CohortContrast: Universal Patient Trajectory Extraction from OMOP CDM

Markus Haug, MSc[1] (markus.haug@ut.ee),
Edward Burn, PhD[2] (edward.burn@ndorms.ox.ac.uk),
Martí Català, PhD[2], (marti.catalasabate@ndorms.ox.ac.uk),
Marta Alcalde-Herraiz, BSc[2], (marta.alcaldeherraiz@ndorms.ox.ac.uk),
Raivo Kolde, PhD[1] (raivo.kolde@ut.ee).
[1]Institute of Computer Science (ICS), University of Tartu, Tartu, Estonia Health Data Sciences,
[2]Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (NDROMS), University of Oxford, United Kingdom

Background

Real-world data enables investigations into the patient treatment pathways and adherence to standard care protocols. Our prior work has focused on developing tools for trajectory construction within the OMOP Common Data Model (CDM) framework and facilitating federated research to investigate these pathways.[1,2,3,4] However, a significant bottleneck in such analyses is defining the relevant events or states of the care process on the data. It requires clinical knowledge about the particular condition and its treatment, as well as about the coding of the data in a particular database. Here, we propose a generic data-driven solution for identifying relevant events for any OMOP cohort through statistical enrichment analysis, accelerating and automating parts of the treatment pathway analysis on OMOP CDM databases.

Methods

All the proposed methods are implemented in a package CohortContrast[5]. The principal input is a target cohort that defines the patients and periods of interest. Prevalence of events recorded in the OMOP database for these periods is compared to the prevalence of the same events within control periods. Several statistical methods, including Z-tests and logistic regression models, are implemented to ensure robust analysis. For defining the control cohort, we have multiple strategies, including self-control and age and sex based matching. The resulting enriched event occurrences can be further characterized by the prevalence, co-occurrence with other events, the demographic details of the patients involved and temporal relations to cohort index events, providing a comprehensive overview of patient treatment in the target cohort.

Depending on the construction of the target and control cohorts it is possible to address wide variety of questions such as:

- What events precede a patient's diagnosis?
- What treatments are administered post-diagnosis?
- What treatment plans are currently practiced?
- How do different treatment regimens compare?
- How have treatment policies evolved over time?
- What are the demographic differences in treatment approaches?
- How do extracted patient clusters differ?
- What are the side effects associated with treatments?

Using patient data from the OMOP CDM and the extracted features, we developed an interactive graphical user interface (GUI) to facilitate further analysis. The GUI allows users to filter concepts by risk ratio and target group prevalence and visualize the prevalence and co-occurrence of the filtered

events. Additionally, users can adjust comparisons over time and invert the cohorts for different perspectives. A user-friendly mapping system within the GUI enables the combination of concepts and the saving of analysis snapshots. We are developing features for utilizing the hierarchy of concept codes for combining events, facilitating the choice of an appropriate level of abstraction. The resulting events and patient trajectories can be exported from the tool for further exploration.

As a case study, we have examined a prostate cancer cohort investigated by Kannus et al.[7] within an OMOP CDM patient trajectory study. For the control cohort, we utilized self-controls from the period before the diagnosis. We identified all drugs, procedures, measurements, conditions, observations and visit types related to patients in the target cohort, highlighting statistically significant prevalence differences through logistic regression analysis. Subsequently, we validated our findings against the standard of care reported by Kannus et al. This study was conducted using the RITA-MAITT database, which includes billing information, claims, prescriptions, and electronic health records for a 10% random sample of the Estonian population.[8]

Results

Our study included a target cohort of 779 newly diagnosed prostate cancer patients. We identified 114 concept IDs with statistically significant differences in prevalence when compared to the control cohort. Among these, 37% were measurements, 25% were procedures, 20% were conditions, 11% were drugs and 6% were observations. A notable portion of these concepts were referenced in the work by Kannus et al.[7]

The identified drugs (Figure 1) were primarily related to hormone therapy (e.g., triptorelin, bicalutamide), symptom management (e.g., tamsulosin, diclofenac), infection control (e.g., amoxicillin, clarithromycin, ciprofloxacin), and anticoagulation (e.g., enoxaparin). The extracted procedures included radiotherapy planning, radiation oncology, teleradiotherapy, diathermy, prostatectomies, and biopsies, most of which align with therapeutic procedures discussed in the referenced study.

