Monograph:
Chang, J.-Y.A. orcid.org/0000-0001-6660-5246, Chilcott, J.B. orcid.org/0000-0003-1231-7817 and Latimer, N.R. orcid.org/0000-0001-5304-5585 (2024) Leveraging real-world data to assess treatment sequences in health economic evaluations: a study protocol for emulating target trials using the English Cancer Registry and US Electronic Health Records-Derived Database. Report. SCHARR HEDS Discussion Papers (24.01). Sheffield Centre for Health and Related Research, University of Sheffield

---

University of Sheffield | Sheffield Centre for Health & Related Research

HEALTH ECONOMICS & DECISION SCIENCE

Discussion Paper Series

HEDS Discussion Paper 24.01

Title: Leveraging Real-World Data to Assess Treatment Sequences in Health Economic Evaluations: A Study Protocol for Emulating Target Trials Using the English Cancer Registry and US Electronic Health Records-Derived Database

Authors:

Jen-Yu Amy Chang ORCID ID: 0000-0001-6660-5246

James B. Chilcott ORCID ID: 0000-0003-1231-7817

Nicholas R. Latimer ORCID ID: 0000-0001-5304-5585

Corresponding author: Jen-Yu Amy Chang, Sheffield Centre for Health and Related Research (SCHARR), Division of Population Health, School of Medicine and Population Health, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA, UK
Tel: +44 (0) 114 222 5454
Email: jy.a.chang@sheffield.ac.uk

Leveraging Real-World Data to Assess Treatment Sequences in Health Economic Evaluations: A Study Protocol for Emulating Target Trials Using the English Cancer Registry and US Electronic Health Records-Derived Database

Jen-Yu Amy Chang[1,*], James B. Chilcott[1], Nicholas R. Latimer[1,2]

[1]Sheffield Centre for Health and Related Research (SCHARR), Division of Population Health, School of Medicine and Population Health, University of Sheffield
[2]Delta Hat Limited, United Kingdom

*Corresponding author: Jen-Yu Amy Chang
Sheffield Centre for Health and Related Research (SCHARR), Division of Population Health, School of Medicine and Population Health, University of Sheffield, Sheffield, UK
Address: Regent Court, 30 Regent Street, Sheffield, S1 4DA, UK
Tel: +44 (0) 114 222 5454
Email: jy.a.chang@sheffield.ac.uk

Running title: Target Trial Emulation for comparing treatment sequences

Key words: Treatment sequence, Health economics, Economic evaluation, Real-world data, Target Trial Emulation, English Cancer Registry, Electronic health records, Causal inference, Benchmarking

Funding: This study was supported by the Wellcome Trust [108903/B/15/Z]. The sponsor had no role in the study design, data collection, analysis, and decision to publish or preparation of the manuscript.

Conflicts of interest: All authors have no conflicts of interest that are directly relevant to the content of this article.

Ethics Approval: This project has been approved by the Ethics Committee of Sheffield Centre for Health and Related Research (SCHARR; formerly known as the School of Health and Related Research (ScHARR)) at the University of Sheffield (Ref. 041331) and the NHS Research Ethics Committee (Ref. 21/PR/0908).

Acknowledgement: The authors greatly appreciate Dr. Carmel Pezaro (Singleton Hospital, Swansea Bay University Health Board, NHS Wales), Professor Derek Rosario (Sheffield Teaching Hospitals NHS Foundation Trust), and Professor Janet Brown (Sheffield Teaching Hospital NHS Foundation Trust) for their invaluable clinical expertise and insights, which have been instrumental in enriching this project. We extend a heartfelt thanks to the team at Flatiron Health, especially Dr. Blythe Adamson and Dr. Philani Mpofu for their expert insights on the Flatiron database and assistance in accessing data. Lastly, we are thankful to Dr. Amanda Loban (University of Sheffield) for her expert guidance with information governance and assistance in fulfilling data management standards.

Project Overview	
Project Title	Leveraging Real-World Data to Assess Treatment Sequences in Health Economic Evaluations: A Study Protocol for Emulating Target Trials Using the English Cancer Registry and US Electronic Health Records-Derived Database • Sub-project 1: Investigating the Application of Causal Inference Methods for Modelling the Impact of Treatment Sequences in Health Economic Evaluations: Utilising Real-world Evidence from the Flatiron Database • Sub-project 2: Investigating the Application of Causal Inference Methods for Modelling the Impact of Treatment Sequences in Health Economic Evaluations: Utilising Real-world Evidence from the English Cancer Registry
Project Objective	This project aims to assess the feasibility of harnessing data from the English Cancer Registry and Flatiron electronic health record (EHR)-derived database to derive unbiased effect estimates for comparing oncology treatment sequences. A key aspect of this investigation is the application of advanced causal inference methods, with a particular focus on Target Trial Emulation.
Principal Investigator (PI)	Jen-Yu Amy Chang (PhD Candidate, Sheffield Centre for Health and Related Research (SCHARR), Division of Population Health, School of Medicine and Population Health, University of Sheffield)
Project Team Members	<u>PhD Supervisors:</u> Professor Nicholas Latimer (Professor of Health Economics, Sheffield Centre for Health and Related Research (SCHARR), School of Medicine and Population Health, University of Sheffield) Professor Jim Chilcott (Professor of Healthcare Decision Modelling, Sheffield Centre for Health and Related Research (SCHARR), School of Medicine and Population Health, University of Sheffield) <u>Clinical Experts:</u> Dr. Carmel Pezaro (Consultant in Oncology, Singleton Hospital, Swansea Bay University Health Board, NHS Wales for Swansea and Neath)

	Professor Derek Rosario (Consultant Urological Surgeon, Sheffield Teaching Hospitals NHS Foundation Trust) Professor Janet Brown (Medical Oncologist/Professor of Translational Medical Oncology, Sheffield Teaching Hospital NHS Foundation Trust)
PI's institution	University of Sheffield
Project Funder	Wellcome Trust [108903/B/15/Z]
Protocol version	2.0

Protocol Version History		
Version	**Date**	**Description**
Sub-project 1. Version 1.3	June 2021	**Protocol for Sub-project 1:** " Investigating the Application of Causal Inference Methods for Modelling the Impact of Treatment Sequences in Health Economic Evaluations: Utilising Real-world Evidence from the Flatiron Database" • Content: This protocol focuses on case studies using the Flatiron Database. • Ethics Approval: Approved by the SCHARR ethics committee at the University of Sheffield. • Database Application: Utilised for applying to access the Flatiron database. • Scientific Review: The project underwent a review by Flatiron's data analytics team to assess the theoretical feasibility for conducting the planned analysis using Flatiron data prior to the data application.
Sub-project 2. Version 1.0	June 2021	**Protocol for Sub-project 2**: "Investigating the Application of Causal Inference Methods for Modelling the Impact of Treatment Sequences in Health Economic Evaluations: Utilising Real-world Evidence from the English Cancer Registry" • Content: This protocol details case studies using the National Cancer Registration and Analysis Service (NCRAS) database. It also includes an overview and key elements of its sister

		project (Sub-project 1), presenting a comprehensive view of their interconnected objectives and methodologies. • Ethics Approval: Granted by the NHS Ethics Committee. • Data Application: Initially submitted in June 2021 to the Office for Data Release, Public Health England (PHE), for access to the NCRAS database. The application was later transferred and re-initiated with NHS Digital in June 2022 due to the dissolution of Public Health England (October 2021). • Scientific Review: Underwent peer-review within the institution for feasibility and scientific merit, and also reviewed by the National Disease Registration Service (NDRS) data analytical team during data application to determine data release for the planned analysis.
Combined protocol. Version 2.0	January 2024	**Combined protocol for Sub-project 1 and 2**: Key Updates: • Content: This protocol merged and refined wording of two sub-project protocols for publication readiness. • Literature update: Incorporated additional references from recent studies. • Enhanced details: Augmented the methods section with more details, including the criteria for assessing the agreement of benchmark trials and real-world evidence. • Team update: updated the project team members.

Contents

Abstract

Background

Considering the sequence of treatments is vital for optimising healthcare resource allocation, especially in cancer care, where sequence changes can affect patients' overall survival and associated costs. A key challenge in evaluating treatment sequences in health technology assessments (HTA) is the scarce evidence on effectiveness, leading to uncertainties in decision making. While randomised controlled trials (RCTs) and meta-analyses are viewed as the gold standards for evidence, applying them to determine the effectiveness of treatment sequences in economic models often necessitates making arbitrary assumptions due to insufficient information on patients' treatment histories and subsequent therapies. In contrast, real-world data (RWD) presents a promising alternative source of evidence, often encompassing details across treatment lines. However, due to its non-randomised nature, estimates of the treatment effectiveness based on RWD analyses can be susceptible to biases if not properly adjusted for confounding factors.

To date, several international initiatives have been investigating methods to derive reliable treatment effects from RWD — by emulating Target Trials that replicate existing RCTs (i.e. benchmarks) and comparing the emulated results against the benchmarks. These studies primarily seek to determine the viability of obtaining trial-equivalent results through deploying specific analytical methodologies and study designs within the Target Trial emulation framework, using a given database. Adopting the Target Trial emulation framework facilitates the analyses to be operated under causal inference principles. Upon validation in a particular database, these techniques can be applied to address similar questions (e.g., same disease area, same outcome type), but in populations lacking clinical trial evidence, leveraging the same RWD source.

Studies to date, however, have predominantly focused on the comparison of individual treatments rather than treatment sequences. Moreover, the majority of these investigations have been undertaken in non-English contexts. Consequently, the use of RWD in evaluating treatment sequences for HTA, especially in an English setting, remains largely unexplored.

Objectives

The goal of this project is to investigate the feasibility of leveraging RWD to produce reliable, trial-like effectiveness estimates for treatment sequences. We aim to assess the capability of two oncology databases: the US-based Flatiron electronic health record and the National Cancer Registration and Analysis Service (NCRAS) database of England. To achieve this, we plan to harness the Target Trial Emulation (TTE) framework for replicating two existing oncology RCTs that compared treatment sequences, with the intent of benchmarking our results against the original

studies. Further, we aim to detail the practicalities involved with implementing TTE in diverse databases and outline the challenges encountered.

Methods

1. We aim to emulate existing RCTs that compare the effect of different treatment sequences by constructing the study design and analysis plan following the TTE framework. Specifically, the following case studies are planned:
   (1) Prostate cancer case study 1 (PC1) - US direct proof-of-concept study (method direct validation): replicating the GUTG-001 trial using Flatiron data
   (2) Prostate cancer case study 2 (PC2) - US-England bridging study (method extension): emulating Target Trials that compare treatment sequences that have been common in England using Flatiron data
   (3) Prostate cancer case study 3 (PC3) - English indirect proof-of-concept study (method indirect validation): emulating the same Target Trial in PC2 using English NCRAS data
   (4) Renal cell carcinoma case study (RCC) - method direct validation in a single-arm setting: emulating the sunitinib followed by everolimus arm in the RECORD-3 trial using English NCRAS data
2. We will compare results of the emulated Target Trials with those from the benchmark trials.
3. We plan to compare different advanced causal inference methods (e.g. marginal structural models using IPW and other g-methods) in estimating the effect of treatment sequences in RWD.

Expected results

This study will provide evidence on whether it is feasible to obtain reliable estimates of the (comparative) effectiveness of treatment sequences using Flatiron data and English NCRAS data. If applicable, we intend to develop a framework that provides a systematic way of obtaining the (comparative) effectiveness of treatment sequences using RWD. It is possible that the data quality is insufficient to emulate the planned Target Trials. In this case, we will report reasons for the implausibility of data analysis. If applicable, we will make suggestions to whether the national health data collection may be enhanced to make the analyses possible. The results of this study will be submitted to peer-reviewed journals and international conferences.

1 Overview

This protocol outlines a series of proof-of-concept case studies focusing on evaluating the use of real-world data (RWD) for making informed decisions in health technology assessment (HTA), particularly in the context of treatment sequences. The protocol begins in Section 2 by underscoring the importance of evaluating treatment sequences in HTA and discussing the opportunities and challenges in leveraging RWD. We describe the Target Trial Emulation (TTE) approach with benchmarking as a means to assess the feasibility of deriving reliable estimates from RWD, and specify RWD sources for investigation. We then lay out the primary objectives of the project.

In Section 3, we elaborate on the project's significance and relevance to pertinent research. Section 4 presents a detailed Analysis Plan for the Target Trial Emulation case studies. Finally, Section 5 details the data requirements necessary for the study.

2 Background

2.1 Evaluating treatment sequences in health technology assessments

With an ever-increasing number of treatment options, the significance of evaluating treatment sequences within HTA has become apparent. Specifically, changing the order of treatments can introduce variability in the overall effectiveness and costs associated with managing a disease, making the assessment of treatment sequences—as opposed to a single line of therapy—vital in HTA wherever relevant.[1-3]

Despite established frameworks for modelling treatment sequences in health economic evaluations, challenges persist, particularly the scarcity of data on the effectiveness of treatment sequences.[1,4-6] Although clinical trials are considered the “gold standard” in evidence, they rarely assess the impact of sequences, focusing instead on the efficacy/effectiveness of a single line of treatment (LOT). Given the scarcity of trials comparing treatment sequences, analysing RWD offers a promising avenue to determine the (comparative) effectiveness of sequential treatment strategies, thereby supporting more informed clinical and economic decision-making.

2.2 Assessing the sequencing effect using rea-world data

The use of RWD is advantageous not only for its capability to capture sequencing information, but also for offering larger, more generalisable sample sizes compared to clinical trials. However, the lack of random treatment allocation in routine practice necessitates careful study design and statistical analysis to avoid biased results[7], notably due to confounding from factors that affect both treatment choices and outcomes, such as disease severity. While existing guidelines discuss RWD's utility for HTA and methods for estimating treatment effectiveness from RWD, none of them explored methods for comparing treatment sequences.[8,9]

To effectively harness RWD for evaluating the causal effects of different treatment sequences—a form of time-related static treatment strategies or dynamic treatment strategies[10]—, it is crucial to employ advanced causal inference methods like marginal structural models with inverse probability weighting (IPW) and other G-methods for ensuring a "fair comparison" across patients receiving different treatment sequences (i.e., achieving balanced patient characteristics between treatment groups, and addressing time-varying confounding).[10-12] Moreover, the successful application of these statistical methods often hinges on the availability of adequate data and relevant variable information.

Our project, in response to the absence of established guidelines, aims to determine the feasibility of applying the aforementioned methods to real-world datasets, especially local ones, to reliably estimate the effectiveness of treatment sequences in the context of supporting decision-making by the National Institute for Health and Care Excellence (NICE). Specifically, The National Cancer Registration and Analysis Service (NCRAS) database for England[13] and the US Flatiron Electronic Health Records (EHR)-Derived database[14] were identified as promising data sources for our initial investigations. More detailed information and the rationale behind choosing these databases are provided in Section 2.4.

2.3 Target Trial Emulation and benchmarking

In addition to advanced statistical methods, the TTE framework, proposed by Hernan et al.[15], emerged as valuable tool for structuring observational studies aimed at answering causal questions. The framework's significance lies in facilitating adequate designs of observational studies, enhancing transparency, thereby further mitigating biases inherent in study designs (which may not be fully rectifiable through statistical methods alone), such as selection bias and immortal time bias. The framework is based on the idea of designing an observational study as a hypothetical Target Trial[15], had such a trial been implementable, and then explicitly emulating this Target Trial using RWD. A standard Target Trial protocol consists of seven key components to resemble the setting of a randomised controlled trial (RCT), including eligibility criteria, intervention strategies being compared, intervention assignment, follow-up period, outcomes of interest, causal contrasts of interest, and analysis plan.

Several initiatives have launched benchmarking studies to determine how effectively the TTE approach can be used within specific real-world datasets to answer causal questions, especially before applying it to other questions in comparable settings.[16,17] These benchmarking studies attempted to replicate the designs and results of existing clinical trials (i.e., benchmark trials) through emulating Target Trials using RWD, including applying the same (or as far as possible) patient inclusion/exclusion criteria and analytical methods to achieve the emulation. Theoretically, if a Target Trial is correctly specified, estimates derived from RWD may be comparable to those from a

benchmark trial, providing a validated means to derive reliable real-world evidence (RWE). However, a review highlights that disagreements between observational studies and RCTs can arise for various reasons, with the specific causes often being indeterminable.[18] Nevertheless, improved benchmarking can be achieved, had an observational study explicitly aimed to emulate a Target Trial.[18]

Building on this concept, our study will employ a similar strategy to assess the applicability of advanced statistical methods in generating reliable RWE for HTA decisions involving treatment sequences. Specifically, in the current study protocol, we outline the design of a series of benchmarking case studies following the TTE framework aiming to replicate the effectiveness estimates from several RCTs comparing treatment sequences[19-22] (Section 4. Analysis Plan). We will compare our findings with those from benchmark RCTs, assessing the potential of RWD to successfully mimic their results. The design of our TTE analyses (Table 1-3) references the published protocol structures of the RCT DUPLICATE case studies.[23,24]

In developing the protocol of our study, we conducted a systematic review to identify candidate benchmark trials, focusing on RCTs that explicitly randomised patients to receive different predetermined treatment sequences. Our study settled on two oncology trials: the prostate cancer trial, GUTG-001, and the renal cell carcinoma (RCC) trial[19,25] RECORD-3[22]. Due to the extensive nature of the review, we will provide the detailed rationale in a separate publication.

