those views then loads them into a temporary source schema in the target OMOP database where they are joined with lookup files and transformed into their associated OMOP tables before being removed.

**LIMITATIONS AND DISCUSSION:** In total, we transformed approximately 27M unique records (incl. derived OMOP tables) for 162,000 individuals affiliated with AZ Damaia hospital encounters. We faced several challenges in this harmonization effort: there were multiple formats of patient identifiers across different source systems that we needed to consolidate and subsequently validate, data access within particular source vendor systems was restricted and required workarounds for indirect access, and source concept descriptions for the procedural (via invoicing) records in particular were highly granular. Building on these experiences, we are in the process of compiling an EHR OMOP module that could be deployed at other hospitals around Belgium with similar health record profiles and source vendors. We will construct a set of SQL-based procedures and associated mapping tables that can execute at regular intervals within a hospital data warehouse. We expect this module and its various components to improve the harmonization efficiency and conformance for hospitals among various Belgian-specific consortia in which we are participating.

**REFERENCE**