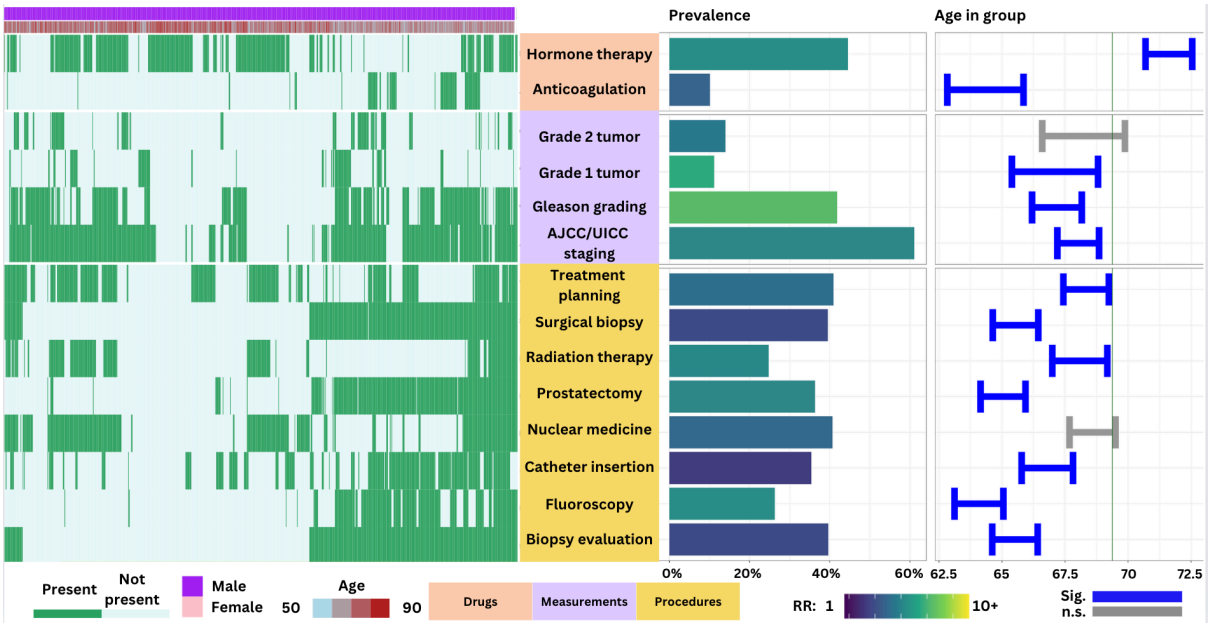

Figure 1. Prevalence of statistically significant drug-groups among the target cohort's patients

Significant findings within measurements and observations included various aspects of prostate cancer staging and diagnostic procedures, such as the Gleason grading system and AJCC/UICC categories.

After generating treatment trajectories using the Cohort2Trajectory[2] package, we were able to analyze the treatment trajectories. As illustrated in Figure 2, patients typically begin their treatment pathway with the measurement of grade 1 or grade 2 tumors, Gleason grading or treatment planning. Some patients with more severe conditions are admitted to intensive care immediately following diagnosis.

The average cycle time for a patient was 1.8 years, with most states exhibiting transition times of less than one month. The average time between initial diagnosis and the initiation of any treatment was approximately four months. Hormone therapy states were recurrent, with new drug eras emerging every four months on average.

It is important to note that we did not define an explicit trajectory end condition, resulting in some patients not completing the entire treatment cycle. This aspect can be regulated once a strict study protocol is established.

