Transforming Breast and Cervical Cancer Screening Data into the OMOP CDM: Early Implementation Insights from Senegal

Authors
Ousmane Diop[1], Rachel Odhiambo[3], Mamadou Lamine Cissé[2], Aminata Mboup[1]

Affiliations
[1] Institut de Recherche en Santé, de Surveillance Épidémiologique et de Formations, Dakar, Senegal
[2] Hôpital des Enfants de Diamniadio, Dakar, Senegal
[3] African Population and Health Research Center (APHRC), Nairobi, Kenya

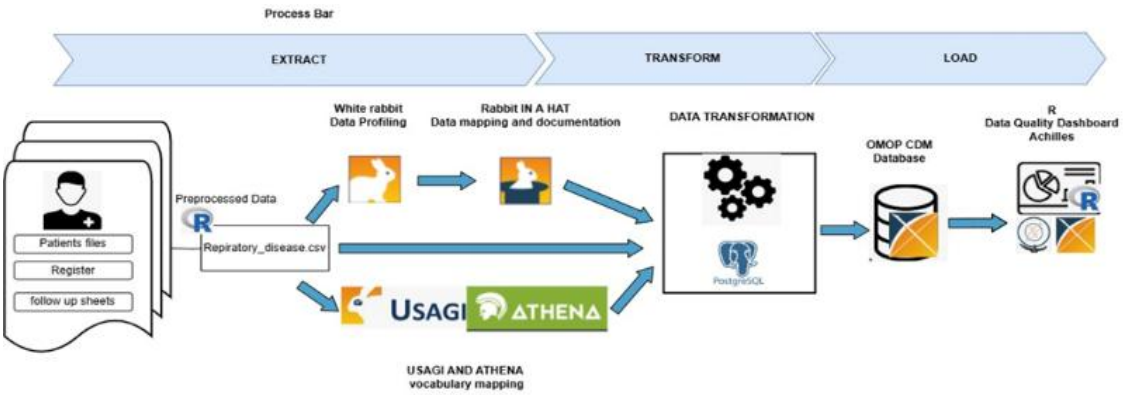
Background

Cancer screening plays a critical role in reducing late-stage diagnoses and improving survival rates, especially in low- and middle-income countries (LMICs). However, the heterogeneity of data collection methods across healthcare institutions in these settings presents challenges for data analysis, integration, and reuse. To facilitate large-scale observational research and enhance interoperability, this study examined the feasibility of converting breast and cervical cancer screening data into the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) as part of a collaborative initiative between research and hospital institutions in West Africa.

Objective

This work aimed to document the implementation process and key challenges encountered while mapping cross-sectional breast and cervical cancer screening data from Senegal into the OMOP Common Data Model (CDM) framework.

Methods



Observational data were collected during annual community-based screening campaign conducted jointly by IRESSEF and Diamniadio Children's Hospital. The dataset included sociodemographic characteristics, screening methods, and clinical outcomes for over 491 women.

As the original data were collected in French, we first conducted a manual translation of all variable names and values into English to ensure compatibility with OHDSI tools and OMOP standard vocabularies.

This step was essential but time-consuming and introduced linguistic challenges, particularly for clinical and locally used terms.

Following the OHDSI ETL framework, we used White Rabbit for data profiling, Usagi, Athena and Jackalope to perform source-to-concept mapping, and Rabbit-in-a-Hat for designing the ETL. We then implemented the transformation and loading process in a PostgreSQL database populated with OMOP vocabulary tables.

Results

An initial assessment of the screening dataset (491 women) was conducted, including sociodemographic data, screening methods, and clinical outcomes. As the data were originally collected in French, all variable names and values were manually translated into English prior to mapping, which added significant time and introduced challenges in interpreting clinical terms. Using OHDSI tools (Usagi, Athena), approximately 80% of the dataset has been mapped to OMOP domains such as person, observation, procedure, and condition. Examples of successfully mapped variables include:

- Age → Person.age
- Married → SNOMED
- Diabetes → LOINC
- family planning → SNOMED

Several variables lacked direct OMOP equivalents, including Quranic education and PCR result test. Local French terminology limited automated mapping, requiring manual curation. The process highlighted resource and infrastructure needs, including translation support and local vocabulary standardization.

Conclusion

Our preliminary demonstrates the implementation of OMOP CDM for cervical and breast cancer screening data in a low-resource setting. While OHDSI tools facilitated the process, challenges included manual translation from French, gaps in vocabulary coverage, and mismatches with local terminology.

These issues highlight the need for improved support for multilingual health datasets. Standardizing data collection during campaigns like Pink October could enhance interoperability and enable collaborative research. This initiative lays the foundation for a standardized health data ecosystem in West Africa.