2.4 The English NCRAS database and the Flatiron database

Our study focuses on implementing the benchmarking studies using two oncology databases: the NCRAS database for England and the Flatiron database. NCRAS, a part of England's National Disease Registration Service (NDRS), coordinates aggregated information from cancer registries across England, Northern Ireland, and Scotland at the UK level, with regional authorities facilitating access to patient-level data. Additionally, NCRAS oversees the patient-level cancer registry data in England.[26] Researchers can access the English Cancer Registry with linkage to a selection of non-cancer specific National Health Service (NHS) England datasets through NHS England's DARS (Data Access Request Service). The application was previously managed by the Office for Data Release, Public Heath England. The English Cancer Registry provides detailed data on NHS England's cancer patients, including important prognostic factors, such as tumour stages, sizes, and patient performance status at the time of diagnosis. Enhancing this, the registry can be linked with other NCRAS datasets, such as the Systemic Anti-Cancer Therapy (SACT) dataset[27], which provides extensive information on cancer treatments, including those under the Cancer Drugs Fund (CDF). Additionally, its viable linkage with NHS hospital records (i.e., Hospital Episode Statistics (HES)), allows for a thorough understanding of patients' medical histories.

The Flatiron database, a US-based EHR-derived database focused on oncology care, provides

detailed diagnostic and treatment records of patient visits as well as laboratory results.[14] It combines structured and machine learning/manually abstracted unstructured data, making it a comprehensive resource for oncological research. Despite being US-based, the Flatiron database was included in our study for its potential of more timely data access, diverse patient demographics (across the US) potentially overlapping with the English population, and its capability to capture treatment sequences relevant to the GUTG-001 trial, which are less common in the UK. This choice facilitates the design of our benchmarking study protocol, tethered to the identified benchmark trials, as detailed in Section 2.5 and Section 4. Furthermore, NICE has partnered with Flatiron Health to explore the use of RWE in improving the assessment of health technologies' clinical and cost effectiveness.[28-30] An example of this is the use of Flatiron's data to supplement clinical trial information in a recent NICE technology appraisal (TA).[31]

2.5 Summary of project aims

In summary, this project aims to examine the feasibility of using the English NCRAS data and the Flatiron data to obtain reliable effectiveness estimates of treatment sequences in prostate cancer and RCC. The aim of this project will be achieved through completing the following objectives.

1) To emulate existing RCTs that compare the effect of different treatment sequences by constructing the analyses plan following the Target Trial framework.
2) To compare results of the emulated Target Trials with those from the benchmark trials.
3) To compare different advanced causal inference methods (e.g. marginal structural models with IPW and other g-methods) in estimating the effect of treatment sequences in RWD.
4) To detail the practical aspects of implementing TTE across different databases and to describe challenges encountered.

Ultimately, our goal is to leverage the insights from this project to create a systematic framework for generating evidence on the effectiveness of treatment sequences using RWD, particularly in the context of health economic evaluations. In the event that benchmarking proves to be infeasible, our focus will shift to providing detailed insights into the practicality of implementing TTE with these databases. This includes an exploration of the challenges encountered, the reasons behind any limitations, and potential areas for improvement and future research.

3 Significance of this study

The significance of this study lies in its potential to expand upon relevant existing research conducted by several initiatives, including the US Food and Drug Administration (FDA) funded RCT DUPLICATE and the pharmaceutical industry-sponsored OPERAND.[16,17,24,32,33] These initiatives have been focusing on replicating clinical trial results using RWD within highly structured frameworks designed to mimic clinical trials.[16,17] RCT DUPLICATE, initiated in 2018 under the 21st

Century Cures Act, seeks to inform the use of RWE studies in regulatory decisions. It leverages the Target Trial Emulation approach to assess the real-world effectiveness of medical products and benchmark them against a large number of RCTs.[24,34] OPERAND, in contrast, explores how treatment effect estimates might differ when the stringent eligibility criteria of RCTs are relaxed. To our knowledge, these studies have primarily focused on cardiovascular diseases and utilised US claims databases (e.g. Medicare), without attempting to assess the feasibility of estimating the comparative effectiveness of treatment sequences.

Our project stands out by seeking to replicate results from sequential treatment trials using data from the English Cancer Registry and the Flatiron EHR-derived database. This approach is expected to enhance the findings from previous large-scale initiatives in several ways. Firstly, it will evaluate different causal inference methods specifically for emulating sequential treatment trials using RWD and benchmark them against existing trials. Secondly, the project will demonstrate the utility of observational data as an alternative source of evidence in health economic evaluations, particularly for modelling treatment sequences. Specifically, it could pave the way for establishing a systematic framework for deriving reliable (i.e. trial-mimicking) comparative effectiveness estimates for sequential treatments, addressing a key challenge in sequence evaluation in HTA. Thirdly, the project aims to improve user experience and enhance the use of English NCRAS data in future HTA.

Furthermore, part of the project (i.e. prostate cancer benchmarking studies) will compare the emulation of the same Target Trials using the US-based Flatiron database versus the English NCRAS database, as outlined in Section 4. Analysis Plan). This comparison will offer insights into the differing treatment patterns for prostate cancer in the US and England, and how they may affect the emulation. For example, sequential use of abiraterone and enzalutamide is not permitted in England[35], but they are key first and second-line treatments in the US[36], as a recently study using Flatiron data shows.[36] The global debate over the benefits of this sequence is ongoing.[37-43]

Based on a recent systematic review focusing on treatment sequences in prostate cancer conducted by the Canadian Agency for Drugs and Technologies in Health (CADTH)[40], it seems that docetaxel-containing treatment sequences with androgen receptor-targeted agents (ATRA) (i.e. abiraterone, enzalutamide) may improve progression-free survival (PFS) compared to sequential therapy with ATRA alone in castration-resistant prostate cancer (CRPC) patients. However, none of the studies included evidence from England and studies included were all retrospective, and therefore, should be interpreted with caution. In addition, no published cost-effectiveness studies were found explicitly comparing different treatment sequences in prostate cancer, despite the interest of decision makers in this. The results of our study will supplement the understanding of these topics using the English cancer registry data to provide English-based effectiveness estimates. Additionally, our analyses will provide insights on whether conducting similar Target Trial analyses using different

observational data sources (i.e. English NCRAS data and US Flatiron data) may require modifications in defining important variables (e.g. definition of progression using retrospective data) and/or result in contrasting final results. We will also explore the strengths of each database in informing HTA treatment sequencing decisions in England, identifying potential areas for improvement, especially within the local database (NCRAS).

4 Analysis Plan

4.1 Overview

This section is structured into four parts. Firstly, we introduce the scope of the section and provide a brief overview of the structure for each set of Target Trial case studies. This is followed by exploring considerations associated with applying causal inference methods for Target Trial analyses using English NCRAS and Flatiron data in Section 4.2. Subsequently, we present detailed plans for our two sets of sequential treatment Target Trial case studies in prostate cancer (Section 4.3: case studies PC1, PC2, and PC3) and RCC (Section 4.4: case studies RCC1 and RCC2).

Each case study set begins with an introduction to the benchmark trials (GUTG-001[22] and RECORD-3[19,25]) and summarises the demographics of cancer patients in the UK/England and the US, assessing the treatment sequences used in NHS and US clinical practice. This is followed by a detailed presentation of the planned Target Trials, featuring a table summarising their seven key components. Primary outcomes in all case studies focus on time-to-event outcomes, particularly the overall survival (OS) of patients receiving specific treatment sequences. At the end of each case study set, we assess whether the NCRAS and Flatiron data offer a sufficient sample size for our planned Target Trials. We will compare the outcomes of our emulated Target Trials with their corresponding counterparts in the benchmark trials. Finally, Section 4.5 details the criteria for determining the agreement between our emulated Target Trials and their corresponding benchmark trials, and Section 4.6 lists the software that will be used for the analyses.

We acknowledge the scarcity of clinical trials comparing treatment sequences that have been conducted in the England, leading to the absence of fully suitable benchmark trials for a proof-of-concept study with “direct benchmarking” using NCRAS data. Nevertheless, we are confident that our systematic review (to be detailed in a forthcoming publication) has identified the best suited benchmark trials for our project, despite their limitations. Specifically, PC1 will emulate an Analogue Target Trial of the GUTG-001 trial using Flatiron data. This serves as a direct proof-of-concept study for comparing treatment sequences using Flatiron data. Depending on PC1's success, PC2 will expand the same approach to a broader population and comparison of alternative treatment sequences in prostate cancer that are prevalent in both the US and England. PC2 will continue to use Flatiron data, functioning as a method extension study. PC3 will then follow, replicating PC2’s design (i.e.,

Analogue Target Trial of PC2) and analysis but using NCRAS data, thus acting as an indirect proof-of-concept study for sequence comparison analysis with NCRAS data. Additionally, the RCC case studies, constrained to replicating a single treatment sequence from RECORD-3 due to the unavailability of the other sequence in England, will nonetheless function as a direct proof-of-concept for single-arm studies using NCRAS data. The interconnections between each planned Target Trial analyses are illustrated in Figure 1 and further elaborated in Section 4.3 and 4.4.

In summary, our case study designs creatively overcome the scarcity of direct benchmark trials for examining the feasibility of sequencing comparisons with NCRAS data. By leveraging and optimising existing, albeit imperfect, benchmarks, we aim to evaluate the feasibility of using NCRAS data to support local HTA sequencing decisions in England.

4.2 Considerations of applying causal inference methods in the NCRAS and Flatiron data

Advanced causal inference methods (i.e. marginal structural models with IPW and other g-methods) will be applied to mimic the effect of randomisation in analysing RWD through the principle of "no unmeasured confounders". Therefore, it is important to understand if all important prognostic factors affecting treatment decisions and outcomes can be well captured in the NCRAS and Flatiron data. For selecting patients, we require basic characteristics (e.g. age, sex) and key tumour prognosis factors (e.g. tumour size, tumour histology, tumour stage) and any factors that might influence survival at diagnosis and the time of treatment switching. For outcome measures, we need the death date of patients (i.e. estimating OS), indication of treatment relapse (i.e. estimating progression free survival (PFS)), and factors influencing patients being lost to follow-up (e.g. moving out of the country). Details of variables required will be described for each case study. We recognise that some desired variables might be unavailable. Thus, a key aspect of the study involves exploring the possibility of extracting necessary information from a blend of other related variables, in cases where direct mapping is absent. The process of variable selection and operational definition, and the use of proxy variables, will be subject to further discussions with clinical experts.

The principal investigator, JYAC, has experience in analysing disease registry and EHR data. Her PhD supervisors, NL and JBC, have extensive experience in research with cancer trial and registry data, oncology HTA and prostate cancer screening programs in the UK. They will help facilitate project collaborations with clinical experts, including Dr Carmel Pezaro, Professor Derek Rosario, and Professor Janet Brown, all part of the project team specialising in prostate and kidney cancer treatments. This project may include other statistical experts specialised in g-methods in the future in the research team or in post-hoc consultation, if necessary.

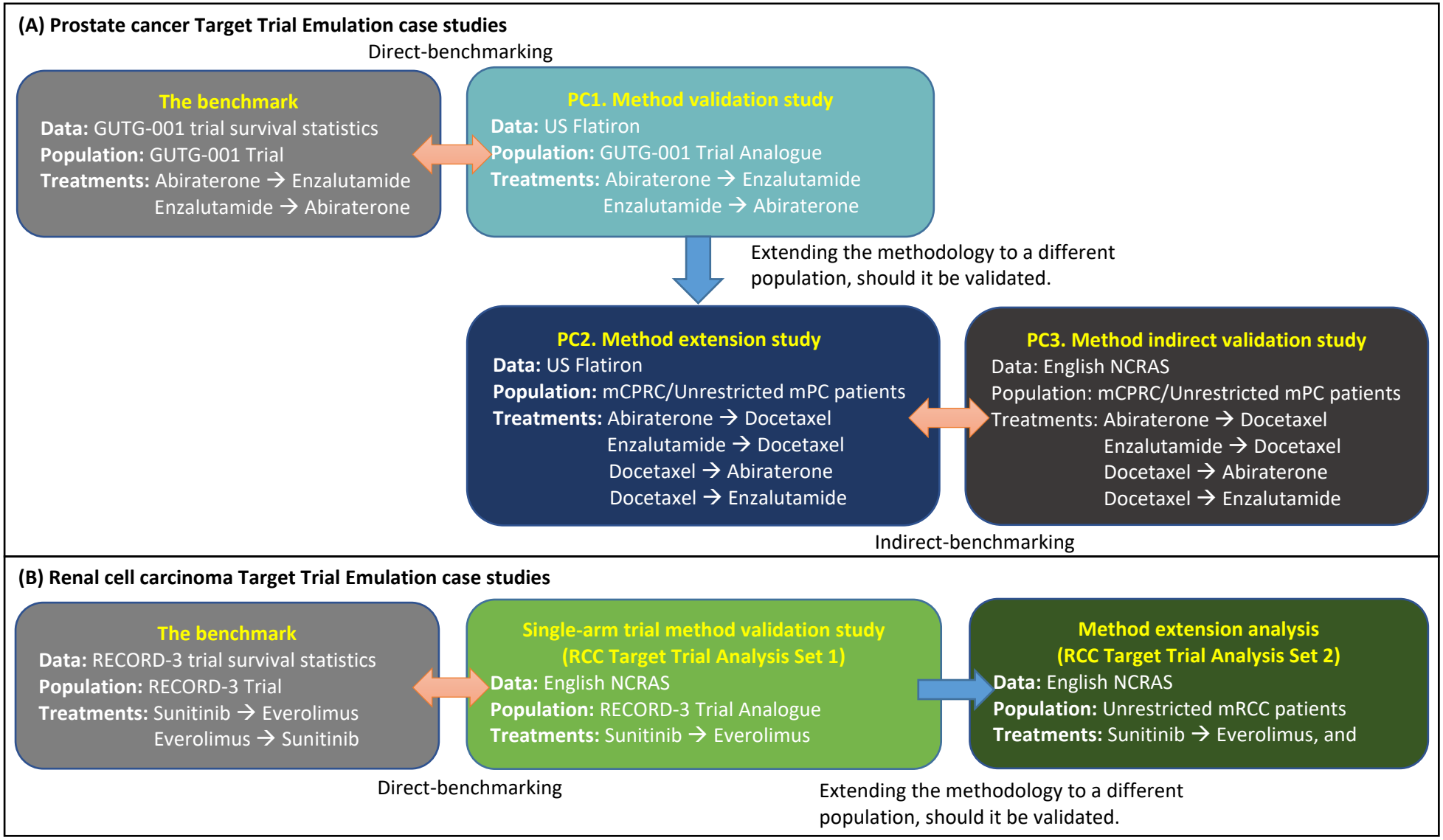

Figure 1. Schematic overview of interrelationships and purposes of each proposed Target Trial Emulation case study

mPC: metastatic prostate cancer; mRCC: metastatic renal cell carcinoma, TTE: Target Trial Emulation.

The orange arrows represent the comparison between benchmarks and their emulated counterparts, while the blue arrows indicate the application of validated emulation methods to a new population.

Furthermore, the application of causal inference methods also relies on the comparison of counterfactual pairs. That is, a certain degree of overlap with respect to patient characteristics between study groups are required to create a reasonable comparison. Thus, we will assess the overlap of patients receiving different treatment sequences and summarise in descriptive statistics. This project is also designed to understand the extent to which the completeness of the data may have an impact on the expected study results (i.e. results deviate from the benchmark trial). It may be possible that we are unable to fully replicate results from the chosen benchmarks. In this case, we will document the potential reasons (e.g. insufficient sample size, incomplete data on time-varying prognostic factors) and provide discussions around how NCRAS and Flatiron data may be enhanced to enable the similar analyses in the future.

4.3 Sequential treatment Target Trial: prostate cancer case studies

4.3.1 Benchmark RCT

In our prostate cancer case study, we identified a phase-2 RCT by Khalaf et al., the GUTG-001 trial[22], as a valuable benchmark. This trial compared the following two treatment sequences in treating treatment-naïve metastatic CRPC (mCRPC) patients:

- abiraterone (plus prednisolone) followed by enzalutamide (n = 101)
- enzalutamide followed by abiraterone (plus prednisolone) (n =101)

Although a recent commentary from some oncologists suggest the use of these two treatment sequences under specific circumstances in the UK, it seems unlikely that we can obtain an adequate sample size (n > 100) for a TTE study involving both sequences using NCRAS data.[39] The reason being these two drugs cannot be used directly after one another within the NHS. Had these treatment sequences been more prevalent in the UK, the GUTG-001 trial would have been a “perfect” benchmark for evaluating the feasibility of comparing treatment sequences with NCRAS data. On the other hand, these sequences have been more commonly used in the US[36], rendering the GUTG-001 trial an effective benchmark when using Flatiron data.

Given the absence of an ideal benchmark trial for a direct proof-of-concept study with NCRAS data, it is challenging to determine if NCRAS data can be used to reliably estimate the effectiveness of treatment sequences (in prostate cancer). To address this, we propose a novel strategy that involves jointly utilising the NCRAS and US Flatiron databases to —indirectly — assess the feasibility of deriving reliable effectiveness estimates from English data for treatment sequence comparisons. Our approach hinges on leveraging analyses with Flatiron data as a benchmark or "bridge" for assessing the analyses performed with NCRAS data. It involves a comparison of emulating identical Target Trials using both NCRAS and Flatiron data. Specifics of this design are elaborated in Section 4.3.4.

