Division: Worldwide Development
Retention Category: [redacted]
Information Type: Worldwide Epidemiology Study Protocol

Title:	5ARI and Prostate Cancer Mortality Study

Compound Number: GI198745

Development Phase IV

Effective Date: 11-15-2012

Description: This retrospective cohort study will assess the association of benign prostatic hyperplasia (BPH) treatment (5-alpha reductase inhibitors (5ARI) and alpha-blocker medications) with the occurrence of prostate cancer related mortality. This study will also assess a number of secondary endpoints including prostate cancer mortality or metastatic prostate cancer, and all cause mortality.

Subject: Prostate cancer mortality, BPH treatment, 5ARI.

Author(s):

[redacted];

[redacted];

[redacted];

[redacted]

[redacted]

SPONSOR SIGNATORY:

[redacted]	**Date**
[redacted]	**Date**
[redacted]	**Date**

SPONSOR INFORMATION PAGE

Worldwide Epidemiology

[redacted]

[redacted]

INVESTIGATOR PROTOCOL AGREEMENT PAGE

Required Standard Wording:

- I confirm agreement to conduct the study in compliance with the protocol.
- I acknowledge that I am responsible for overall study conduct. I agree to personally conduct or supervise the described study.
- I agree to ensure that all associates, colleagues and employees assisting in the conduct of the study are informed about their obligations. Mechanisms are in place to ensure that site staff receives the appropriate information throughout the study.

Investigator Name: _____

Investigator Signature	Date

TABLE OF CONTENTS

PAGE

LIST OF ABBREVIATIONS

AE	Adverse Event
GSK	GlaxoSmithKline
HGT	High grade tumors
BPH	Benign prostatic hyperplasia
5ARI	5-alpha reductase inhibitor
AUR	Acute urinary retention
PSA	Prostate-specific antigen
PCPT	Prostate Cancer Prevention Trial
REDUCE	Reduction by Dutasteride of Prostate Cancer Events
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]

Trademark Information

Trademarks of the GlaxoSmithKline group of companies	Trademarks not owned by the GlaxoSmithKline group of companies
AVODART	

PROTOCOL SUMMARY

Rationale

The potential role of 5-alpha-reductase inhibitors (5ARI) to reduce the risk of prostate cancer by blocking the conversion of testosterone to dihydrotestosterone was the basis for two large randomized controlled trials. The results of the Prostate Cancer Prevention Trial (PCPT) and the Reduction by Dutasteride of Prostate Cancer Events (REDUCE) trial demonstrated that finasteride and dutasteride, respectively, significantly reduce the risk of prostate cancer compared to placebo. However, there was a greater proportion of high grade tumors (HGT) observed in the 5ARI groups in both studies compared to placebo. In REDUCE (Andriole et al., 2011), 29 (0.9%) patients were diagnosed with Gleason 8-10 tumors in the dutasteride arm, compared with the placebo arm (n = 19, 0.6%; p = 0.15) over 4 years, with the most pronounced treatment group difference in REDUCE in years 3 and 4 (0.5% versus <0.1%, p = 0.003).

GSK is committed to further investigations to advance the understanding of the benefits and risks of dutasteride. Over the past several months, the GSK AVODART™ Team has evaluated the appropriateness of an observational study among men who use 5ARIs for symptoms of benign prostatic hyperplasia (BPH) to measure the risk of incident HGTs compared to risk in non-users. In consultation with internal and external experts, the team subsequently concluded that such a study would not be feasible: there is no other drug as a comparator drug that has similar effects on screening, detection, and biology as 5ARIs, resulting in several rate-limiting biases:

- 5ARIs lower PSA levels, potentially differentially affecting prostate cancer screening patterns;
- 5ARIs lower prostate volume, potentially differentially affecting biopsy accuracy;
- 5ARIs are reported to preferentially affect Gleason pattern 3 over pattern 4 and 5 prostate cancer, potentially differentially affecting Gleason score read-out and distribution.

Since the concern regarding HGTs is their higher agressiveness potential and risk of poorer disease related outcomes, this study will provide evidence as to whether patients who use 5ARIs have a greater likelihood of dying from prostate cancer. The outcome of prostate cancer mortality allows for an assessment of whether the relative frequency of high grade tumors among 5ARI versus alpha-blocker users translates into a difference in prostate cancer related mortality. This study will assess the association between 5ARI use with or without alpha-blockers and prostate cancer mortality in men treated with BPH medications.

Objective(s)

Primary Objective:

Prostate cancer mortality

1. To assess the risk of prostate cancer mortality associated with use of 5ARIs, with or without alpha-blockers, compared to alpha-blockers in men treated with BPH medications.

Secondary Objectives:

Prostate cancer mortality or metastatic prostate cancer

2. To assess the risk of prostate cancer mortality or metastatic prostate cancer associated with use of 5ARIs, with or without alpha-blockers, compared to alpha-blockers in men treated with BPH medications.

All cause mortality

3. To assess the risk of all cause mortality associated with use of 5ARIs, with or without alpha-blockers, compared to alpha-blockers in men treated with BPH medications.

Long term exposure to BPH medications

4. To assess the risk of prostate cancer mortality, prostate cancer mortality or metastatic prostate cancer, and all cause mortality associated with long-term exposure (2 or more years cumulative exposure) to 5ARIs, with or without alpha-blockers, compared to alpha-blockers in the subset of men using long term BPH treatment.

Additional descriptive analyses:

5. To evaluate the validity of classifying prostate cancer deaths with the developed electronic algorithm as compared to those identified based on cause of death coding and physician chart abstraction (see section 4.3: outcome definitions for further details).

6. To evaluate the validity of identifying cases of metastatic prostate cancer with the developed electronic algorithm as compared to those identified based on medical record abstraction.

7. To describe the occurrence of prostate cancer in men treated with BPH medications and across treatment groups (5ARIs and alpha-blockers).

8. To describe the occurrence of metastatic prostate cancer in men treated with BPH medications and across treatment groups.

9. To describe the occurrence of cardiovascular related mortality in men treated with BPH medications and across treatment groups.

10. To describe PSA testing patterns across treatment groups after treatment initiation and over the course of the study period.

11. To describe across treatment groups the frequency of biopsy, the number of cores per biopsy, and the number of positive cores per biopsy.

12. To describe across treatment groups Gleason Score and reclassification of Gleason Score between initial diagnosis biopsy and radical prostatectomy biopsy among those receiving radical prostatectomies.

Study Design

A retrospective cohort study from 1992-2010 will be conducted using data from 4[redacted] sites: [redacted] Southern California [redacted], [redacted] Northern California [redacted], [redacted] Northwest [redacted]), and [redacted] Colorado [redacted]. Men treated with BPH medications, 5ARIs (with and without concomitant and/or previous alpha-blocker use) will be compared to men treated with alpha-blockers. A matched design will be used with each man treated with 5ARIs being matched with 5 or 6 men treated with alpha-blockers. Men 50 years or older at the time of their first prescription for a study defined BPH medication, initiating treatment between 1992 and 2008 with at least 1-year of coverage in the healthcare system before the first prescription for BPH medication and at least 3 consecutive prescriptions (90 days of supply) for a BPH medication will be eligible for inclusion in the study. Men with a diagnosis of prostate cancer any time before the first prescription for BPH medication, having a diagnosis of prostate cancer within 3 months after initiation of their first BPH medication, and those treated with finasteride 1mg prior to their BPH medication will be excluded from the study. 5ARI initiators will be matched to alpha-blocker users in a ratio of 1:5 or 1:6 to yield an overall matching ratio of 1:5.4. Matching factors include age (+/- 1 year), timing of BPH treatment initiation (+/- 1 year), race, and duration of prior use of alpha-blockers. Based on the feasibility study from [redacted] we expect there to be approximately 284,000 men treated with BPH medications meeting eligibility criteria for inclusion in the study sample.

Study Endpoints/Assessments

The following outcomes will be assessed: prostate cancer mortality, prostate cancer mortality or metastatic cancer, and all cause mortality.

1. INTRODUCTION

Avodart (dutasteride), launched in 2003, is approved in over 90 countries as a monotherapy or in combination with tamsulosin to treat the symptoms of benign prostatic hyperplasia (BPH) in men with an enlarged prostate and, in some countries, to reduce the risk of acute urinary retention (AUR) and BPH-related surgery. Dutasteride has been extensively studied in clinical trials involving 7,500 men with BPH. The total estimated post-marketing exposure to Dutasteride from product launch through March 2012 is 8.9 million patient years. The estimated exposure to Combodart/Duodart/Jalyn from product launch through March 2012 is 461,627 patient years.

1.1. Background

The potential role of 5-alpha-reductase inhibitors (5ARI) to reduce the risk of prostate cancer by blocking the conversion of testosterone to dihydrotestosterone was the basis for two large randomized controlled trials. The results of the Prostate Cancer Prevention Trial (PCPT) demonstrated that finasteride taken for 7 years compared with placebo reduced the risk of prostate cancer by 25% in men with normal digital rectal exams (DRE) and prostate-specific antigen (PSA) levels < 3.0 ng/ml at baseline. Recent findings from the Reduction by Dutasteride of Prostate Cancer Events (REDUCE) trial again demonstrated that dutasteride, compared to placebo, taken for 4 years significantly reduced the risk of biopsy detectable prostate cancer by 23% in men without prostate cancer at baseline (confirmed by a negative biopsy within 6 months before enrollment) who were at increased risk of developing the disease by virtue of an elevated PSA (2.5-10 ng/ml).

However, there was a greater proportion of high grade tumors (HGT) observed in the 5ARI groups in both studies compared to placebo (Table 1). In PCPT (Thompson et al., 2003), where biopsies were conducted for cause and an end of study biopsy was offered to participants, tumors of Gleason 7-10 were more common in the finasteride group (n = 280, 6.4%) than in the placebo group (n = 237, 5.1%; P=0.005). (A higher rate of Gleason 8-10 tumors was also observed in PCPT; see Table 1). In REDUCE (Andriole et al., 2010), where biopsies were protocol mandated at 2 and 4 years as well as conducted for cause, 29 (0.9%) patients were diagnosed with Gleason 8-10 tumors in the dutasteride arm, compared with the placebo arm (n = 19, 0.6%; p = 0.15) over 4 years, with the most pronounced treatment group difference in REDUCE in years 3 and 4 (0.5% versus <0.1%, p = 0.003).

Table 1. Results from the two randomized control trials: PCPT and REDUCE: 5ARI use and risk of HGT

PCPT*	(7 years)		Gleason 7-10***	Gleason 8-10***
		Finasteride (n=4368)	280 (6.4%)	90 (2.1%)
		Placebo (n=4692)	237 (5.1%)	53 (1.1%)
		P	0.005	Not reported
REDUCE**	**(4 years)**	**Dutasteride (n=3298)**	220 (6.7%)	29 (0.9%)
		Placebo (n=3406)	233 (6.8%)	19 (0.6%)
		P	0.81	0.15
	Years 1-2	**Dutasteride (n=3239)**	144 (4.4%)	17 (0.5%)
		Placebo (n=3345)	175 (5.2%)	18 (0.5%)
		P	0.15	1.00
	Years 3-4	**Dutasteride (n=2446)**	76 (3.1%)	12 (0.5%)
		Placebo (n=2342)	58 (2.5%)	1 (<0.1%)
		P	0.19	**0.003****

*N=number of men evaluated with for cause biopsies or end of study biopsy.
**N=number of men receiving study mandated biopsies at years 2 and 4 or for cause biopsies.
***N=count of Gleason score 7-10 and 8-10 tumors among men biopsied. Percentages= percent 7-10 and 8-10 Gleason score tumors among men receiving biopsies based on PCPT and REDUCE study definitions for biopsied populations.
****p-value not adjusted for multiple comparisons.

