Jani Järvinen

OMOP COMMON DATA MODEL AND ATLAS IN PROSTATE CANCER RESEARCH

Syventävien opintojen kirjallinen työ
Kevätlukukausi 2025

Jani Järvinen

OMOP COMMON DATA MODEL AND ATLAS IN PROSTATE CANCER RESEARCH

Kliininen laitos
Kevätlukukausi 2025
Vastuuhenkilö: Otto Ettala

TURUN YLIOPISTO
Lääketieteellinen tiedekunta

JÄRVINEN, JANI: OMOP Common Data Model and Atlas in prostate cancer research

Syventävien opintojen kirjallinen työ, 15 s.
Kirurgia
Huhtikuu 2025

Prostate cancer is the second most common cancer globally, with Finland reporting 5 514 new cases in 2022 and a 5-year prognosis of 94%. Treatment options range from passive to active, influenced by patient preferences and other factors. New medications typically emerge from Randomized Controlled Trials (RCT), but these populations often lack diversity compared to typical clinical settings. The Observational Medical Outcomes Partnership (OMOP) aimed to standardize medical data through a Common Data Model (CDM). The Observational Health Data Sciences and Informatics (OHDSI) collaboration was established to support observational, evidence-based research. This project assessed the incidence and treatment of prostate cancer at TYKS and evaluated the usability of the OHDSI tool, Atlas, for data analysis.

Using patient data in OMOP CDM format, the study began in summer 2021 with TAYS and HUS, forming a cohort diagnosed between January 1, 2010, and December 31, 2022. Outcomes included annual new diagnoses and Atlas's usability compared to source database figures.

Atlas identified 7 376 new cases, while R script yielded 7 291. Patient age averaged 71, with common, most often metabolic comorbidities. Among diagnosed patients, 67% received treatment. There’s been a decline in Androgen Deprivation Therapy as monotherapy and a rise in combination therapies. Rates of radiotherapy and surgery have remained constant.

While Atlas is a promising tool for Real World Data research, further investigation of discrepancies between OMOP-CDM and source data is necessary to optimize its use in future studies.

Grammarly AI was utilized in condensing the abstract into its final form.

OMOP Common Data Model and Atlas in Prostate Cancer research

Abstract

Introduction
Prostate cancer is the second most common cancer worldwide, particularly prevalent in Western societies, with Finland reporting 5 514 new diagnoses in 2022 and a 5-year prognosis of 94%. Treatment options vary from passive to active, depending on patient preferences and factors. While new medications are often introduced through Randomized Controlled Trials (RCT), these patient populations can be too homogeneous compared to typical clinical settings.

The Observational Medical Outcomes Partnership (OMOP) aimed to standardize medical data through a Common Data Model (CDM) over a five-year project in the US. To further this goal, the volunteer based Observational Health Data Sciences and Informatics (OHDSI) collaboration was established to facilitate observational, evidence-based research.

This project aimed to assess the incidence and treatment of prostate cancer in TYKS and evaluate the usability of the OHDSI tool Atlas for data analysis.

Materials and Methods
The study utilized patient data in OMOP CDM format, by project starting in the summer of 2021 with TAYS and HUS to map local terminology. A cohort of patients diagnosed with prostate cancer between 1.1.2010-31.12.2022, was formed, incorporating variables such as Charlson Comorbidity Index (CCI) and treatments. The primary outcome was the annual number of new diagnoses, while secondary outcomes examined Atlas's usability compared to OMOP-CDM and source database figures, presented as absolute counts or percentages.

Results
Atlas identified 7 376 new diagnoses, while R script yielded 7 291 and original data showed 7 384. The average patient age was 71, with comorbidities like diabetes, hypertension, and myocardial infarctions being common. Among the diagnosed patients, 67% received treatment. Notably, Androgen Deprivation Therapy (ADT) as monotherapy has declined, whereas combination therapies are on the rise, and surgery and radiotherapy rates have remained consistent.

Discussion
Patients from Vaasa, Satakunta and Åland regions getting treated in Southwest Finland's wellbeing cervices county may introduce bias in patient counts and age demographics. The distribution of treatments and diagnoses generally aligns with other Nordic countries. However, Atlas demonstrated some reliability issues, as results varied from R script findings. Limited guidance and tutorials for Atlas created challenges, particularly with CCI analysis, where technical issues compromised reliability by requiring alterations to some diagnoses.

Conclusions
Atlas proves to be an effective tool for Real World Data research but necessitates further investigation into discrepancies between OMOP-CDM and source data to fully exploit its potential in future studies.

Grammarly AI was utilized in condensing the abstract into its final form.