4.3.2 Demographics of metastatic prostate cancer

England

In England, around 6,000 new cases of metastatic prostate cancer are diagnosed annually.[44] Our oncology expert Dr. Carmel Pezaro noted that prostate cancer patients often start receiving castration therapy post-prostate cancer diagnosis, encompassing either surgical castration (i.e., bilateral orchiectomy) or medical castration (i.e., life-long androgen deprivation therapy, ADT). Professor Derek Rosario added that surgical castration accounts for only a very small fraction of these cases. Life-long ADT involves luteinising hormone-releasing hormone (LHRH, also known as gonadotrophin-releasing hormone (GnRH)) agonists or antagonists, such as padeliporfin and degarelix.[45] If the cancer progresses despite castration, the condition is termed castration-resistant prostate cancer (CRPC). The time of developing castration-resistance (i.e. hormone-relapse) from being castration-sensitive (i.e. hormone-sensitive) varies among patients, with an English study indicating that about 28% of prostate cancer patients may develop castration-resistance.[46] Patients with prostate cancer are typically managed by oncologists, while those on long-term ADT alone may be overseen by GPs post-initial treatment.

Docetaxel has been accessible as a first-line treatment for treatment-naïve patients with mCRPC within the NHS since 2006, whereas abiraterone and enzalutamide, two ARTAs, have been available for the same indication through The Cancer Drugs Fund (CDF) since 2016.[47-49] Since 2012 and 2014, respectively, abiraterone and enzalutamide have been introduced as second-line therapy options in the NHS (through CDF) for the treatment of patients with mCRPC who have previously undergone docetaxel treatment.[35,50,51] Prior to the introduction of abiraterone and enzalutamide, docetaxel therapy served as the sole standard treatment for patients with mCPRC. Since 2016, Cabazitaxel has been included as an alternative second-line treatment option in the treatment pathway (through CDF), exclusively for patients who have previously received docetaxel therapy.[52] Additional treatment options for patients at a later phase of mCRPC are available, including Radium-223 being approved since 2016 for patients with bone metastases (through CDF).[53] In May 2023, Olaparib was approved for mCRPC patients with breast cancer gene (BRCA) mutations through CDF[54], while the use of Lutetium-177 Vipivotide Tetraxetan for PSMA-positive patients after two or more prior treatments was not recommended in a recent NICE TA.[55]

Importantly, the English standard practice does not allow for the sequential use of abiraterone and enzalutamide, as mentioned earlier.[35] However, Dr. Carmel Pezaro confirms that if patients experience severe adverse events (such as toxicity) with either drug, they can switch to the other agent without it being considered as disease progression at that time. Such switching typically occurs within three months of treatment initiation, while switching after three months may indicate disease

progression.

Prostate cancer treatment has shifted, now integrating additional systematic treatments alongside ADT at earlier stages before developing metastasis or castration-resistance. Notably, darolutamide[56] and apalutamide[57] have been approved for high-risk non-metastatic castration-resistant prostate cancer (nmCRPC) patients, available through the CDF since 2021 and 2022[58], respectively. For newly diagnosed metastatic hormone-sensitive prostate cancer (mHSPC), the treatment options have expanded to include docetaxel[59], enzalutamide (available through CDF since 2021)[60], apalutamide (available through CDF since 2022)[61], and darolutamide (available through CDF since 2023 for patients ineligible for chemotherapy with docetaxel).[58,62] Key clinical trials comparing ADT alone to ADT combined with docetaxel in treating mHSPC were conducted from 2004 to 2013, including the GETUG-AFU15 (France)[63,64], the CHAARTED (US)[65], and the STAMPEDE (UK).[66,67] The results of these studies, published between 2013 and 2019, were inconsistent regarding the benefits of upfront docetaxel and the specific mHSPC patient subgroups that might benefit from it, leading to varied adoption timelines across different medical practices for the use of docetaxel in treating mHSPC.

Reflecting on these recent treatment advancements, our oncologist Dr. Pezaro noted that patients now receiving abiraterone or enzalutamide as first-line treatments for mCRPC might have previously undergone other treatments like docetaxel when their cancer was castration-sensitive. Dr. Pezaro suggested that this trend in treatment strategy likely became more apparent after 2016-2017, following publication of UK-based STAMPEDE trial results. The widespread use of medications in England depends not only on the UK Medicines and Healthcare Products Regulatory Agency (MHRA) approval but also on recommendations from NICE. Consequently, treatments other than docetaxel as upfront treatment prior to the development of mCRPC did not become widely adopted until after 2021. In rare cases, docetaxel may be re-administered upon disease relapse while receiving abiraterone or enzalutamide (i.e. docetaxel (castration-sensitive prostate cancer) → abiraterone/enzalutamide (mCRPC) → docetaxel (mCRPC)).

Identifying castration-resistant patients in NCRAS data can be challenging due to the absence of a variable documenting the date of patients becoming castration-resistant. Dr. Pezaro suggested that the emergency use of certain drugs in England upon the onset of castration resistance could serve as a potential proxy indicator, although it is uncommon. These drugs include non-standard treatment options like maximal ADT[68], which involves adding an additional androgen receptor (e.g., bicalutamide) to the standard ADT (e.g., leuprorelin, cetrorelix). Another option is adding low-dose dexamethasone. Professor Derek Rosario seconded these statements but noted the uncertainty surrounding the complete capture of emergency drug usage in NCRAS data. Inclusion of such information, even if available, may result in the identification of a distinct group of patients, given

the variation in treatment preferences among different physicians without a defined standard practice.

Given this protocol was developed after the onset of Covid-19 pandemic, it is important to acknowledge certain limitations associated with utilising retrospective data from the relevant period. Professor Rosario noted that docetaxel was not recommended for prostate cancer patients during the Covid-19 outbreak since around mid-2020, while abiraterone emerged as a preferred treatment option. This shift implies that treatment patterns during the pandemic might differ from other periods. Therefore, sensitivity analyses may be needed to investigate the adequacy of overlaps between patients receiving comparator treatment sequences, factoring the impact of including patients from different periods. This is crucial because the propensity of a patient receiving a certain treatment sequence may be influenced not only by their personal characteristics but also by the changing nature of treatment paradigms over time.

In summary, directly replicating the benchmark trial GUTG-001 using English NCRAS data is unfeasible due to the specific treatment patterns of mPC patients in England. Particularly, GUTG-001 investigated the effects of sequential treatments with abiraterone followed by enzalutamide, and the reverse sequence, but such sequential use is not permitted in England.

US (Flatiron population)

In the US, the treatment options for prostate cancer are generally similar to those in England, with one significant difference being the ability to use abiraterone and enzalutamide in a sequential manner. A publication based on Flatiron data indicated that the primary treatment sequences for prostate cancer in the US during 2013-2017 were abiraterone followed by enzalutamide or the reverse sequence.[36] In contrast to England, the US FDA granted approval for abiraterone and enzalutamide in castration-sensitive prostate cancer treatments, in 2018 and 2019, respectively[69,70], with adoption influenced by the US National Comprehensive Cancer Network (NCCN) guidelines[71], local practices, and individual insurance coverage.

Flatiron data experts were consulted during the data application process. Their preliminary analysis indicated that the Flatiron database could capture approximately 600 mCRPC patients who were treated with abiraterone followed by enzalutamide, as well as around 400 mCRPC patients who received enzalutamide followed by abiraterone. As of the data cut-off on March 31, 2019, a total of 4,000 metastatic prostate cancer patients were identified as having received a first-line treatment, and among them, 1,700 patients had undergone a second-line therapy. These statistics suggest the potential feasibility of replicating the GUTG-001 trial with Flatiron data.

4.3.3 Target Trial Emulation (TTE)

Our planned TTE analyses in prostate cancer involves a series of interconnected and progressive components. The specifics of these steps are outlined in Figure 2 and further explained in the

following paragraphs.

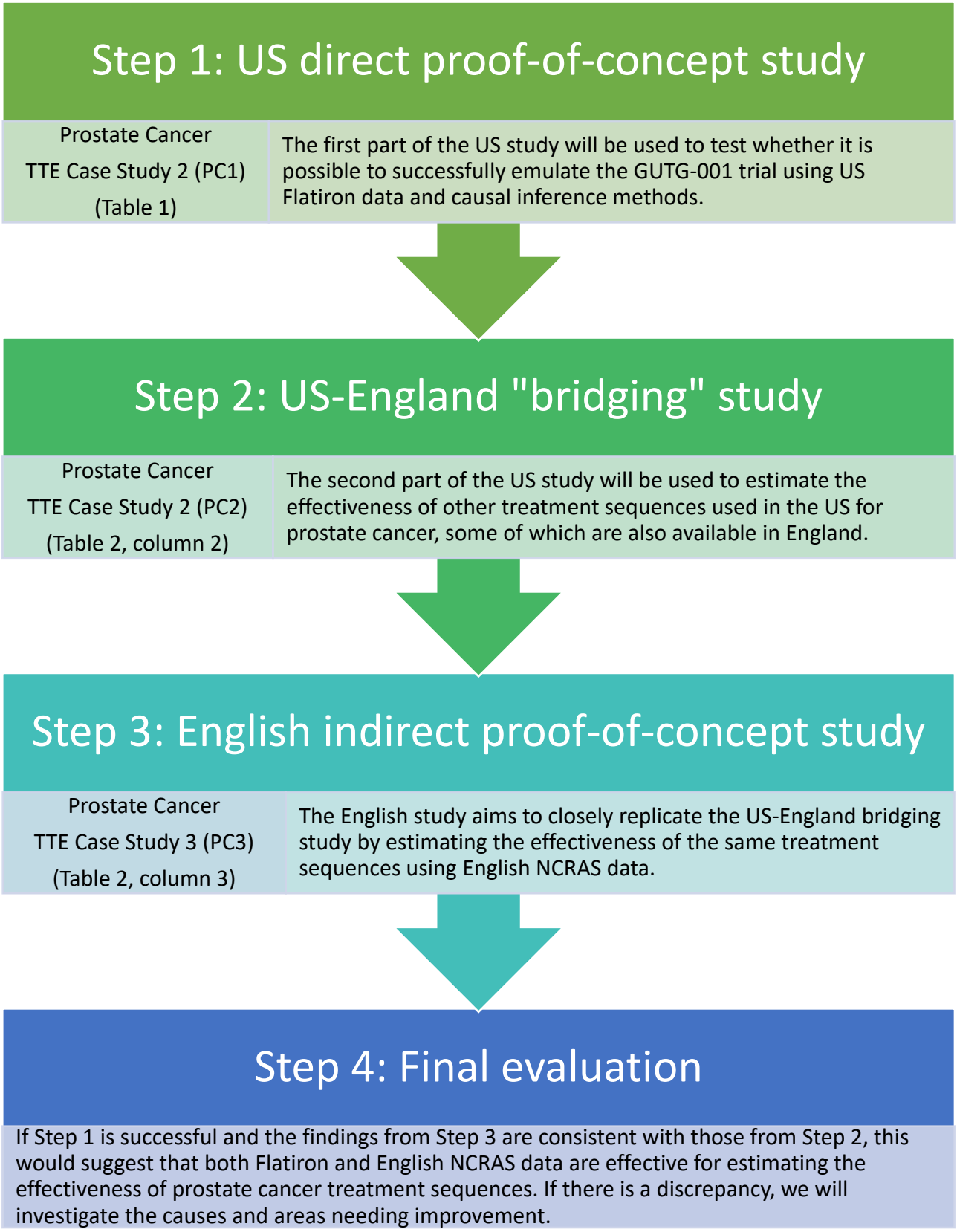

Figure 2. Flow chart of the US-England coupled sequential treatment Target Trial emulation studies in prostate cancer

National Cancer Registration and Analysis Service (NCRAS); PC: prostate cancer; TTE: Target Trial Emulation; US, United States

4.3.3.1 Prostate cancer TTE 1 (PC1)

The first step involves conducting a TTE (PC1), aiming to assess the feasibility of emulating the GUTG-001 trial (i.e., GUTG-001 Analogue) with US Flatiron data, leveraging causal inference methods. This serves as a direct critical proof-of-concept study to identify appropriate statistical methods for obtaining reliable effectiveness estimates of treatment sequences from RWD. Success in replicating the GUTG-001 results using US Flatiron data will imply the potential for the same methods to be applied in comparable scenarios, including comparisons of other treatment sequences across broader populations. Table 1 outlines the specifics of the PC1 TTE analyses.

4.3.3.2 Prostate cancer TTE 2 (PC2)

If the benchmarking in the US proof-of-concept study (PC1) prove successful, Step 2 (PC2) will expand these methods for TTE in broader populations, focusing on comparing alternative treatment sequences prevalent in both the US and England using Flatiron data. These first- and second-line treatment sequences for treating treatment-naïve mCPRC include:

- docetaxel followed by enzalutamide upon disease progression
- docetaxel followed by abiraterone upon disease progression
- docetaxel followed by cabazitxel upon disease progression
- enzalutamide followed by docetaxel upon disease progression
- abiraterone followed by docetaxel upon disease progression

Subsequently, results from PC2 could serve as an "emulated benchmark trial" (i.e., a bridge), providing a basis for comparison with the Target Trial analyses in PC3 (Section 4.3.3.3). Such comparisons aim to indirectly validate the applicability of the same methods for comparing treatment sequences in the NCRAS database. In addition to functioning as an emulated benchmark, PC2 also aims to examine how the estimates of PC1 will change when the restrictions on the patient population are relaxed. PC2's detailed Target Trial design is presented in Table 2, column 2.

Dr. Pezaro notes that the exact date of developing castration-resistance may be unavailable in the database, particularly the NCRAS data, given its nature as a disease registry. If NCRAS data lacks specific timing for mCRPC diagnosis an alternative could involve including all newly diagnosed metastatic patients, irrespective of their hormone status. In such cases, treatment sequences for metastatic prostate cancer (mPC) patients could be relevant, which include:

- docetaxel (plus ADT) followed by enzalutamide upon disease progression
- docetaxel (plus ADT) followed by abiraterone upon disease progression
- ADT alone followed by abiraterone upon disease progression
- ADT alone followed by enzalutamide upon disease progression
- ADT alone followed by docetaxel upon disease progression

Table 1. Prostate cancer case study 1: Target Trial using US Flatiron Data to replicate the GUTG-001 Trial

	Original benchmark RCT (GUTG-001)[22]	PC1: GUTG-001 Analogue Target Trial
Purpose of the Target Trial	N/A	A direct proof-of-concept study: replicating the GUTG-001 trial using Flatiron data
Eligibility criteria	The eligibility criteria presented here are an abridged version from Khalat et al.'s 2019 publication, with further details in GUTG-001's protocol (Version 6.0) on ClinicalTrials.gov.[72] Patients who were aged 18 years or older and had newly diagnosed prostate adenocarcinoma without evidence of neuroendocrine differentiation, with metastatic disease on CT scan, MRI, or bone scan, and a rising PSA (PSA progression per PCWG2 criteria) with castrate concentrations of testosterone (≤ 1·7 nmol/L) with ongoing medical castration or previous bilateral orchiectomy. Patients were required to receive LHRH agonist or antagonist therapy for the duration of study treatment if not surgically castrated. Eligible patients were required to have adequate organ function, defined as absolute neutrophil count $1{\cdot}5\times10^9$ cells/L or higher, platelet count 100×10^9/L or higher, haemoglobin 80 g/L or higher, creatinine clearance 30 mL/min or higher, serum potassium higher than lower limit of normal range, total bilirubin 1.5 times upper limit of normal or less, and alanine aminotransferase and aspartate aminotransferase five times upper limit of normal or less. Patients who were previously treated with any CYP17A1 inhibitors (e.g. abiraterone, enzalutamide or experimental androgen receptor inhibitors) were excluded, while previous use of docetaxel for castration-sensitive disease was allowed. Patients who had contraindications to abiraterone and enzalutamide were excluded per manufacturer's label. Other exclusion criteria were ECOG performance status more than 2, brain metastases, active epidural disease, severe concurrent illness or comorbid disease, active concurrent malignancy, history of seizures or cerebrovascular events, major surgery within 4 weeks of starting study treatment, gastrointestinal disorders affecting absorption, and life expectancy of less than 6 months. The presence of visceral metastasis and pain requiring opioid analgesia were allowed.	Matching the eligibility criteria of the GUTG-001 trial as far as possible, following GUTG-001's protocol version 6.0.[72]