In order to answer the question of whether there was a differential risk for HGT (Gleason 8-10) among those REDUCE subjects who also had BPH at baseline, various post hoc analyses in the REDUCE database have been conducted. The totality of the data resulting from these post hoc analyses is aligned with the overall REDUCE population results and would not suggest a differential risk with dutasteride treatment in those with BPH compared to those with no BPH. Gleason 8-10 cancers rates were constant over

time in the dutasteride group however an imbalance in Gleason 8-10 cancers between dutasteride and placebo was observed in yrs 3-4 (higher rates in dutasteride) driving the overall 4 yrs study differences.

However, data from the REDEEM trial, a 3-year study assessing the efficacy and safety of dutasteride in extending the time to progression of prostate cancer, did not show an imbalance in high grade Gleason score tumors at 18 month and 3 year biopsy in the placebo and dutasteride treated groups. There were 2 Gleason score 8-10 cancers in the placebo group at 18 months versus 0 in the dutasteride group and 3 versus 2 Gleason score 8 tumors in the placebo and dutasteride treated groups, respectively, at the final biopsy [redacted]

Observational studies have also been conducted examining prostate cancer risk and HGT risk among users of 5ARIs. A cohort of men participating in the screening arm of the Finnish Prostate Cancer Screening Trial exposed to finasteride or alpha-blockers (tamsulosin and alfuzosin) for the treatment of BPH were followed for the occurrence of prostate cancer. Overall no significant decreased risk for prostate cancer occurrence was seen in users of finasteride or alpha-blockers in this study, however, the risk of low grade tumors (Gleason 2-6) was significantly decreased among finasteride users. In contrast to PCPT no increased risk of high grade tumors (7-10) was found. However, among men using finasteride for $\geq$ 4 years there was an increased risk of high grade (7-10) tumors compared to non-users of finasteride and alpha-blockers with a borderline significant p-for-trend across increasing duration of exposure categories (p=0.057) (Murtola et al, 2009). The current study will allow for an analysis of duration of exposure to extend these findings.

In a recent study by Pinsky et al prostate cancer survival rates from the Prostate, Lung, Colorectal, and Ovarian (PLCO) trial were used to project prostate cancer mortality based on prostate cancer incidence patterns seen in the REDUCE and PCPT trials. The overall relative risk of prostate cancer mortality comparing treatment to placebo arms was not significantly increased based on projections for patients in the REDUCE or PCPT trials, however mortality rates for those with Gleason scores of 2-6 were modestly decreased while mortality rates for those with Gleason scores of 7 and 8-10 were modestly increased. These results suggest at most a small increase and potentially a modest decrease in prostate cancer related mortality in the treatment arms of these studies (Pinsky et al, 2012). The current study will allow for the real world follow-up of men exposed to these medications over time for the assessment of mortality and will extend these results beyond the framework of hypothetical modelling.

GSK reviews the safety and performance of all its medicines on a continuous basis. Following the decision of the Oncology Drug Advisory Committee in early December 2010 not to recommend the use of dutasteride for prostate cancer prevention, GSK announced in March 2011 that it will no longer pursue global approval (marketing authorisation) for the use of Avodart for this indication. http://www.gsk.com/media/pressreleases/2011/2011_pressrelease_10043.htm. GSK has since added new safety information to the Avodart label concerning the risk of high grade prostate cancer and effects on PSA monitoring.

GSK is committed to further investigations to advance the understanding of the benefits and risks of dutasteride. Over the past several months, the GSK Avodart Team has evaluated the appropriateness of an observational study among men who use 5ARIs for symptoms of BPH to measure the risk of incident high grade tumors (HGT) compared to risk in non-users. In consultation with internal and external experts, the team subsequently concluded that such a study would not be feasible: there is no other drug as a comparator drug that has similar effects on screening, detection, and biology as 5ARIs, resulting in several rate-limiting biases:

- 5ARIs lower PSA levels, potentially differentially affecting prostate cancer screening patterns;
- 5ARIs lower prostate volume, potentially differentially affecting biopsy accuracy;
- 5ARIs are reported to preferentially affect Gleason pattern 3 over pattern 4 and 5 prostate cancer, potentially differentially affecting Gleason score read-out and distribution.

Additionally, based on evidence from REDUCE and PCPT, 5ARIs reduce the risk of prostate cancer occurrence. The outcome of prostate cancer mortality allows for an assessment of whether the relative frequency of high grade tumors among 5ARI versus alpha-blocker users translates into a difference in prostate cancer related mortality. This study will assess the association between 5ARI use and prostate cancer mortality in men treated with BPH medications.

1.2. Rationale

The primary purpose of this study is to assess whether there is an increased risk of prostate cancer mortality associated with the use of 5ARIs in men treated with BPH medications when compared to the use of alpha-blockers in a real-world setting. Since the concern regarding HGTs is their higher agressiveness potential and risk of poorer disease related outcomes, this study will provide evidence as to whether patients who use 5ARIs have a greater likelihood of dying from prostate cancer. Based on a feasibility assessment performed by [redacted] Southern California (see section 4.2, study design) there are very few users of dutasteride in the target study population (4% of 5ARI users). We will therefore pool dutasteride and finasteride into one exposure group and consider a class effect in this study. The study will support the ongoing team efforts regarding the assessment of both the benefits and risks of 5ARIs.

2. OBJECTIVE(S)

Primary Objective:

Prostate cancer mortality

1. To assess the risk of prostate cancer mortality associated with use of 5ARIs, with or without alpha-blockers, compared to alpha-blockers in men treated with BPH medications.

Secondary Objectives:

Prostate cancer mortality or metastatic prostate cancer

2. To assess the risk of prostate cancer mortality or metastatic prostate cancer associated with use of 5ARIs, with or without alpha-blockers, compared to alpha-blockers in men treated with BPH medications.

All cause mortality

3. To assess the risk of all cause mortality associated with use of 5ARIs, with or without alpha-blockers, compared to alpha-blockers in men treated with BPH medications.

Long term exposure to BPH medications

4. To assess the risk of prostate cancer mortality, prostate cancer mortality or metastatic prostate cancer, and all cause mortality associated with long-term exposure (2 or more years cumulative exposure) to 5ARIs, with or without alpha-blockers,
compared to alpha-blockers in the subset of men using long term BPH treatment.

Additional descriptive analyses:

5. To evaluate the validity of classifying prostate cancer deaths with the developed electronic algorithm as compared to those identified based on cause of death coding and physician chart abstraction (see section 4.3: outcome definitions for further details).

6. To evaluate the validity of identifying cases of metastatic prostate cancer with the developed electronic algorithm as compared to those identified based on medical record abstraction.

7. To describe the occurrence of prostate cancer in men treated with BPH medications and across treatment groups (5ARIs and alpha-blockers).

8. To describe the occurrence of metastatic prostate cancer in men treated with BPH medications and across treatment groups.

9. To describe the occurrence of cardiovascular related mortality in men treated with BPH medications and across treatment groups.

10. To describe PSA testing patterns across treatment groups after treatment initiation and over the course of the study period.

11. To describe across treatment groups the frequency of biopsy, the number of cores per biopsy, and the number of positive cores per biopsy.

12. To describe across treatment groups Gleason Score and reclassification of Gleason Score between initial diagnosis biopsy and radical prostatectomy biopsy among those receiving radical prostatectomies.

3. TARGET AUDIENCE

Global safety board, Avodart project team, External scientific community.

4. METHODOLOGY

The proposed study will use a retrospective cohort design with data collected from electronic health records.

Data Source

This study will identify a cohort of patients in the [redacted] closed health care systems with long-term follow-up and low attrition rates. Individual patient data from the four participating sites (Southern California, Northern California, Northwest, and Colorado) will be pooled and analyzed with [redacted] acting as the central data-coordinating center [redacted] will also perform all data analyses for this study. The inclusion of patient data from multiple sites is to ensure a sufficient number of prostate cancer deaths to obtain a suitably precise and clinically meaningful estimate of risk.

[redacted] Regional Overviews

[redacted]) is an integrated health care system that provides comprehensive health services in 8 regions nationwide. Data from four of the larger [redacted] sites: Southern California [redacted]), Northern California [redacted]), [redacted] Colorado [redacted], [redacted] Northwest ([redacted]) will be used in this study. Taken together, these four sites bring racial, ethnic, and geographic diversity to this study. They serve approximately 3.25 million residents of Southern California, 3.3 residents in Northern California, 535,000 residents in Colorado and 477,932 residents in the metropolitan Portland, OR area. These are the sites with the most extensive experience in identifying and classifying cancer patients and they have the most experience in extracting and analyzing cancer-related treatments, costs and outcomes as they all participate in the NCI funded Cancer Research Network (CRN). [redacted] and [redacted] are part of the SEER Tumor Registry system and participate in their respective state tumor registries [redacted] and [redacted] have accredited tumor registries covering their memberships, and participate in their respective state tumor registries. The majority of adult ambulatory cancer care services, including surgery, infused/injected chemotherapy services and radiotherapy are provided at health plan owned facilities, and all four sites purchase pharmaceutical products under the same national contract and employ similar formularies.
Complete data on chemotherapy (including oral as well as injected and infused agents) is available and extractable at all of the participating sites. [redacted] has successfully transitioned to an electronic health record system to support further population and clinical care management at the bedside. The information contained in these electronic records is routinely extracted for research purposes. Thus, information on virtually all aspects of care delivered is captured and retrievable for the proposed research. All of the sites employ the EpicCare® electronic medical record (EMR) system. The EMR systems capture electronically all patient care contacts in the outpatient setting using a controlled medical terminology (based upon ICD-9 and CPT-4 code systems) to document patient assessments, services ordered, and services provided.

Data Availability from Electronic Health Records and Medical Charts

The primary data source for the [redacted] sites will be extracted from the Virtual Data Warehouse (VDW). (See section 7, Study Management, for details on data identification and protection of personally identifiable information). The VDW was originally established by the CRN with funding from NCI for efficient collaboration and pooling of automated data across sites using a standard data formats and definitions. The VDW is virtual in that it is a distributed data system where the health plans retain local control of their data, but a programmer at one site can write a program than can be run at all sites. Using the VDW, standardized automated data on patient characteristics including demographics (age, race, ethnicity, vital status), health plan enrollment, SEER-compatible tumor characteristics (stage and histology), and variables noting the type and date of treatment received (surgery, radiation, chemotherapy, and hormone treatment obtained from tumor registries) can be extracted. In addition, recurrence (local or distal) and date of recurrence are abstracted fields in [redacted] and [redacted] tumor registry files. For [redacted] and [redacted], natural language processing will be used to search electronic chart

information to find evidence of metastases (see Outcome definition section). Utilization files include automated clinical data associated with inpatient and outpatient encounters, types of procedure and diagnostic codes including International Classification of Diseases (ICD-9) codes for procedures and diagnoses, or HCFA Common Procedure Coding System (HCPCS) codes for procedures. Measures of socioeconomic status are also available via the VDW Census data files where enrollees' residential addresses are mapped to census block data using geocoding software.

