1. Title Page

Title	Comparative effectiveness and safety in patients with prostate cancer who received radical prostatectomy vs. radiotherapy: target trial emulation
Research question & Objectives	This study aims to compare the effectiveness and safety outcomes in patients with early stage prostate cancer (PCa) who received radical prostatectomy or radiotherapy using the target trial emulation framework across a network of databases in the male population of Europe. In detail, the main objectives of the study are: 1. To estimate the comparative effectiveness and safety of surgical versus radiological treatment of PCa in the target trial population 2. To estimate the comparative effectiveness and safety of surgical versus radiological treatment of PCa in the real world data population
Protocol version	1
Last update date	<Text>

Contributors	**Primary investigator contact information:** Eng Hooi (Cheryl) Tan, University of Oxford **cheryl.tan@ndorms.ox.ac.uk** **Contributor names:** Daniel Prieto-Alhambra, University of Oxford Danielle Newby, University of Oxford Cecelia Campanile, University of Oxford Marti Catala Sabate, University of Oxford Daan Nieboer, Erasmus Medical Center Nora Tabea Sibert, DKG Christoph Kowalski, DKG
Sponsor	**Organization:** OPTIMA is funded through the Innovative Medicines Initiative Phase 2 (IMI2) Joint Undertaking and is listed under grant agreement No. 101034347. IMI2 receives support from the European Union's Horizon 2020 research and innovation programme and the European Federation of Pharmaceutical Industries and Associations (EFPIA).
Conflict of interest	<Text>

Table of contents

2. Abstract

This is a study protocol for an observational health data analysis as part of OPTIMA, a novel project of the Innovative Medicine Initiative (IMI) with the mission to generate evidence through real world data (RWD) to update cancer treatment guideline recommendations. This study aims to estimate treatment effects of early stage prostate cancer (PCa) patients across a network of databases in the male populations in Europe and subgroups of patients identified by individual disease characteristics, demographics, and comorbidities. The study will rely on large observational data, namely population-based registries, electronic health records, quality assurance and insurance claims data. The study will be an observational cohort study based on routinely collected health care data which has been mapped to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM).

3. Amendments and updates

Version date	Version number	Section of protocol	Amendment or update	Reason

4. Milestones

Table 1 Milestones

Milestone	Date
Protocol and ethics approval	01 Oct 2024
Completion of data analysis	31 May 2026
Final study report	31 Aug 2026

5. Rationale and background

What is known about the condition?

Prostate cancer (PCa) is the most commonly diagnosed cancer and the second leading cause of cancer mortality in men worldwide (1). The age-standardised incidence rates were estimated to range from 46.4 to 83.4 per 100,000 men in Europe. Despite the high incidence, the age-standardised mortality rates were lower, ranging from 7.8 to 13.7 per 100,000 men across Europe (1), possibly due to advancement in treatment and earlier detection through increased prostate-specific antigen (PSA) screening (2). Risk factors for PCa include age, family history of PCa, African descent, and germline mutation of BRCA1/2, HOXB13 (3). Evidence for risk moderation by lifestyle factors is sparse and inconclusive. Body mass index (BMI) has been associated with negligible increase of overall PCa risk and a small increase of advanced PCa risk (4). However, there was strong evidence of a non-linear relationship between BMI and surrogate markers of PCa. Compared to normal weight men, PSA levels in overweight men were on average 3.4% lower, whereas PSA levels in obese men were 12.9% lower (4). The association between BMI and PCa may be biased if missed diagnoses were not accounted for.

What is known about the exposures of interest?

The European Association of Urology (EAU) clinical guidelines defined risk stratification groups based on clinical staging, PSA level, and Gleason score for biochemical recurrence (BCR) of localised and locally advanced PCa (3). For localised PCa, patients may be offered active surveillance, radical prostatectomy (RP), or radical or local radiotherapy (RT) (5). The Prostate Testing for Cancer and Treatment (ProtecT) trial which investigated these three interventions found that overall mortality in the three arms were low, RP and RT reduced disease progression compared to active surveillance (6). There is a trade-off between encountering higher risk of disease progression with no treatment versus adverse effects of radical treatment on urinary, bowel, and sexual function, and longer follow-up (5 to 10 years) is essential to evaluate this trade-off (6). In locally advanced PCa, RP or RT should be considered.

Gaps in knowledge:

The EAU Prostate Cancer Guideline panel and other prostate cancer Key Opinion Leaders were consulted to propose the most critical questions in the field of PCa to be answered using big data and these questions were prioritised using a Delphi consensus building process in the PIONEER consortium (7). To avoid duplication in efforts, the relevant PCa questions not addressed by PIONEER will be addressed by the OPTIMA consortium. A recent systematic review found that surgery had negative effects on urinary and sexual function when compared with active surveillance and EBRT whereas EBRT had worse effects on bowel function when compared with active surveillance and surgery (8). The review mostly featured studies from the US, with a lack of

representation of patients in Europe (2 studies from Germany with limited sample size <100 per treatment).