Treatment strategies	• Group A: patients received abiraterone 1000 mg orally once daily plus prednisone 5 mg orally twice daily as first study treatment until confirmed PSA progression, wide-field radiotherapy of symptomatic bone metastases, unacceptable treatment-related toxicity or withdrawal of consent. They then crossed over to receive enzalutamide 160 mg orally once daily until symptomatic or clinical progression, unacceptable treatment-related toxicity, or withdrawal of consent. • Group B: patients received enzalutamide and abiraterone plus prednisone in a reverse sequence until confirmed PSA progression, wide-field radiotherapy of symptomatic bone metastases, unacceptable treatment-related toxicity, or withdrawal of consent. Patients then crossed over to receive enzalutamide 160 mg orally once daily until symptomatic or clinical progression, unacceptable treatment-related toxicity, or withdrawal of consent. Dose modification for treatment-related adverse events was allowed.	• Group A: patients receiving abiraterone plus prednisolone followed by enzalutamide • Group B: patients receiving enzalutamide followed by abiraterone plus prednisolone For all treatment sequences, patients may switch to second-line treatment in cases of disease relapse* or unacceptable treatment-related toxicity. Additionally, patients may discontinue their first-line treatment without proceeding to subsequent treatment, based on clinical/patient decisions.
Assignment procedures	Eligible patients were randomly assigned (1:1) to receive abiraterone + prednisolone followed by enzalutamide or the reverse. Investigators and participants were not masked to treatment assignment.	Same as in GUTG-001 To effectively emulate the randomisation, we need to adjust for all measurable confounding factors to ensure the comparability of two treatment arms (counterfactual) at baseline (i.e., screening visit prior to randomisation). To align with GUTG-001, we plan to use the initiation date of first-line treatment as the reference point for assessing patients' baseline characteristics (i.e., time zero). This aligns with GUTG-001's tracking of time-to-event outcomes from first-line treatment commencement, which was within five days post-randomisation. The randomisation emulation will be performed using inverse probability weighting or other g-methods (e.g. standardisation). Important prognostic factors will be used to derive propensity score using a multivariable regression model. These important prognostic factors include age, tumour status, ECOG performance status, prior treatments, comorbidities, and PSA level. The final covariate selection will be based upon discussion with

		clinicians, and will be based upon attempting to satisfy the "no unmeasured confounding" assumption.
Follow-up period	Patients were followed up since the initiation of their first-line treatment, which began within five days of randomisation, until either death, data-cut off or lost-to-follow-up, whichever occurred first. The median duration of follow-up in the GUTG-001 trial was 30.7 months (IQR 25.1-36.2) of the data cut-off (May 31, 2018). Given that the final enrolment in the GUTG-001 trial occurred on December 13, 2016, the minimum follow-up period would have been approximately 17 months, had no patients been lost to follow-up.	The follow up begins with the initiation of the first-line therapy until the occurrence of death, loss to follow-up, or data cut-off, whichever occurs first. Our analysis will target patients who could have a theoretical minimum follow-up of 17 months, matching the GUTG-001 trial's follow-up duration as closely as possible. For example, for a data cut-off date of May 31, 2018, we will include all patients who were eligible for enrolment before the end of 2016, regardless of their actual follow-up period. Patients with less than 17 months of actual follow-up will be marked as lost to follow-up (censored), ensuring alignment with the GUTG-001 trial without introducing selection bias. The criterion of 17 months may be relaxed if the sample size is insufficient.
Outcomes	<u>Primary endpoints:</u> 1. Time to second PSA progression: time from the start of first-line therapy to PSA progression on second-line therapy, or death from prostate cancer before crossover, whichever occurred first. 2. The proportion of patients with PSA response on second-line therapy. <u>Secondary endpoints:</u> 1. Time to PSA progression on first-line therapy: time from the start of first-line therapy to confirmed PSA progression on first-line therapy (Preliminary results of this endpoint were reported in Annala et al. 2019[73]) 2. Time to PSA progression on second-line therapy: time from crossover to confirmed PSA progression 3. Overall survival: time from the start of first-line therapy to time of death from any cause, or last follow-up (censored); 4. Time on treatment for second-line therapy: time from crossover to end of second-line treatment or death 5. Time to clinical progression on second-line therapy: time from crossover	<u>Primary endpoints:</u> 1. OS, measured as the time from the start of first-line therapy until death from any cause, or last follow-up (censored). <u>Secondary endpoints:</u> 1. Time to second progression, defined as time from the start of first-line therapy to progression* on second-line therapy, or death from prostate cancer before crossover, whichever occurred first. 2. Time to progression on first-line therapy, defined as the time from the start of first-line therapy to any type of progression*, including death from prostate cancer. <u>Exploratory endpoints (</u>PSA-related endpoints are contingent on the availability and quality of PSA levels in the Flatiron database): 1. Time to progression on second-line therapy, defined as the time from crossover to any type of progression*, including death from prostate cancer. Since both the GUTG-001 trial and the TTE are designed to evaluate the effectiveness of treatment sequences from the initial baseline (i.e., the start of first-line treatment, analyses that use the time

to clinical progression on second-line therapy, including death from prostate cancer (This endpoint was not analysed because the endpoint was subject to variability in individual physician decision making of local study investigators).
6. Safety of second-line abiraterone and enzalutamide
7. Change in Montreal Cognitive Assessment score on first-line and second-line therapy (Results of this endpoint was reported elsewhere).[74]
8. Correlation of cell-free DNA biomarkers with PSA response after first-line and second-line treatment.

Post-hoc analysis:

1. Time to progression on first-line therapy: time from treatment initiation to confirmed PSA progression, radiographic progression (PCWG2 criteria), clinical progression, or prostate cancer-related death, whichever occurred first (preliminary results was reported in Annala et al. 2019[73]).
2. Time to progression on second-line therapy: time from crossover to confirmed PSA progression, radiographic progression (PCWG2 criteria), clinical progression, or prostate cancer-related death, whichever occurred first.
3. Time to second progression: time from treatment initiation to confirmed PSA progression, radiographic progression (PCWG2 criteria), clinical progression on second-line therapy, or or prostate cancer-related death, whichever occurred first.
4. Comparison of second-line PSA responses between groups using Pearson's chi-square test
5. Clinical correlates of time to PSA progression and PSA response in patients receiving second-line enzalutamide
6. Comparison of crossover clinical characteristics between groups
7. Sensitivity analysis of time to second PSA progression (primary endpoint), excluding patients with delayed crossovers, > 2 weeks
8. Comparison between groups of time from first progression of any kind to crossover
9. Subgroup analysis to determine whether second-line enzalutamide was better than second-line abiraterone in all patient subgroups

of treatment crossover as a secondary baseline are prone to bias, and such analyses should be adjusted for prognostic characteristics. Adjustments were not made in the published analyses of the GUTG-001 trial and this is not the focus of our analysis. Therefore, we regard this as an exploratory endpoint and will only present naïve exploratory analyses for this endpoint.

2. Time to second PSA progression: time from the start of first-line therapy to PSA progression on second-line therapy, or death from prostate cancer before crossover, whichever occurs first.
3. Time to PSA progression on first-line therapy: time from the start of first-line therapy to PSA progression, including death from prostate cancer.
4. Time to PSA progression on second-line therapy: time from crossover to PSA progression, including death from prostate cancer.
5. The proportion of patients with a PSA response on first-line therapy.
6. The proportion of patients with a PSA response on second-line therapy.

Causal contrasts of interest	All endpoints were analysed using the intention-to-treat principle, with first-line or combined treatment endpoints evaluated in all randomised patients, and second-line treatment endpoints assessed in those who switched treatments.	<u>Main analysis:</u> Analogue of per-protocol effect: estimating the hypothetical effect had all patients adhered to the treatment strategy to which they are assigned in our analyses <u>Exploratory:</u> 1. Analogue of intention-to-treat effect: estimating the effect according to the first-line therapy. 2. Analogue of as-treated effect: estimating the effect restricted to those who received the specific treatments sequences outlined in our analysis. First-line or combined treatment endpoints will be evaluated in all randomised patients, and endpoints for second-line treatments will be evaluated specifically in those patients who crossover.
Analysis plan	• Time-to-event outcomes: KM survival curves and log-rank tests were used. Hazard ratios and 95% CI were estimated from Cox proportional hazard models for PFS, OS and combined PFS, stratified by the MSKCC risk criteria. • Proportion of PSA response: compared between groups using Pearson's chi-square test. • Comparison of crossover clinical characteristics between groups: Continuous-valued characteristics were compared using the rank-sum test, and Boolean characteristics were compared using Fisher's exact test. All Cox regression analyses, associated confidence intervals, and Kaplan-Meier curves were calculated using R (version 3.6.0) with the survival package (version 2.44.1.1). Confidence intervals for PSA response, Pearson's chi-square tests, rank-sum tests and Fisher's exact tests were calculate using Julia (version 1.1.0) with the HypothesisTests package (version 0.8.0).	• Descriptive analyses will be conducted to understand the overall treatment pattern in the data and to estimate the sample size. • T-tests and chi-square tests will be conducted to compare patient characteristics at treatment initiation and at cross-over, and compared with those in the GUTG-001 trial • KM survival curves, survival probability (e.g. median survival)/event incidence, and cox proportional hazard ratios (and/or risk ratios using pooled logistic regression), will be conducted for all time-to-event outcomes. Time-to-event outcome analyses will be performed for the per-protocol analogue, intention-to-treat analogue, and as-treated analogue effect. Marginal structural models with inverse probability weighting and other G-methods will be used to emulate the randomisation process of the Target Trial and account for time-varying confounders.

ALT, alanine aminotransferase; AST, aspartate aminotransferase; CI: confidence interval; CYP17A1, Cytochrome P450 Family 17 Subfamily A Member Enzyme; ECOG, Eastern Cooperative Oncology Group performance status; HTA: health technology assessment; KM, Kaplan-Meier; IQR, interquartile range; LHRH, maintain luteinising hormone-releasing hormone, mCRPC, metastatic castration-resistant prostate cancer; MSKCC, Memorial Sloan Kettering Cancer Center; NCRAS, National Cancer Registration and Analysis Service; PCWG2, Prostate Cancer Working Group 2; PFS, progression-free survival; PSA, prostate-specific antigen test; OS, overall survival; RCT, randomised controlled trials; US, United States

* The final operational definition of disease progression can vary across databases and will be determined through discussions with clinical experts and data experts, due to the uncertainty in data quality. If a specific progression date is unavailable, we might use a composite of variables to obtain a proxy date of disease progression, such as treatment discontinuation.

Table 2. Prostate cancer case study 2 & 3: US-England bridging study and English indirect proof-of-concept study

	PC2: An emulated benchmark using Flatiron data	PC3: PC2 Analogue Target Trial using NCRAS data
Purpose of the Target Trial analysis	• US-England bridging study: comparing prostate cancer treatment sequences common in England using Flatiron data to serve as an emulated benchmark for PC3 • Investigate how estimates from PC1 might differ when patient population restrictions are relaxed.	An indirect proof-of concept study: replicating the "emulated benchmark trial(s)" in PC2 using English NCRAS data.
Eligibility criteria	Analysis Set 1: All patients who were aged 18 years or older with mCRPC will be included. Analysis Set 2: According to Dr. Pezaro, the exact date of a patient becoming castration-resistant prostate cancer may be unknown or only available as a proxy in the database. Should the data quality of NCRAS be inadequate for determining the timing of mCRPC diagnosis, the inclusion of all newly diagnosed metastatic patients who were aged 18 years or older will be considered as an alternative (i.e., including mHSPC patients). For a detailed justification, please refer to the "patient inclusion/exclusion criteria" in Section 4.3.6.	Matching PC2 as far as possible.
Treatment strategies	• Group A: docetaxel followed by enzalutamide if disease relapse* or unacceptable treatment-related toxicity. • Group B: enzalutamide followed by docetaxel Alternative treatment strategies may be chosen for comparison based on the sample size of each treatment sequence in Flatiron data and NCRAS data: • docetaxel followed by abiraterone • abieraterone followed by docetaxel • abieraterone followed by cabazitaxel • abiraterone followed by enzalutamide • enzalutamide followed by abiraterone For Analysis Set 2, we will consider comparison of treatment sequences that either start with a systematic treatment (e.g. docetaxel) or no treatment (i.e., with only baseline ADT).	Same as in PC2

	For all treatment sequences, patients may switch to second-line treatment in cases of disease relapse* or unacceptable treatment-related toxicity. Additionally, patients may discontinue their first-line treatment without proceeding to subsequent treatment, based on clinical/patient decisions.	
Assignment procedures	Participants are randomly assigned to one of two strategies at baseline. To effectively emulate the randomisation, we need to adjust for all measurable confounding factors to ensure the comparability of two treatment arms (counterfactual) at baseline. The randomisation emulation will be performed using propensity score matching, inverse probability weighting or other g-methods (e.g. standardisation). Important prognostic factors will be used to derive propensity score using a multivariable regression model. These important prognostic factors include age, tumour status, ECOG performance status, prior treatments, comorbidities, and PSA level†. The final covariate selection will be based upon discussion with clinicians, and will be based upon attempting to satisfy the "no unmeasured confounding" assumption.	Matching PC2 as far as possible.
Follow-up period	The follow-up period starts from the time of treatment initiation (sensitivity analysis: starts from the time of diagnosis (mCPRC for Analysis Set 1, mPC for Analysis Set 2)) and continues until the event of death, loss to follow-up, or the data cut-off date, whichever comes first. In PC 1, follow-up begins on the date of first-line mCRPC treatment initiation (baseline), aligned with the GUTG-001 trial, implying that all patients included in the trial have survived to receive their first-line mCRPC treatment. However, this could potentially lead to immortal time bias in RWD analysis if the time from mCPRC diagnosis to first-line treatment initiation differs significantly between the two study groups (which might not be fully adjustable with statistical methods). While this concern might be less significant in PC1, as abiraterone and enzalutamide are often interchangeable according to our oncology experts, we would like to examine such design's impact on estimating the comparative treatment effectiveness from RWD in the sensitivity analyses for PC2. Specifically, we will begin follow-up at diagnosis instead, and use techniques like cloning[75,76] and/or TTE with sequential eligibility criteria[77] to assist in assigning treatment groups, contrasting it with setting the first-line treatment initiation as baseline (time zero).	Matching PC2 as far as possible. However, it is crucial to acknowledge that for sensitive analysis, the dates of castration-resistance and metastasis might be unavailable in NCRAS data.

Outcomes	Primary endpoints: 1. Overall survival: the time from treatment initiation (for primary analysis)/diagnosis (for sensitivity analysis) until death from any cause, or last follow-up (censored). Secondary endpoints 1. Time to second progression: time from diagnosis to progression* on second-line therapy, or death from prostate cancer before crossover, whichever occurred first. 2. Time to progression on a first-line therapy, defined as the time from diagnosis to any type of progression*, including death from prostate cancer. Exploratory endpoints: 1. Time to progression on second-line therapy, defined as the time from crossover to any type of progression*, including death from prostate cancer. Since the GUTG-001 trial and the TTE are designed to evaluate the effectiveness of treatment sequences from the initial baseline (i.e., the start of first-line treatment, analyses that use the time of treatment crossover as a secondary baseline are prone to bias, and such analyses should be adjusted for prognostic characteristics. Adjustments were not made in the published analyses of the GUTG-001 trial and this is not the focus of our analysis. Therefore, we regard this as an exploratory endpoint and will only present naïve exploratory analyses for this endpoint.	Matching PC2 as far as possible.
Causal contrasts of interest	Main analysis: Analogue of per-protocol effect: estimating the hypothetical effect had all patients adhered to the treatment strategy to which they are assigned in our analyses Exploratory: 1. Analogue of intention-to-treat effect: estimating the effect according to the first-line therapy. 2. Analogue of as-treated effect: estimating the effect restricted to those who received the specific treatments sequences outlined in our analysis. First-line or combined treatment endpoints will be evaluated in all randomised patients, and endpoints for second-line treatments will be evaluated specifically in those patients who crossover.	Matching PC2 as far as possible.

Analysis plan	• Descriptive analyses will be conducted to understand the overall treatment pattern in the data and to estimate the sample size. • T-tests and chi-square tests will be conducted to compare patient characteristics at treatment initiation and at cross-over. • KM survival curves, survival probability (e.g. median survival)/event incidence, and cox proportional hazard ratios (and/or risk ratios using pooled logistic regression), will be conducted for all time-to-event outcomes. Time-to-event outcome analyses will be performed for the intention-to-treat analogue, as-treated analogue, and per-protocol analogue effect. Marginal structural models using inverse probability weight and other G-methods will be used to emulate the randomisation process of the Target Trial and account for time-varying confounders.	Matching PC2 as far as possible.

ECOG, Eastern Cooperative Oncology Group performance status; HTA: health technology assessment; KM, Kaplan-Meier; mCRPC, metastatic castration-resistant prostate cancer; mHSPC, metastatic hormone-sensitive prostate cancer; mPC: metastatic prostate cancer; NCRAS, National Cancer Registration and Analysis Service; PSA, prostate-specific antigen test; RCT, randomised controlled trials; US, United States
N/A: not applicable

* The final operational definition of disease progression can vary across databases and will be determined through discussions with clinical experts and data experts, due to the uncertainty in data quality. If a specific progression date is unavailable, we might use a composite of variables to obtain a proxy date of disease progression, such as treatment discontinuation.
† PSA level is unavailable in the NCRAS data, and therefore will not be included in the English NCRAS analysis.

4.3.3.3 Prostate cancer TTE 3 (PC3)

Step 3 (PC3): The design of PC2 allows us to conduct an indirect proof-of-concept study assessing if the quality of NCRAS data is adequate for reliably assessing the effectiveness of treatment sequences using causal inference methods. PC3 aims to emulate the same Target Trial as PC2, but using English NCRAS data. Differing from the PC1, PC3 will only indirectly reference information from the original benchmark RCT (i.e., GUTG-001 trial). PC3's detailed Target Trial design is presented in Table 2, column 3.

Comparison between similar analyses in NCRAS data (PC3) and Flatiron data (PC2) relies on the existence of common treatment sequences in both datasets. There are several common treatment sequences in treating prostate cancer in the US and the UK as described in 4.3.3.2. Docetaxel followed by enzalutamide and its reverse sequence represent one of the most prevalent treatment sequences available in both the US and the UK,[35,36] and therefore these sequences are chosen as the main comparators. However, alternative sequences like docetaxel-abiraterone (and vice versa) or other non-symmetric pairs remain viable (e.g. abiraterone-docetaxel versus enzalutamide-docetaxel, abiraterone-docetaxel versus abiraterone-enzalutamide, enzalutamide-docetaxel versus enzautamide-abiraterone), subject to the sample sizes in both the Flatiron and the NCRAS data.