Extensive data are available through electronic medical records and historical electronic health plan databases which are comprised in the VDW. Specific to this proposal, in the VDW, the following is available: cancer diagnoses going back to 1988, outpatient and emergency room diagnoses and procedures from 1994, laboratory data from 1994 on, mortality data from 1988 on, outside claims from 1990 on, inpatient and outpatient diagnoses and procedures from 1994 to current and pharmacy records from 1992 on. The dates of available data in the virtual data warehouse shared across [redacted] sites are shown in the table below. There is the availability to go further back in historical databases across sites and any data that is needed that is not available electronically can be found in paper medical charts through abstraction as outlined below. Diagnoses are also routinely verified by abstracting a random sample of medical charts and comparing them to electronic medical records.

Data currently available in the [redacted] Virtual Data Warehouse

Data System	[redacted]	[redacted]	[redacted]	[redacted]
Pharmacy	1992-	1995[1]-	1986-	1992-2012
Mortality			1990-current[2]	1990-2011[2]
Underlying cause	1988-2010	1966-2010		
Multiple causes	1988-2009	1966-2009	1990-current[2]	1990-2011
Laboratory	1992-	1988-	1993[3]-	1994 *
Tumor Registries	1988-2009	1974-June 2011	1960-	1987-2011
Utilization, Diagnoses and Procedures	1981-	1995[1]-	1994-	1994-2012

[1]Phased in starting in 1991

[2]Based on tumor registry variable identifying men died of prostate cancer

[3]Earlier data available via archived files and paper records.

*At [redacted] Lab procedures are available for 1994 forward. Lab results are available from 2000 forward. PSAs are available back through 1997. However, there is a capability to get older values from archived files and paper records.

Mortality Data

Death information is derived from several sources. Membership files track notices of active members' deaths. This is most complete when the death occurs at a facility. Two external sources are used in addition, the state death records and Social Security Index records. Standard matching algorithms are used that take into account name, last known address, date of birth and Social Security Numbers (when available) to identify members in these two files. Cause of death codes are assigned by the State mortality registries and follow standard National Center for Health Statistics algorithms. The state death records lag about 14 months for date and underlying cause of death and between two and three years for all causes listed on the death certificate. The mortality files contain, but are not limited to, the following fields: medical record number, date of death, place of death, and ICD-9 coded underlying cause of death.

National Death Index

The National Death Index (NDI) mortality data will be used to supplement cause of death information where is unable to capture it through their existing systems. NDI is a central computerized index of death record information on file in the state vital statistics offices maintained by the Centers for Disease Control. Working with these state offices, the National Center for Health Statistics established the NDI as a resource to aid epidemiologists and other health and medical investigators with their mortality ascertainment activities. The NDI includes a national file of identifying death record information (beginning with 1979 deaths) compiled from computer files submitted by State vital statistics offices. Death records are added to the NDI file annually, approximately 12 months after the end of a particular calendar year. NDI Plus searches provide cause of death codes and are available through 2010 currently (2011 will become available Spring 2013). For known descendants, the CDC charges a flat fee of $5.00 per record for processing, regardless of the number of years processed.

will first submit a NDI application form to NCHS. CDC states that applicants should allow about 2 months for their applications to be reviewed and approved. Once approved, will submit the data from all regions which will include subjects' names, social security numbers, dates of birth, and related information to NCHS on diskette or CD-ROM. Once received, the NDI will mail a password protected CD with the match results to along with the original CD. Analysts at will then follow the NDI's suggested algorithm for determining true matches and merge this data with the existing analytic data set from all regions.

Based on estimates from the pilot study in , it is expected that 30% of the study eligible population will die during the study. Of these men who died, 25% will not have cause of death information and will need to be matched through the NDI. So, approximately 7% of the overall eligible population will be matched. Because and are similar in population size, if an assumption is made that a similar eligible population will be identified in each region, approximately 18000 men will need to be sent to NDI match from these regions (9000 per region). Applying these numbers to the population sizes of NW and CO, it is expected that 1500 deaths will need to be validated

from each of those regions through NDI. So in total, an estimated maximum of **21,000** records will need to be sent to the CDC for cause of death information.

Pharmacy Data

Drug exposure data will be derived from the Pharmacy Information Management System or PIMS, as similar systems are used in all [redacted] sites. All [redacted] facilities pharmacies are now connected via a central data processing operation to a pharmacy database. Records include the patient's unique medical record number, the drug name and strength, NDC code, the drug class, treatment regimen, date dispensed, prescribing physician and department and days' supply. Information enters the system at the time prescriptions are processed by the pharmacy and virtually 100% of dispensed prescriptions are captured, as it is a "real-time" system used to generate the label attached to packaging (e.g., tube, drug container) provided the patient. Information from PIMS is available in the VDW, but may not go back to 1991 for all sites. So data available in the VDW will be supplemented with electronic files (and potentially paper records).

Cancer Registries

[redacted] and [redacted] are part of the SEER Tumor Registry system and participate in their respective state tumor registries. [redacted] and [redacted] have accredited tumor registries covering their memberships, and participate in their respective state tumor registries. The cancer registry database in the VDW contains information on patients who were diagnosed at [redacted] hospitals, or who received at least part of their first course of treatment for cancer at a [redacted] hospital, for all reportable cancers. Variables found in the Regional database can be grouped in the following general categories: 1) Patient Identification (name, sex, DOB, race, MRN, SSN, etc.); 2) Case Identification (type of case, hospital of diagnosis, date of diagnosis); 3) Tumor Information (site, histology, stage, size, lymph node involvement, markers, etc.); 4) Treatment Methods (types and dates of treatment); and 5) Follow-up (vital status, date of last contact or death, tumor status, and cause of death). In addition, [redacted] collects survival information acquired through an aggressive follow-up program that achieves a more than 95 percent retention rate. Because the [redacted] and [redacted] Cancer Registries report to SEER and the [redacted] and [redacted] tumor registries are certified by the NCI, the registry data are 99% complete for both inpatient and outpatient admissions for the diagnosis of new and prevalent cancers. While cancer diagnosis information is available through both the cancer registry and electronic medical records, information regarding the diagnosis and initial treatment of cancer are pulled from the cancer registry. This is because it is a more complete data source, as cancer diagnoses are not systematically recorded in the electronic charts.

Laboratory data

Laboratory data from all four participating [redacted] sites is available in the VDW going back to at least the mid-1990s for all sites. This system tracks both inpatient and outpatient laboratory orders and results. Records include the patient's unique medical record number, the laboratory test name and procedure codes, date drawn, units, references ranges, whether the test was abnormal, ordering physician information, and facility information and the lab test results.

4.1 Study Population

All men age 50 years and older treated with a BPH medication (5ARI and/or alpha-blocker) will be eligible for inclusion. Men using alpha-blocker medications were selected as the comparison group for this study as alpha-blockers are the most common treatment for BPH and men with BPH or lower urinary tract symptoms that are untreated for their condition are likely to be very different from those with severe enough symptoms to seek medical care and take medication. Furthermore, there is no published literature to suggest that alpha-blockers are associated with the risk of prostate cancer occurrence. Participants are not required to have a BPH diagnosis at the time of their first 5ARI or alpha-blocker prescription as based upon data from the feasibility report approximately half of all men received their first recorded BPH diagnosis after initiating treatment. Furthermore, approximately 25% of participants treated with 5ARIs or alpha-blockers did not have a diagnosis code for BPH in their medical record. BPH diagnosis codes were not used in a consistent way historically in the [redacted] data. Sensitivity analyses will be conducted in the analysis phase of the study to examine the effects of BPH diagnosis on the exposure disease relationship. Men should have coverage within the healthcare system for at least 1-year before the first BPH medication prescription. Men with a history of prostate cancer or who develop prostate cancer within <3 months of starting their first BPH medication are not eligible for inclusion.

INCLUSION CRITERIA:

- Male
- A new prescription for BPH medication (5ARI and/or alpha-blocker) in 1992 or later that is identified as appropriate treatment for BPH/LUTS from the [redacted] National Pharmacy guidelines (see section 4.4 Exposure definition).
- Treatment with BPH medication must be initiated prior to Jan1, 2008.
- Age 50 years or older at time of treatment with 5ARI or alpha-blocker.
- At least 1-year of coverage in the healthcare system before the first prescription for BPH medication (5ARI and/or alpha-blocker).
- At least 3 consecutive prescriptions (90 days of supply) for a BPH medication (5ARI and/or alpha-blocker).

EXCLUSION CRITERIA:

- Diagnosis of prostate cancer any time before the first prescription for BPH medication (5ARI and/or alpha-blocker).
- Diagnosis of prostate cancer within 3 months after first BPH medication (5ARI and/or alpha-blocker)
- Patients treated with finasteride 1mg prior to BPH medication. Finasteride 1mg is the dose approved for androgenic alopecia and as the target population for this study is men with treated BPH, we will exclude all men treated with the 1mg dose. Patients treated with 1mg Finasteride will be characterized in terms of which study exposure group they would have transitioned into (5ARI or alpha-blocker) had they been included in the study population, and basic baseline demographic factors.

4.2 Study Design

The study will use a retrospective cohort design with data from four [redacted] sites collected from electronic health records and data abstracted from paper records from 1992-2010.
[redacted] conducted a feasibility study using data from their site only (see Appendix 1). The following inclusion criteria were used in the feasibility study:

- Male
- A new prescription for BPH medication (5ARI and/or alpha-blockers) in 1992 to 2008.
- Prescription for a BPH medication identified as appropriate treatment for BPH/LUTS from the [redacted] National Pharmacy guidelines.Treatment with BPH medication initiated prior to Jan1, 2008.
- Age 50 years or older at time of treatment with 5ARI or alpha-blocker.
- At least 1-year of coverage in the healthcare system before the first prescription for BPH medication (5ARI and/or alpha-blocker).
- No history of prostate cancer diagnosis before first prescription for BPH medication.

The feasibility study identified 123,503 men in [redacted] Southern California exposed to BPH medications and meeting the inclusion criteria; 104,272 men were exposed to alpha-blockers only, 1,619 to 5ARIs only, and 17,612 to combination therapy. Among combination therapy users, the vast majority (92%) used an alpha-blocker and then switched to or added a 5ARI over the course of the study period. As most 5ARI patients have a history of alpha-blocker use or are currently using alpha-blockers, the 5ARI exposure group in this study will include combination therapy users and prior use of alpha-blockers will be a matching factor (see Data analysis section 4.9). Additional

exploratory analyses will be conducted among those using 5ARIs and alpha-blockers concurrently versus those using 5ARIs only (see section 4.9.2 exploratory analyses).

Among those using 5ARI medications in the feasibility study, 96% were exposed to finasteride, with only 4% exposed to dutasteride. Therefore we will combine dutasteride and finasteride users in this study and consider the class effect of 5ARI medications. The ratio of 5ARI users to alpha-blockers users was 1:5.4 in the feasibility study. Assuming that this same ratio will be maintained in the overall study sample, we will match 5ARI:alpha-blocker patients in a 1:5.4 ratio.

