

# Development of a Framework for Cancer Profiling and Visualization

Soobeen Seol<sup>1</sup>, Jimyung Park<sup>1</sup>, Chungsoo Kim<sup>1</sup>, Rae Woong Park<sup>1,2</sup>

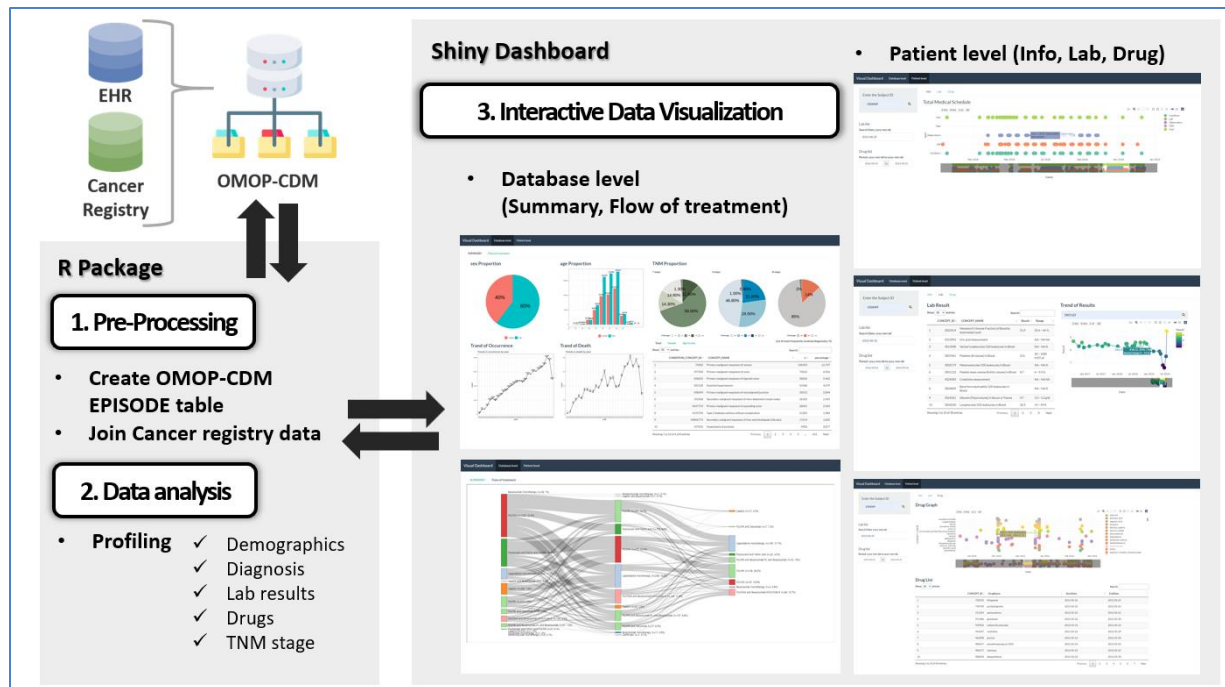
<sup>1</sup> Department of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, Gyeonggi-do, Republic of Korea; <sup>2</sup>Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Gyeonggi-do, Republic of Korea

## Background

Treatment of cancer requires regular and frequent follow-up and interventions, particularly within the initial period after diagnosis. Furthermore, the patient's information should be comprehensively searched and viewed for the clinical decision making. However, the patient's information usually fragmented across Electronic Health Records (EHRs) from structured to unstructured form<sup>1</sup>. Fragmented data is difficult to process quickly and accurately, and cognitive effort is required to structure clinical decisions. Several clinical decision supporting systems have been developed to facilitate informed decision making, but these systems are still difficult to be applied to other institutions. Hence, it is necessary to develop a scalable data exploratory system while profiling patient information from both structured and unstructured data. In this study, we aimed to develop a patient data profiling application based on the standardized medical database.