Additionally, Professor Rosario highlighted the possibility of assuming no systematic difference between abiraterone and enzalutamide, given their similar mechanisms of action. Consequently, for PC2 and PC3, we can also consider including the following two sequences to increase the sample size if necessary:

- docetaxel followed by abiraterone or enzalutamide upon disease progression
- abiraterone or enzalutamide followed by docetaxel upon disease progression

The final selection of sequence pairs for PC2 & PC3 will be based on ensuring larger sample sizes than in the GUTG-001 trial for both Flatiron and English NCRAS data (guided by analyses of treatment patterns analyses from the actual data), and finalised after discussions with clinicians regarding their clinical interests. In the exploratory analysis, sequence pairs with varying rates of crossover/treatment regimen violations may be compared. This may be used to assess the performance of methods potentially sensitive to a small percentage of patients adhering to the assigned treatment strategy.

4.3.3.4 Final evaluation

If Step 1 is successful and the findings from Step 3 are consistent with those from Step 2, this would suggest that both Flatiron and English NCRAS data are effective for estimating the effectiveness of prostate cancer treatment sequences. If there is a discrepancy, we will investigate the causes and areas needing improvement. We intend to assess the extent of discrepancy using the matrix

outlined in Section 4.5

4.3.4 Patient inclusion/exclusion criteria

The study targets patients aged 18 and over with mCRPC. Depending on the quality of Flatiron and NCRAS data, as well as clinical implications, this focus may broaden to include all metastatic prostate cancer (mPC) or all advanced prostate cancer cases. Specifically, Prof Rosario and Dr. Pezaro are concerned about the databases' ability to accurately identify when patients become castration-resistant, a status typically determined by prostate-specific antigen (PSA) levels. For the English population, this information is unavailable in English NCRAS data and currently only accessible in CPRD data.[78,79] Additionally, as docetaxel is increasingly used in mHSPC patients prior to castration-resistance, distinguishing between patients who received docetaxel post-castration resistance and those treated while hormone-sensitive in the databases poses a significant challenge.

The 2020 UK National Prostate Cancer Audit Annual Report indicates an increase in the use of docetaxel with standard ADT for new metastatic prostate cancer cases, from 27% in 2019 to 36% in 2020.[80] This usage varied widely across different NHS providers in England, ranging from 0% to 47%.[80] Nevertheless, it may still be possible to capture mCRPC patients among all metastatic patients using NCRAS data if we have patients' full treatment trajectory and time-varying prognostic factors not only limited to those upon the emergence of metastases. A crucial part of the study process is to investigate whether a proxy of patient's disease status (i.e. castration-resistant) can be defined by specifying an algorithm when a direct variable is lacking. This rationale supports our requests for specific data periods and relevant variable information outlined in section 5 Data Requirements.

In summary, identifying mCRPC patients in RWD, especially within the English NCRAS, may be challenging. We aim to collaborate with clinicians to identify these patients using a combination of variables if necessary. If we cannot develop an ideal algorithm for this purpose, we will relax the restriction regarding mCRPC patients. Specifically, we will analyse newly diagnosed metastatic patients regardless of their castration-status. Most importantly, the same criteria will be applied to both PC2 & PC3 to enable a fair comparison between these "bridging-studies". Although the study population may slightly differ from what was initially planned, it remains clinically relevant. Notably, Professor Rosario highlighted a shift in the treatment approach for metastatic prostate cancer, increasingly favouring the early use of systematic anti-cancer therapy alongside ADT, even before patients develop castration-resistance.

4.3.5 Sample size estimation

The Flatiron dataset is expected to provide a sufficient sample size for our case studies (PC1 & PC2), as indicated by a US treatment pattern study from 2013-2017. This study showed 227 patients receiving docetaxel followed by enzalutamide (including docetaxel-only patients) and 414 receiving

enzalutamide followed by docetaxel (including enzalutamide-only patients).[36] Furthermore, preliminary data evaluations by Flatiron specialists have confirmed the number of prostate cancer patients receiving relevant treatments appears to be sufficient for our study (i.e., preferably exceeding the sizes in the GUTG-001 trial[22] and previous observational studies of prostate cancer treatment sequences, aiming for at least 100-250 patients[38,81]).

On the other hand, while precise estimation of patients receiving docetaxel and/or abiraterone/enzalutamide annually in NCRAS data is challenging, it is likely to be sufficient. This assessment is based on the published UK epidemiology data described Section 4.3.2.[44] Specifically, according to a recent NICE TA, there are approximately 5,500-5,800 CRPC patients who may be eligible for a first-line treatment in England and Wales every year.[49] Further, abiraterone, enzalutamide, and docetaxel are among the most frequently administered therapies.[48-51]

4.4 Sequential treatment Target Trial: renal cell carcinoma case studies

4.4.1 Benchmark RCT

RECORD-3, an international phase-2b trial, has been identified as a valuable benchmark in evaluating treatment sequences for metastatic treatment-naïve metastatic RCC (mRCC) patients. This trial compares the efficacy of the two following treatment sequences in treating metastatic treatment-naïve mRCC patients: [19,25]

- everolimus as first-line therapy followed by second-line sunitinib (n = 238)
- sunitinib as first-line therapy followed by second-line everolimus (n = 233)

Despite being a potential benchmark, only one of the treatment sequences in RECORD-3 has been available in the NHS. Everolimus was never recommended as first-line therapy and sunitinib was not recommended as second-line therapy in the NHS.[82] This statement has been confirmed with our medical oncologist, Professor Janet Brown and further supported by recent observational data from three UK hospitals.[83] Everolimus accounts for only 0.6% of all treatments among 652 mRCC patients in 2008-2015.[83] Further, only 0.5% of all patients who ever received a second-line treatment received sunitinib as a second-line therapy.[83] On the contrary, sunitinib accounts for 60.7% of first-line therapies, and everolimus accounts for 41.9% of second-line therapies. However, it is likely that the percentage of patients using sunitinib followed by everolimus as first- and second-line therapy has been decreasing because other newer agents for treating mRCC have been available since 2015. Particularly, everolimus has gradually been shifted to be used as a later-line therapy (e.g. 3rd or 4th line) in recent years, while sunitinib remains as a common first-line therapy.[84] Despite these changes, historical data from the NCRAS, especially for the RCC incident cohort in 2015 and 2016 (at the beginning of acquisition of newer treatments in the NHS), may still provide insights into the sequence of sunitinib followed by everolimus.

Similar to the prostate cancer study series (Section 4.4), there is no "perfect" benchmark trial for the RCC. However, at least one of the treatment sequences in RECORD-3 is likely to exist in the NCRAS data. Thus, RECORD-3 remains a valuable reference, enabling a direct comparison between the benchmark trial results and the Target Trial analysis conducted using NCRAS data. In summary, the primary aim of this RCC case study is to explore the feasibility and reliability of using NCRAS data to replicate results of a single arm in the RECORD-3 trial (i.e. sunitinib → everolimus).

4.4.2 Demographics of renal cell carcinoma (RCC) in England

In England, approximately 13,000 new kidney cancer cases are reported annually (2015-2017)[85], with more than 80-90% being RCC.[44] Approximately one-fourth of these patients present with advanced-stage cancer and 25-34% have metastases at diagnosis.[85,86] About 75% of advanced-stage RCC patients are eligible for a first-line systematic therapy.[87]

Current treatment for mRCC primarily includes targeted therapy (e.g., tyrosine kinase inhibitors, TKIs) and immunotherapy (e.g., PD-1 and PD-L1 inhibitors).[83,88] In the NHS, sunitinib and pazopanib have been first-line therapies for mRCC since 2009 and 2011, respectively.[89] Tivozanib and cabozantinib (both through CDF) were introduced as additional first-line treatment options for mRCC starting in 2018.[58] Subsequently, more treatments became exclusively available as first-line therapies for mRCC through the CDF[58], including nivolumab plus ipilimumab (since 2019 and further passed NICE CDF review in 2022), avelumab-axitinib combination (since 2020), lenvatinib-pembrolizumab combination (since 2023). The NHS has adopted the following treatments for mRCC patients who are in need of second-line or later-line therapies: axitinib[90] (from 2015), nivolumab[87] (from 2016), everolimus (from 2017 through CDF, previously through CDF for more restricted indication), cabozantinib (from 2017 through CDF), and the lenvatinib-everolimus combination (from 2018 through CDF).[58]

Professor Janet Brown stated that in England, before targeted therapy and immunotherapy became widespread, approximately 10-15% of patients with less advanced cancer were treated with interferon-alpha before receiving any other systemic anti-cancer therapies. Additionally, everolimus has been shifted from second-line treatment to later stages of treatment (third or fourth-line treatment) in recent years for patients with an Eastern Cooperative Oncology Group (ECOG) score below 2. This adjustment reflects the lack of substantial survival benefits of everolimus and instead emphasises its role in improving patients' quality-of-life.

Professor Brown clarified that there is no standardised adjuvant therapy for mRCC patients, indicating that systematic therapy may or may not accompany surgical interventions before metastases appear.[91] She added that patients who are suitable for operations like nephrectomy, and those who do not rapidly develop metastases post-nephrectomy, usually have a more favourable

prognosis. Specifically, these patients are typically deemed healthy enough to undergo surgical interventions at the first place. Conversely, elevated calcium levels in patients typically indicates a poorer prognosis.

Professor Brown further suggested that patients with mRCC who remain in good health conditions (i.e. desirable performance status) may receive up to 5 or 6 lines of treatment with each subsequent relapse. However, a significant proportion (roughly 40-50%) of mRCC patients receive only three lines of treatment before death. This can be attributed to either their frailty preventing further treatment upon disease progression or mortality occurring before subsequent treatments. The introduction of newer treatment options, particularly cancer immunotherapies since 2016, has significantly improved the overall survival (OS) of mRCC patients.[92] Previously, patients who received solely TKIs had an OS range of 1-2 years up to 4-5 years after being diagnosed.

4.4.3 Target Trial Emulation

This section presents the design of a single-arm Target Trial to assess the viability of using NCRAS data to replicate the sunitinib to everolimus sequence results from the RECORD-3 trial. Table 3 specifies these details.

In routine clinical practice, unlike controlled sequential treatment trials, patients may receive various second-line treatments based on the outcomes of their first-line therapy. Additionally, some patients might not receive any second-line treatment, and the reasons for this can differ from those seen in clinical trials. For instance, in the NCRAS data, patients initially treated with sunitinib might have different second-line treatments, though everolimus could have been an option. In our emulation of the RECORD-3 trial's sunitinib to everolimus arm, solely analysing NCRAS data patients who completed this treatment sequence could be problematic. Such an approach implies selecting patients based on a non-random future decision: the progression to second-line everolimus is contingent on the outcome of the first-line sunitinib treatment, rather than a predetermined treatment plan. Selection conditioning on a post-treatment variable (i.e., as-treated effect) could lead to immortal-time bias. Nevertheless, such an approach is not uncommon in observational studies and one should be cautious in interpreting the results of these analyses.[93,94]

In our RCC study, we aim to highlight the strengths of causal inference in estimating treatment sequence effectiveness in a single-arm Target Trial by contrasting the following three effects:

- As-treated effect: include only patients who received sunitinib as their first-line treatment and proceeded to everolimus as their second-line therapy for time-to-event outcome assessment
- Standard per-protocol effect: include all patients who received sunitinib as their first-line therapy for time-to-event outcome assessment and censor those who did not proceed to

everolimus as their second-line therapy by the time of treatment-switching

- Hypothetical per-protocol effect assuming complete adherence to treatment assignment: include all patients who received sunitinib as a first-line therapy for time-to-event outcome assessment and adjust for treatment-switching with causal inference methods if they did not receive everolimus as a second-line therapy

While the as-treated effect clearly faces the risk of immortal time bias, the standard per-protocol effect could encounter issues if censoring is informative. On the other hand, the hypothetical per-protocol effect, assuming complete adherence to the assigned treatment, somewhat mirrors the approach used in RCTs to address treatment-switching.[95] This involves adjustments for unintended switches, such as when patients initially assigned to standard therapy subsequently move to a new drug. Implementing this approach could help mitigate biases in estimating the effects of treatment sequences. While there are similarities in the causal inference methods applicable, our single-arm Target Trial analysis, which uses RWD, presents additional complexities. Specifically, it lacks a randomisation baseline for reference and patients in real-world settings often switch between multiple drugs. Part of our study involves assessing if varying methods for the hypothetical per-protocol effect yield discrepancies, and how to interpret these differences.

The challenges in replicating the RECORD-3 trial single-arm results could stem from several factors. A primary concern is confounding by subsequent treatments beyond second-line. About half of patients in the RCORD-3 trial received additional treatments after the second-line, but detailed information on these is unavailable. It may be possible that the options of subsequent treatments are different or the rate of receiving a subsequent treatment is incomparable to the clinical practice in England. Another potential issue is the difference in patient demographics, including age and adherence to treatment, despite matching the same inclusion/exclusion criteria of the trial as far as possible. If feasible within our project's limited timeline, we might consider undertaking exploratory analyses to see whether additional causal inference methods can aid in resolving these issues. These methods may include adjustments for variations in treatment exposure due to non-adherence (e.g. delay of treatment), or methods of extending inferences and formulating external control arms using RWD (i.e., emulating the sunitinib → everolimus arm as an external control for the trial's everolimus → sunitinib arm).[96,97] However,

4.4.4 Patient inclusion/exclusion criteria

The RCC case study mainly focuses on patients with mRCC. In this study, Target Trial Analysis Set 1 strives to closely match the eligibility criteria of the RECORD-3 trial, essentially creating a RECORD-3 trial-mimicking population (Table 3: Eligibility criteria). Conversely, the Target Trial Analysis Set 2 aims to extrapolate the findings in a more generalised population, including all mRCC

patients aged 18 or older.

4.4.5 Sample size estimation

In the UK, it is estimated that each year, 2,500 to 3,000 new cases of mRCC are diagnosed. Approximately three-quarters of these patients (est.n = 1,875-2,250), might be eligible for first-line therapy. According to a recent UK study, prior to 2015, sunitinib may account for roughly 61% of first-line therapies before 2015.[83] This suggests that around 500 to 600 patients in the NCRAS data (2012-2018) could be eligible for our study. Given these figures, pooling mRCC incident cohorts from 2012 to 2018 (or later) should provide us with a sufficiently large sample size for the single arm Target Trial analyses.

Table 3. A single-arm Target Trial replicating sunitinib-everolimus arm in the RECORD-3 trial using NCRAS data

	Benchmark RCT: RECORD-3 trial[19,25]	A single-arm Target Trial: Analogue of the RECORD-3 sunitinib-everolimus arm using NCRAS data
Eligibility criteria	The eligibility criteria presented here are an abridged version from Knox et al.'s 2017 publication, with further details in RECORD-3's protocol on ClinicalTrials.gov.[98] Patients aged 18 years or older with measurable mRCC as per RECIST v1.0 were included. Prior nephrectomy was not a prerequisite. Key eligibility criteria included no previous systemic therapy, a KPS score of 70% or higher, adequate hematologic, liver, and kidney function, and a normal left ventricular ejection fraction. Patients with brain metastases were excluded.	Target Trial Analysis Set 1: Matching the eligibility criteria in the RECORD-3 as far as possible. The KPS score information is not available in the NCRAS data; however, it may be converted into an ECOG score.[99] Target Trial Analysis Set 2: All patients who were aged 18 years or older with mRCC will be included.
Treatment strategies	• Group A: first-line everolimus 10 mg/day until PD followed by sunitinib (4 weeks on, 2 weeks off as second-line therapy (n = 238) • Group B: first-line sunitinib 50 mg/day (4 weeks on, 2 weeks off) until PD followed by everolimus as second-line therapy (n = 233) The crossover period is defined as the interval between the end of the first-line treatment and the start of the second. The crossover (initiation of the second-line treatment) should occur within 35 days of disease progression. Patients had a minimum 2-week period after discontinuation of the first-line drug because of progression before beginning the second-line drug. Dose modifications were permitted for adverse events.	Single group: first-line sunitinib until disease progression* followed by everolimus as second-line therapy. Treatment crossover (the initiation of second-line therapy) should occur within 35 days of progression, and patients should have a minimum 2-week period after discontinuation of the first-line drug because of progression before beginning the second-line drug. Dose modifications were permitted for adverse events. Professor Brown noted that patients switching treatments due to toxicity, rather than disease progression (thus not advancing in treatment lines), typically did so within 3-6 months of starting therapy. In such cases, patients often moved to another drug within the same class, like from one TKI to another TKI. For those who started with everolimus before 2015, a reduced dosage sequence (e.g., 50 mg/day → 37.5 mg/day → 25 mg/day) upon disease progression was common due to the lack of alternative later-line therapies at that time.
Assignment procedures	Eligible patients were randomly assigned (1:1) to receive either everolimus followed by sunitinib or the reverse sequence. The random assignment was stratified by MSKCC risk criteria (favorable, intermediate, or poor risk). Patients received the first-line drug until disease progression (according to RECIST v1.0), discontinuation due	All patients will be receiving sunitinib as first-line therapy and followed with a second-line everolimus upon disease progression*. In this single-arm Target Trial, randomisation is not required. Our focus is on replicating the condition where patients were assigned a treatment sequence of

	to unacceptable toxicity, or for any other reason. Upon disease progression, patients were eligible to switch to the second-line drug until further progression.	sunitinib followed by everolimus. Although the RECOTD-3 trial stratified patient recruitment using the MSKCC risk criteria, this information is unavailable in NCRAS data. Professor Brown indicated that ECOG might not be an ideal substitute for MSKCC criteria, and therefore, we will conduct further investigations into alternative variable combinations that could serve as better proxies.
Follow-up period	Patients were followed up since the initiation of their first-line treatment, until either death, data-cut off, lost-to-follow-up, whichever occurred first. Since the RECORD-3 trial's final enrolment was in June 2011 and the data cutoff for the final analysis was in June 2014, the minimum follow-up period would have been approximately 3 years, had no patients been lost to follow-up and all remain alive.	The follow up begins with the initiation of the first-line therapy until the occurrence of death, loss to follow-up, or data cut-off, whichever occurs first. Our analysis will target patients who could have a theoretical minimum follow-up of 3 years, matching the RECORD-3 trial's follow-up duration as closely as possible. For example, for a data cut-off date of December 2021, we will include all patients who were eligible for enrolment before the end of 2018, regardless of their actual follow-up period. Patients with less than 3 years of actual follow-up will be marked as lost to follow-up (censored), ensuring alignment with the RECORD-3 trial without introducing selection bias. The criterion of 3 years may be relaxed if the sample size is insufficient.
Outcome	<u>Primary endpoints:</u> 1. PFS of first-line therapy: time from the first date of first-line treatment to progression during first-line treatment or death from any cause. Patients without progression* or death at data cut-off for the analysis or at the time of receiving additional anticancer therapy, including the second-line drug were censored at their last date of adequate tumour evaluation. <u>Secondary endpoints:</u> 1. Combined first- and second-line PFS: the time from randomisation to progression after second-line treatment or death from any cause. Patients who did not crossover to second-line therapy or who did not experience progression after the start of second-line treatment or who were alive at data cut-off for the analysis or at the time of receiving an additional anticancer therapy, were censored at last date of tumour evaluation.	<u>Primary endpoints:</u> 1. OS: time from the first date of first-line treatment to death <u>Secondary endpoints:</u> 1. Combined first- and second-line PFS: the time from the first date of first-line treatment to progression* after second-line treatment or death from any cause. Patients who did not crossover to second-line therapy or who did not experience progression* after the start of second-line treatment or who were alive at data cut-off for the analysis or at the time of receiving an additional anticancer therapy were censored at their last date of structural activity plus grace period. 2. PFS of first-line therapy: the time from the first date of first-line treatment to progression* during first-line treatment or death from any cause. Patients without progression* or death at data cut-off for the analysis or at the time of receiving additional anticancer therapy, including the second-line drug, were censored at their last date of structural activity plus grace period.