Cohort Selection

5ARI users will be selected from the available pool of eligible men and will be matched in a ratio of 1:5 or 1:6 with alpha-blocker users to yield an overall matching ratio of 1:5.4. Matching factors include age (+/- 1 year), timing (calendar year) of BPH treatment initiation (+/- 1 year), race, and duration of prior use of alpha-blockers (see Appendix 2: matching diagram). The purpose of matching on prior use of alpha-blockers is to control for potential confounding factors associated with being treated for BPH and accessing the medical care system such as increased screening and treatment for medical conditions. Additionally, matching on prior use of alpha-blockers will help to control for the use of alpha-blockers in the 5ARI exposure group.

5ARI patients initiating BPH treatment with a 5ARI (i.e. having no prior use of alpha-blockers) will be matched to alpha-blocker users having the same date (+/- 1 year) of alpha-blocker treatment initiation. Categories for the prior use of alpha blockers will be defined based on the distribution of the number of years of alpha-blocker use in the [redacted] data. One year strata for history of alpha-blocker use will be used if the number of available matches across strata is sufficient. Matching will be conducted within Kaiser site ([redacted]). When a suitable match cannot be found within a study site consideration will be given to matching across sites to maximize the available sample for the study. Additionally, if it is difficult to find matches for particular participants, matching criteria may be relaxed to allow for inclusion of all eligible 5ARI men. For example, matching on year of initiation of treatment may be relaxed to (+/- 2 years) to allow for more possible matches.

Two approaches for matching will be explored:

1. Eligible matches for 5ARI users will be defined based on their medication exposure at the time (+/- 1 year) of 5ARI initiation. For example, a patient initiating 5ARI medication in 1995 would have a pool of potential matches that includes all men not previously exposed or currently taking 5ARIs in 1995. Therefore a patient taking an alpha-blocker in 1995 who adds a 5ARI in 1999 would be an eligible match for the 5ARI user as in 1995 they have not yet been exposed to 5ARIs. When this alpha-blocker patient initiates 5ARI therapy in 1999, their follow-up time will be censored.

2. A group of patients exposed to alpha-blockers only over their duration of follow-up will be defined and will serve as potential matches to patients exposed to 5ARI medications.

Matching approach #2 will allow for a cleaner comparison group and for longer duration of follow-up time to be included for alpha-blocker users as follow-up time will not be truncated due to initiation of a 5ARI. However, this approach could introduce bias due to the selection of patients based on post-treatment initiation information. Based on the feasibility study, it is expected that approximately 92% of patients will use alpha-blockers only and 8% will use 5ARIs with or without alpha-blockers. So, there is a low probability that 5ARI users will be matched to alpha-blocker users who eventually start therapy with a 5ARI. Therefore both approaches should yield similar patient groups and we will examine the characteristics of the resulting patient groups under matching schemes #1 and #2.

Follow-up time for each matched group begins at the date of first prescription of a 5ARI medication. For prostate cancer mortality, and all cause mortality, follow-up ends with death, loss to follow-up or the end of the study (defined as last available date in the database, December 2010). For prostate cancer mortality or metastatic prostate cancer follow-up ends with death, occurrence of metastatic prostate cancer, loss to follow-up or the end of the study (defined as last available date in the database, December 2010). Additionally, under matching scenario 1, for alpha-blocker users who are matched to 5ARI users and initiate therapy with a 5ARI, their follow-up time will be censored at the date when they add or switch to a 5ARI medication.

4.3 Outcome definitions

Metastatic Prostate Cancer:
Metastatic Prostate Cancer at the time of diagnosis will be identified using data recorded in the cancer registries. The occurrence of metastatic prostate cancer over the course of the study period will be ascertained using an algorithm to be developed jointly by GSK and [redacted] Based on previous [redacted] studies, the ICD-9 code for metastatic cancer (196.x,197.x, 198.x and 199.x) is underutilized and therefore lacks the sensitivity needed to comprehensively capture cases of metastatic cancer. Using electronic medical records a partly natural language processing algorithm (see natural language processing section below) will be developed. Information from available electronic medical records (procedure codes and results, patient medications, and lab values) including medical chart notes will be included in the algorithm. This information will be used along with the ICD-9 code for metastatic cancer to identify newly occurring cases of metastatic cancer. (see section 4.5 for additional discussion concerning detection bias and metastatic prostate cancer).

Prostate Cancer Mortality:

Summary of the overall process of identification and validation of prostate cancer related deaths

Prostate Cancer deaths will be identified among men with a diagnosis of prostate cancer. First, cause of death codes from state death information or NDI death records will be reviewed for any mention of prostate cancer, CVD, or other causes thought to be related to prostate cancer, as a cause of death. Many previous investigations have relied solely on cause of death information from death certificates to determine whether deaths are related to prostate cancer (Eifler et al, 2012, Abdollah et al, 2012, Jorgensen et al 2012, Kim et al 2012, Scosyrev et al, 2012). The agreement between prostate cancer related cause of death information from death certificates compared to the gold standard of physician chart review has been found to fall in the range of 87% to 97% (Albertsen et al, 2000, Penson et al, 2001). In this study an electronic algorithm will be used to improve the accuracy of the determination of cause of death from death certificates.

Once deaths based on cause of death codes from death certificates have been identified, an algorithm developed by [redacted] NW, that will be adapted for use in this study, will be applied to validate the cause of death as related to prostate cancer based on death certificate coding (see natural language processing and cause of death classification sections below). Algorithms have been used in cancer screening clinical trials to aid in determining cause of death. In the Prostate, Lung, Colorectal, and Ovarian cancer screening trial (PLCO), an algorithm was used to determine which deaths were likely due to cancer and required further review by physicians, and the adjudication committee then determined the cause of death based upon clinical judgement (Miller et al, 2000). In the European Randomized Screening for Prostate Cancer trial (ERSPC), physicians reviewed information abstracted from patient medical records and using pre-determined decision points defined in a cause of death algorithm, classified deaths as related or not related to prostate cancer. Adjudication by a committee of three reviewers was then used to determine final cause of death (De Koning et al, 2003). In this study we will use an electronic algorithm with pre-defined decision points to classify cause of death among men with prostate cancer. A random stratified sample chart review will then be performed to further validate cause of death as determined by the algorithm. (see Appendix 3 flow diagram).

[redacted] *NW algorithm*

[redacted] and [redacted], investigators at [redacted] NW have developed an algorithm for determining the cause of death among prostate cancer patients. This algorithm was developed for use in a multi-site study using data on cause of death from the [redacted] tumor registries. [redacted] has reviewed the charts of prostate cancer patients for several past studies and this algorithm was developed based on her expertise and is a documentation of the process she used to determine cause of death. Additionally, as part of the development of the algorithm, all participating sites sent a list of the causes of death listed for all patients who died and had prostate cancer. This list was reviewed by [redacted] and causes of death deemed to be potentially misclassified and likely due to prostate cancer and not the recorded cause of death, were included in question 2 b. of the algorithm (see Appendix 4). Charts were reviewed for all patients from the time of cancer diagnosis until death for the necessary elements included

in the algorithm. For patients included in the [redacted] NW tumor registry, who die, chart review is performed routinely to determine cause of death and there is a variable available in the tumor registry electronic data indicating whether or not a patient's death is due to cancer. The algorithm correctly captured 98% of the cases identified as prostate cancer deaths by this tumor registry variable. This algorithm will be reviewed by [redacted] and the study sponsor and changes may be made to the algorithm in an attempt to improve its ability to classify deaths correctly. The revised algorithm will be tested with deaths known to be related to prostate cancer identified from the [redacted] NW registry, to ensure that the elements included in the algorithm are sufficient to capture prostate cancer related deaths. Many of the data elements needed for this algorithm can be abstracted from electronic medical records. For the remaining algorithm items for which review of the notes or free text section of electronic medical records would be required, natural language processing will be used.

Natural Language Processing

Approximately 85% of the world's data are unstructured. With the rapid development of computer, internet, and EMR system, the amount of unstructured data are being created at an increasing speed. [redacted] EPIC based EMR system's user-friendly design allows the caregivers to dictate and/or enter free text easily. Those unstructured free text fields contains a plethora of rich information on a patient's status and granular detail of their current therapy(ies).

Natural language processing (NLP) is a field of computer science and linguistics that aims to understand human (natural) languages and facilitate the interaction between human and computer. NLP is a multidisciplinary technology and typically uses statistically based machine learning technology to learn the patterns of human languages. One way NLP is utilized is to identify and extract valuable information from the unstructured 'free text'. Compared to human chart review of medical records, NLP is more efficient and the results are consistent; it can also match and surpass the accuracy of human chart review (Denny et al, 2012).

In [redacted] Research and Evaluation Department, NLP has successfully been used on a series of research projects across different disease areas, such as identifying potential lung cancer patients based on the radiology reports and comparing these results with ICD/CPT codes (Danforth et al, 2012), patient identification and extraction of their related disease characteristics for prostate and breast cancer patients.

For this study, we will utilize NLP in two ways: 1) To help in determining among men with prostate cancer who died, whether their deaths were attributable to prostate cancer and 2) to aid in identifying men with metastatic prostate cancer. We will leverage clinical information including (but not limited to) pathology reports, radiology reports, and clinical notes, to find pre-specified free-text information based on a modified version of the [redacted] NW algorithm mentioned above.

Classification of cause of death among men with prostate cancer

All deaths among men with prostate cancer will be divided into 3 groups based on coding from death certificates (see Appendix 3):

1) CVD related
2) Prostate cancer related
3) Other causes likely related to prostate cancer (to be defined based on the [redacted] algorithm and clinical expertise).

These groups will be defined using underlying cause of death only, and using underlying and contributing causes.

The modified [redacted] NW algorithm, using variables from the EMR and natural language processing (see Appendix 4 for a list of potential variables), will then be applied to a subset of the data and will be tested and validated using random sample chart review in Southern California before being applied to each region to identify potential prostate cancer deaths. As part of this process, the algorithm will be applied to known cases of metastatic disease and deaths known to be related to prostate cancer (based on physician chart review) to ensure that the algorithm can detect these cases. The testing and validation of the algorithm will utilize clinical input from the clinician-investigators, including urologists, epidemiologists, and oncologists as well as NLP specialized programmers, and as needed, chart review to determine whether further tailoring of the algorithm is required. These processes may be repeated as necessary until an agreed upon level of validity of the algorithm is reached (see section 4.9.1 essential analyses). If a suitable sensitivity and specificity is not reached, discussions with GSK and [redacted] investigators regarding other alternatives or iterations needed will take place. Also, a sub-sample chart review within each region's cases will be performed to check for any systematic differences in the algorithm's performance across regions that may arise from within region variability in the reporting of metastatic disease. There will be an assessment of the concordance of the classification of deaths using coding from death certificates alone versus the algorithm.

Once validated, the algorithm will be applied to all deaths among men diagnosed with prostate cancer in all regions to identify men who died from prostate cancer. Based on the results of the algorithm, men will then be categorized as follows:

- Men who died of prostate cancer
- Men who died of cardiovascular disease:
    - Likely due to prostate cancer
    - Not likely due to prostate cancer
- Men who died of other causes:
    - Likely due to prostate cancer
    - Not likely due to prostate cancer

Validation of the electronic algorithm with chart review

Further validation of the cause of death algorithm will be performed on a random sample of 200 records. A case report form will be created including relevant information from patients medical histories to be defined by GSK and [redacted] and chart review by two trained abstractors will be performed to validate the cause of death. A random sample, stratified as follows will be performed:

- Prostate cancer deaths (as identified by algorithm)
- Cardiovascular deaths (as identified by algorithm)
- Other significant deaths (causes of death thought to be potentially related to prostate cancer such as cerebrovascular disease, suicide, etc.) Exact list of causes of death is to be determined.