Introduction

Prostate cancer

Prostate cancer is the second most common cancer in the world and its occurrence is especially high in industrialized western societies. In Finland there were 5 514 new cases of prostate cancer in 2022 totalling 28% of all new cancer diagnoses in Finnish male population. Despite having a high occurrence rate, prostate cancer has a very good prognosis; 94% of patients are alive 5-year after the diagnosis. Prostate cancer is most common in the age group over 70-years of age and incidence in this group was 3 302 new cases (823.7 / 100 000 people). It possessed the second highest mortality rate in cancer deaths totalling 920 deaths in 2022.[1,2]
Treatments vary greatly depending on the stage of the disease, patient derived factors, and patient preferences. While in asymptomatic patients with a life expectancy less than 10 years and a local or locally advanced prostate cancer watchful waiting is a viable option, in patients with life expectancy exceeding 10 years an active surveillance, a radical prostatectomy or radiation therapy is chosen[2]. Metastatic cancer is usually treated with androgen deprivation therapy (ADT) monotherapy or in combination of androgen receptor targeted agent (ARTA), abiraterone and/ or chemotherapy (docetaxel of cabazitaxel). Surveillance of the utilization of different treatment modalities is the basis for controlling the quality of care but also to know how new treatment modalities and especially new medical interventions are utilized in prostate cancer care in real clinical practice. While new treatments are introduced through randomised controlled trials (RCT), a commonly acknowledged issue is that RCTs utilize very homogenous cohort of patients not representative of patients in normal clinical practice. Therefore, studies utilizing retrospective real world data (RWD)-studies complementing data from RCTs have gained

popularity in recent years. RWD allows amassing larger data pools both nationally and internationally and thus provides valuable insight into more variable patient material and in prostate cancer, to these novel treatments.[3] However, the challenges with RWD-studies lie in data interoperability and privacy-preservation.

OMOP CDM & OHDSI and the standard data utilisation

The Observational Medical Outcomes Partnership (OMOP) was a 5-year US based partnership that aimed at studying and learning suitable methods for observing and handling medical information. During the project Observational Medical Outcomes Partnership Common Data Model (OMOP CDM), a model to represent dispersed health data from different sources in a uniform, standardised manner was developed. [4–6] From this OMOP collaboration, Observational Health Data Sciences and Informatics (OHDSI), a volunteer based international collaborative comprised of academics, scientists and other health industry professionals, was created. OHDSI's purpose is to continue with these standardised vocabularies to further support international research with observational, evidence-based data. The data in electronic health records (EHRs) are transformed organisationally into OMOP-CDM after which the data can be utilised cross-organisationally and cross-border as the data is cohesive and comparable to other sources alike. Since the data are stored, maintained and analysed locally and only analyses results are shared, privacy is preserved. [4] In addition to common data model, the OHDSI community provides a variety of open-source tools to access and analyse the data such as scripts for SQL, R and Python as well Atlas, a web-based analytical tool with intuitive graphic interface[7–9].

Objectives of the project

The aim of this project was to assess the yearly rate of new prostate cancer diagnoses and how the treatment patterns have changed over the years. In addition, the aim was to evaluate the feasibility and reliability of Atlas tool to do such analyses.

Materials and Methods

OMOP common data model

All of the observed urologic health care information in this study was processed in the form of The Observational Medical Outcomes Partnership Common Data Model (OMOP CDM).

Data mapping

Starting in 2021, a finOMOP vocabulary was established in collaboration with Turku (TYKS), Helsinki (HUS) and Tampere (Tays) university hospitals. In Turku the project was carried out by Auria Clinical Informatics representing the local healthcare district. Most common terminology was mapped through automated process and the remaining, less frequently used terminology was mapped manually with the OHDSI provided Usagi tool; in

Turku and Tampere by two medical students in each location, while in Helsinki the work was carried out by the hospital haematology department.

Study cohort

A cohort of patients with new prostate cancer diagnosis between 1.1.2010 and 31.12.2022 was studied. Patients presenting any occurrence of a prostate cancer diagnosis prior to 1.1.2010 were excluded. The prostate cancer diagnosis included a list of standard concepts with their descendants: non-metastatic prostate cancer; malignant tumor of prostate; carcinoma in situ of prostate; adenocarcinoma in situ, NOS, of prostate gland; adenocarcinoma in situ in villous adenoma of prostate gland and adenocarcinoma in situ in tubulovillous adenoma of prostate gland. As an example, malignant tumor of prostate has descendants such as primary malignant neoplasm of prostate or hormone sensitive prostate cancer. There were 195 descending concepts to these 6 original concepts selected into forming this cohort.

Variables

For every subject included, age, comorbidities included in the Charlson Comorbidity Index (CCI) and treatment modality were collected (Table 1)[10]. The comorbidities must have occurred before or at the time of the first occurrence of prostate cancer diagnosis. The treatment modalities considered were radical prostatectomy, radiation therapy with or without ADT, ADT monotherapy, antiandrogens, chemotherapy, and new androgen modulating agents (Table 2). To be selected for the analysis, the modality must have occurred 180 days prior and 365 days after the diagnosis. By this it was ensured that also those modalities that occurred before the first occurrence of prostate cancer diagnosis were included, as at times prostate cancer diagnoses may have been registered to EHRs only after the treatment has already been started. Every concept included its descendants.

Outcomes

The primary outcome is the yearly numbers of new prostate cancer diagnoses assessed using the Atlas tool. Secondary outcomes are yearly numbers of new prostate cancer diagnoses assessed using the OMOP-CDM database and the source database with R, yearly numbers of different treatment modalities and a qualitative estimation of the feasibility of Atlas tool.

Statistical analysis

Figures under observation were presented as either absolute counts or numbers (%). Age distribution of the studied population was presented in blocks of 5 years.

Ethical considerations

Research on urology patient data was conducted under an institutional research approval, T286_2016.