	2. OS: time from randomisation to death (no formal power calculation was made, and the expected number of deaths was 300) <u>Exploratory endpoints:</u> 1. Second-line PFS: time from the start of second-line treatment to progression or death	Professor Janet Brown suggested that the judgement of disease progression in routine practice (i.e. a combination of clinical and radiology progression assessment) may vary between clinicians and may differ from clinical trials, which often involve independent central using RECIST. Patients remain on a treatment or radiology showing < 20% increase may be an indicator of disease remains stable without progression.
Causal contrasts of interest	Per-protocol effect was estimated for first-line PFS, OS and combined PFS. A notable proportion of patients were censored in estimating the combined PFS and OS in RECORD-3, owing to delayed crossover to second-line therapy, with 57% patients in the sunitinib-everolimus arm and 56% patients in the everolimus-sunitinib arm. The crossover, which is the start of receiving second-line treatment, should occur within 35 days of progression. The period between the end of first-line treatment and the beginning of second-line therapy is the crossover period.[19]	<u>Main analysis:</u> 1. Hypothetical per-protocol effect: estimating the hypothetical effect had all patients adhered to the treatment strategy to which they are assigned in our analyses. This approach includes all patients who received sunitinib as a first-line therapy for time-to-event outcome assessment and adjust for treatment-switching with causal inference methods if they did not receive everolimus as a second-line therapy or did not have a timely cross-over. <u>Exploratory:</u> 1. Analogue of intention-to-treat effect: estimating the effect according to the first-line therapy. 2. As-treated effect: estimating the effect according to the actual treatment sequences. This approach only includes patients who received sunitinib as their first-line treatment and proceeded to everolimus as their second-line therapy for time-to-event outcome assessment. 3. Standard per-protocol effect: This approach includes all patients who received sunitinib as their first-line therapy for time-to-event outcome assessment and censor those who did not proceed to everolimus as their second-line therapy by the time of treatment-switching or did not have a timely cross-over without any adjustment for informative censoring
Analysis plan	Kaplan-Meier survival curves and log-rank tests. Hazard ratios and 95% CI were estimated from a Cox proportional hazard models for PFS, OS and combined PFS, stratified by the MSKCC risk criteria.	• Descriptive analyses will be conducted to understand the overall treatment pattern in the data and to estimate the sample size. • Patient characteristics at treatment initiation and at cross-over will be examined and compared with those in the RECORD-3 trial • KM survival curves, survival probability (e.g. median survival)/event incidence, and cox proportional hazard ratios (and/or risk ratios using pooled logistic regression), will be conducted for all time-to-event outcomes.

	Time-to-event outcome analyses will be performed for the as-treated, and standard per-protocol effect, the hypothetical per-protocol effect assuming complete adherence to treatment assignment. Marginal structural models using inverse probability weight and other G-methods will be used to facilitate the emulation of the hypothetical per-protocol effect assuming complete adherence to treatment assignment and account for time-varying confounders.

PD, progressive disease; ECOG, Eastern Cooperative Oncology Group performance status; KM: Kaplan-Meier; KPS, Karnofsky performance status; mRCC, metastatic renal cell carcinoma; MSKCC, Memorial Sloan-Kettering Cancer Center; NCRAS, National Cancer Registration and Analysis Service; OS: overall survival; PFS: progression-free survival; RCT, randomised controlled trials; RECIST, Response Evaluation Criteria in Solid Tumors; TKI, tyrosine kinase inhibitors

* The final operational definition of disease progression can vary across databases and will be determined through discussions with clinical experts and data experts, due to the uncertainty in data quality. If a specific progression date is unavailable, we might use a composite of variables to obtain a proxy date of disease progression, such as treatment discontinuation.
† PSA level is unavailable in the NCRAS data, and therefo%re will not be included in the English NCRAS analysis.

4.5 Benchmark trial-RWE agreement assessment

This section delves into the methods to assess the alignment between benchmark clinical trials and their emulated counterparts, which are crucial for evaluating the extent of and discussing potential reasons for any discrepancies observed. The matrix primarily comprises three components, with the first three assessment criteria adapted from those used in the RCT DUPLICATE studies.[24,34,100]

(1) Regulatory agreement: This component assesses whether the RWE replicates its benchmark's results (such as hazard ratio (HR) and risk ratio (RR)) in terms of both direction and statistical significance observed in the benchmark trials. While RCT DUPLICATE established criteria for evaluating regulatory agreement in both superiority and non-inferiority trials[34], our chosen benchmarks (such as GUTG-001 and RECORD-3) were not intended for regulatory use and hence did not specify any non-inferiority margin. Therefore, our primary focus is to compare RWE and RCT data and treated them as superiority trials, evaluating whether the direction and significance of RWE's estimates matched those of the benchmarks. Endpoints with non-significant effects in RCTs should also show no significant effect in RWE.

(2) Estimate agreement: Considering the potentially disproportionately large sample size of RWE, achieving statistical significance might be easier compared to the benchmarks (and thus not easier to meet the first criteria). Therefore, this step examines whether the point estimate of RWE's effect sizes falls within the 95% confidence intervals (CIs) of the benchmark trial. Furthermore, we added an extra procedure to include the comparison of non-relative effect estimates for time-to-event outcomes. For example, it examines whether the point effect of median survival estimates falls within the 95% confidence interval (CI) of the trial.

(3) Exploratory - standardised differences: The third criterion is exploratory in nature. It involves computing the standardised difference to compare the relative effect estimates from the benchmark and the RWE, to determine whether there is a statistically significant difference in the estimated effects, as indicated in the RCT DUPLICATE study, using the formula below.[34]

$Z = \frac{\hat{\theta}_{RWE} - \hat{\theta}_{RCT}}{\sqrt{\sigma^2{}_{RWE} + \sigma^2{}_{RCT}}}$, where an absolute Z-value less than 1.96 indicates no significant difference between the estimates from RWE and RCT.

(4) Exploratory - survival curve comparison: This additional criterion, specific to our study, involves comparing RWE survival curves with those from the benchmark trial for time-to-event outcomes. The key aspect here is assessing whether the point estimates of the RWE survival curve for each treatment-sequence group fall within the 95% CI of the benchmark trial. Since RCTs typically present Kaplan-Meier (KM) curves, without patient-level data, we aim to reconstruct patient-level survival data from benchmark RCTs using Guyot et al.'s digitisation method for extracting information from published KM curves[101], wherever possible. We introduced this extra criterion

beyond what was included in RCT DUPLICATE, emphasising the importance of verifying whether absolute outcomes (like survival times) in our emulation match those in the RCT, and not solely focusing on relative effect estimates. This is crucial because there is a possibility that, even with similar relative effect estimates, the absolute effect may significantly vary, indicating a less ideal emulation. This holds particular importance in scenarios where RWE is employed to form external control arms, notably in our RCC case study (Section 4.4) which focuses on the emulation of a single arm.

This assessment matrix is specifically designed for comparing the benchmark-RWE pair in the PC1 case study (GUTG-001 versus its emulation using Flatiron data). However, its application to the benchmark-RWE pairs in the PC2, PC3, and RCC case studies can be limited in terms of interpretation. For example, PC2 utilises an emulated benchmark instead of a traditional RCT. Additionally, in PC3's single-arm study, only the fourth criterion and certain aspects of the second criterion, particularly the comparison of non-relative effects, are pertinent.

4.6 Software for analysis

The data will be analysed in R and STATA to conduct Target Trial emulations in different cancer types to explore the feasibility of applying causal inference methods for estimating the effectiveness of different treatment sequences using RWD. All analyses will be documented in R script and Stata. Do files.

5 Data acquisition

5.1 Overview

This section offers an overview of the specification of datasets and variables we originally planned to request from the NCRAS and Flatiron databases for the analyses detailed in Section 4 (Analysis Plan). For transparency, we reported our original specifications and included feedback received from analysts of both databases during the data application process. Our main objective is to gather detailed patient information to identify eligible patients who fit the inclusion/exclusion criteria of our Target Trial case studies. This includes two specific patient groups: 1) individuals with prostate cancer (ICD-10: C61x) in both the Flatiron and NCRAS databases, and 2) those with RCC (ICD-10: C64x) in the NCRAS database. Importantly, we also need longitudinal patient data for measuring time-varying treatments, covariates and outcomes, such as overall survival.

Patient selection and emulation taking into account time-varying treatment exposures will be based on various factors including basic characteristics (age, sex, date of diagnosis, cancer stage, performance status), tumour prognostic factors (histology, morphology, tumour size, co-morbidities), and treatment details (anti-cancer treatments, pre-diagnosis treatment history, treatment duration).

This information is needed at diagnosis, first-line treatment, and, if available, at the time of treatment switching. For evaluating outcomes, we require information on the dates of patients' deaths (to estimate overall survival), signs of treatment relapse (for assessing progression-free survival), and factors leading to follow-up loss (e.g., relocating out of the country), wherever available.

For the Flatiron data application, we submitted our study protocol and variable requirements, as outlined above and in Section 5.3, specifically for prostate cancer patients, to the Flatiron scientists for evaluation. This approach was necessary because Flatiron's data dictionary is not publicly available. In response, Flatiron informed us of potential limitations in using their database for our planned analysis. Contrary to NCRAS, which supplies only the variables specifically requested, Flatiron grants access to their entire standard disease-specific database, such as the metastatic prostate cancer database, upon approval of the data application.

For the NCRAS database application, we employed the most recent ODR NCRAS data dictionary template (v4.4)[102] available at the time when we initially applied. This helped us identify relevant variables and customise our data request, including justifications for each variable we requested. A complete list of datasets, variables, and cohort definitions for the NCRAS data application can be found in the Appendix. This list, along with the protocol, has been under review by the NHS ethics committee and the ODR/NHSD, and has been amended based on their feedback.

The subsequent paragraphs (Section 5.2-5.5) provide specifics on the datasets and variables we requested from Flatiron and ODR/NHSD, including detailed original content specifications.

5.2 Relevant datasets required

5.2.1 Flatiron datasets required

All relevant datasets in the standard Flatiron mPC database, containing necessary variable information as outlined in Sections 5.1 and 5.3, will be used wherever available.

5.2.2 NCRAS datasets required

- English Cancer Registry
- SACT dataset
- Radiotherapy dataset (RTDS)
- HES admitted care
- HES outpatient
- HES accident and emergency
- Cancer Waiting Time

5.3 Relevant variable information required

Our primary analyses (Figure 1) only planned to include patients that match with the inclusion/exclusion criteria in the benchmark RCTs (case study PC1 and RCC Analysis Set 1). However, we will conduct further "real-world effect" analyses on a wider population compared to the restricted trial population (case study PC2, PC3 and RCC Analysis Set 2). Consequently, our data application requests records of all patients diagnosed with prostate cancer or RCC, not just a specific subset. We requested all available pre- and post-diagnosis patient records, wherever available. Comprehensive data collection is crucial for our sequential treatment Target Trial analyses, especially for integrating time-varying exposures in assessing treatment sequence effectiveness. Table 4 presents the operational definitions of the required variable/information, with detailed justifications for each specific requested variable provided separately to ODR/NHS Digital using the ODR NCRAS data dictionary version 4.4[102].

Time-to-event outcomes like overall survival (OS) will be calculated from treatment initiation to the patient's death. In contrast, defining PFS requires careful consideration due to potential unavailability of exact progression dates. Our approach involves integrating proxy variables to estimate these dates, such as instances of treatment discontinuation. Patients who encounter adverse events or dropouts, when identifiable and pertinent to the protocol, may be censored as appropriate.

Operational definitions defined in Table 4 may be refined following discussions with clinical experts and database analysts, and upon discovering more relevant algorithms in the literature, such as those for comorbidity definition.

Table 4. Summary of variable requirements and operational definitions

	Prostate cancer case studies	**Renal cell carcinoma case studies**
Cancer type	ICD-10: C61x	ICD-10: C64x
Basic characteristics	Age, sex, date of diagnosis, date of first-line treatment, date of second-line treatment, cancer stage, ECOG performance status, tumour size, tumour histology, tumour morphology	
Comorbidities considered for replicating the inclusion and exclusion criteria of the benchmark trials in PC1 and RCC Analysis Set 2	• Brain metastasises (ICD-10: C79.3) • Contraindications of abiraterone and enzalutamide - pregnancy (ICD-10: O00x, O01x, O02x, O03x, O04x, O05x, O06x, O07x, O08x)[103] (It is unlikely that men have any diagnosis of pregnancy, but this can be used to test the quality of the NCRAS/HES data) • Active epidural disease (G95x other and unspecified diseases of spinal cord) • Active concurrent malignancy (ICD-10: C00x-C43x, C45x-C96x, D00x-D05x, D07x-49x) • History of seizures or cerebrovascular events (ICD-10: G40x, I60x-69x) • Gastrointestinal disorders affecting absorption (ICD-10: K90x)	• Metastasises (ICD-10: C79x) (of other cancers) • Brain metastasises (ICD-10: C79.3): Professor Janet Brown suggested that roughly 1 in 20 mRCC patients may have de novo brain metastatses. • End-stage renal disease (ICD-10: N18.6)

	• End-stage renal disease (ICD-10: N18.6) • Major surgery within 4 weeks of starting study treatment	
Comorbidities considered for adjusting prognostic factors	We planned to use the Charlson Comorbidity Index (CCI) from NCRAS data as an indicator of disease severity for adjusting baseline characteristics. Given the CCI's nature as a summary score with various algorithm versions, we may perform sensitivity analyses with alternative algorithms for defining comorbidities. A study utilising HES data validated a CCI ICD-10 translation (based on the Deyo and Dartmouth-Manitoba ICD-9 adaptations) for predicting in-hospital mortality in urological cancer surgery patients.[104] Although not specific to prostate cancer or RCC, this algorithm developed by Sundararajan et al. (outlined below) may be relevant as it has been validated in NHS cancer patients. Other validated tools, such as Elixhauser scores or individual comorbidity variables may also be considered subject to discussion with clinicians.[105] For example, for the US population, the Combined Comorbidity Score developed by the Harvard group may be particularly relevant.[106,107]	
Treatments	• All patients whoever received any systematic anti-cancer therapy for their prostate cancer.	• All patients whoever received any systematic anti-cancer therapy for their RCC.