What is the expected contribution of this study?

We aim to answer the prioritised research questions as identified by healthcare professionals and patients, using real world data in an international federated network across multiple data sources. We will systematically generate evidence to compare the effectiveness and safety of prostatectomy and radiation. Our study aims to perform target trial emulation (9) using routinely collected data, using the ProtecT study (10) as a benchmark. We also aim to expand the study to patients who received RP or RT outside the eligibility criteria of the trial, as real world outcomes of older patients, and those with more comorbidities may differ from the trial population.

6. Research question and objectives

Table 2 Primary and secondary research questions and objective

Objective:	1. To estimate the effectiveness and safety of surgical versus radiological treatment of PCa in real world data (RWD) population 2. To estimate the effectiveness and safety of surgical versus radiological treatment of PCa in emulated target trial (TT) population
Hypothesis:	The incidence of long-term side effects differs between surgical and radiological treatment of PCa
Population *(mention key inclusion-exclusion criteria):*	Male adults (age ≥ 18) who received first surgical or radiological treatment for PCa (index date) RWD population • At least 365 days of observation pre index date • Clinically localised/locally advanced PCa (T1-T2, NX, M0) or Stage 1-2 diagnosed with no M1 or Stage 3-4 diagnosis Target trial (TT) population • First clinically localised prostate cancer (T1-T2, NX, M0) or Stage 1-2 diagnosed in the 180 days pre-index date • At least 365 days of observation before PCa diagnosis • Age 50-69 years on the date of PCa diagnosis • PSA in the range 3.0-19.99 ng/ml in the 180 days pre-index date • No concomitant or past malignancies (other than non-melanoma skin cancer) pre-index date • No serious cardiac or respiratory problems i.e. stroke, MI,

	heart failure, COPD in the previous 365 days of the index date, • No kidney dialyses or transplantation pre-index date • No bilateral hip replacement pre-index date • No prior treatment for prostate malignancy (androgen deprivation therapy within 180 days of index date allowed for radiotherapy cohort)
Exposure:	Surgical treatment (radical prostatectomy including open, laparoscopic, robotic-assisted with or without LN dissection)
Comparator:	Radiological treatment (EBRT, brachytherapy, focal therapy)
Outcome:	Adverse events, T grading, Quality of Life
Time *(when follow up begins and ends):*	From index date until death or end of observation period
Setting:	Ambulatory and inpatient
Main measure of effect:	Hazard ratio

7. Research methods

7.1. Study design

Research design (e.g. cohort, case-control, etc.): Cohort

Rationale for study design choice: One key advantage of cohort studies is their ability to establish temporal relationships between exposures and outcomes. Because participants are followed over time, researchers can observe the development of outcomes and identify when and how exposures may have contributed to the outcome. This temporal sequence is critical for establishing a causal relationship between exposures and outcomes, which is often essential for developing effective interventions and policies.

Cohort studies also allow for the measurement of multiple exposures and outcomes, which can help identify complex and multifactorial relationships. By collecting data on a wide range of potential risk factors and outcomes, researchers can investigate the interplay between different exposures and outcomes and identify possible interactions.

Additionally, cohort studies can provide valuable information about the natural history of diseases and their risk factors. By observing participants over time, researchers can track the development and progression of disease and identify key risk factors and predictors of disease.

Finally, cohort studies can be useful for investigating rare or long-term outcomes that may not be detected in other study designs. Because participants are followed over a long period of time, researchers can observe and collect data on outcomes that may take years or even decades to develop.

In conclusion, cohort studies are an important tool in observational health research because they allow researchers to establish temporal relationships between exposures and outcomes, investigate complex and multifactorial relationships, track the natural history of diseases, and observe rare or long-term outcomes.