A panel of clinicians, including the two urologists and the medical oncologist who are [redacted] co-investigators on this study, will adjudicate the cause of death by reviewing the available information on the case report forms abstracted from the stratified random sample chart review and applying clinical judgment to determine if the death was due to prostate cancer. Death information and supporting documentation will be sent to each clinician and blinded to treatment status. Including clinicians from three different sites allows for further insight into any regional differences that may be present in the medical record data. Each reviewer will then assign the cause of death into the three categories as used with the algorithm:

- Men who died of prostate cancer
- Men who died of cardiovascular disease:
    - Likely due to prostate cancer
    - Not likely due to prostate cancer
- Men who died of other causes:
    - Likely due to prostate cancer
    - Not likely due to prostate cancer

The results from the adjudication will then be compared to the cause of death algorithm based on electronic and NLP data and the concordance and discordance will be quantified. The kappa statistic, sensitivity and specificity will be calculated, using the adjudication review panel as the gold standard. In instances where the algorithm and reviewer do not agree or the reviewers do not agree with each other, the case will be discussed among the panel until a consensus is reached regarding the cause of death. The cases will be discussed via teleconference and the majority decision will be used to determine final cause of death.

Detailed documentation concerning the development and validation of the algorithm will be maintained and shared with the study sponsor.

All-Cause Mortality:

All cause mortality in the overall study cohort and among men with prostate cancer will be defined as any death regardless of cause. Death information is derived from several sources. Membership files track notices of active members' deaths. This is most complete when the death occurs at a [redacted] facility. Two external sources are used in addition, the state death records and Social Security Index. In addition, an NDI match will be performed on members who are known to be deceased but are lacking cause of death information. Membership files and the external sources listed above will be searched to identify all deaths occurring in the study cohort.

4.4 Exposure definitions

Exposure will be based upon the treatment administered for BPH. Patients must have at least 3 consecutive prescriptions (90 days of supply) for either a 5ARI or alpha-blocker product to be considered exposed. The following medications will be included in the analysis and the data will be queried using medication name (brand and generic):

EXPOSURES:

- 5 ARI(s): **Finasteride (5mg) or Dutasteride**
    - **Including** Combination Treatment: **Jalyn (fixed dose combination with tamsulosin)**, or any combination of 5ARI & alpha-blocker either concurrently or consecutively. Jalyn was not launched until 2010. So, this medication will be included to capture any patients who have initiated BPH therapy prior to 2008 and switch to fixed dose combination therapy over the course of the study.
    - **Use of medications indicated for androgenic alopecia (Finasteride 1mg) will be identified in the data source to allow for exclusion of patients treated for alopecia.**
- Selective alpha 1 blocker (s)/alpha-adrenergic blocking agent(s): **Alfuzosin, Doxazosin, , Silodosin, Tamulosin, Terazosin.** Prazosin was indicated for BPH treatment according to [redacted] formulary during the study period, but is not currently indicated. We will examine the utilization of Prazosin over the study period and determine whether patients taking this medication should be excluded from the analysis.

Alpha-Blockers		5-alpha reductase inhibitors		Combination therapy
GENERIC NAME	**BRAND NAME**	**GENERIC NAME**	**BRAND NAME**	DUTASTERIDE-TAMSULOSIN HC CAP
ALFUZOSIN HCL ER TAB 10MG	ALFUZOSIN HCL	DUTASTERIDE 0.5MG	AVODART CAP	JALYN
ALFUZOSIN HCL 10MG	UROXATRAL TAB	FINASTERIDE 5MG	FINASTERIDE TAB	
DOXAZOSIN MESYLATE TABLET	CARDURA 1MG	FINASTERIDE TABLET	PROSCAR 5MG	
DOXAZOSIN MESYLATE TABLET	CARDURA 2MG	FINASTERIDE	PROSCAR TAB 5MG	
DOXAZOSIN MESYLATE TABLET	CARDURA 4MG	FINASTERIDE UD	PROSCAR TAB 5MG	
DOXAZOSIN MESYLATE TABLET	CARDURA 8MG	*FINASTERIDE (ALOPECIA)**	*PROPECIA 1MG TABLET*	
DOXAZOSIN MESYLATE 1MG UD	CARDURA TAB	*FINASTERIDE (ALOPECIA)**	*PROPECIA PRO-PAK*	
DOXAZOSIN MESYLATE 2MG UD	CARDURA TAB	*FINASTERIDE (ALOPECIA)**	*PROPECIA TAB 1MG*	
DOXAZOSIN MESYLATE 4MG UD	CARDURA TAB			
DOXAZOSIN MESYLATE 8MG UD	CARDURA TAB			
DOXAZOSIN MESYLATE MESYLATE TAB 1MG	DOXAZOSIN			
DOXAZOSIN MESYLATE	DOXAZOSIN			

MESYLATE TAB 2MG			
DOXAZOSIN MESYLATE MESYLATE TAB 4MG	DOXAZOSIN		
DOXAZOSIN MESYLATE MESYLATE TAB 8MG	DOXAZOSIN		
PRAZOSIN HCL CAPS UD	MINIPRESS 1MG		
PRAZOSIN HCL CAPSULE	MINIPRESS 1MG		
PRAZOSIN HCL CAPSULES	MINIPRESS 1MG		
PRAZOSIN HCL CAPS UD	MINIPRESS 2MG		
PRAZOSIN HCL CAPSULE	MINIPRESS 2MG		
PRAZOSIN HCL CAPS UD	MINIPRESS 5MG		
PRAZOSIN HCL CAPSULE	MINIPRESS 5MG		
PRAZOSIN HCL 1MG	MINIPRESS CAP		
PRAZOSIN HCL 1MG UD	MINIPRESS CAP		
PRAZOSIN HCL 2MG	MINIPRESS CAP		
PRAZOSIN HCL 2MG UD	MINIPRESS CAP		
PRAZOSIN HCL 5MG	MINIPRESS CAP		
PRAZOSIN HCL 5MG UD	MINIPRESS CAP		
PRAZOSIN HCL CAP	PRAZOSIN 1MG		
PRAZOSIN HCL CAPSULE	PRAZOSIN 1MG		
PRAZOSIN HCL STARTER KIT	PRAZOSIN 1MG		
PRAZOSIN HCL CAP	PRAZOSIN 2MG		
PRAZOSIN HCL CAPSULE	PRAZOSIN 2MG		
PRAZOSIN HCL CAP	PRAZOSIN 5MG		
PRAZOSIN HCL CAPSULE	PRAZOSIN 5MG		
PRAZOSIN HCL CAP 1MG	PRAZOSIN HCL		
PRAZOSIN HCL CAP 1MG UD	PRAZOSIN HCL		
PRAZOSIN HCL CAP 2MG	PRAZOSIN HCL		
PRAZOSIN HCL CAP 2MG UD	PRAZOSIN HCL		
PRAZOSIN HCL CAP 5MG	PRAZOSIN HCL		
PRAZOSIN HCL CAP 5MG UD	PRAZOSIN HCL		
PRAZOSIN HCL PINK CAP 2MG U	PRAZOSIN HCL		
PRAZOSIN HCL WHITE CAP 1MG	PRAZOSIN HCL		
SILODOSIN 4MG	RAPAFLO CAP		
SILODOSIN	RAPAFLO CAP		

8MG			
TAMSULOSIN HCL	FLOMAX 0.4MG		
CAP UD			
TAMSULOSIN HCL	FLOMAX CAP		
0.4MG			
TAMSULOSIN HCL	TAMSULOSIN HCL		
CAP 0.4MG			
TAMSULOSIN HCL	TAMSULOSIN HCL		
CAP 0.4MG UD TERAZOSIN HCL			
TERAZOSIN HCL	HYTRIN 1MG		
TABLET			
TERAZOSIN HCL	HYTRIN 1MG		
TABS UD			
TERAZOSIN HCL	HYTRIN 2MG		
TABLET			
TERAZOSIN HCL	HYTRIN 2MG		
TABS UD			
TERAZOSIN HCL	HYTRIN 5MG		
TABLET			
TERAZOSIN HCL	HYTRIN 5MG		
TABS UD			
TERAZOSIN HCL	HYTRIN CAP		
10MG			
TERAZOSIN HCL	HYTRIN CAP		
10MG UD			
TERAZOSIN HCL	HYTRIN CAP 1MG		
TERAZOSIN HCL	HYTRIN CAP 1MG		
UD			
TERAZOSIN HCL	HYTRIN CAP 2MG		
TERAZOSIN HCL	HYTRIN CAP 2MG		
UD			
TERAZOSIN HCL	HYTRIN CAP 5MG		
TERAZOSIN HCL	HYTRIN CAP 5MG		
UD			
TERAZOSIN HCL	HYTRIN TAB 10MG		
UD			
TERAZOSIN HCL	HYTRIN TAB 1MG		
TERAZOSIN HCL	HYTRIN TAB 1MG		
UD			
TERAZOSIN HCL	HYTRIN TAB 2MG		
TERAZOSIN HCL	HYTRIN TAB 2MG		
UD			
TERAZOSIN HCL	HYTRIN TAB 5MG		
TERAZOSIN HCL	HYTRIN TAB 5MG		
UD			
TERAZOSIN HCL	TERAZOSIN 1MG		
CAP			
TERAZOSIN HCL	TERAZOSIN 1MG		
CAP STARTER K			
TERAZOSIN HCL	TERAZOSIN 1MG		
CAP STARTER P			
TERAZOSIN HCL	TERAZOSIN 2MG		
CAP			
TERAZOSIN HCL	TERAZOSIN HCL		
2MG CAP			
TERAZOSIN HCL	TERAZOSIN HCL		
CAP 10MG			
TERAZOSIN HCL	TERAZOSIN HCL		
CAP 10MG UD 1			
TERAZOSIN HCL	TERAZOSIN HCL		
CAP 1MG			
TERAZOSIN HCL	TERAZOSIN HCL		
CAP 1MG UD			
TERAZOSIN HCL	TERAZOSIN HCL		

CAP 2MG TERAZOSIN HCL CAP 2MG UD TERAZOSIN HCL CAP 5MG TERAZOSIN HCL CAP 5MG UD TERAZOSIN HCL STARTER CARD 1MG/	TERAZOSIN HCL TERAZOSIN HCL TERAZOSIN HCL TERAZOSIN		

*Drugs for male-pattern baldness will also be collected in order to separate out later

EXPOSURE CATEGORIES:

1) Any 5ARI use (at least 3 prescriptions (90 day supply)) including patients with concomitant or consecutive alpha blocker use.
2) Alpha-blocker users (at least 3 prescriptions (90 day supply)).

DURATION OF EXPOSURE

Duration of exposure among all 5ARI or alpha blocker users with at least 3 consecutive prescriptions (90 days supply) will begin at the start of the follow up time for each patient and will be categorized as:

1) Any exposure
2) < 1 year
3) 1 - <2 years
4) 2 or more years

Cumulative Exposure:

Define **drug era** for each eligible man:

- When the days of drug supply is available, use this info to determine whether drug use between two adjacent prescriptions are continuous;
- When the days of drug supply is NOT available, assume average prescription cycle as 30 d *(This was learned to be the case from previous projects – the median gap between any two adjacent prescriptions was 30 d. This assumption will be confirmed in the [redacted] data before analysis begins).*
- If the gap between two prescriptions is <= 30 d, count it as continuous, as one drug era. Otherwise, assume there was a gap, and start another drug era after the gap.

