

## Background

RWD regimen analysis is highly needed in prostate cancer studies because treatment patterns are heterogeneous and not necessarily aligned with treatment guidelines. The OMOP episode table provides an integrated, longitudinal view of patient history at a glance, allowing users to trace key cancer-related events without querying multiple event tables.

ARTEMIS, an orchestrated regimen-mapping and post-processing program, uses the Temporal Smith–Waterman (TSW) algorithm and is provided by the OHDSI community. However, when we applied the default settings to prostate cancer regimens, the resulting output was not readily analyzable. Key challenges included the following: (1) regimens used for only one day were filtered out because of the default setting of  $\text{totAlign} > 1$ ; and (2) long administration intervals for androgen-deprivation therapy (ADT), as well as daily steroid dosing, were easily penalized in long treatment histories and therefore received lower scores.

## Methods

Regimen definitions were curated by cleaning the HemOnc sigs file, which served as the regimen reference file. All regimens containing “androgen-deprivation therapy” as a component were removed except for ADT monotherapy. A variant was added to each regimen, in which the alldays value for every component was set to 1. Missing important or new regimens were added as needed. For daily components combined with cyclic component(s), cycle-anchored alldays was used instead of daily alldays. For ADT monotherapy, we prepared various interval variants. For example, for a 3-month cycle, 90-day and 84-day (12-week) variants were added.

For regimens containing prednisone or prednisolone, those steroid components were removed. Topical drugs were excluded, whereas nasal ADT formulations were retained. The cython version of the ARTEMIS TSW algorithm was run separately for ADT and non-ADT regimens, followed by separate post-processing, and the outputs were appended at the end.

Post-TSW filtering was modified in several ways. When adjustedS was 1,  $\text{totAlign} = 1$  was permitted. For each candidate regimen era, we checked whether all components actually existed throughout the era (i.e., to the end of the era). If not,

the mapped regimen was filtered out for that era. Regimen eras for the same regimen were merged if the interval was 180 days or less for ADT and 63 days or less for all other regimens. In addition, when both monotherapy and combination therapy sharing the same component appeared during overlapping but non-identical eras, we prioritized combination therapy over monotherapy.

## **Results**

With the default configuration, approximately 15% of patients who received chemotherapy had no chemotherapy-containing regimens identified. After the proposed modifications, this proportion decreased to 0%.

Regimens matched for only 1 day, such as orchiectomy, were also retained rather than being dropped. In addition, the problem with prolonged combination regimen eras was resolved, resulting in a correct combination-therapy era followed by a monotherapy era.

By design, there are no ADT-containing regimens, and line of therapy (LoT) is not readily available.

## **Conclusions**

In principle, over-customization by cancer type should be avoided to ensure better maintainability and reproducibility, especially when developing an episode table for multiple cancer types. However, in this case, separating ADT from other regimens substantially resolved multiple problems at once and proved to be a necessary customization.

These findings suggest that cancer-type-specific adaptations may be justifiable when the root cause of the problem is specific to a particular cancer type. Future work will focus on extending this approach to LoT derivation.