7.2. Study design diagram

Figure 1a Study design diagram for RWD

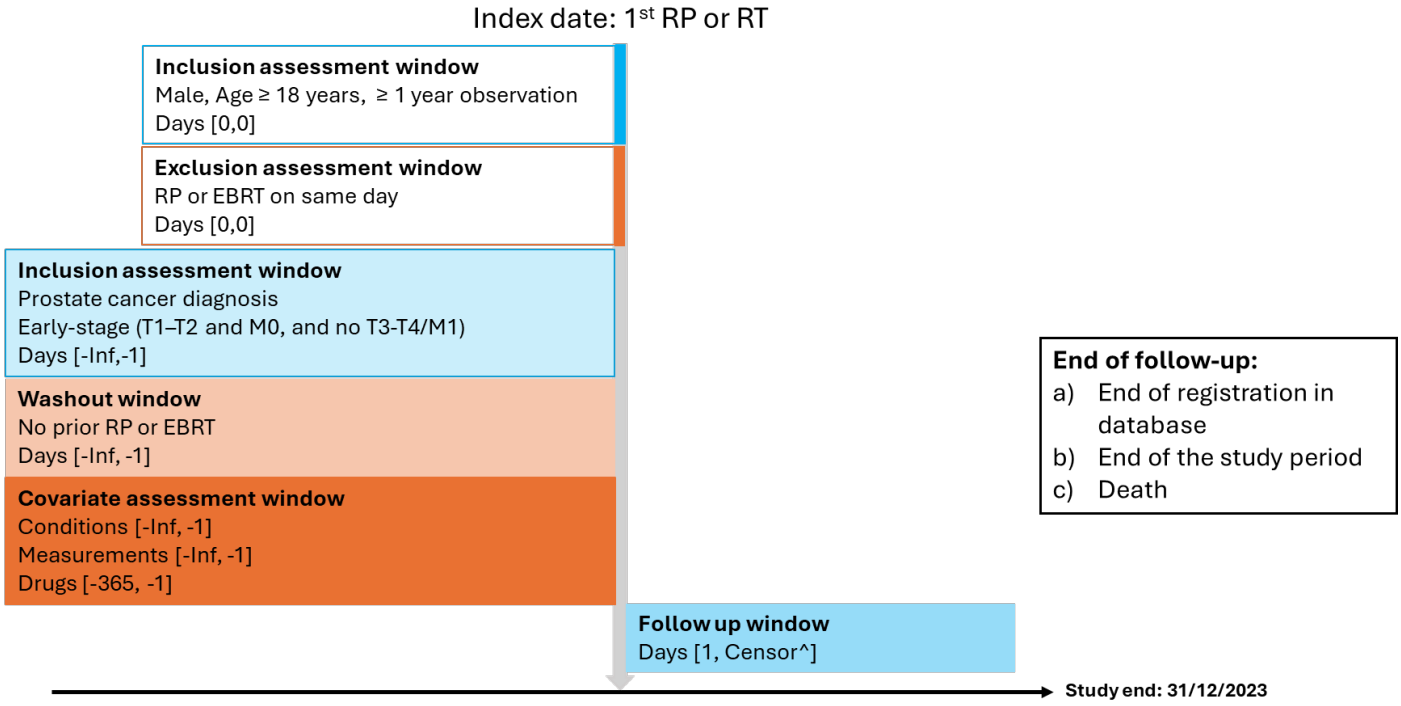
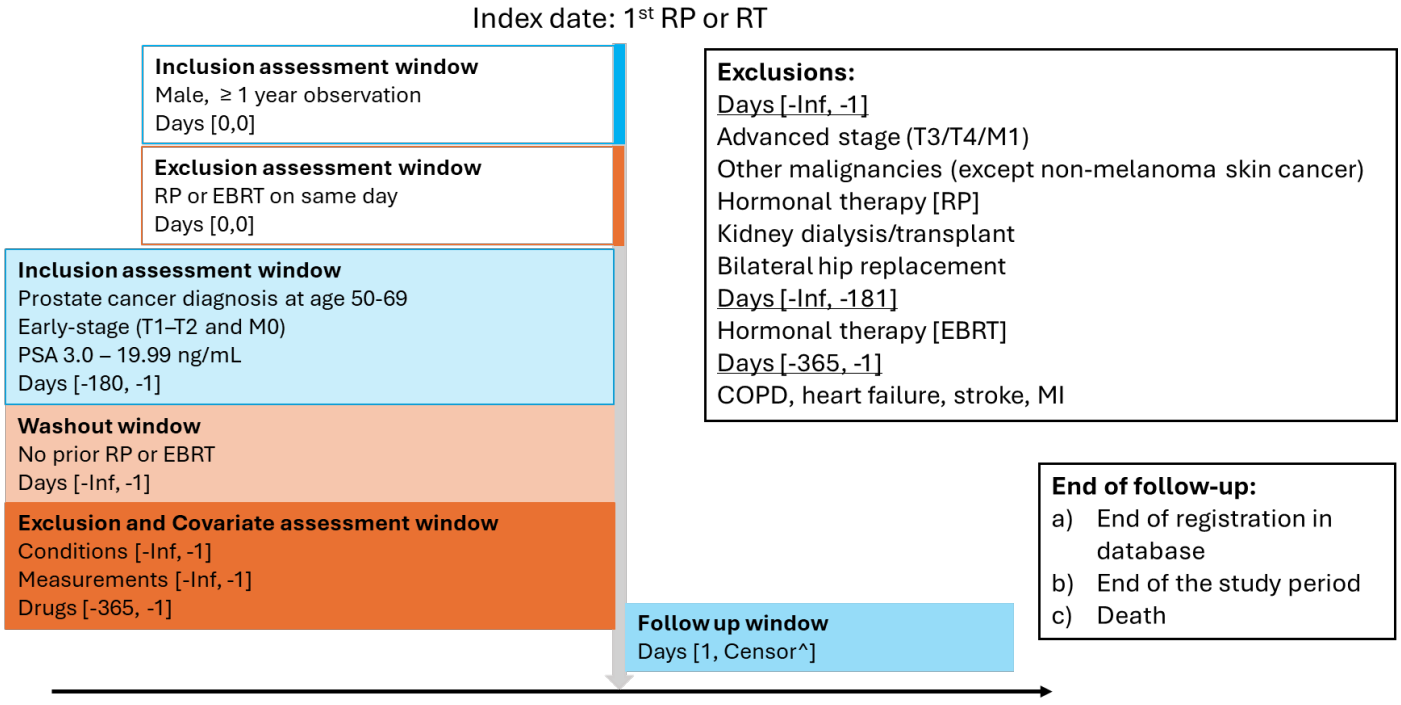