Duration of one/each drug era (for each drug or drugs)

- If the adjacent prescription periods do NOT overlap: drug era = From 1st prescription date to "last prescription date + days of drug supply for the last prescription (30 d, if not available)" or the end of observational period, whichever comes first.
- If the adjacent prescription periods DO overlap: drug era = From 1st prescription date to "last prescription date + days of drug supply for the last prescription (30 d, if not available)" or the end of observational period, whichever comes first + days of overlap.

Duration of cumulative exposure = Duration of exposure for drug era 1 + Duration of exposure for drug era 2 +…. + last drug era

4.5 Other Variables (Including potential Confounders and effect modifiers)

Variable definitions for descriptive analyses:

Prostate cancer:
Prostate cancer diagnosis will be identified in all participating [redacted] sites through the tumor registries. These registries contain information on patients who were diagnosed at [redacted] hospitals, or who received at least part of their first course of treatment for cancer at a [redacted] hospital, for all reportable cancers. These diagnoses are based on the established and validated methods that SEER uses (based on ICD-03 codes) by cancer site. C619 identifies prostate cancer as the primary site. Detailed information regarding the SEER coding and staging methodology can be found at:http://seer.cancer.gov/tools/codingmanuals/.

Cardiovascular related mortality:
Cardiovascular related mortality is defined as deaths due to cardiovascular-related causes as coded by State mortality registries or NDI. The state registries follow the National Center for Health Statistics algorithm for assigning CVD death. Detailed instructions regarding the NCHS methods for cause of death determination can be found at: http://www.cdc.gov/nchs/nvss/instruction_manuals.htm. All deaths that occur among men with prostate cancer that have cardiovascular disease mentioned will be captured and these deaths will similarly be validated using the [redacted] algorithm. Cardiovascular deaths

will be captured as a mutually exclusive category from prostate cancer related deaths. Therefore care will be taken in determining cause of death in patients with advanced prostate cancer whose underlying cause of death is listed as cardiovascular disease based on state death certificate information. Additionally, all deaths among the overall cohort where cardiovascular disease is listed as the underlying cause of death will be identified as part of descriptive analyses when assessing all cause mortality.

PSA testing patterns:

Laboratory data from all four participating [redacted] sites is available in the VDW going back to at least the mid-1990s for all sites. This system tracks both inpatient and outpatient laboratory orders and results. Records include the patient's unique medical record number, the laboratory test name and procedure codes, date drawn, units, references ranges, whether the test was abnormal, ordering physician information, and facility information and the lab test results. [redacted] has rich PSA testing data for the study period. In the past 3 years alone in [redacted], we have approximately 25,000 men per year over the age of 50 who have received at least one PSA test, and we have administered almost 1 million PSA tests in this time period. PSA tests for all 4 [redacted] sites are currently housed within the VDW and have gone through quality assurance checks for previous studies. CPT codes which apply for finding PSA tests are [redacted].

Biopsies:

Prostate biopsies will be identified within each region according to either CPT or internal [redacted] codes. Care will be taken with the change in coding practices over time within region. Potential CPT codes that will be used include [redacted], [redacted] and [redacted]

Gleason Scores:

Gleason scores are only available in more recent years from the cancer registry (approximately 2007 forward depending on region). Gleason scores are however, available using natural language processing from the pathology reports of both biopsy specimens and surgical specimens, when available. Changes in Gleason score definition over time will be taken into account when abstracting the data. The primary and secondary patterns are added together to create a score. Primary pattern is doubled when there is no secondary pattern. Tertiary pattern is not used to determine Gleason score.

Prostate cancer treatment paradigms:

Prostate cancer treatment will be identified through the cancer and tumor registries at each site. The registries contain information on surgical, chemotherapy, hormone therapy, brachytherapy, external beam radiation and active surveillance treatment choices. These are based on codes for treatment as outlined by SEER. This is the most reliable source of treatment information within the [redacted] system.

Confounders and effect measure modifiers

Information on potential confounders will be collected in the 1 year pre-treatment initiation period for all outcome analyses. Age, race, and family history of prostate cancer are the only consistent risk factors for prostate cancer incidence and prostate cancer death based on the published literature. While obesity, diet, exercise, and socioeconomic status have been associated with risk of prostate cancer and prostate cancer death, a clear causal relationship between these risk factors and prostate cancer has not been established (Brawley, 2012). Matching by age and race in the design phase of this study will help to address potential confounding by these important risk factors. It is not possible in this study to control for family history of prostate cancer as this variable is not systematically collected by [redacted] The relative frequency of variables know to be independent risk factors for the outcomes of interest will be compared between the 5ARI and alpha-blocker treatment groups. If an imbalance is found in any of the independent risk factors across the two treatment groups, these variables will be considered for inclusion in the Cox proportional hazards models or as stratification variables.

Potential confounding factors:

Prostate cancer mortality, prostate cancer mortality or metastatic prostate cancer occurrence:

Potential confounding factors at baseline:

- Race/Ethnicity
  - Race and ethnicity is collected across sites and based on a race and ethnicity variable that is categorized as: Non-Hispanic Whites, African Americans, Hispanic, Asian/Pacific Islander, Other and Missing.

- Age
  - Age is routinely collected and calculated from date of birth.

- Socioeconomic status (SES)
  - Aggregate SES measures for [redacted] members were calculated via geocoding in 2002 using 2000 US census estimates at the block, block-group, tract and ZIP level and include race, income and education.

- PSA
  - PSA is available through our laboratory data in the VDW and is measured in ng/mL.

- BMI
    - BMI was not routinely collected as a vital sign until the implementation of the electronic medical record in each site. Therefore, only a few years of complete data may be available. It is calculated using standard formula and measured in kg/m2.

- Smoking status
    - Smoking status information also was collected as part of patient's history during office visits starting with the implementation of the electronic medical record system (mid-2000s). Information on tobacco use (ever, former, never), pack-years, type of tobacco product used (cigarettes, chewing, pipe, cigar) is available.

- Charlson Comorbidity Index
    - Charlson comorbidity Index is collected based on ICD-9 codes of the included conditions/complications and a standard algorithm and coding macro are used to derive the score which is usually categorized into (0, 1 or 2+).

- CV endpoints
    - Data from diagnosis, lab, and pharmacy are available to define any of these endpoints.
        - History of cardiovascular disease
        - History of high blood pressure
        - History of diabetes
        - History of hyperlipidemia

- History of cancer other than prostate
    - History of cancer other than the prostate will be determined through the cancer registry and tumor registries at each site. A standard SEER coding system is used to identify cancer diagnosis, diagnosis date and site.

Potential confounding factors over the follow-up period:

Additionally information on several variables occurring over the follow-up period will be collected including:

- PSA testing patterns after treatment initiation and over the course of the study period.
- The frequency of biopsy, the number of cores per biopsy, and the number of positive cores per biopsy.
- Gleason Score at diagnosis and reclassification of Gleason Score between initial diagnosis biopsy and radical prostatectomy biopsy among those receiving radical prostatectomies.

- Frequency of prostate cancer treatment paradigms (radical prostatectomy, external beam radiation, brachytherapy, cryotherapy, etc.).

Imbalances in these variables across treatment groups will be examined as they may help to explain some of the association between the exposure groups and occurrence of prostate cancer mortality and metastases.

Potential effect measure modifiers for consideration:

Race/ethnicity

Age

SES

Additional considerations for ascertainment of metastatic prostate cancer cases at diagnosis and over the follow-up period

Care will be taken to assess any detection bias across treatment groups in the ascertainment of metastatic cancer cases at the time of diagnosis and over the course of the follow-up. At the time of diagnosis we will examine the frequency of occurrence of tests to ascertain metastatic prostate cancer across the treatment groups including bone scans, PSA tests, and other relevant tests. Over the course of the follow-up we will also document the frequency of occurrence of these tests for metastatic cancer across the treatment groups as part of the development of the natural language processing algorithm for metastatic prostate cancer. This information can be used to help to quantify any detection bias that might be present in the study and potentially help to control for this bias in the analysis phase of the study.

All cause mortality:

Potential confounding factors at baseline:

- Age
- Race/Ethnicity
- SES
- BMI
- Smoking status
- Charlson Comorbidity Index
- History of cancer other than prostate
- Other medications patients may be taking during the study period (to be determined as part of the statistical analysis plan using data from [redacted]

4.6 Adverse drug experience/event measures

Information on adverse events will not be collected in this retrospective analysis. Adverse events will be reported according to [redacted] Safety Reporting from Epidemiology Studies and Analyses of Epidemiology Databases.

4.7 Data collection and management

We will use data from four [redacted] sites: [redacted] Southern California, [redacted] Northern California, [redacted] North West, and [redacted] Colorado. The majority of the data is available electronically. Data for some variables will be abstracted from paper based medical records (see section 4.1 Data source). [redacted] Southern California will serve as the coordinating site for the study with all other sites sending de-identified data to the Southern California site for compilation into one large data set [redacted] Southern California will also act as a central IRB with all other study sites ceding IRB authority so that one central IRB approval for the study will be obtained.

4.8 Validation procedures

See outcome definitions (section 4.3)

4.9 Data analysis

Assessment of the success of matching and description of the sample

The first step for these analyses is to identify and quantify the amount of imbalance that existed between the two groups prior to matching. This involves calculating a series of statistics that identify which baseline characteristics are different between the treated and control groups. We now describe the statistics that we will examine for continuous and categorical variables.

Continuous variables will be compared using two-sample *t*-statistics, variance ratios, and /or standardized differences in percent for each variable. The standardized difference in percent is defined as the mean difference as a percentage of the average standard deviation:

$$\frac{100(\bar{x}_c - \bar{x}_t)}{\sqrt{\frac{(s_c^2 + s_t^2)}{2}}}$$

Where for each covariate $\bar{x}_c$ and $\bar{x}_t$ are the sample means in the alpha-blocker and 5ARI groups, respectively, and the s_c^2 and s_t^2 are the corresponding sample variances. The variance ratio is defined as s_c^2 / s_t^2.

Significant differences based on the two sample *t*-statistics, variance ratios that are larger than 1.5 or standardized differences larger than 20% would indicate that there is substantial imbalance between the two groups for that variable. Ideally, there would be no significant *t*-statistics, all variance ratios would be between .8 and 1.2 and standardized differences in percent would be less than 10% if the two groups were well balanced (as would be the expectation if they had been randomized).

For categorical/binary variables we will compare chi-square statistics and observed proportions by treatment group. Significant differences based on chi-square statistics or large differences in observed proportions would suggest that there is an imbalance between the groups on the variables being examined.

The next step is to compare the matched "treated" (5ARI) and "control" (alpha-blocker) patients on the four matching variables (age, timing of treatment initiation, race, and duration of prior use of alpha-blockers) to confirm that indeed the groups were successfully matched on these characteristics. The success of the matching will be assessed based on whether balance between the treated and control groups is achieved in the matched samples.

