Development of a Framework for Cancer Profiling and Visualization

Soobeen Seol[1], Jimyung Park[1], Chungsoo Kim [1], Dong Yun Lee[2], Byungjin Choi[1], Rae Woong Park[1, 2]

[1]Department of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, Gyeonggi-do, Republic of Korea; [2]Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Gyeonggi-do, Republic of Korea

Background

Patients with cancer require regular clinical interventions with frequent follow-up visits. Clinicians should be able to access patient data for a more accurate clinical decision-making process. However, patient data is usually fragmented in the electronic health records (EHRs) from structured to unstructured form[1]. This can lead to unnecessary cognitive effort, interfering with a timely and accurate clinical process. However, prior works are lacking studies of data exploration systems that can reduce unnecessary cognitive effort by summarizing patient data.

In this study, we introduce a framework that can profile patient data using fragmented data in EHRs. The profiled data is shown as visualized information that can be understood at a glance about patients with cancer. We report the result of initial efforts to develop a framework with a focus on oncology.

Methods

The framework was developed using the Ajou University School of Medicine (AUSOM) database which is a tertiary hospital database in South Korea, containing $\geq$ 2.9M patients in Observational Medical Outcome Partnership - Common Data Model (OMOP-CDM) format. The framework consists of preprocessing, data analysis, interactive data visualization, and cohort generation. Figure 1 describes the workflow of the framework development.

We extracted the chemotherapy information as regimen-level using the Tool for Regimen-level Abstraction of Chemotherapy Episode Records (TRACER)[2]. We updated the TRACER with 66 newly introduced chemotherapy. Generated regimen-level data was stored in the EPISODE table of OMOP-CDM. The TNM stage from cancer registry data, which was operated separately in the subject hospital, was linked to the MEASUREMENT table of the OMOP-CDM.

Subsequently, we developed a dashboard profiling and visualizing cancer data. The data profiling included demographics, diagnoses, laboratory tests, medications, procedures and TNM stages. The visualization of profiled data was developed using the R Shiny. The dashboard consists of three main parts: 1) cohort-level visualization, 2) individual-level visualization and 3) cohort generation.

The cohort-level visualization was designed with 'Summary' and 'Treatment' tabs to provide a brief overview of the cohort. The 'Summary' tab includes information such as proportion of gender and age group, trends of occurrence of disease, mortality and TNM staging distribution. In 'Treatment' tab, not only appears a flow of chemotherapy regimens generated by the TRACER but also statistical information on admissions and visits of the cohort.

The individual-level visualization was designed to give a detailed understanding of each patient by presenting trends of diagnoses, test results and prescribed medications when entering subject id. This visualization was extended from a tool for exploring individual longitudinal observation data named profiles in ATLAS, an interactive analysis platform. While both the profile in ATLAS and the individual-level visualization show a summary of individual patient information, individual-level visualization includes

graphs over time as well as a simple list. For example, in the 'Lab' tab, a list of tests performed on a selected date was shown and a graph presented trends of the selected test result. Also, in the 'Drug' tab, a graph presented trends of all the drugs prescribed and a list of drugs prescribed was shown within the selected period.

The cohort generation was designated to generate new cohorts explored. It enables eliciting a subset from the original cohort by gender, age, and TNM stage.

As a proof-of-concept study, we defined a target cohort as the patients diagnosed with colorectal cancer (CRC), which showed a rapid increase in incidence in South Korea in recent decades[3].

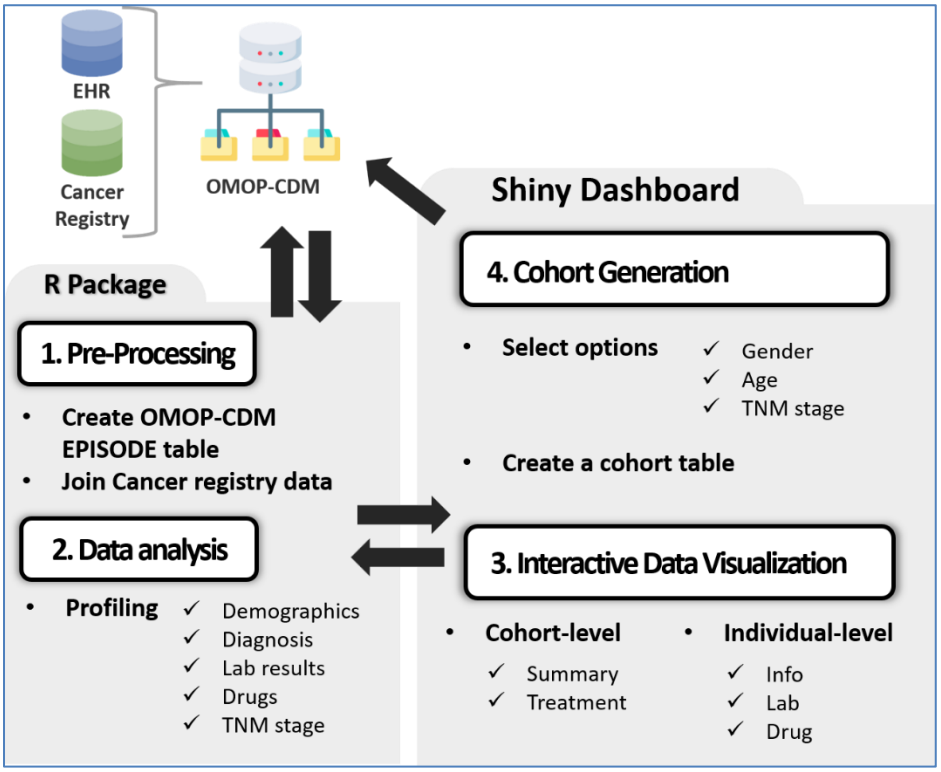

Figure 1. The schematic workflow of the framework. EHR: electronic health record; OMOP-CDM: Observational Medical Outcome Partnership-Common Data Model.