Figure 1b Study design diagram for TT



7.3. Setting

7.3.1 Context and rationale for definition of time 0 (and other primary time anchors) for entry to the study population

Time 0 is the date of first record of treatment after PCa diagnosis.

Table 3 Operational Definition of Time 0 (index date) and other primary time anchors

Study population name(s)	Time Anchor Description (e.g. time 0)	Number of entries	Type of entry	Washout window	Care Setting	Code Type	Diagnosis position	Incident with respect to…
RWD TT	time 0 is the date of first record of RP or RT after PCa diagnosis	Single entry	Incident	[-Inf, -1]	Ambulatory or inpatient	SNOMED	Any	RP or RT for PCa

7.3.2 Context and rationale for study inclusion criteria:

Patients with clinically localised/locally advanced PCa were included as patients who receive treatment for advanced or metastatic disease would have different prognosis and outcomes. The ProtecT trial eligibility criteria was emulated for the TT population to assess comparability of patient characteristics and outcomes.

Table 4. Operational Definitions of Inclusion Criteria

Criterion	Details	Order of application	Assessment window	Care Settings	Code Type	Diagnosis position	Applied to study populations:
Male	NA	Before	[-Inf, 0]	Ambulatory or inpatient	SNOMED	Any	RWD, TT
Age ≥ 18	NA	After	[0, 0]				RWD
Age 50 to 69 at first diagnosis of PCa	NA	After	[-180, 0]				TT
At least 365 days of prior observation	Before treatment (RWD) Before PCa diagnosis (TT)	After	[-365, 0]				RWD, TT
Clinically localised/locally advanced PCa (T1-T2, NX, M0) or Stage 1-2	NA	After	[-Inf, 0] [-180, 0]				RWD TT
PSA in the range 3.0-19.99 ng/ml	NA	After	[-180,0]				TT

7.3.3 Context and rationale for study exclusion criteria

The exclusion criteria emulate that in the ProtecT trial.

Table 5. Operational Definitions of Exclusion Criteria

Criterion	Details	Order of application	Assessment window	Care Settings	Code Type	Diagnosis position	Applied to study populations:
No M1 or Stage 3-4 diagnosis	NA	After	[-Inf, 0]	Ambulatory or inpatient	SNOMED	Any	RWD, TT
No concomitant or past malignancies (other than non-melanoma skin cancer)	NA	After	[-Inf, 0]		SNOMED		TT
No serious cardiac or respiratory problems (stroke, MI, heart failure, COPD)	NA	After	[-365, 0]		SNOMED		TT
No kidney dialyses or transplantation	NA	After	[-Inf, 0]		SNOMED		TT
No bilateral hip replacement pre-index date	NA	After	[-Inf, 0]		SNOMED		TT
No prior treatment for prostate malignancy (androgen deprivation therapy within 180 days of index date allowed for radiotherapy cohort)	NA	After	[-Inf, 0]		SNOMED RxNorm		TT
No treatment for RP and RT in the same date	NA	After	[0, 0]		SNOMED		TT

7.4. Variables

7.4.1 Context and rationale for exposure(s) of interest

The exposures of interest are surgical treatment (open, laparoscopic, robotic-assisted with or without LN dissection) whereas the comparator is radiological treatment (EBRT, brachytherapy, focal therapy)

Table 6. Operational Definitions of Exposure

Exposure group name(s)	Details	Washout window	Assessment Window	Care Setting	Code Type	Diagnosis position	Applied to study populations:	Incident with respect to…
RP	NA	[-Inf, 0]	[-Inf, -1]	Ambulatory or inpatient	SNOMED	Any	RWD, TT	RP for PCa
RT	NA	[-Inf, 0]	[-Inf, -1]	Ambulatory or inpatient	SNOMED	Any	RWD, TT	RT for PCa