Next, background characteristics (including cumulative exposure time, follow-up time, follow-up time after cancer diagnosis, and potential pre-treatment initiation confounding

factors) will be compared between the groups to identify whether there exist variables that would need to be considered in the outcome analyses as covariates

4.9.1 Essential analyses

Primary analysis (prostate cancer mortality)

Once the cohort has been selected and the groups are confirmed to have been adequately matched on age, date of treatment initiation, race, and duration of prior use of alpha-blocker, then the outcome analyses will be performed. As described above, descriptive statistics will be summarized overall and according to treatment group for age at first prescription, race, cumulative exposure time, follow-up time, other co-morbid conditions, and clinical variables of interest. Each patient included in the analysis will have their survival outcome determined as follows. All patients who die from prostate cancer will be considered as events. Patients who die from other causes or who are lost to follow-up (leave the [redacted] system) will be coded as censored at the time of death or loss to follow-up. We will search the NDI for deaths occurring within 2 years of patients leaving [redacted] Any deaths determined to be associated with prostate cancer based on information available in NDI will be assigned to the appropriate treatment group based on treatment information available in [redacted] up until patients left the system. Patients who are alive at their last follow-up visit will be censored at that time.

An initial Kaplan Meier curve will be estimated comparing the 5ARI vs alpha-blocker users for prostate cancer mortality without any additional adjustments. Additionally, a plot of cumulative incidence, adjusting for competing risks of death, will be constructed allowing for the investigation of the effect of competing risks on the Kaplan-Meier probability estimates.

Next, we will examine pre-treatment assignment characteristics such as prior co-morbidities (if available) for each treatment group. Then we will examine post-treatment assignment characteristics such as cumulative exposure to the treatment, latency (time since last treatment), and diagnosis of prostate cancer (yes/no). Among those patients who develop prostate cancer we will then compare between the two groups the duration of prostate cancer, the prostate cancer stage at diagnosis and Gleason score at diagnosis.

To better understand the pattern of changes in absolute mortality rates and hazard ratios, by duration of exposure, the data will then be stratified by duration of 5ARI/alpha-blocker exposure: any exposure >3months, < 1 year, 1 - <2 years, 2+ years and the crude mortality rates and hazard ratio will be estimated within strata of exposure. Given power limitations due to the likely number of deaths across exposure duration strata, the data will be pooled for use in regression analyses.

The data will then be stratified using different lag times to be defined based on the distribution of time from BPH treatment initiation to prostate cancer mortality seen in the

data set. The distribution of time from treatment initiation to death in the cohort will be examined.

Additionally, the mortality rates for those using 5ARIs only and those using 5ARIs with an alpha-blocker will be examined separately. Given the likely small size of patients using 5ARIs alone without alpha-blockers these groups will likely be combined in the analysis, however, combining groups assumes that the mortality rate among patients exposed to 5ARIs only is equivalent to that of patients using 5ARIs with an alpha-blocker.

Next, the data will be pooled across exposure and lag time strata and Cox proportional hazard regression models will be fit in the overall data set to compare prostate cancer mortality between groups while adjusting for the pre-treatment characteristics identified above. Of interest will be testing whether there are any potential interactions among the pre-treatment patient characteristics and the treatments (5ARI vs alpha-blocker) and if such interactions are found, this may suggest that the Cox proportional hazards models may need to be fit stratified by the characteristics of interest. Post-treatment characteristics will not be adjusted for in the primary analysis as they may be treatment effects or in the causal path between the exposures and the outcome of prostate cancer mortality. If large imbalances exist in post-treatment variables of interest, stratified analyses may be considered.

An issue that we will confront in analyses is that the outcome of interest is prostate cancer mortality, yet the groups will be selected for analysis prior to the diagnosis of prostate cancer and in fact many patients may not develop prostate cancer at all during this study. Therefore, although the primary outcome of interest is prostate cancer mortality, prostate cancer diagnosis will be examined descriptively across treatment groups.

Secondary analyses:

A similar approach in terms of classifying follow-up time, constructing Kaplan Meier curves, examination of pre- post-treatment assignment characteristics, and stratification by exposure duration and lag time, as detailed above for the primary analysis for prostate cancer related mortality, will be used for the following analyses:

- Prostate cancer mortality or metastatic prostate cancer occurrence
- All cause mortality
- Long terms users of BPH medications

Among users of 5ARIs and alpha-blockers, who use these medications for 2 years or more, the association between 5ARI exposure and prostate cancer related mortality will be estimated.

Cox proportional hazard regression models will then be fit adjusting for imbalances in pre-treatment assignment characteristics. Characteristics that are balanced between the groups may also be included in regression models to increase precision in the treatment effect estimates.

Descriptive analyses:

Prostate cancer

The percent of patients diagnosed with cancer and the incidence rate per 1,000 person years of prostate cancer will be calculated, overall and by treatment group. The data will then be stratified and the incidence of prostate cancer by duration of 5ARI/alpha-blocker exposure will be estimated: any exposure, >3months, < 1 year, 1 - <2 years, 2 or more years. The stage and grade distribution of patients diagnosed with prostate cancer will be determined, overall and by treatment group.

Additionally, information on several variables occurring over the follow-up period will be collected, including:

- BPH related surgery and hospitalizations
    - BPH surgeries will be captured using CPT codes and [redacted] internal codes for older data. Procedures for BPH included most commonly in our system include: transurethral resection of the prostate (TURP) (CPT: 52601, 52612, 52614) and visual laser ablation of the prostate (VLAP) (CPT: 52647 and 52648). However, we will also collect information on: transurethral needle ablation (TUNA) (CPT: 53850 and 53852), high-intensity focused ultrasound (HIFU) (53850), Prostatectomy for benign disease (55801, 55821, 55831).
    - Hospitalizations due to BPH will be captured by abstracting the discharge codes for any inpatient stay and determining whether BPH was one of the conditions coded. The ICD-9 diagnosis code for BPH (600.x) will be used as well as internal [redacted] codes for older data where the ICD-9 code for BPH was not reliably assigned.

- PSA testing patterns after BPH treatment initiation and before cancer diagnosis occurring over the course of the study period..
- Frequency of biopsy, the number of cores per biopsy, and the number of positive cores per biopsy, after BPH treatment initiation and before cancer diagnosis..

Imbalances in these variables across treatment groups will be examined as they may help to describe potential detection bias in the diagnosis of prostate cancer

Metastatic prostate cancer
The percent of patients with metastatic prostate cancer and the incidence rate per 1,000 person years of metastatic prostate cancer will be calculated overall and by treatment group among patients who do not die during the study period. The data will then be stratified and the incidence of metastatic prostate cancer by duration of 5ARI/alpha-blocker exposure will be estimated: any exposure, >3months, < 1 year, 1 - <2 years, 2 or more years. The occurrence of metastatic disease by stage and grade at initial diagnosis will also be examined.

Cardiovascular related mortality
The percent of deaths overall and by treatment group, among men with prostate cancer, due to cardiovascular disease will be estimated as part of the validation process for prostate cancer related deaths. Additionally, the percent of deaths in the overall cohort due to cardiovascular disease will be estimated, overall and by treatment group.

Validation of prostate cancer mortality and metastatic prostate cancer algorithms
The sensitivity, specificity, positive and negative predictive values for the prostate cancer mortality and metastatic prostate cancer algorithms will be assessed in comparison to medical record abstraction. GSK and [redacted] will agree on an adequate level of sensitivity, specificity and predictive value for the algorithms based on the results from the validation process. If numerous iterations of NLP algorithm application and medical chart abstraction cannot produce an adequate level of validity for the algorithms, chart abstraction may need to be used for a larger sample of medical records and GSK and [redacted] will determine how best to proceed and how this will impact the scope of work and timelines for the study.

[redacted] will have 2 abstractors responsible for abstracting medical records, reviewing, and determining the cause of death (or metastatic disease status) for patients included in the validation studies. A third [redacted] person will be responsible for adjudicating cause of death (or metastatic disease status) when disagreements arise between abstractors. The kappa statistic will be calculated for agreement between abstractors.

PSA testing patterns across treatment groups after treatment initiation and over the course of the study period
The percent of patients with a PSA test and the mean and median number of PSA tests over the follow-up will be estimated overall and by treatment group.

Frequency of biopsy, the number of cores per biopsy, and the number of positive cores per biopsy

The percent of patients with a biopsy and the mean and median number of biopsies over the follow-up will be estimated overall and by treatment group. Similarly the mean and median number of cores per biopsy and number of positive cores per biopsy will be estimated.

The percent of patients with a negative biopsy in the 1-year pre-baseline period will be calculated overall and by treatment group. Additionally, the percent of patients with a negative biopsy within their entire [redacted] medical history pre-baseline will be calculated overall and by treatment group. A biopsy is usually performed in response to a signal related to the possible presence of prostate cancer such as increasing PSA or detection of a lump during digital rectal examination. History of negative biopsy will be collected as a way to describe the prostate cancer risk among patients in the sample before the start of the study period.

Gleason Score and reclassification of Gleason Score between initial diagnosis biopsy and radical prostatectomy biopsy among those receiving radical prostatectomies

The distribution of Gleason score at prostate cancer diagnosis will be determined overall and by treatment group. Changes in Gleason score definition over time will be taken into account when abstracting the data. Among those patients who have a radical prostatectomy, Gleason scores will again be determined and any reclassification of Gleason score from initial diagnosis biopsy based on a radical prostatectomy biopsy will be noted.

4.9.2 Exploratory analyses

Primary analysis:

- Sensitivity analyses will be conducted including the following:

- The data will be stratified by cumulative dose of 5ARI medication and the association between 5ARI exposure and prostate cancer related mortality will be estimated within each stratum. Categories for cumulative dose will be defined based on the distribution of cumulative dose in the data and independently from outcome status. Preliminary categories will be defined as part of the development of the statistical analysis plan.

- After examining the treatment data independently from the mortality data we will define groups of patients based on gaps in their exposure history and will conduct sensitivity analyses to examine how the exposure-disease relationship might vary based on these groups. For example, participants with > 9 month gaps between drug eras and a maximum duration of drug era of 3 months over the course of follow-up might constitute one group while those with cumulative exposure of >2 years with <3 months of time between drug eras (time off of treatment) over the course of the study period may constitute another. In this way, we will examine the validity of the assumption that exposure time can be summed across treatment eras without accounting for gaps in treatment (main study analysis).

- The data will be stratified by time since the end of BPH treatment to prostate cancer related mortality (< 1 year, 1- <2years, 2 to 3 years, 3+ years) to see how the main association differs within each of these groups.

- When patients drop out of the [redacted] system they will be censored at the time they leave the system for the primary analysis of prostate cancer mortality. The NDI will be searched for deaths occurring within the study period among men who dropped out of [redacted]. Any deaths determined to be associated with prostate cancer based on information available in NDI will be assigned to the appropriate treatment group based on treatment information available in [redacted] up until patients left the system. A sensitivity analysis will be conducted including these deaths to see if the primary association differs with the inclusion of these additional cases.

- Deaths due to prostate cancer occurring in the first year after treatment initiation will be excluded from the analysis as it is not biologically plausible that BPH treatments could cause prostate cancer and subsequent death within the first year of treatment exposure.

- Men with a history of cancer other than prostate will be excluded from the analysis to see how the main association differs when this groups is excluded.

- Men without a BPH diagnosis will be excluded from the analysis to see how the main association differs when this groups is excluded.