Condition	Weights	ICD-10-AM
Acute myocardial infarction	1	I21, I22, I252
Congestive heart failure	1	I50
Peripheral vascular disease	1	I71, I790, I739, R02, Z958, Z959
Cerebral vascular accident	1	I60, I61, I62, I63, I65, I66,G450, G451, G452, G458, G459, G46, I64, G454, I670, I671, I672, I674, I675, I676, I677 I678, I679, I681, I682, I688, I69
Dementia	1	F00, F01, F02, F051
Pulmonary disease	1	J40, J41, J42, J44, J43, J45, J46, J47, J67, J44, J60, J61, J62, J63, J66, J64, J65
Connective tissue disorder	1	M32, M34, M332, M053, M058, M059, M060, M063, M069, M050, M052, M051, M353
Peptic ulcer	1	K25, K26, K27, K28
Liver disease	1	K702, K703, K73, K717, K740, K742, K746, K743, K744, K745
Diabetes	1	E109, E119, E139, E149, E101, E111, E131, E141, E105, E115, E135, E145
Diabetes complications	2	E102, E112, E132, E142 E103, E113, E133, E143 E104, E114, E134, E144
Paraplegia	2	G81 G041, G820, G821, G822
Renal disease	2	N03, N052, N053, N054, N055, N056, N072, N073, N074, N01, N18, N19, N25
Cancer	2	C0, C1, C2, C3, C40, C41, C43, C45, C46, C47, C48, C49, C5, C6, C70, C71, C72, C73, C74, C75, C76, C80, C81, C82, C83, C84, C85, C883, C887, C889, C900, C901, C91, C92, C93, C940, C941, C942, C943, C9451, C947, C95, C96
Metastatic cancer	3	C77, C78, C79, C80
Severe liver disease	3	K729, K766, K767, K721
HIV	6	B20, B21, B22, B23, B24

	• Common treatments: abiraterone, enzalutamide, docetaxel, carbizitaxel, sipuleucel-T, radium-223	• Common treatments: sunitinib, everolimus, pazopanib, sorafenib, temsirolimus, axitinib, nivolumab, cabozantinib, tivozanib, lenvatinib with everolimus, nivolumab with ipilimumab, avelumab with axitinib
Outcome measurements	Death date, disease status (i.e., disease progression, treatment discontinuation, metastasis (e.g. ICD-10: C79x), wherever available)	

5.4 Data time period

The full duration of the Flatiron mPC database is not publicly available, but the data curation lag is typically minimal. As of the March 31, 2019 cut-off, Flatiron's analysts provided initial sample size estimates for our protocol population. They identified about 4,000 mCRPC patients who underwent first-line treatment, with 1,700 advancing to second-line therapy. This includes about 600 patients treated first with abiraterone then enzalutamide, and 400 with the reverse sequence, suggesting a sufficient sample size for the study, especially with further inclusion of patients who received abiraterone or enzalutamide as first-line therapy without subsequent treatments. Based on this confirmation, we were able to confidently proceed with the data application for Flatiron's mPC database.

Patients diagnosed before April 2012 might lack comprehensive treatment data, essential for our treatment sequence analysis. The NDRS analysts further recommended focusing on patients treated from 2014 onwards due to the improved completeness of SACT data. Moreover, explicitly emulating the theoretical minimum follow-up periods of approximately 2 and 3 years, as observed in the GUTG-001[22] and RECORD-3[19] trials (i.e., from the last patient's enrolment to the data cut-off date), promises fairer comparisons. With data access expected to cover up to the end of 2019/2020 by the time of application, we aimed to include patients diagnosed or starting their first-line treatment between 2014-2017 for prostate cancer, and 2014-2016 for RCC case studies, taking into account both the respective minimum follow-up periods and SACT data maturity. However, to accommodate sensitivity analyses assessing the impact of treatment data and the inclusion of patients not meeting the theoretical minimum follow-up periods, we requested data for all patients diagnosed with prostate cancer (C61) or renal cell carcinoma (C64) between 01/01/2011 and 31/12/2020 (See Appendix). For these patients, we also requested an extended period of records—six years prior to their diagnosis and all records post-diagnosis. This is necessary to accurately define time-varying covariates, treatments, and outcomes as specified in Section 5.3 (See Appendix).

5.5 Geography criteria

For Flatiron data, we requested records from the entire population associated with Flatiron Health's US-affiliated providers. Our NCRAS data request was exclusively for patients in England.

References

1. Zheng Y, Pan F, Sorensen S. Modeling Treatment Sequences in Pharmacoeconomic Models. *Pharmacoeconomics.* 2017;35(1):15-24.
2. Viola MG, Diamantopoulos A. Determining the Baseline Strategy in a Cost-Effectiveness Analysis with Treatment Sequences. *Appl Health Econ Health Policy.* 2020;18(1):17-29.
3. Tosh J, Stevenson M, Akehurst R. Health Economic Modelling of Treatment Sequences for Rheumatoid Arthritis: A Systematic Review. *Curr Rheumatol Rep.* 2014;16(10):447.
4. Lewis RA, Hughes D, Sutton AJ, Wilkinson C. Quantitative Evidence Synthesis Methods for the Assessment of the Effectiveness of Treatment Sequences for Clinical and Economic Decision Making: A Review and Taxonomy of Simplifying Assumptions. *Pharmacoeconomics.* 2020:1-37.
5. Al MJ, Jin H, Robins J, Sculpher M. Virtual ISPOR Europe 2020: Novel Methods for Modeling Treatment Sequencing (Spotlight session 3). 2020; https://www.ispor.org/conferences-education/conferences/past-conferences/ispor-europe-2020/program/spotlight-sessions. Accessed Nov 19, 2020.
6. Huang M, Ramsey S, Xue W, Xie J, Pellissier J, Briggs A. Conceptual Framework and Methodological Challenges for Modeling Effectiveness in Oncology Treatment Sequence Models. *Pharmacoeconomics.* 2021:1-12.
7. Berger ML, Sox H, Willke RJ, et al. Good practices for real-world data studies of treatment and/or comparative effectiveness: recommendations from the joint ISPOR-ISPE Special Task Force on real-world evidence in health care decision making. *Value Health.* 2017;20(8):1003-1008.
8. Bell H, Wailoo A, Hernandez M, et al. The use of real world data for the estimation of treatment effects in NICE decision making (report by the NICE Decision Support Unit). 2016; http://nicedsu.org.uk/wp-content/uploads/2018/05/RWD-DSU-REPORT-Updated-DECEMBER-2016.pdf. Accessed June 2, 2020.
9. Faria R, Hernadez Alava M, Manca A, Wailoo A. NICE DSU Technical Support Document 17: The use of observational data to inform estimates of treatment effectiveness in technology appraisal: methods for comparative individual patient data. 2015; http://nicedsu.org.uk/wp-content/uploads/2016/03/TSD17-DSU-Observational-data-FINAL.pdf. Accessed 2020, June 2.
10. Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: FL. Chapman & Hall/CRC; 2020.

11. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology.* 2000;11(5):550-560.
12. Hernán MA, Lanoy E, Costagliola D, Robins JM. Comparison of dynamic treatment regimes via inverse probability weighting. *Basic Clin Pharmacol Toxicol.* 2006;98(3):237-242.
13. England PH. A guide to NCRAS data and its availability. 2020; https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1073359/NCRAS_data_and_its_availability-October_2020.pdf. Accessed Apr 30, 2022.
14. Flatiron Health. Flatiron Health: Publications. 2020; https://flatiron.com/publications/. Accessed June 2, 2020.
15. 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):758-764.
16. RCT Duplicate. RCT duplicate: randomized controlled trials duplicated using prospective longitudinal insurance claims: applying techniques of epidemiology 2020; www.rctduplicate.org. Accessed Jan 15, 2021.
17. OptumLabs. Using RWD in regulatory decision-making. 2020; https://www.optumlabs.com/work/data-regulatory-decision.html. Accessed Jan 15, 2021.
18. Forbes SP, Dahabreh IJ. Benchmarking Observational Analyses Against Randomized Trials: a Review of Studies Assessing Propensity Score Methods. *J Gen Intern Med.* 2020:1-9.
19. Knox JJ, Barrios CH, Kim TM, et al. Final overall survival analysis for the phase II RECORD-3 study of first-line everolimus followed by sunitinib versus first-line sunitinib followed by everolimus in metastatic RCC. *Ann Oncol.* 2017;28(6):1339-1345.
20. Chirgwin JH, Giobbie-Hurder A, Coates AS, et al. Treatment Adherence and Its Impact on Disease-Free Survival in the Breast International Group 1-98 Trial of Tamoxifen and Letrozole, Alone and in Sequence. *J Clin Oncol.* 2016;34(21):2452-2459.
21. Derks MGM, Blok EJ, Seynaeve C, et al. Adjuvant tamoxifen and exemestane in women with postmenopausal early breast cancer (TEAM): 10-year follow-up of a multicentre, open-label, randomised, phase 3 trial. *The Lancet Oncology.* 2017;18(9):1211-1220.
22. Khalaf DJ, Annala M, Taavitsainen S, et al. Optimal sequencing of enzalutamide and abiraterone acetate plus prednisone in metastatic castration-resistant prostate cancer: a multicentre, randomised, open-label, phase 2, crossover trial. *The Lancet Oncology.* 2019;20(12):1730-1739.

23. Wang SV, Schneeweiss S, Franklin JM, et al. Emulation of randomized clinical trials with nonrandomized database analyses: results of 32 clinical trials. *JAMA.* 2023;329(16):1376-1385.
24. Franklin JM, Patorno E, Desai RJ, et al. Emulating Randomized Clinical Trials with Nonrandomized Real-World Evidence Studies: First Results from the RCT DUPLICATE Initiative. *Circulation.* 2020.
25. Motzer RJ, Barrios CH, Kim TM, et al. Phase II randomized trial comparing sequential first-line everolimus and second-line sunitinib versus first-line sunitinib and second-line everolimus in patients with metastatic renal cell carcinoma. *J Clin Oncol.* 2014;32(25):2765-2772.
26. NCRAS. About the National Cancer Registration and Analysis Service. http://www.ncin.org.uk/about_ncin/. Accessed June 2, 2020.
27. The SACT Team. The SACT Systemic Anti-Cancer Therapy Dataset website. 2020; http://www.chemodataset.nhs.uk/home. Accessed June 2, 2020.
28. Flatiron Health. NICE Partners with Flatiron Health to Develop Real-World Evidence Research Methodologies. 2020; https://flatiron.com/resources/press-release/nice-partnership-2020. Accessed June 30, 2023.
29. Flatiron Health. Flatiron Health Presents Research at ISPOR Europe. 2021; https://flatiron.com/resources/press-release/flatiron-health-presents-research-at-ispor-europe. Accessed June 30, 2023.
30. Mpofu P, Kent S, Jónsson P, et al. Evaluation of US oncology electronic health record real-world data to reduce uncertainty in health technology appraisals: a retrospective cohort study. *BMJ open.* 2023;13(10):e074559.
31. NICE. Dabrafenib plus trametinib for treating BRAF V600 mutation-positive advanced non-small-cell lung cancer (TA 898). 2023; https://www.nice.org.uk/guidance/ta898. Accessed Oct 20, 2023.
32. Crown WH, Bierer BE. Real-World Evidence: Understanding Sources of Variability Through Empirical Analysis. *Value Health.* 2020.
33. Crown W, Dahabreh IJ, Li X, Toh S, Bierer B. Can Observational Analyses of Routinely Collected Data Emulate Randomized Trials? Design and Feasibility of the Observational Patient Evidence for Regulatory Approval Science and Understanding Disease Project. *Value Health.* 2023;26(2):176-184.
34. Franklin JM, Pawar A, Martin D, et al. Nonrandomized real-world evidence to support regulatory decision making: process for a randomized trial replication project. *Clin Pharmacol Ther.* 2020;107(4):817-826.

35. NICE. NICE Pathways: Prostate cancer overview. 2021; https://pathways.nice.org.uk/pathways/prostate-cancer. Accessed Feb 5, 2021.
36. George DJ, Sartor O, Miller K, et al. Treatment patterns and outcomes in patients with metastatic castration-resistant prostate cancer in a real-world clinical practice setting in the united states. *Clin Genitourin Cancer.* 2020.
37. Lorente D, Mateo J, Perez-Lopez R, de Bono JS, Attard G. Sequencing of agents in castration-resistant prostate cancer. *The Lancet Oncology.* 2015;16(6):e279-e292.
38. Mori K, Miura N, Mostafaei H, et al. Sequential therapy of abiraterone and enzalutamide in castration-resistant prostate cancer: a systematic review and meta-analysis. *Prostate Cancer Prostatic Dis.* 2020;23(4):539-548.
39. McGrane J, Rowe M, Hidayat A, et al. The Impact of Treatment Crossover and Prostate-specific Antigen Flare with Enzalutamide and Abiraterone in Metastatic Castrate Resistant Prostate Cancer. *Clin Oncol (R Coll Radiol).* 2021;33(1):e92.
40. Tran K, McGill S. Treatment Sequences of Androgen Receptor–Targeted Agents for Prostate Cancer. *Canadian Journal of Health Technologies.* 2021;1(3).
41. Tordrup D, Chouaid C, Cuijpers P, et al. Priorities for Health Economic Methodological Research: Results of an Expert Consultation. *Int J Technol Assess Health Care.* 2017;33(6):609-619.
42. Cassinello J, Dominguez-Lubillo T, Gomez-Barrera M, et al. Optimal treatment sequencing of abiraterone acetate plus prednisone and enzalutamide in patients with castration-resistant metastatic prostate cancer: A systematic review and meta-analysis. *Cancer Treat Rev.* 2021;93:102152.
43. Cattrini C, España R, Mennitto A, et al. Optimal sequencing and predictive biomarkers in patients with advanced prostate cancer. *Cancers (Basel).* 2021;13(18):4522.
44. NCRAS. CancerData. https://www.cancerdata.nhs.uk/. Accessed Jan 4, 2021.
45. NICE. Padeliporfin for untreated localised prostate cancer (TA 546). 2018; https://www.nice.org.uk/guidance/ta546. Accessed Apr 30, 2021.
46. Hirst C, Cabrera C, Kirby M. Epidemiology of castration resistant prostate cancer: a longitudinal analysis using a UK primary care database. *Cancer Epidemiol.* 2012;36(6):e349-e353.
47. NICE. Docetaxel for the treatment of hormone-refractory metastatic prostate cancer (TA 101). 2006; https://www.nice.org.uk/guidance/ta101. Accessed Apr 30, 2021.
48. NICE. Abiraterone for treating metastatic hormone-relapsed prostate cancer before chemotherapy is indicated (TA 387). 2016; https://www.nice.org.uk/guidance/ta387. Accessed Apr 30, 2021.

49. NICE. Enzalutamide for treating metastatic hormone-relapsed prostate cancer before chemotherapy is indicated (TA 377). 2016; https://www.nice.org.uk/guidance/ta377. Accessed Apr 30, 2021.
50. NICE. Abiraterone for castration-resistant metastatic prostate cancer previously treated with a docetaxel-containing regimen (TA 259). 2012; https://www.nice.org.uk/guidance/ta259. Accessed Apr 30, 2021.
51. NICE. Enzalutamide for metastatic hormone-relapsed prostate cancer previously treated with a docetaxel-containing regimen (TA 316). 2014; https://www.nice.org.uk/guidance/ta316. Accessed Apr 30, 2021.
52. NICE. Cabazitaxel for hormone-relapsed metastatic prostate cancer treated with docetaxel (TA 391). 2016; https://www.nice.org.uk/guidance/ta391. Accessed Apr 30, 2021.
53. NICE. Radium-223 dichloride for treating hormone-relapsed prostate cancer with bone metastases (TA 412). 2016; https://www.nice.org.uk/guidance/ta412. Accessed Apr 30, 2021.
54. NICE. Olaparib for preOlaparib for previously treated BRCA mutation-positive hormone-relapsed metastatic prostate cancer (TA 887). 2023; https://www.nice.org.uk/guidance/ta887. Accessed Dec 3, 2023.
55. NICE. Lutetium-177 vipivotide tetraxetan for treating PSMA-positive hormone-relapsed metastatic prostate cancer after 2 or more treatments (TA 930). 2023; https://www.nice.org.uk/guidance/ta930. Accessed Dec 3, 2023.
56. NICE. Darolutamide with androgen deprivation therapy for treating hormone-relapsed non-metastatic prostate cancer (TA 660). 2020; https://www.nice.org.uk/guidance/ta660. Accessed Apr 30, 2021.
57. NICE. Apalutamide with androgen deprivation therapy for treating high-risk hormone-relapsed non-metastatic prostate cancer (TA 740). 2021; https://www.nice.org.uk/guidance/ta740. Accessed Dec 2, 2023.
58. NHS England. National Cancer Drugs Fund (CDF) list. 2023; https://www.england.nhs.uk/publication/national-cancer-drugs-fund-list/. Accessed Dec 3, 2023.
59. Vale CL, Fisher DJ, Godolphin PJ, et al. Which patients with metastatic hormone-sensitive prostate cancer benefit from docetaxel: a systematic review and meta-analysis of individual participant data from randomised trials. *The Lancet Oncology.* 2023;24(7):783-797.
60. NICE. Enzalutamide for treating hormone-sensitive metastatic prostate cancer (TA 712). 2021; https://www.nice.org.uk/guidance/ta712. Accessed Dec 2, 2023.

61. NICE. Apalutamide with androgen deprivation therapy for treating hormone-sensitive metastatic prostate cancer (TA 741). 2021; https://www.nice.org.uk/guidance/ta741. Accessed Dec 2, 2023.
62. NICE. Darolutamide with androgen deprivation therapy and docetaxel for treating hormone-sensitive metastatic prostate cancer (TA 903). 2023; https://www.nice.org.uk/guidance/ta903. Accessed Apr 30, 2021.
63. Gravis G, Boher J-M, Joly F, et al. Androgen deprivation therapy (ADT) plus docetaxel versus ADT alone in metastatic non castrate prostate cancer: impact of metastatic burden and long-term survival analysis of the randomized phase 3 GETUG-AFU15 trial. *Eur Urol.* 2016;70(2):256-262.
64. Gravis G, Fizazi K, Joly F, et al. Androgen-deprivation therapy alone or with docetaxel in non-castrate metastatic prostate cancer (GETUG-AFU 15): a randomised, open-label, phase 3 trial. *The lancet oncology.* 2013;14(2):149-158.
65. Kyriakopoulos CE, Chen Y-H, Carducci MA, et al. Chemohormonal therapy in metastatic hormone-sensitive prostate cancer: long-term survival analysis of the randomized phase III E3805 CHAARTED trial. *J Clin Oncol.* 2018;36(11):1080.
66. Clarke NW, Ali A, Ingleby F, et al. Addition of docetaxel to hormonal therapy in low-and high-burden metastatic hormone sensitive prostate cancer: long-term survival results from the STAMPEDE trial. *Ann Oncol.* 2019;30(12):1992-2003.
67. James ND, Sydes MR, Clarke NW, et al. Addition of docetaxel, zoledronic acid, or both to first-line long-term hormone therapy in prostate cancer (STAMPEDE): survival results from an adaptive, multiarm, multistage, platform randomised controlled trial. *The Lancet.* 2016;387(10024):1163-1177.
68. Group PCTC. Maximum androgen blockade in advanced prostate cancer: an overview of the randomised trials. *The Lancet.* 2000;355(9214):1491-1498.
69. FDA. FDA approves abiraterone acetate in combination with prednisone for high-risk metastatic castration-sensitive prostate cancer. 2018; https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-abiraterone-acetate-combination-prednisone-high-risk-metastatic-castration-sensitive. Accessed June 21.
70. FDA. FDA approves enzalutamide for metastatic castration-sensitive prostate cancer. 2019; https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-enzalutamide-metastatic-castration-sensitive-prostate-cancer. Accessed June 21.
71. Schaeffer EM, Srinivas S, Adra N, et al. NCCN Guidelines® insights: prostate cancer, version 1.2023: featured updates to the NCCN guidelines. *J Natl Compr Canc Netw.* 2022;20(12):1288-1298.