CCI Comorbidities	Used standard concepts
Myocardial Infarction	Subsequent myocardial infarction of anterior wall Subsequent myocardial infarction Old myocardial infarction Myocardial Infarction Acute ST segment elevation myocardial infarction of inferior wall Acute ST segment elevation myocardial infarction due to right coronary artery occlusion Acute ST segment elevation myocardial infarction due to occlusion of anterior descending branch of left coronary artery Acute ST segment elevation myocardial infarction Acute non-ST segment elevation myocardial infarction Acute myocardial infarction of inferior wall Acute myocardial infarction of anterior wall Acute myocardial infarction due to right coronary artery occlusion Acute myocardial infarction due to left coronary artery occlusion Acute myocardial infarction Acute anterior ST segment elevation myocardial infarction
CHF - Congestive Heart Failure	Congestive heart failure Hypertensive heart disease with congestive heart failure
Peripheral Vascular Disease	Peripheral vascular disease Peripheral vascular disease associated with another disorder
CVA or TIA - Cerebrovascular Accident or Transient Ischemic Attack	Cerebrovascular disease
Dementia	Vascular dementia Subcortical vascular dementia Primary degenerative dementia of the Alzheimer type, senile onset Primary degenerative dementia of the Alzheimer type, presenile onset Multi-infarct dementia, uncomplicated Multi-infarct dementia due to atherosclerosis Multi-infarct dementia Mixed dementia Mixed cortical and subcortical vascular dementia Dementia due to Parkinson's disease Dementia associated with another disease Dementia associated with alcoholism Dementia Delirium co-occurrent with dementia
Chronic Pulmonary Disease	Chronic pneumonia Chronic obstructive lung disease Chronic bronchitis Asthma
Connective Tissue Disease	Vasculitis Systemic lupus erythematosus

	Sjögren's syndrome Sarcoidosis Rheumatoid arthritis Polymyositis Polymyalgia rheumatica
Peptic Ulcer Disease	Perforated peptic ulcer closure Peptic ulcer with perforation Peptic ulcer with hemorrhage Peptic ulcer Chronic peptic ulcer with perforation Chronic peptic ulcer with hemorrhage Chronic peptic ulcer Acute peptic ulcer with perforation Acute peptic ulcer with hemorrhage Acute peptic ulcer
Liver Disease	Transplanted liver present Ruptured varicose veins Cirrhosis of liver Chronic type B viral hepatitis Chronic hepatitis C
Diabetes Mellitus	Diabetes mellitus
Hemiplegia	Paraplegia Hemiplegia
Moderate to severe CKD - Chronic Kidney Disease	Chronic kidney disease
AIDS	AIDS

Table 1. CCI comorbidities. Every concept includes its descendants. Abbreviations: CCI, Charlson Comorbidity Index; AIDS, acquired immunodeficiency syndrome.

Treatments	Used standard concepts
RALP	Robot assisted surgery
Excluded concepts	Retropubic prostatectomy Radical retropubic prostatectomy Prostatectomy Laparoscopic excision of pelvic lymph node Ilioinguinofemoral lymphadenectomy Excision of inguinal lymph nodes Excision of iliac lymph nodes
Radiotherapy	Radiation oncology AND/OR radiotherapy
Excluded concepts	Stereotactic radiotherapy Radiation Therapy (Procedure) Palliative course of radiotherapy Combined chemotherapy and radiation therapy
ADT (inc. descendants)	triptorelin leuprolide

	Hormone therapy goserelin 10.8 MG Drug Implant goserelin degarelix
Antiandrogens (inc.descendants)	flutamide bicalutamide
ARTA (inc. descendants)	enzalutamide darolutamide apalutamide abiraterone
Chemotherapy	docetaxel cabazitaxel

Table 2. Given treatments. Abbreviations: RALP, robot assisted laparoscopic radical prostatectomy; ADT, androgen deprivation therapy; ARTA, androgen receptor-targeted agents.

Results

From the prostate cancer population totalling 7 376 people, it can be seen, that the yearly incidence of new carcinoma diagnoses revolves around 500 and 650 as can be seen in Figure 1. A slight variation of the number of new diagnoses is constant throughout the perceived years and no clear pattern of increase or decrease in new cases emerge from the data. New cases of cancer retrieved directly from OMOP-CDM database and the source database with R script were 7 291 and 7 384 cases, respectively.