Results

The colorectal cancer was identified as 10,258 subjects in the AUSOM database. We extracted the information of chemo-regimen and inserted 212,327 episodes into the EPISODE table of the OMOP-CDM. The framework provides a dashboard which visualized the profiled patients' data (Figure 2).

On the summary tab of the cohort-level page, among the colorectal cancer, the ratio of men was 59.10%, and the main age group was in 60s (27.38%). The T3, N0 and M0 was the highest proportion for each stage. On the treatment tab of the cohort-level page, the mean (SD) observation period was 1521.24 (1682.96) days, and the mean (SD) duration of hospitalization was 13.26 (45.98) days.

The detailed treatment information of each patient was found on the individual-level page. It shows the

patient's overall treatment history as well as the treatment history for a specific period such as lab results, list of prescribed drugs.

On the cohort generation page, the preview displays the selected cohort that will be created in the cohort database via the generation button. The age range in the sidebar is displayed from the minimum age to the maximum age of the cohort. The preview displays cohort definition id, subject id, cohort start date, cohort end date, TNM stage and organ code.

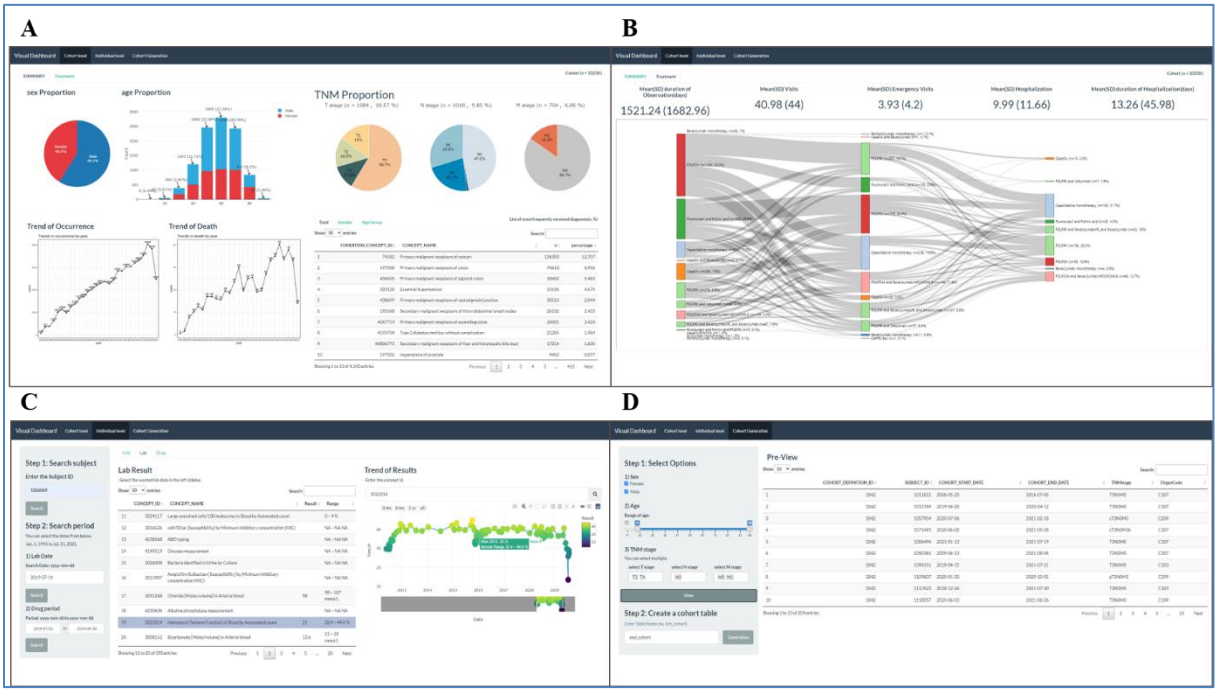

Figure 2. Snapshots of the dashboard that visualize the profiled patients' data. A, summary tab on cohort-level page. B, treatment tab on cohort-level page. C, lab tab on the individual-level page. D, cohort generation page.

Conclusion

In this study, we developed a standardized framework, which seeks to support clinical decisions by representing the longitudinal status of patients with cancer. For comprehensive information, we integrated the cancer registry data into the OMOP-CDM. The dashboard not only presents the aggregated summary statistics of cancer but patient-centric information through visualization. This visualized information can help clinicians to monitor or analyze patients with cancer more in-depth. Additionally, the function which generates a cohort and inserts it into the database can be the basis of subsequent studies.

Acknowledgment

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
3. Khil, H., Kim, S.M., Hong, S. et al. Time trends of colorectal cancer incidence and associated lifestyle factors in South Korea. Sci Rep 11, 2413 (2021).