7.4.2 Context and rationale for outcome(s) of interest

The following outcomes will be estimated in this study

- Incidence of clinical progression (including metastasis, biochemical recurrence, initiation of long term hormone therapy)
- Treatment complications (rectal/bowel injury/symptoms, sexual dysfunction, urethral stricture, lower urinary tract symptoms and treatment, incontinence and obstruction)
- Depression and psychological state
- Metabolic syndrome (hypertension, hypercholesterolaemia, type 2 diabetes)
- Fracture
- Cardiovascular outcomes (stroke, MI, DVT, PE)
- Death (all-cause, PCa, cardiovascular)

Table 7. Operational Definitions of Outcome

Outcome name	Details	Primary outcome?	Type of outcome	Washout window	Care Settings	Code Type	Diagnosis Position	Applied to study populations:
Incidence of clinical progression (including metastasis, biochemical recurrence, initiation of long term hormone therapy) Treatment complications (rectal/bowel injury/symptoms, sexual dysfunction, urethral stricture, lower urinary tract symptoms and treatment, incontinence and obstruction) Depression and psychological state Metabolic syndrome (hypertension, hypercholesterolaemia, type 2 diabetes) Fracture Cardiovascular outcomes (stroke, MI, DVT, PE) Death (all-cause, PCa, cardiovascular)	NA	All primary	Time to event	[-Inf, 0]	Ambulatory or Inpatient	ICD10 or SNOMED or RxNORM	Any	TT, RWD

7.4.3 Context and rationale for follow up

Patients will be followed up until the earliest of outcome of interest, death or end of observation period.

Table 8. Operational Definitions of Follow Up

Follow up start	Day 1	
Follow up end[1]	**Select all that apply**	**Specify**
Date of outcome	Yes	
Date of death	Yes	
End of observation in data	Yes	
Day X following index date *(specify day)*	No	
End of study period (specify date)	No	
End of exposure *(specify operational details, e.g. stockpiling algorithm, grace period)*	No	
Date of add to/switch from exposure *(specify algorithm)*	No	
Other date *(specify)*	No	

[1] Follow up ends at the first occurrence of any of the selected criteria that end follow up.

7.4.4 Context and rationale for covariates (confounding variables and effect modifiers, e.g. risk factors, comorbidities, comedications)

A preliminary list of covariates is listed below to be included as descriptive characteristics in the characterisation domain or potential confounders in the comparative effectiveness/safety domain.

- Age
- TNM staging
- Gleason score
- PSA
- Comorbidities
- Comedications

Table 9. Operational Definitions of Covariates

Characteristic	Details	Type of variable	Assessment window	Care Settings	Code Type	Diagnosis Position	Applied to study populations:
Age	Age in years defined by (time 0 – year of birth)/365)	Continuous	Years from time 0	NA	NA	Any	TT, RWD
TNM staging	NA	Binary	[-180, 0], [-Inf, 0]	Ambulatory or inpatient	SNOMED		TT RWD
Gleason score	NA	Ordinal	[-Inf, 0]	Ambulatory or inpatient	SNOMED		TT, RWD
PSA	NA	Continuous	[-180, 0], [-Inf, 0]	Ambulatory or inpatient	SNOMED		TT RWD
Comorbidities	NA	Counts	[-Inf, 0]	Ambulatory or inpatient	SNOMED		TT, RWD
Comedications	NA	Counts	[-365, 0]	Ambulatory or inpatient	RxNORM		TT, RWD

7.5. Data analysis

7.5.1 Context and rationale for analysis plan

Hazard ratios will be calculated to compare PCa patients receiving surgical (open, laparoscopic, robotic-assisted with or without LN dissection) versus radiological (EBRT, brachytherapy) treatments in risk of effectiveness and safety/quality of life outcomes.

Table 10. Primary, secondary, and subgroup analysis specification

Hypothesis:	The incidence of long-term side effects differs between surgical and radiological treatment of PCa
Exposure contrast:	Surgical treatment (open, laparoscopic, robotic-assisted with or without LN dissection) vs. Radiological treatment (EBRT, brachytherapy)
Outcome:	Incidence of clinical progression (including metastasis, biochemical recurrence, initiation of long term hormone therapy) Treatment complications (rectal/bowel injury/symptoms, sexual dysfunction, urethral stricture, lower urinary tract symptoms and treatment, incontinence and obstruction) Depression and psychological state Metabolic syndrome (hypertension, hypercholesterolaemia, type 2 diabetes) Fracture Cardiovascular outcomes (stroke, MI, DVT, PE) Death (all-cause, PCa, cardiovascular)
Analytic software:	R
Model(s): ***(provide details or code)***	MatchIt https://cran.r-project.org/web//packages//MatchIt/
Confounding adjustment method	***Name method and provide relevant details, e.g. bivariate, multivariable, propensity score matching (specify matching algorithm ratio and caliper), propensity score weighting (specify weight formula, trimming, truncation), propensity score stratification (specify strata definition), other.***
	Matching: variable ratio matching, 0.2 standard deviations on the logit scale Weighting: trim to equipoise, only subjects with a preference score between 0.25 and 0.75 will be retained Stratification: 5 strata based on the propensity score of the overall population
Missing data methods	***Name method and provide relevant details, e.g. missing indicators, complete case, last value carried forward, multiple imputation (specify model/variables), other.***
	Not applicable
Subgroup Analyses	***List all subgroups***
	RWD: age <70, ≥70 Calendar period: 2010-2020