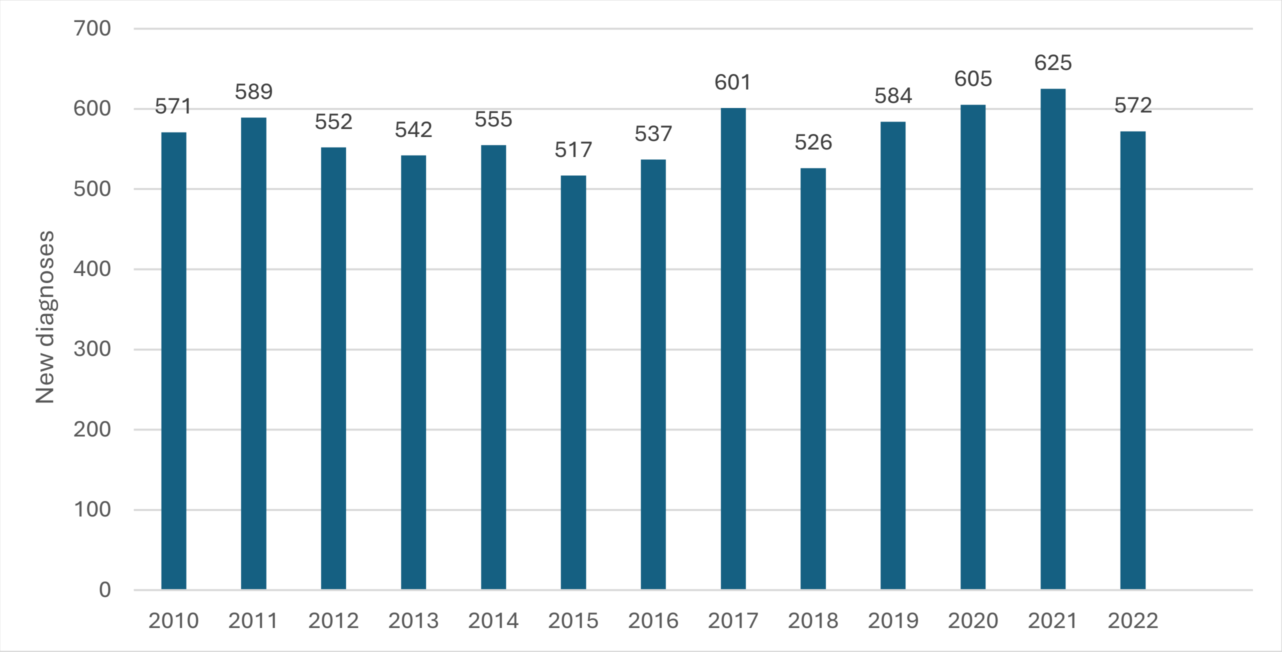

Figure 1. Prostate cancer incidence in TYKS between 2010 and 2022.

The highest rate of patients falls between 70-74 years of age as depicted in Figure 2. The average age of patients with a new prostate carcinoma diagnosis was 71. Youngest patient in the population was 20 years old whilst oldest being 102. Median age for the population was 71, standard deviation 9.0.

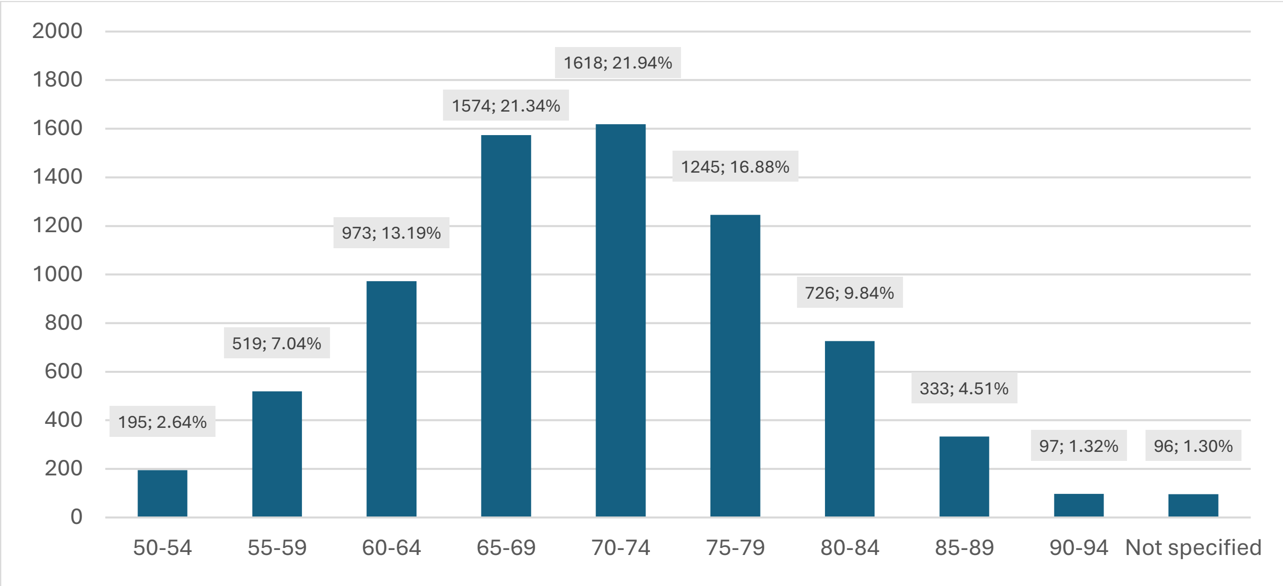

Figure 2. Age distribution in inspected population. Atlas described the largest age groups, excluding youngest and oldest outliers in the data. These patients are described in the chart as "not specified".

Average age being quite high in the studied population, various comorbidities are common in this group of patients with emphasis being on vascular and pulmonary diseases as shown in Table 3.