72. Chi KN, Eigl B, Seto L, et al. A Randomized Phase II Study of Sequencing Abiraterone Acetate and Enzalutamide in Metastatic Castration-Resistant Prostate Cancer: GUTG-001 (Veresion 6.0, 16Jul2018). 2018; https://classic.clinicaltrials.gov/ProvidedDocs/57/NCT02125357/Prot_SAP_000.pdf. Accessed June 30, 2023.
73. Annala M, Vandekerkhove G, Khalaf D, et al. Circulating tumor DNA genomics correlate with resistance to abiraterone and enzalutamide in prostate cancer. *Cancer Discov.* 2018;8(4):444-457.
74. Khalaf DJ, Sunderland K, Eigl BJ, et al. Health-related quality of life for abiraterone plus prednisone versus enzalutamide in patients with metastatic castration-resistant prostate cancer: results from a phase II randomized trial. *Eur Urol.* 2019;75(6):940-947.
75. Maringe C, Benitez Majano S, Exarchakou A, et al. Reflection on modern methods: trial emulation in the presence of immortal-time bias. Assessing the benefit of major surgery for elderly lung cancer patients using observational data. *Int J Epidemiol.* 2020;49(5):1719-1729.
76. Zhao SS, Lyu H, Yoshida K. Versatility of the clone-censor-weight approach: response to "trial emulation in the presence of immortal-time bias". *Int J Epidemiol.* 2021;50(2):694-695.
77. Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J Clin Epidemiol.* 2016;79:70-75.
78. Wolf A, Dedman D, Campbell J, et al. Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum. *Int J Epidemiol.* 2019;48(6):1740-1740g.
79. Young GJ, Harrison S, Turner EL, et al. Prostate-specific antigen (PSA) testing of men in UK general practice: a 10-year longitudinal cohort study. *BMJ open.* 2017;7(10):e017729.
80. NICE. Prostate cancer: diagnosis and management (NG 131). 2019; https://www.nice.org.uk/guidance/ng131. Accessed Apr 30, 2021.
81. Pereira-Salgado A, Kwan EM, Tran B, Gibbs P, De Bono J, IJzerman M. Systematic review of efficacy and health economic implications of real-world treatment sequencing in prostate cancer: Where do the newer agents enzalutamide and abiraterone fit in? *European urology focus.* 2021;7(4):752-763.
82. NICE. NICE Pathways: Renal cancer overview. 2021; https://pathways.nice.org.uk/pathways/renal-cancer. Accessed Jan 5.

83. Hawkins R, Fife K, Hurst M, et al. Treatment patterns and health outcomes in metastatic renal cell carcinoma patients treated with targeted systemic therapies in the UK. *BMC Cancer.* 2020;20(1):670.
84. NICE. Avelumab with axitinib for untreated advanced renal cell carcinoma (TA 645). 2020; https://www.nice.org.uk/guidance/ta377. Accessed Apr 30, 2021.
85. Cancer Research UK. Kidney cancer statistics. 2021; https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/kidney-cancer. Accessed Apr 30, 2021.
86. NICE. Cabozantinib for untreated advanced renal cell carcinoma (TA 542). 2018; https://www.nice.org.uk/Guidance/TA417. Accessed Apr 30, 2021.
87. NICE. Nivolumab for previously treated advanced renal cell carcinoma (TA 417). 2016; https://www.nice.org.uk/Guidance/TA417. Accessed Apr 30, 2021.
88. Iacovelli R, Ciccarese C, Bria E, et al. Immunotherapy versus standard of care in metastatic renal cell carcinoma. A systematic review and meta-analysis. *Cancer Treat Rev.* 2018;70:112-117.
89. Jonasch E. NCCN guidelines updates: management of metastatic kidney cancer. *J Natl Compr Canc Netw.* 2019;17(5.5):587-589.
90. NICE. Axitinib for treating advanced renal cell carcinoma after failure of prior systemic treatment (TA 333). 2015; https://www.nice.org.uk/guidance/ta333. Accessed Mar 23, 2020.
91. Martinez Chanza N, Tripathi A, Harshman LC. Adjuvant Therapy Options in Renal Cell Carcinoma: Where Do We Stand? *Curr Treat Options Oncol.* 2019;20(5):44.
92. Labriola MK, George DJ. Setting a new standard for long-term survival in metastatic kidney cancer. Vol 128: Wiley Online Library; 2022:2058-2060.
93. Géczi L, Bodoky G, Rokszin G, Fábián I, Torday L. Survival benefits of second-line axitinib versus everolimus after first line sunitinib treatment in metastatic renal cell carcinoma. *Pathology & Oncology Research.* 2020;26(4):2201-2207.
94. Iacovelli R, Cartenì G, Milella M, et al. Clinical outcomes in patients with metastatic renal cell carcinoma receiving everolimus or temsirolimus after sunitinib. *Canadian Urological Association Journal.* 2014;8(3-4):E121.
95. Latimer NR, Henshall C, Siebert U, Bell H. Treatment switching: statistical and decision-making challenges and approaches. *Int J Technol Assess Health Care.* 2016;32(3):160-166.
96. Thorlund K, Dron L, Park JJH, Mills EJ. Synthetic and External Controls in Clinical Trials - A Primer for Researchers. *Clin Epidemiol.* 2020;12:457-467.

97. Dahabreh IJ, Steingrimsson JA, Robins JM, Hernán MA. Randomized trials and their observational emulations: a framework for benchmarking and joint analysis. *arXiv preprint arXiv:220314857.* 2022.
98. ClinicalTrials.gov. Efficacy and Safety Comparison of RAD001 Versus Sunitinib in the First-line and Second-line Treatment of Patients With Metastatic Renal Cell Carcinoma (RECORD-3) (NCT00903175). 2011; https://clinicaltrials.gov/study/NCT00903175. Accessed March 31, 2021.
99. ESMO. Performance scales: Karnofsky & ECOG scores. https://oncologypro.esmo.org/oncology-in-practice/practice-tools/performance-scales. Accessed Mar 15, 2021.
100. Heyard R, Held L, Schneeweiss S, Wang SV. Design differences explain variation in results between randomized trials and their non-randomized emulations. *medRxiv.* 2023.
101. Guyot P, Ades AE, Ouwens MJ, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. *BMC Med Res Methodol.* 2012;12:9.
102. NCRAS. NCRAS Cancer Registration Data Dictionary v4.4. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1054575/NCRAS_ODR_data_dictionary_v4.4.xlsx. Accessed June 2, 2021.
103. Minassian C, Williams R, Meeraus WH, Smeeth L, Campbell OM, Thomas SL. Methods to generate and validate a Pregnancy Register in the UK Clinical Practice Research Datalink primary care database. *Pharmacoepidemiol Drug Saf.* 2019;28(7):923-933.
104. Nuttall M, van der Meulen J, Emberton M. Charlson scores based on ICD-10 administrative data were valid in assessing comorbidity in patients undergoing urological cancer surgery. *J Clin Epidemiol.* 2006;59(3):265-273.
105. Austin SR, Wong Y-N, Uzzo RG, Beck JR, Egleston BL. Why summary comorbidity measures such as the Charlson comorbidity index and Elixhauser score work. *Med Care.* 2015;53(9):e65.
106. Gagne JJ, Glynn RJ, Avorn J, Levin R, Schneeweiss S. A combined comorbidity score predicted mortality in elderly patients better than existing scores. *J Clin Epidemiol.* 2011;64(7):749-759.
107. Sun JW, Rogers JR, Her Q, et al. Adaptation and Validation of the Combined Comorbidity Score for ICD-10-CM. *Med Care.* 2017;55(12):1046-1051.

Appendix: Specification for NCRAS data extraction

1. Cohort definition

- All patients who have been diagnosed with prostate cancer (C61) or renal cell carcinoma (C64) between 01/01/2011 and 31/12/2020.
- Limited to England.
- Prostate cancer: males. Kidney cancer: males and females.
- Dataset extraction:
  (1) Cancer Registration data, linked via tumourid for prostate cancer and renal cell carcinoma diagnoses as above.
  (2) Data extraction for the cohort from 2005 (6 years prior) for the above patients to the latest available data, from the following datasets: SACT, CWT, HES (APC), HES (OP), HES (A&E) and RTDS records, all linked at patient level (therefore comprising treatment data for the concurrent tumours below too). For the selected fields within each of the selected tables, as shown in the table below. Data is released in separate tables for each dataset.
  (3) A customized separate data table for concurrent tumour records: For the patients in the aforementioned cohort, the following fields from the Cancer Registry are required for all other cancer diagnoses (all ICD-10 C and D codes) diagnosed from 2009 to 2020: pseudonymised patientid, pseudonymised tumourid (ensure different pseudo tumourids to those used for cohort above), DIAGNOSISDATE1, DIAGNOSISDATE2, DIAGNOSISDATEBEST, DIAGNOSISDATEFLAG, SITE_ICD10_O2, SITE_ICD10_O2_3CHAR, MORPH_ICD10_O2, STAGE_BEST, T_BEST, N_BEST, M_BEST, GLEASON_PRIMARY, GLEASON_SECONDARY, GLEASON_TERTIARY, GLEASON_COMBINED.

2. Selected data tables (datasets) and fields (variables)

Data table (datasets within NCRAS)	Requested fields
Cancer Registry Data	PATIENTID (project specific pseudonymised)
	TUMOURID (project specific pseudonymised)
	SEX
	ETHNICITY
	ETHNICITYNAME
	AGE
	DIAGNOSISDATE1
	DIAGNOSISDATE2
	DIAGNOSISDATEBEST
	DIAGNOSISDATEFLAG
	BASISOFDIAGNOSIS
	SITE_ICD10_O2
	SITE_ICD10_O2_3CHAR
	MORPH_ICD10_O2
	BEHAVIOUR_ICD10_O2
	SITE_CODED
	SITE_CODED_DESC
	SITE_CODED_3CHAR
	CODING_SYSTEM
	CODING_SYSTEM_DESC
	MORPH_CODED
	BEHAVIOUR_CODED
	BEHAVIOUR_CODED_DESC
	HISTOLOGY_CODED
	HISTOLOGY_CODED_DESC
	GRADE

	TUMOURSIZE
	NODESEXCISED
	NODESINVOLVED
	TUMOURCOUNT
	BIGTUMOURCOUNT
	ROUTE_CODE
	FINAL_ROUTE
	STAGE_BEST
	T_BEST
	N_BEST
	M_BEST
	STAGE_BEST_SYSTEM
	T_IMG
	N_IMG
	M_IMG
	STAGE_IMG
	STAGE_IMG_SYSTEM
	T_PATH
	N_PATH
	M_PATH
	STAGE_PATH
	STAGE_PATH_SYSTEM
	STAGE_PATH_PRETREATED
	CHRL_TOT_27_03
	CHRL_TOT_78_06
	HES_LINKED
	GLEASON_PRIMARY
	GLEASON_SECONDARY
	GLEASON_TERTIARY
	GLEASON_COMBINED
	LATERALITY
	DCO
	VITALSTATUS
	VITALSTATUSDATE
	DEATHDATEBEST
	DEATHDATEFLAG
	EMBARKATION
	EMBARKATIONDATE
	DEATHCAUSECODE_1A
	DEATHCAUSECODE_1B
	DEATHCAUSECODE_1C
	DEATHCAUSECODE_2
	DEATHCAUSECODE_UNDERLYING
	DIAG_HOSP
	DIAG_HOSP_NAME
	FIRST_HOSP
	FIRST_HOSP_NAME
	FIRST_HOSP_DATE
	DIAG_TRUST
	DIAG_TRUST_NAME
	FIRST_TRUST
	FIRST_TRUST_NAME
	CCG_CODE
	CCG_NAME
	COUNTY_CODE
	COUNTY_NAME
	GOR_CODE
	GOR_NAME
	CTRY_CODE
	CTRY_NAME
	INCOME_QUINTILE (Income domain, based on version most appropriate for timing of tumour diagnosis)
	IMD_QUINTILE (Full IMD, based on version most appropriate for timing of tumour diagnosis)
	RT_FLAG
	CT_FLAG
	SG_FLAG
	EVENTID (project specific pseudonymised)
	NUMBER_OF_TUMOURS
	EVENTCODE
	EVENTDESC
	EVENTDATE
	PROVIDERCODE
	PROVIDERDESC
	TRUST_CODE
	TRUST_NAME

	PRACTITIONERCODE (Pseudonymised)
	WITHIN_SIX_MONTHS_FLAG
	SIX_MONTHS_AFTER_FLAG
	OPCS4_CODE
	OPCS4_NAME
	RADIOCODE
	RADIODESC
	IMAGINGCODE
	IMAGINGDESC
	LESIONSIZE
	CHEMO_ALL_DRUGS
	CHEMO_DRUG_GROUP
	MULTIFOCAL
	EXCISIONMARGIN
Systemic Anti-Cancer Therapy Dataset (SACT)	PATIENTID
	TUMOURID
	NHS_Number_Status
	Ethnicity
	Consultant_GMC_Code (pseudonymised)
	Consultant_Speciality_Code
	Organisation_Code_of_Provider
	Primary_Diagnosis
	Morphology_clean
	Stage_at_Start
	Programme_Number
	Regimen_Number
	Intent_of_Treatment
	Adjunctive therapy
	Analysis_Group
	Benchmark_Group
	Height_At_Start_of_Regimen
	Weight_At_Start_of_Regimen
	Performance_Status_at_Start_of_Regimen_Clean
	Comorbidity_Adjustment
	Date_Decision_To_Treat
	Start_Date_of_Regimen
	Clinical_Trial
	Chemo_Radiation
	Number_of_Cycles_Planned
	Cycle_Number
	Start_Date_of_Cycle
	Weight_At_Start_Of_Cycle
	Performance_Status_At_Start_Of_Cycle_Clean
	OPCS_Procurement_Code
	Drug_Group
	Actual_Dose_Per_Administration
	Administration_Route
	Administration_Date
	OPCS_Delivery_Code
	Date_of_Final_Treatment
	Regimen_Modification_Dose_Reduction
	Regimen_Modification_Time_Delay
	Regimen_Modification_Stopped_Early
	Regimen_Outcome_Summary
	regoutsum_cur_not_com_plan
	regoutsum_non_curat
	regoutsum_toxic
	regoutsum_cur_com_plan
Radiotherapy Dataset (RTDS)	PATIENTID (project specific pseudonymised)
	RADIOTHERAPYEPISODEID (project specific pseudonymised)
	APPTDATE
	DECISIONTOTREATDATE
	EARLIESTCLINAPPROPRIATEDATE
	RADIOTHERAPYPRIORITY
	TREATMENTSTARTDATE
	RADIOTHERAPYDIAGNOSISICD
	RADIOTHERAPYINTENT
	RTTREATMENTREGION
	RTTREATMENTANATOMICALSITE
	PRIMARYPROCEDUREOPCS
	PROCEDUREDATE
NDRS Linked HES A&E	PATIENTID (project specific pseudonymised)
	ethnos
	aeattendcat
	aeattenddisp

	arrivaldate (date YYYY-MM-DD)
	diag_n
	diag2_n
	diaga_n
	diags_n
	treat_n
	treat2_n
NDRS Linked HES APC	PATIENTID (project specific pseudonymised)
	startage
	ethnos
	admidate (date YYYY-MM-DD)
	elecdate (date YYYY-MM-DD)
	admimeth
	firstreg
	elecdur
	disdate (date YYYY-MM-DD)
	bedyear
	spelbgin
	epiend (date YYYY-MM-DD)
	epistart (date YYYY-MM-DD)
	speldur
	spelend
	epidur
	epiorder
	epitype
	diag_4n
	diag3_3n
	opertn_nn
	opdate_nn
	operstat
	intmanig
	mainspef
	tretspef
NDRS Linked HES OP	PATIENTID
	ethnos
	apptdate
	attended
	outcome
	priority
	diag_nn
	diag_4
	diag3
	opertn_nn
	opertn_01
	opertn3
	operstat
	mainspef
	tretspef
Cancer Waiting Times (Treatments Only)	PATIENTID (project specific pseudonymised)
	TUMOURID (project specific pseudonymised)
	TREAT_PERIOD_START
	SITE_ICD10
	METS_SITE
	LATERALITY
	TREAT_START
	WTA_TREAT
	WTA_TREAT_REASON
	CTE_TYPE
	MODALITY
	CLIN_TRIAL
	RADIO_PRIORITY