Table 11. Sensitivity analyses – rationale, strengths and limitations

	What is being varied? How?	Why? (What do you expect to learn?)	Strengths of the sensitivity analysis compared to the primary	Limitations of the sensitivity analysis compared to the primary
Clone censor weighting	Addition of active surveillance Index date is date of PCa diagnosis	Outcomes of patients who did not receive PCa treatment	Avoid immortal time bias when no index date for treatment	

7.6. Data sources

7.6.1 Context and rationale for data sources

CPRD GOLD

Reason for selection: The Clinical Practice Research Datalink GOLD (CPRD) is an ongoing primary care database of anonymised medical records from general practitioners, with coverage of over 20 million patients from 985 practices in the UK (as of June 2022) with a median follow up time of 13 years and with ~ 9 million patients available for linkage. Contains demographic characteristics, diagnoses and symptoms, drug exposures, vaccination history, laboratory tests and referrals to hospital and specialist care as well as linkage to hospital, death and cancer registries etc. Approximately 6.9% of the UK population are included and patients are broadly representative of the UK general population in terms of age, sex and ethnicity

Strengths of data source(s): Covers all of the UK.

Limitations of data source(s): Requires linkage to cancer registries for specific questions relating to cancer staging and treatments. Information on cancer treatment only available 2014-18 and National Radiotherapy Dataset data is only available April 2009 to December 2018. Linkage is only available for patients in England.

Data source provenance/curation: CPRD GOLD is a widely used primary care database used for epidemiology research including cancer. CPRD carries out curation and quality control and provides users with information about what data is acceptable for research (patient level) and what data is up to standard at the GP practice level. Additional QC of data is performed at mapping and before analysis.

CPRD AURUM

Reason for selection: The Clinical Practice Research Datalink Aurum (CPRD) is an ongoing primary care database of anonymised medical records from general practitioners, with coverage of over 40 million patients from 1491 practices in the UK (as of may 2022) with a median follow up time of 8.74 years and with ~ 38 million patients available for linkage. Contains demographic characteristics, diagnoses and symptoms, drug exposures, vaccination history, laboratory tests and referrals to hospital and specialist care as well as linkage to hospital, death and cancer registries etc.

Strengths of data source(s): Large primary care database of > 40 million patients cover England and can be linked with cancer registries.

Limitations of data source(s): Requires linkage to cancer registries for specific questions relating to cancer staging and treatments. Information on cancer treatment only available 2014-18 and National Radiotherapy Dataset data is only available April 2009 to December 2018. Linkage is only available for patients in England.

Data source provenance/curation: CPRD Aurum is the largest primary care database, however, a relatively recent data resource compared to GOLD. Nevertheless, since the recent launch of CPRD Aurum a number of approved research projects is already underway – including pharmacovigilance, drug prescribing patterns, health services and policy evaluation, and disease risk factors. QC of data is performed at mapping and before analysis.

Estonian Biobank

Reason for selection: The Estonian Biobank is a database that covers more than 200 thousand voluntary gene donors in Estonia, which is about 15% of Estonian population. All the participants are genotyped, and the dataset includes also all their medical records and prescriptions going back to 2005 covering all levels of healthcare services. The data is also linked with the cancer and death registry information allowing to concentrate on validated cancer cases and study cause specific mortality.

Strengths of data source(s): Covers significant part of the whole population and includes comprehensive medical data from all levels of healthcare. Genetic profile of every donor is available in addition to health data.

Limitations of data source(s): Database is comprised of voluntary donors and might not be entirely representative.

Data source provenance/curation: The data is curated and linked by Estonian Biobank and University of Tartu computer science department teams. The teams have implemented thorough data cleaning, extraction and standardisation pipelines, with automated quality checks. Still, for each research questions additional quality control must be done by researchers.

DKG

Reason for selection: The DKG (German Cancer Society) database for prostate cancer patients is based on the multi-centre, prospective PCO (“Prostate Cancer Outcomes”) study combining patient-reported outcomes (PRO) surveys before and 12 months after beginning of treatment with quality assurance data from the certification process of the prostate cancer centres. PCO is an on-going study and part of the worldwide TruenNTH Global Registry initiative to measure and compare clinical and PROs of local and locally advanced prostate cancer. The PCO data is a large dataset with more than 40,000 prostate cancer patients from more than 130 different in-patient health care providers from Germany, Austria and Switzerland. Besides clinical, procedural and structural patient-specific data, socioeconomic as well as PRO data is available for all patients.