## Methods

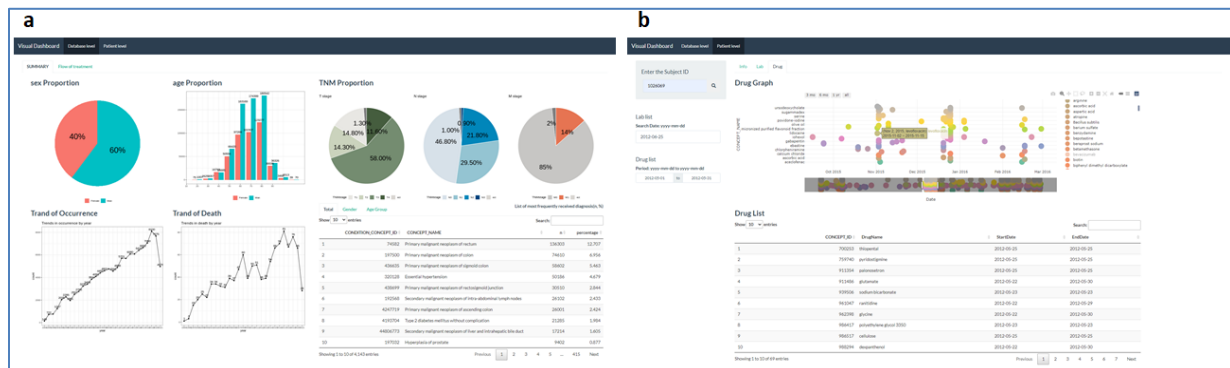
This study used the Ajou University School of Medicine (AUSOM) database that is a tertiary hospital database in South Korea, containing  $\geq 3.4$ M patients. The AUSOM database has been converted into Observational Medical Outcome Partnership - Common Data Model (OMOP-CDM). As a proof-of-concept study, we defined the target cohort as the patients diagnosed with colorectal cancer and used it for the application development. Figure 1 shows the workflow. We added 66 chemotherapy regimens to the Tool For Regimen-level Abstraction of Chemotherapy Episode Records (TRACER)<sup>2</sup>. TRACER is an algorithm extracting individual chemo-drugs from OMOP-CDM and generating an OMOP-CDM EPISODE table. In the preprocessing, we created an EPISODE table using TRACER. Additionally, since cancer registry data in the subject hospital is separately existing from EHRs, we identified the target patients' cancer registry data and linked it to the OMOP-CDM. The overall comprehensive data was profiled and visualized in the application. Profiling is included general demographics, disease diagnoses, laboratory test results, medications, procedures, and TNM stages as a new data structure that represents in time series representation of the individual's entire medical schedule. Finally, the application was developed with the interactive RShiny application. This dashboard visualizes database-level profiling based on total cohort and patient-level profiling based on individual patients' data.



**Figure 1.** The general workflow of this study consists of preprocessing, data analysis, and interactive data visualization.

## Results

The OMOP-CDM EPISODE table was created from the EHR data of 10,258 patients with colorectal cancer, and combined that table with cancer registry data. We profiled the data as time series and analyzed it. The dashboard was implemented based on data analysis results from profiled patients (Figure 2). The dashboard consists of a database-level and a patient-level. The database-level is designed to understand briefly the cohort-based demographics, trends of incidence, mortality, TNM stage distribution, frequency of diagnosis, and flow of treatment. The flow of treatment does not simply show an individual chemo-drugs, but takes the form of treatment trajectories of chemotherapy regimens. The patient-level is designed to understand immediately a patient's overall medical visit, test results and trends, and a list and trends of medications taken. The overall medical visit marked each event with a dot plot and provides a brief description. In the 'Lab' tab, these are shown tests performed on a selected date and trends of the test results that entered in the test code. Also, in the 'Drug' tab, these are shown trends of all the drugs prescribed to the patient and a list of drugs prescribed to the patients within the selected period.



**Figure 2.** A cancer patient-centric visualization Dashboard based on profiled in time series and analyzed data in the AUSOM database. a, Summary tab on Database level page. b, Drug tab on Patient level page.

## **Conclusion**

This study reports the results of initial efforts to develop a standardized framework that systematically and visually represents the current state of patient with cancer and supports clinical decisions. We have created a new data table structured from the integration of cancer registry data and OMOP-CDM EPISODE table. Comprehensive health data including the new data table was profiled in time series and analyzed in an integrated manner. Based on aggregated summary statistical data, we implemented a patient-centric visualization dashboard that can monitor and analyze individuals and groups with cancer. Consequently, a patient data profiling application based on standardized medical database was created and it will provide insights into patient with cancer.

## **Acknowledgement**

This research was supported by a grant of the project for Infectious Disease Medical Safety, funded by the Ministry of Health, Republic of Korea (grant number: HG22C0024). This research was also funded by the Bio Industrial Strategic Technology Development Program (20003883, 20005021) funded By the Ministry of Trade, Industry & Energy (MOTIE, Korea) and a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HR16C0001).

## **References**

1. Morin, O., Vallières, M., Braunstein, S. et al. An artificial intelligence framework integrating longitudinal electronic health records with real-world data enables continuous pan-cancer prognostication. *Nat Cancer* 2, (2021);709–722
2. Jeon H, You S, Kang S, Seo S, Warner J, Belenkaya R, Park R, Characterizing the Anticancer Treatment Trajectory and Pattern in Patients Receiving Chemotherapy for Cancer Using Harmonized Observational Databases: Retrospective Study, *JMIR* 2021;9(4):e25035