Comorbidity	n (%)
Diabetes Mellitus	609 (8.26)
Chronic Pulmonary Disease	425 (5.76)
Myocardial Infarction	413 (5.60)
Connective Tissue Disease	339 (4.60)
Cerebrovascular Disease	282 (3.82)
Peripheral Vascular Disease	191 (2.59)
Peptic Ulcer Disease	149 (2.02)
Moderate to severe renal disease	133 (1.80)
Dementia	112 (1.52)
Congestive Heart Failure	62 (0.84)
Liver Disease	29 (0.39)
Hemiplegia / paraplegia	21 (0.28)
AIDS	0 (0.00)

Table 3. Charlson Comorbidity Index-comorbidity occurrence. Abbreviations: AIDS, acquired immunodeficiency syndrome.

Of all patients with new prostate cancer diagnosis, 4 942 (67%) received treatment for their cancer and in the remaining 2 434 (33%) no treatment occurred during the given time frame, and they were categorised as patients with no immediate treatment (NIT).

Most of the treatments have remained on a solid level through the observation period but there are also few changes that have transpired over time. Figure 3 depicts treatments for localized or locally advanced prostate cancer and Figure 4 for metastasised cancer. Both figures portray a total number of each treatment carried out over the years counting them as individual occurrence thus including given radiotherapy both in radiotherapy and ADT+radiotherapy curves.
Androgen deprivation therapy (ADT) as a monotherapy has seen a reduction in usage but at the same time a clear rise in numbers has appeared in ADT dual therapies shown in chemotherapy and ARTA (always combined with ADT) as well as radiotherapy. Of these three dual therapies, chemotherapy, seems to have stabilized to a steady level of use since 2014. Conversely to ADT treatments, antiandrogen treatments appear to have been on a declining trend for the last five years. With non-metastasised cancer, both robot assisted laparoscopic radical prostatectomy (RALP), and radiotherapy appear to have reached a constant level (RALP 140-170; radiotherapy 150-180 treated patients yearly) around 2013 and remained steady since.

Atlas as an interface for conducting the study was direct and visually consistent to use. Side bar helps to navigate between different sections of the tool and thus makes using different functions efficient. Forming concept sets for this study was straightforward as the previously defined concepts were easy to search through a search bar-function and descending concepts could be tagged along when necessary by checking the choice box. Defining cohorts was more complex than drawing up concept sets as the function has multiple variables in form of configurable time frames and dropdown menus with multiple, not always intuitive options. In Cohort definition, results generation occasionally malfunctioned providing incorrect figures thus making this feature a bit unreliable and requiring a good insight on the expected results from the person making the analysis. Characterizations-menu generates clear visual analyses of the patient data and was utilised in this study to provide information about the age distribution of the formed cohort. There are plenty of possibilities to generate these visual reports about different data, but the provided Charlson Index-report was not able to be generated of this patient material.

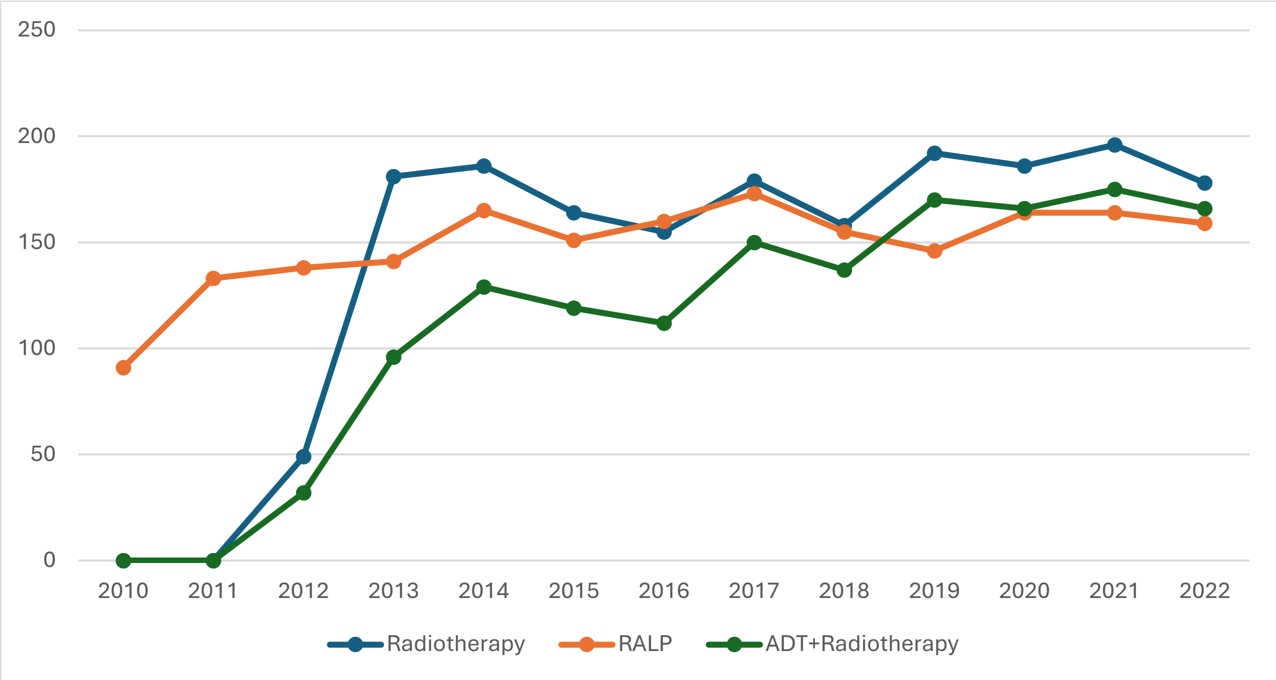

Figure 3. Trend pattern of treatments for non-metastasised prostate cancer. Abbreviations: ADT, androgen deprivation therapy; RALP, robot assisted laparoscopic radical prostatectomy

