Presentation Abstracts

022

Paving the way to Precision Medicine in ICU: Biobank integration with OMOP-CDM. Challenges, opportunities, and insights from *Hospital Israelita Albert Einstein, Brazil*

Gabriel Mesquita De Souza[1], Maria Tereza Fernandes Abrahão[1], Uri Adrian Prync Flato[1], Mateus de Lima Freitas[1], Gabriela Chiuffa Tunes[1], Etienne Duim[1], Amanda Gomes Rabelo[1], Cesar Truyts[1], Diogo Patrão[1], Edson Amaro Junior[1], Tatiana Ferreira de Almeida[1], Adriano José Pereira[1]

[1] Hospital Israelita Albert Einstein, São Paulo, SP, Brazil.

Category: Safety / Quality / Management

Gabriel Mesquita De Souza - https://orcid.org/0000-0003-0586-5145
Maria Tereza Fernandes Abrahão - https://orcid.org/0000-0003-2701-670X
Uri Adrian Prync Flato - https://orcid.org/0000-0002-8381-8830
Mateus de Lima Freitas - https://orcid.org/0000-0001-5819-7342
Gabriela Chiuffa Tunes - https://orcid.org/0000-0002-8701-4258
Etienne Duim - https://orcid.org/0000-0003-0459-4363
Amanda Rabelo - https://orcid.org/0000-0002-2620-1404
Cesar Truyts - https://orcid.org/0000-0001-5334-0496
Diogo Patrão - https://orcid.org/0000-0002-1129-4337
Edson Amaro Junior - https://orcid.org/0000-0002-5889-1382
Tatiana Ferreira de Almeida - https://orcid.org/0000-0001-9897-6658
Adriano José Pereira - https://orcid.org/0000-0002-9467-6516

Corresponding author
e-mail: gabriel.mesquita@einstein.br

Introduction: Integration of biobank information into the OMOP Common Data Model (CDM) presents both opportunities and challenges in the field of healthcare data management. However, the process of mapping biobank information to the OMOP CDM poses various challenges, including data harmonization, standardization of terminologies, ensuring data quality, and addressing privacy and ethical considerations.[1]

Objective: To demonstrate the challenges encountered in mapping the biobank patients of *Hospital Israelita Albert Einstein* (HIAE) to the OMOP CDM, as well as the opportunities and insights that arised in the context of critically ill patients.

Methods: Qualitative and descriptive study (Conversion to CDM uses SQL and Python scripts). The study utilizes SQL and Python for data mapping and employs Extract, Transform, and Load (ETL) processes. The OHDSI Atlas tool is utilized to visualize mapped data and create cohorts. The Impala engine is used to retrieve data from files stored in Hadoop Distributed File System in OMOP CDM format.

Results: The architecture of the HIAE OMOP Biobank is built upon a local Hadoop cluster boasting a total of 168 cores, 655 GB of memory, and 32.4 TB of storage. This cluster is shared with additional applications and can be accessed through Impala or Spark. Currently, a total of 4,788 biobank patients have been successfully mapped (Figure 1), with their information spread across 11 OMOP-CDM Domain tables, resulting in a grand total of 35.093.493 records and 5,942 unique terms. Of those, 186 patients are critically ill and 2,790 samples of them are stored. By aligning biobank data with the standardized structure of the OMOP-CDM, researchers and healthcare providers can leverage the vast potential of biobank resources for real-world evidence generation, observational research, and personalized medicine. Precision medicine-based interventions could be proposed and tested in clinical trials, using this structure, in near future, in a scenario in ICUs of many negative/neutral trials and generic diagnosis (syndromes).[2] The project encountered several challenges, some of which are commonly faced by projects adopting this model. These challenges include a shortage of experienced teams knowledgeable in

OMOP, the requirement for specialized computational resources, the need for ETL work, and the integration of diverse databases.[1,3] However, working with genetic data introduces additional complexities. The standard OMOP-CDM certification is still being developed for genetic data, resulting in a lack of appropriate concepts for detailed mapping of information.[2]

Conclusion: The initial experience on integrating biobank data and OMOP-CDM in HIAE here reported is believed to be the way to launch the conditions to bring precision to the ICU. Allowing collaborative research using this infrastructure created can support Intensive Care Medicine to overcome the challenge of treating complex and almost unique patients.

▌ REFERENCES:

1. Reinecke I, Zoch M, Reich C, Sedlmayr M, Bathelt F. The Usage of OHDSI OMOP - A Scoping Review. Stud Health Technol Inform. 2021;283:95-103.

2. Peng Y, Nassirian A, Ahmadi N, Sedlmayr M, Bathelt F. Towards the Representation of Genomic Data in HL7 FHIR and OMOP CDM. Stud Health Technol Inform. 2021;283:86–94.

3. Pinto Junior EPP, Normando P, Flores-Ortiz R, Afzal MU, Jamil MA, Bertolin SF, Oliveira VA, Martufi V, de Sousa F, Bashir A, Burn E, Ichihara MY, Barreto ML, Salles TD, Prieto-Alhambra D, Hafeez H, Khalid S. Integrating real-world data from Brazil and Pakistan into the OMOP common data model and standardized health analytics framework to characterize COVID-19 in the Global South. J Am Med Inform Assoc. 2023;30(4):643-55.
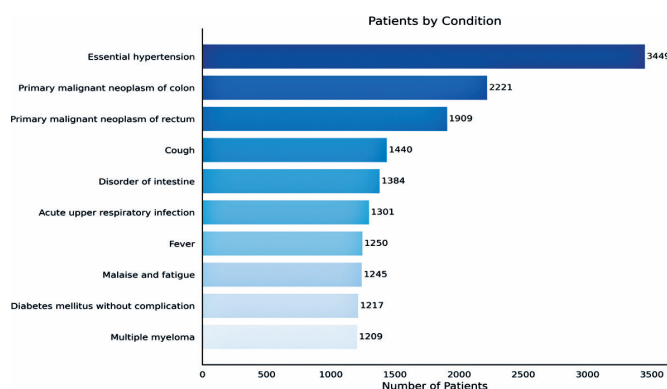
SGPP number: 5231-22

CAAE: 60696722.5.0000.0071

Figure 1. Biobank patients grouped by condition