



# Comparing EHR Data and a Research Database of Myeloma using i2b2 and OMOP Common Data Models

Miguel Pedrera Jiménez<sup>1</sup>, M<sup>a</sup> Teresa García Morales<sup>1</sup>, Pablo Serrano Balazote<sup>2</sup>, David Pérez del Rey<sup>3</sup>, Laura Meloni<sup>4</sup>, Joaquín Martínez López<sup>2</sup>, Agustín Gómez de la Cámara<sup>2</sup>

<sup>1</sup>Instituto de Investigación Sanitaria del Hospital 12 de Octubre, <sup>2</sup>Hospital Universitario 12 de Octubre, <sup>3</sup>Universidad Politécnica de Madrid, <sup>4</sup>Custodix NV

## Background

EHR contain health data intended to improve, register and provide the continuity of patient care, that is, responds to a primary use of information. Coexisting with the main EHR, numerous databases are designed and developed for specific research projects and patient cohorts.

- ❖ Ideal scenario: research databases should not reproduce the content of the EHR, but only contain additional data for research.
- ❖ Real scenario: research databases replicate data already registered in the EHR, and even, extend the scope of a primary use of health data.

This work defines the first steps in the transformation of models to compare and evaluate the agreement and completeness of EHR and a database of patients diagnosed with myeloma, using the OMOP (Observational Medical Outcomes Partnership) model as the common data model for comparison.

## Methods

### i2b2 repository of Hospital 12 de Octubre

Due to the participation of the Hospital 12 de Octubre in the InSite Platform, certain data selected from the EHR and other primary departmental systems is periodically loaded into a centralized and standardized repository according to the i2b2 model.

- ❖ The scope of the i2b2 repository covers: demographic information, encounters, diagnosis, procedures, medication, laboratory test results, anatomic pathology and other clinical findings.
- ❖ All observations must be recorded in a standardized coded format. For this purpose, different terminologies and clinical taxonomies are used depending on the type of the clinical concept that is being observed.

### Research database of patients with myeloma

The Hematology Department of Hospital 12 de Octubre has compiled in a database, the information of all the patients included in their clinical trials (Phases I to IV), with a diagnosis of Smoldering Myeloma or Multiple Myeloma, since 2011.

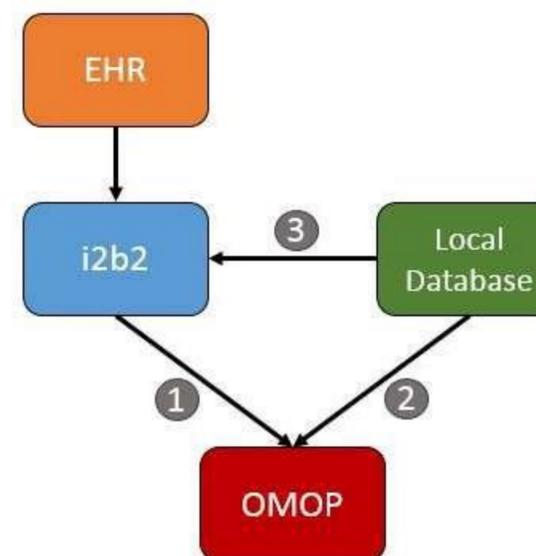
- ❖ The variables collected include the dates of inclusion and discontinuation of the study, reason for the discontinuation, maximum response achieved, and serious toxicity condition.
- ❖ The variables that allow estimating the overall survival of the patient, both inside and outside the study, are also collected.
- ❖ Currently, the database includes information from 21 Phase I-II clinical trials, and 20 Phase III-IV clinical trials, with a total recruitment of 110 and 222 patients respectively.

## Results

This project consists on transforming a subset of the i2b2 repository, and research database of patients with myeloma, into an OMOP instance.

The first step is to define the set of data elements of interest for the clinical case of myeloma. Domain experts, with support from the modeling and terminology experts, must do this. Once the theoretical catalog of elements has been defined, it is possible to carry out the extraction and loading of the data.

Different Extraction Transformation and Load (ETL) processes are required in this project. Thus, the data from the i2b2 repository will be loaded into the OMOP repository (ETL 1). On the other hand, data from the research database for patients with myeloma will be loaded in the OMOP repository ad hoc (ETL 2). Finally, the transformation rules between the research database and the i2b2 repository will be defined (ETL 3).



## Conclusions

In this project, an OMOP repository will be implemented from data of an i2b2 repository and a local database for research of patients with myeloma.

This OMOP repository will allow future work to compare and evaluate both sources of information: EHR and a specific research database.

This work wants to demonstrate that our mapping and transformation methodology is robust and reusable. Preliminary experiments suggests that we might get such results due to the flexible design of i2b2

We expect that such capabilities will facilitate the new uses cases proposed in this challenging project.