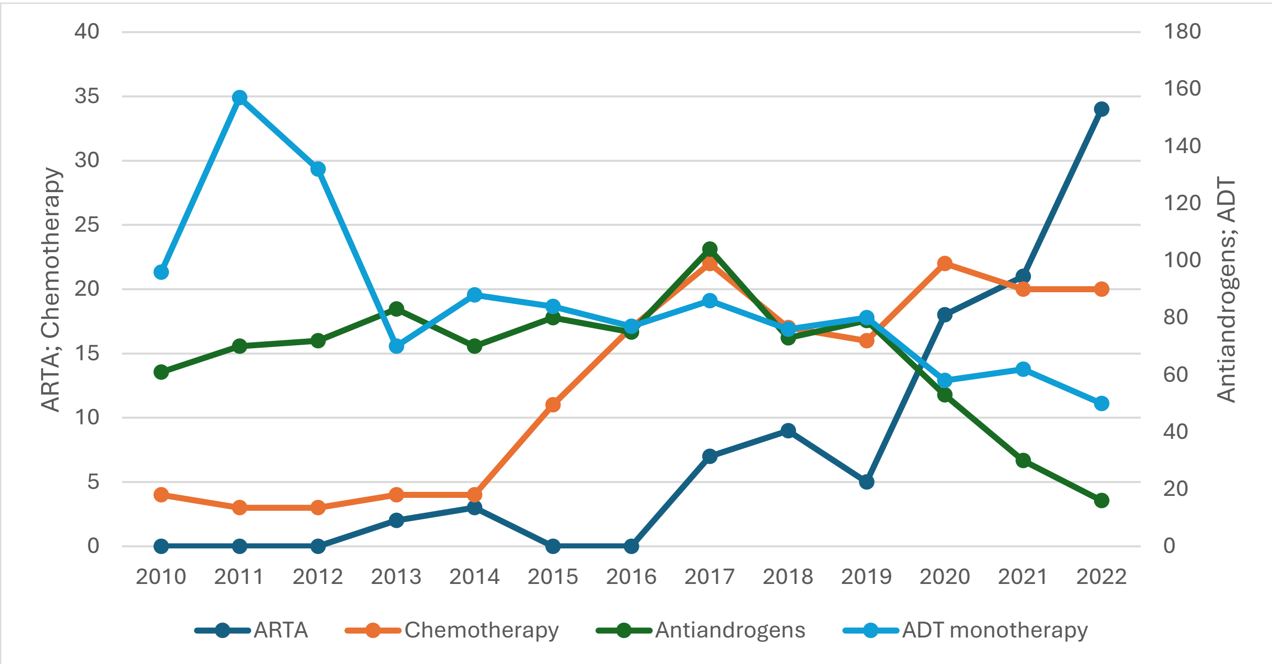

Figure 4. Trend pattern of treatments for metastasised prostate cancer. Abbreviations: ADT, androgen deprivation therapy; ARTA, androgen receptor-targeted agents.

Discussion

As a key result of this study, there were around 500-650 new diagnoses of prostate cancer in TYKS patient registry annually during the studied time period. Most common comorbidities among studied patients represented those part of metabolic syndrome[11]. There has been a slight shift in treatments patterns over the years emphasizing ARTA and chemotherapy becoming more prevalent during recent years in metastatic patients whilst surgery and radiotherapy have remained on a constant basis and both ADT monotherapy and antiandrogens have lost some foothold among other treatments.

From the Finnish cancer registry, it can be stated that during the last 5-year observation period (2018-2022) of provided statistics there were 4489 new prostate cancer diagnoses in the entire Western Finland collaborative area and 2335 in Southwest Finland wellbeing services county (Varha) respectively(1). In this study the count was 2912 over this time period. This variance could be explained by the factor, that some cancer patients have their treatment begun in TYKS (eg. patients from Åland, Satakunta and Vaasa regions being operated in Turku) despite belonging to a different wellbeing services county.

The fact, that Atlas cohort produced 7 376 new cancer diagnoses within the inspected timeframe, R script yielded 7 291 cases from OMOP-CDM database is most likely due to slightly different concepts and their descendants selected in Atlas and in the R script. However, this does not explain the 7 384 new cases found in the hospital source database, which might be explained through mapping technicalities leading to the script identifying slightly different pieces of data as part of requested information. In the scope of time of this project, a deeper understanding was not able to be attained.