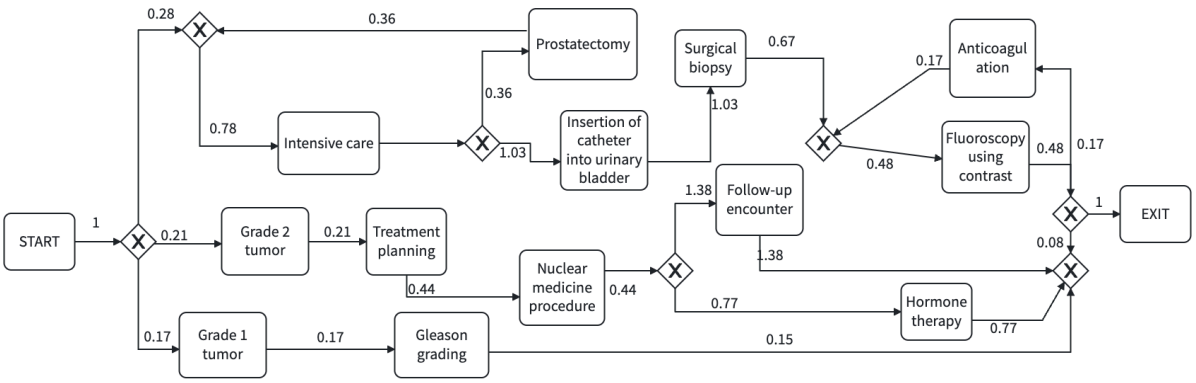

Figure 2. Average case treatment trajectory with states extracted from CohortContrast.

These insights serve as a foundation for further investigation into prostate cancer patient pathways, potentially offering avenues for personalized approaches to prevention and treatment.

Conclusion

We have developed an effective workflow for investigating cohort differences and establishing initial patient treatment trajectories using OHDSI CDM instances. This methodology provides a foundation for advanced research into treatment trajectories, facilitating analyses such as treatment modeling, cost-effectiveness evaluation, outcome prediction, adherence analysis and comparative effectiveness studies. Our approach simplifies these analyses by decreasing the need for a deep understanding of medical processes.

Validation against a study on newly diagnosed prostate cancer patients' trajectories demonstrated the potential efficacy of our methods. Our future objectives include scaling these studies across multiple databases using federated learning and validation techniques. This will allow us to encompass various phenotype cohorts and address a broad spectrum of research questions.

References

1. Haug M, Oja M, Pajusalu M, Mooses K, Reisberg S, Vilo J, Giménez AF, Falconer T, Danilović A, Maljkovic F, Dawoud D, Kolde R. Markov modeling for cost-effectiveness using federated health data network. Journal of the American Medical Informatics Association 2024. doi: 0.1093/jamia/ocae044
2. Haug M. Cohort2Trajectory, v1.1.7. GitHub. Accessed June 21, 2024. https://github.com/HealthInformaticsUT/Cohort2Trajectory
3. Haug M. TrajectoryMarkovAnalysis, v1.0.5. GitHub. Accessed June 21, 2024. https://github.com/HealthInformaticsUT/TrajectoryMarkovAnalysis
4. Pajusalu M. TrajectoryViz, v1.0. GitHub. Accessed June 21, 2024. https://github.com/HealthInformaticsUT/TrajectoryViz
5. Haug M. CohortContrast, v1.0. GitHub. Accessed June 21, 2024. https://github.com/HealthInformaticsUT/CohortContrast
6. Athena. Accessed June 21, 2024. http://athena.ohdsi.org/search-terms/terms.
7. Kannus A, Oja M, Tamm S, Kolde R. Prostate cancer treatment trajectories in Estonia based on electronic health records, 2012-2019. University of Tartu. 2022. Accessed June 21, 2024. https://dspace.ut.ee/items/556acc8b-25bd-44c8-874e-68113bac811e
8. Oja M, Tamm S, Mooses K, Pajusalu M, Talvik HA, Ott A, Laht M, Malk M, Lõo M, Holm J, Haug M, Šuvalov H, Särg D, Vilo J, Laur S, Kolde R, Reisberg S, Transforming Estonian health data to the Observational Medical Outcomes Partnership (OMOP) Common Data Model: lessons learned, JAMIA Open. 2023:4. doi: 10.1093/jamiaopen/ooad100