A comprehensive **ETL framework**, based on OHDSI suite and open-source tools, to support mapping data from patients with **Hematological Malignancies** to OMOP format

**Title:** Mapping Data of Patients with Hematological Malignancies to the OMOP Common Data Model: A Case Study of Chronic Lymphocytic Leukemia

**Background:**

Chronic lymphocytic leukemia (CLL)
- A paradigmatic hematological malignancy of mature B cells
- The most prevalent adult leukemia in western countries [1]
- Highly heterogeneous clinical course and outcome

Need for large scale collaborations
- to facilitate multi-center research
- to guide the selection of therapeutic protocols towards precision medicine [2]

In this study, we aimed to explore the provided solutions and face the challenges involved in mapping CLL data to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM)

**Results**

Unveiling Transformative Results

- ETL framework for Relational Database with CLL data
  - OHDSI tools
  - In-house code (SQL and Python)
  - Docker Container
- Mapping process
  - 12 table-to-table transformation
  - 17 OMOP vocabularies
- Transformation assessment
  - Applied to a representative subset of anonymized clinical and laboratory data, and the current pass rate of the DQD is 99%, indicating successful execution

Handling challenges in OMOP ETL process of research data

- Treatment description: Mapping treatment lines and corresponding regimens for CLL to EPISODE vs DRUG EXPOSURE vs DRUG ERA tables
- Reporting adverse events: as described for each treatment line
- Positive vs Negative reports: include rule-out conditions and observations
- Molecular information in standardized vocabularies: Missing suitable concepts in order to describe Immunoglobulin (IG) analysis results
- Free text: Refinement of source data prior to transformation, reducing missing values
- Missing dates: Converting description of timepoints to specific date reporting

**Methods**

**Dataset:**
We used retrospective and observational, anonymized data coming from multiple research collaboration between a wide network of hematology centers in Greece.

**Framework:**

- Extract Transform Load (ETL)
- Quality Control

**Conclusions:**
- We highlighted the challenges and provided a solution framework involved in mapping data coming from patients with hematological malignancies to the OMOP CDM and demonstrated the feasibility of this approach using as a pilot the example of CLL.
- This study revealed the importance of collaboration and quality assurance measures in ensuring the accuracy and reliability of data in medical research.
- We provide a foundation for future work in this area and network collaborating studies and highlight the potential benefits of using a common data model to support clinical and translational research.

**References**


Evangelia Mingi [1], Dimitra Chomou, Thomas Chatzikonstantinou, Pantelis Natsiavas, Kostas Stamatopoulos, Evangelos Handakas [2,3], Anastasia Chatzidimitriou [*]