Strengths of data source(s): Large dataset from oncological care provided in specialised prostate cancer centres with a high amount of quality assurance and data checks and a rigorous methodology. PRO data is available as well. Detailed information about diagnosis and treatment.

Limitations of data source(s): No long-term (> one year) reliably available for all patients.

Data source provenance/curation: Clinical data is documented by health care providers and undergoes several internal as well as external quality control checks (including on-site audits) since it is used for certification purposes. PROs and sociodemographic data are collected within a scientific study setting and thus follows a strict study protocol.

Table 12. Metadata about data sources and software

	Data 1	Data 2	Data 3
Data Source(s):	CPRD GOLD	CPRD Aurum	Estonian Biobank
Study Period:	1995- to date This will depend on dates available for linkage to cancer registry and subsets	1995- to date This will depend on dates available for linkage to cancer registry and subsets	2006- to date This will depend on dates available for linkage to cancer registry and subsets
Eligible Cohort Entry Period:	See above	See above	See above
Data Version (or date of last update):	will get this when extracted	will get this when extracted	<Text>will get this when extracted
Data sampling/extraction criteria:	people with prostate cancer broad	people with prostate cancer broad	people with prostate cancer broad
Type(s) of data:	Primary care, HES (if applicable) linkage to cancer registry	Primary care, HES (if applicable) linkage to cancer registry	Primary care, Secondary care, Linkage to cancer registry
Data linkage:	Will require linkage to Cancer registries but this is external to CPRD	Will require linkage to Cancer registries but this is external to CPRD	All are linked
Conversion to CDM*:	Yes 5.3.1	Yes 5.3.1	Yes 5.3.1
Software for data management:	Postgres	Postgres	Postgres

*CDM = Common Data Model

7.7. Data management

A federated network analysis will be conducted. Sites will run the study analysis package locally on their data mapped to the OMOP CDM. Only aggregate results will be shared with the study coordinator. Result files will be automatically stored into a ZIP file that can be transmitted through a site's preferred SFTP client using a site-specific key provisioned by the Study Coordinator. Local data stewards are encouraged to review study parameters to ensure minCellCount function follows local governance. At a minimum, it is encouraged to keep this value to >5 to avoid any potential issues with re-identification of patients.

7.8. Quality control

Data partners will have run the OHDSI Data Quality Dashboard tool (https://github.com/OHDSI/DataQualityDashboard) adapted for OPTIMA. This tool follows Kahn's framework (11) and provides numerous checks relating to the conformance, completeness, and plausibility of the mapped data. Conformance focuses on checks that describe the compliance of the representation of data against specified standards and formats, in terms of recorded data values, relational constraints, or computational definitions. Completeness focuses on quantifying missingness, or the absence of data. Plausibility seeks to determine the believability of data values in terms of uniqueness and temporality. Each component are evaluated in two contexts: validation and verification. Validation relates to how well data align with external benchmarks with expectations derived from known true standards, while verification relates to how well data conform to local knowledge, metadata descriptions, and system assumptions.

7.9. Study size and feasibility

This study is undertaken using routinely collected data, all patients meeting the eligibility criteria above are included. We estimate the number of prostate cancer patients in each database using diagnosis codes.

The following number of patients were studied in previous publications:

Surgical and radiological treatment adverse events: 750 patients received surgery, 680 patients received radiotherapy (12)

Our sample size estimations suggest that 19,081 participants are required to detect a small hazard ratio of 1.2 between the two treatment groups for experiencing a rare adverse event (with a frequency of 5% in the PCa population as reported previously (13) with 80% power and alpha of 0.05.

Table 14. Feasibility counts

Data source	**Number of prostate cancer patients**
CPRD linked dataset (UK)	69,109
Estonian Cancer	18,030
DKG	40,000

8. Limitation of the methods

Observational health research has several limitations that can affect the quality and reliability of the findings. Some of the major limitations include:

Confounding: Observational studies cannot control all the factors that may affect the outcome of interest. There may be confounding variables that are associated with both the exposure and the outcome, which can make it difficult to determine the true relationship between the two. Where relevant, statistical methods are applied to account for confounding (such as Propensity Score methods).

Bias: Observational studies are prone to various types of bias, such as selection bias and measurement bias. These biases can lead to overestimation or underestimation of the true effect size. Selection bias might arise if data are missing for some confounders and the analyses are limited to complete cases. Information bias may occur if exposures, outcomes, or covariates are not correctly measured. Lack of recording (i.e., incompleteness) of exposures and outcomes may lead to misclassification of the variables. Medical conditions may be misclassified due to erroneous entries as they were defined based on the presence of diagnostic or procedural codes, or binarisation of covariates, with the absence of records indicative of absence of disease. Misclassification may be differential between the comparison groups however, this will be minimized by ensuring the prevalence and incidence of outcomes and prevalence of covariates correspond with clinical knowledge and published literature.