Based on TYKS patient material, it appears that of all yearly prostate cancer diagnoses, approximately 25-30% received radical prostatectomy, 30-35% radiotherapy and over the last year of the study in 2022, roughly5,9% of the patients received ARTA. The national prostate cancer registries and annual reports were studied to compare the local treatment patterns to Denmark, Norway and Sweden. Radical prostatectomy was conducted approximately to a similar portion of newly diagnosed patients in every inspected country and almost all the surgery was conducted robot aided. These proportions were 26,5% in Denmark (in 2022), 30,2% in Norway (2023) and 25,7% in Sweden (2023). Radiotherapy rates showed more inconsistency and appeared lower in Denmark (615 primary treatments, 13,8%) and Norway (847 / 16,1%) while Sweden presented slightly higher values than in prostatectomy (2 906 / 27,4%). The way the data is presented differs between countries and thus makes the comparison slightly more challenging. With medical treatments, different levels of analysis were offered between countries, but the main message appeared to compare well with findings of local patient data. ARTA has been the “go-to” medication that is recommended to initiate alongside ADT in metastasised cancers in recent years, thus comparing well to Figure 4 portraying a clear rise in use of ARTA. [12–15]

There is a slight overestimation of yearly diagnosis and surgery statistics in the data as few patients from other wellbeing services counties visited TYKS for treatment. Home municipality information of patients is not recoverable through Atlas and thus could not be taken into account when assembling this study. In future, it is vital information to consider when conducting similar research in order to retrieve more exact results. Treating additional patients from other wellbeing services counties most likely also explain the difference between patient age distribution of this study compared to national data. Patients under 60-years of age constitute around 10% of the studied patient material while generally this group represents only about 5% of the diagnoses[16].

Active surveillance and watchful waiting (NIT patients) were inspected as a separate line of treatment (one entity counting both groups), but no further information could be attained of these patients. This was because there is a local practice not to classify these as “treatments” as nothing is being done at the time of diagnosis or during follow-up. Therefore, there was no patient data to match these standard concepts when using Atlas. Danish, Norwegian and Swedish reports include variable information about NIT patients. In Denmark, 721 of 4 443 (16,2%) new patients were assigned to watchful waiting while 659 of (14,8%) were in active surveillance. Norwegian report did not include watchful waiting and only offers very limited information without exact current numbers about active surveillance. In general, the portion of actively surveilled patients seem to have risen steadily as a choice of treatment over 2010’s having been 1 128 patients in 2020. Swedish reports also exclude watchful waiting but have a total of 2 709 (26%) patients under active surveillance. Comparing the local data having 2 434 (33%) NIT patients, it appears to be quite well in line with Danish prostate cancer population.[12–15]

Considering Atlas’ usability, performing this study proved to be challenging what comes to navigating through patient data with Atlas. OHDSI being a volunteer-based community, there appears to be plenty of discussion about its functionalities between individuals over several forums about Atlas, but readily available written and video tutorials about the software’s usage appear to be quite modest and cover only the basics through set examples[8,17,18]. The Cohort Definition-function in Atlas has a way to include wanted concepts, but no simple method to exclude a concept was found. This produced a challenge in creating the cohorts for treatments as every treatment had to be analysed separately on a yearly basis instead of being able to treat them as a group. Inclusion criteria-function also presented with some issues and somewhat lacked easy, intuitive usage. A group of patients having received some single treatment could be picked but attempting to combine several different inclusion criteria into a single analysis showed no success. Only way Atlas would count these patients in was that if they met all the set criteria instead of one of treatments. Thus, a total amount of patients treated with any treatment could be defined but how many of those patients received which treatment was left unclear. A clear “OR-function” could not be found in selecting different treatments. When creating a cohort about a single treatment, it acknowledged all the patients having received the selected treatment for the first time in their history, but it would leave out a prior treatment if a patient had received any, thus skewing the total amount of first-time treatments given from the actual value (4 942 vs. sum of all treatments during the observation period).

There was also an issue with defining the CCI-analysis. In Atlas there exists a pre-established function for calculating the CCI-index, but despite selecting this function, such report could not be formed. The issue is most likely connected to the local OMOP-CDM since using the CCI R script also gave zero result. These challenges resulted in manually constructing a CCI-report like analysis of presented comorbidities. It portrayed some challenges as there was no clear way of forming a coherent concept set of both kidney diseases and malignancies listed in the CCI Index thus resulting in modifying kidney concepts and leaving malignant comorbidities out of this manual analysis.

Main challenges of the project revolved around the usability of Atlas as the technical implementation of defining cohorts proved to challenging with the knowledge on hand. As discussed, patient data could not be interpreted in Atlas on a level of filtering a home municipality, thus leading to excess amounts of yearly diagnosed and operated patients. This resulted in somewhat unreliable comparison to other Scandinavian data on treatments. Some reliability issues rose up during gathering of cohort data. Patient quantities varied unexpectedly giving faulty information when running the analysis. One incidence of this rose while inspecting radiotherapy yearly occurrence. An expected shift in quantity would have been around a 150 patient rise in total treated patients but instead Atlas deducted 1 100 from the total amount. Running the few second long analysis again resolved this issue correcting the numbers but emphasised that it is important to pay attention to the numbers and have a preliminary understanding of how the numbers should evolve or appear when running the analysis. Total number of treated patients also varied slightly (4 942 vs. 4 931) depending on the method of running the Cohort Definition. Despite these small underlying inconsistencies, the established OMOP-database presents truly a population-based cohort of men including every prostate cancer patient treated in the wellbeing society of Southwest Finland. Technical difficulties and nuances could be addressed with consistent and broad use of the database.

Conclusions

With minor issues, the established prostate cancer OMOP-database seems to be representative of prostate cancer patients in wellbeing service county of Southwest Finland. The level of immersion of doing research with Atlas appeared not satisfactory enough to meet all the goals set for sought results. This was because of limited resources within the project and mostly self-directed education over the management of Atlas software. In a broader context, Atlas appears to be a multifaceted tool suitable for various needs of RWD-based academic research and by having a more precise strategy over the induction of its use within the frames of this study, the produced findings could have been even closer the desired outcomes of matching R script results. It is of high priority to investigate the differences in OMOP-CDM and source data before full utilization of the OMOP-CDM database in future RWD studies.

References

1. The Finnish Cancer Registry [Internet]. [cited 2025 Mar 5]. The Finnish Cancer Registry. Available from: https://cancerregistry.fi/

2. Eturauhassyöpä. Current Care Guidelines. Working group set up by the Finnish Medical Society Duodecim and the FUA Finnish Urological Society. Helsinki: The Finnish Medical Society Duodecim, 2017. [Internet]. [cited 2025 Mar 1]. Available from: https://www.kaypahoito.fi/hoi11060#s7

3. Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. Trials. 2015 Nov 3;16:495.

4. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. Stud Health Technol Inform. 2015;216:574–8.

5. Ahmadi N, Peng Y, Wolfien M, Zoch M, Sedlmayr M. OMOP CDM Can Facilitate Data-Driven Studies for Cancer Prediction: A Systematic Review. Int J Mol Sci. 2022 Oct 5;23(19):11834.

6. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. J Am Med Inform Assoc JAMIA. 2012;19(1):54–60.

7. Observational Health Data Sciences and Informatics. Chapter 1 The OHDSI Community | The Book of OHDSI. November 1st, 2021. [Internet]. [cited 2025 Apr 7]. Available from: https://ohdsi.github.io/TheBookOfOhdsi/

8. Observational Health Data Sciences and Informatics. Chapter 8 OHDSI Analytics Tools | The Book of OHDSI. November 1st, 2021. [Internet]. [cited 2025 Mar 23]. Available from: https://ohdsi.github.io/TheBookOfOhdsi/

9. Observational Health Data Sciences and Informatics. Chapter 9 SQL and R | The Book of OHDSI. November 1st, 2021. [Internet]. [cited 2025 Apr 7]. Available from: https://ohdsi.github.io/TheBookOfOhdsi/

10. MDCalc [Internet]. [cited 2025 Feb 13]. Charlson Comorbidity Index (CCI). Available from: https://www.mdcalc.com/calc/3917/charlson-comorbidity-index-cci

11. Li S, Wen CP, Tu H, Wang S, Li X, Xu A, et al. Metabolic syndrome including both elevated blood pressure and elevated fasting plasma glucose is associated with higher mortality risk: a prospective study. Diabetol Metab Syndr. 2025 Feb 25;17(1):72.

12. DaProCa: Dansk Prostata Cancer Database. Årsrapport 2022. Aarhus N: Regionernes Kliniske Kvalitetsudviklingsprogram; 2023 June 12th. [Internet]. [cited 2025 Apr 4]. Available from:

https://ducg.dk/fileadmin/ingen_mappe_valgt/Prostata_rapport_2022_version_til_offentliggoerelse_12062023_final.pdf

13. Nasjonalt kvalitetsregister for prostatakreft. Årsrapport 2023 Resultater og forbedringstiltak fra Nasjonalt kvalitetsregister for prostatakreft. Oslo. [Internet]. [cited 2025 Apr 4]. Available from: https://www.kreftregisteret.no/globalassets/publikasjoner-og-rapporter/arsrapporter/publisert-2024/arsrapport-2023-nasjonalt-kvalitetsregister-for-prostatakreft.pdf

14. Nationella prostatacancerregistret (NPCR). Prostatacancer Nationell kvalitetsrapport för 2023. [Internet]. [cited 2025 Apr 4]. Available from: https://npcr.se/wp-content/uploads/2024/09/20240924_npcr_nationell_rapport_2023.pdf

15. National Prostate Cancer Register (NPCR) of Sweden. RATTEN - Interactive On Line Report from NPCR. [Internet]. [cited 2025 Apr 1]. Available from: https://statistik.incanet.se/npcr/

16. Tammela T. Eturauhassyöpä. In: Evidence-Based Medicine Guidelines [online]. Helsinki: Duodecim Medical Publications Ltd, 2025 [Updated 22.06.2022]. Available from Internet (password required). [Internet]. [cited 2025 Mar 23]. Available from: https://www.terveysportti.fi/apps/dna/ltk/article/ykt00294/search/eturauhassy%C3%B6p%C3%A4

17. Observational Health Data Sciences and Informatics. ATLAS Tutorials. 2025. [Internet]. [cited 2025 Mar 23]. Available from: https://ohdsi.org/resources/tutorials/

18. Schüttler C, Prokosch HU, Sedlmayr M, Sedlmayr B. Evaluation of Three Feasibility Tools for Identifying Patient Data and Biospecimen Availability: Comparative Usability Study. JMIR Med Inform. 2021 Jul 21;9(7):e25531.