Causality: Observational studies can establish an association between an exposure and an outcome, but they cannot establish causality. There may be other factors that are responsible for the observed association.

Generalisability: Observational studies are often conducted in specific populations, which may not be representative of the broader population. The findings may not be applicable to other populations.

Limited control over exposure: Observational studies rely on natural variation in exposure, rather than experimental manipulation. This can limit the ability to control exposure levels and may result in exposure misclassification.

Overall, while observational health research can provide valuable insights into the relationship between exposures and outcomes, it is important to interpret the findings with caution and consider the limitations of the study design.

9. Protection of human subjects

Where frequency counts of patients are less than five, data will be censored to maintain patient confidentiality.

10. Reporting of adverse events

According to the guidelines for good pharmacovigilance practice by the European Medicines Agency (EMA/873138/2011 Rev 2) (14) and International Society of Pharmacoepidemiology (15), there is no requirement for expedited reporting of adverse drug reactions from studies with secondary use of data (such as electronic health care databases).

11. References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209-49.
2. Tsodikov A, Gulati R, Heijnsdijk EAM, Pinsky PF, Moss SM, Qiu S, et al. Reconciling the Effects of Screening on Prostate Cancer Mortality in the ERSPC and PLCO Trials. Ann Intern Med. 2017;167(7):449-55.
3. Mottet N, van den Bergh RCN, Briers E, Van den Broeck T, Cumberbatch MG, De Santis M, et al. EAU-EANM-ESTRO-ESUR-SIOG Guidelines on Prostate Cancer-2020 Update. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent. Eur Urol. 2021;79(2):243-62.
4. Harrison S, Tilling K, Turner EL, Martin RM, Lennon R, Lane JA, et al. Systematic review and meta-analysis of the associations between body mass index, prostate cancer, advanced prostate cancer, and prostate-specific antigen. Cancer Causes Control. 2020;31(5):431-49.
5. National Institute for Health and Care Excellence. NICE guideline [NG131] Prostate cancer: diagnosis and management 2019 [updated 15 December 2021; cited 2023 12 Jan]. Available from: https://www.nice.org.uk/guidance/ng131.
6. Hamdy FC, Donovan JL, Lane JA, Mason M, Metcalfe C, Holding P, et al. 10-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Localized Prostate Cancer. New England Journal of Medicine. 2016;375(15):1415-24.
7. Omar MI, Roobol MJ, Ribal MJ, Abbott T, Agapow P-M, Araujo S, et al. Introducing PIONEER: a project to harness big data in prostate cancer research. Nature Reviews Urology. 2020;17(6):351-62.
8. Lardas M, Liew M, van den Bergh RC, De Santis M, Bellmunt J, Van den Broeck T, et al. Quality of Life Outcomes after Primary Treatment for Clinically Localised Prostate Cancer: A Systematic Review. Eur Urol. 2017;72(6):869-85.
9. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. Am J Epidemiol. 2016;183(8):758-64.
10. Hamdy FC, Donovan JL, Lane JA, Metcalfe C, Davis M, Turner EL, et al. Fifteen-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Prostate Cancer. N Engl J Med. 2023;388(17):1547-58.
11. Kahn MG, Callahan TJ, Barnard J, Bauck AE, Brown J, Davidson BN, et al. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. EGEMS (Wash DC). 2016;4(1):1244.
12. Neal DE, Metcalfe C, Donovan JL, Lane JA, Davis M, Young GJ, et al. Ten-year Mortality, Disease Progression, and Treatment-related Side Effects in Men with Localised Prostate Cancer from the ProtecT Randomised Controlled Trial According to Treatment Received. Eur Urol. 2020;77(3):320-30.
13. Porcaro AB, Tafuri A, Rizzetto R, Amigoni N, Sebben M, Shakir A, et al. Predictors of complications occurring after open and robot-assisted prostate cancer surgery: a retrospective evaluation of 1062 consecutive patients treated in a tertiary referral high volume center. Journal of Robotic Surgery. 2022;16(1):45-52.
14. European Medicines Agency. Guideline on good pharmacovigilance practices (GVP) Module VI – Collection, management and submission of reports of suspected adverse reactions to medicinal products (Rev 2) 2017 [updated 02 Aug 2017. Available from: https://www.ema.europa.eu/en/human-regulatory/post-authorisation/pharmacovigilance/good-pharmacovigilance-practices.
15. Public Policy Committee ISOP. Guidelines for good pharmacoepidemiology practice (GPP). Pharmacoepidemiol Drug Saf. 2016;25(1):2-10.

12. Appendices