- A subgroup of men will be defined who had an increase in their PSA level. Some of the men will have received a 5ARI in response to their PSA increase and some will not. 5ARI patients will be matched to alpha-blocker patients and we will examine prostate cancer mortality in these groups. This analysis will allow us to examine the effect of 5ARI initiation when it is triggered by PSA increases..

- We will stratify 5ARI exposed patients into 2 groups: those using 5ARIs as monotherapy and those using 5ARIs concurrently with alpha-blockers. We will examine descriptively whether there are any differences between patients using these different treatment regimens with respect to prostate cancer mortality.

- If the results from stratified analyses indicate heterogeneity across strata of post-treatment characteristics of interest (PSA testing patterns after treatment initiation and over the course of the study period, frequency of biopsy, the number of cores per biopsy, the number of positive cores per biopsy, and frequency of prostate cancer treatment paradigms), marginal structural models (Hernan et al, 2000) may be used to explore the potential effects of post-treatment characteristics on the relationship between BPH medication use and prostate cancer related mortality.

Secondary analyses:

Prostate cancer mortality or metastatic prostate cancer occurrence, all cause mortality

Depending upon the results from the exploratory analyses for the primary outcome of prostate cancer related mortality we will consider repeating some of the exploratory analyses listed above, as appropriate, among long term users of BPH treatment and for the outcomes of:

- prostate cancer mortality or metastatic prostate cancer occurrence

- all cause mortality

Descriptive analyses:

Prostate cancer

- We will stratify the data by cumulative dose of 5ARI medication and estimate the incidence of prostate cancer overall and within each stratum. Categories for cumulative dose will be defined based on the distribution of cumulative dose in the data and independently from outcome status.

  We will stratify the data using different latency periods to be defined based on the distribution of time from BPH treatment exposure to detection of cancer seen in the data set.

4.9.3 General considerations for data analyses

Matching on pre-treatment characteristics, considered major potential confounders based on the literature and expert opinion, will be conducted to minimize potential confounding by these covariates. Additionally, covariate adjustment and stratification will be used in the analysis phase of the study to adjust for imbalances in baseline characteristics between exposure groups.

4.9.4 Data handling conventions

N/A

5. SAMPLE SIZE AND POWER/PRECISION CALCULATIONS

Assuming a hazard ratio of 1.0, 90% power and a 0.05 two sided alpha we would need approximately 850 deaths to rule out a hazard ratio of 1.25 or higher. This calculation is based on equation 2 in the article by Saville et al (Saville et al, 2011) and assumes a 1:1 ratio of patients across treatment groups. A 25% increase in risk of prostate cancer is thought to be a signal of concern. While this study is designed to detect any difference (increase or decrease) in risk between 5ARI and alpha-blocker users, given the potential limitations of power in this study, if a difference in risk is not found between the exposure groups, this will not be interpreted to rule out an increased risk of prostate cancer related mortality associated with the use of 5ARIs. If no increased risk is found for 5ARI users compared to alpha-blocker users, this result will add to the body of evidence used to interpret the findings from the REDUCE trial of an increased number of HGTs in the dutasteride arm compared to the placebo arm. Further consideration for the possible results from the analysis (e.g. a hazard ratio of 1.15 with a confidence interval not including 1) and how they will be interpreted will be included in the statistical analysis plan.

Based on the feasibility assessment using [redacted] data, the allocation of patients across treatment groups (alpha-blocker:5ARI) is expected to be 5.4:1. With this unequal allocation of patients in the exposed and unexposed groups an adjustment must be made to the sample size calculation. Using the adjustment suggested by Saville et al with the same assumptions for power, 1,500 prostate cancer related deaths (214 in the 5ARI exposed group and 1,286 in the alpha-blocker only group) are needed to rule out a HR of 1.25 or higher. This calculation assumes a 6:1 allocation of unexposed (alpha-blocker) vs. exposed (5ARI) and equal risk in both exposure groups.

The target of 1,500 deaths should be reached by pooling data across four [redacted] sites. 846 deaths were observed in the feasibility study. Roughly the same number of deaths from [redacted] can be expected and about 15% of this amount from [redacted] and [redacted], yielding an estimated total number of deaths of [redacted]. When these prostate cancer related deaths are validated it is estimated that up to 25% of them will be reclassified as non-prostate cancer related which will yield approximately 1,500 prostate cancer related deaths.

Based on feasibility data from [redacted] we expect there to be approximately 284,000 men eligible for inclusion in the study. As mentioned above, there were 846 prostate cancer related deaths observed in the feasibility study resulting from a population of 123,503 men yielding an event rate of 0.7% (846/123,503). Dividing the expected number of deaths, 1500, by the event rate results in a target sample size of 218,977 men. If we assume that the ratio of alpha-blocker:5ARI users will be the same in the overall study sample (5.4:1) as in the feasibility study, then we would expect to have 34,215 5ARI users and 184,762 alpha-blocker users in our study.

There will be a pool of approximately 249,785 alpha-blocker users from which 184,762 matches will be selected for the 5ARI users in this study. As 74% of the available matches will need to be included in the study sample to meet the target sample size of

218,977 men, there may be some difficulties in finding 5 to 6 appropriate matches for each 5ARI patient. Therefore some deaths among alpha-blocker users may not be included in the study. To address this potential issue matching criteria may be relaxed to increase the windows for age, timing of treatment initiation, and duration of alpha-blocker (e.g. from ± 1 year to ± 2 years).

Additionally, sample size calculations have been performed assuming power of 80% and 85%. Under the scenario of 85% power, and a 0.05 two sided-alpha we would need approximately 1262 deaths to rule out a hazard ratio of 1.25. Under the scenario of 80% power, and a 0.05 two sided-alpha we would need approximately 1099 deaths to rule out a hazard ratio of 1.25.

Power	Allocation ratio of 5ARI users to alpha blocker users	Number of prostate cancer related deaths
90	1/5	1314
	1/6	1474
85	1/5	1125
	1/6	1262
80	1/5	980
	1/6	1099

6. STUDY LIMITATIONS

Limitations

- There were changes in how patients are followed and screened for prostate cancer over the study period (1992-2008). These changes may affect the likelihood of detecting prostate cancer, and how diagnosed prostate cancer is treated. To help minimize this bias we are matching patients on calendar time of treatment initiation and will describe the frequency of use of different cancer treatment paradigms across the treatment groups

- Based on the feasibility study from [redacted] there are very few patients using 5ARIs only for the treatment of BPH. Most patients use an alpha-blocker and then switch to or add 5ARI therapy. For this reason we are matching patients on their duration of alpha-blocker use at the time that the index patient (5ARI user) initiates treatment. Matching on prior use of alpha-blockers will help to control for the use of alpha-blockers in the 5ARI exposure group.

- There may be inherent differences in patients who switch therapy from alpha-blocker to 5ARIs or add 5ARIs to their alpha-blocker treatment versus those who remain on alpha-blocker only therapy. These differences may be related to the likelihood of developing and being diagnosed with prostate cancer and also to prostate cancer related mortality. We will attempt to control for some of the potential differences between study groups by matching in the design phase of the study, and by controlling for confounding through regression analysis and stratification in the analysis phase of the study.

- There is likely to be differential surveillance for prostate cancer for patients using 5ARIs versus alpha-blockers. Treatment with 5ARIs lowers PSA levels so any increase in PSA while on treatment is a signal for medical intervention or increased surveillance and/or testing. To examine the potential effect of this bias we will describe PSA testing patterns and biopsy frequency over the course of the study period.

- The primary endpoint in this analysis is the occurrence of mortality due to prostate cancer. There may be a long duration of follow-up time between a patient's diagnosis of prostate cancer and occurrence of prostate cancer related death with many potential events occurring. For example, patients may be treated differently for their cancer across exposure groups and treatment modality may affect the likelihood of prostate cancer mortality. Adjusting for variables that occur once a patient has started treatment may bias the mortality ratio across exposure groups as it is difficult to disentangle the effects of various potential causal factors in this observational study. Therefore we will likely not control for variables occurring after the index date or beginning of time at risk for the outcome. We will attempt to balance all potential confounders of interest at the index date and will consider all events occurring after the index date to be part of

the treatment effect. We will describe some events including type of prostate cancer treatment and if these variables are unbalanced across treatment groups we will consider different methods for controlling for the imbalances.

- [redacted] pharmacy data captures fills of prescriptions, but we will not know if patients actually took their medications or took them as they were prescribed. There is no reason to think that this exposure misclassification would be differential across exposure groups.

- We will be developing two algorithms based on natural language processing to identify metastatic prostate cancer cases and deaths due to prostate cancer. These algorithms will misclassify some deaths and cases of metastatic cancer and we will minimize this bias by validating the algorithms with medical record abstraction. Misclassification of these outcomes should be independent of exposure status as those classifying the outcomes will be blinded to the exposure status of the patients.

- [redacted] data represents patients with commercial insurance and those enrolled in a [redacted] Medicare plan. Several regions within the US are represented within the study, but the study population will not be representative of the US population with respect to insurance coverage or distribution of race/ethnicity. Therefore the results from the study may not be generalizable to all US men with treated BPH.

- Some important confounders such as family history of prostate cancer are not systematically collected and will therefore not be adjusted for in the analysis.

- As this is a retrospective observational study there are limitations in terms of information available for all patients in the study. The study period was designed to coincide with the launch of the prescription medications under study. Therefore patients will have varying amounts of medical history available depending upon when they start treatment and meet inclusion criteria for the study. To minimize this bias we are ascertaining information on potential confounders over the one year pre-treatment period for all patients. However, information on co-morbidities and disease duration may not be complete for all patients.

Strengths

- Pooling data across 4 [redacted] sites allows for a large study population representing many person years of exposure to BPH medications.

- [redacted] data contains information from electronic medical records including lab values and procedure results which will be necessary for this study. Additionally patients can be linked to state or national death record information. Information from patient medical records can be abstracted to validate cause of death information from death certificates.

- [redacted] and [redacted] participate in the Surveillance Epidemiology and End Results (SEER) program of the National Cancer Institute and [redacted] and [redacted] participate in the NIC funded Cancer Research Network (CRN) therefore the cancer diagnosis data collected across [redacted] sites must meet the quality standards to be included as part of SEER and CRN.

- There is a low turn over in [redacted] data compared to commercial claims data. Based on the feasibility study using [redacted] data approximately 4% of patients were lost to follow-up due to disenrollment over the study period of 1992 to 2008.

- Over 90% of patients who turn 65 years of age and are eligible for Medicare and are currently enrolled in [redacted] plans choose to remain in [redacted] and enrol in one of their Medicare plans. As our study population in 50 years and over this is a major strength of the [redacted] data in answering our study question related to mortality as few patients are lost to follow up.

- The NLP algorithm developed as part of this study to validate prostate cancer related deaths will increase the validity of cause of death information in this study beyond the more common method of using cause of death from death certificates.

7. STUDY MANAGEMENT

7.1 Ethical approval and subject consent

This study will be reviewed and approved by the IRB in Southern California and approved by the other participating regions. No patient contact will be made as part of this study as only previously collected data will be utilized.

7.2 Subject confidentiality

Data used in this study will be de-identified at each [redacted] site before the data is sent to [redacted], the coordinating site, for data analysis. Patients are identified by unique IDs in the data set that will not be linked to personal identifiable information. GSK will not have access to the data and will receive results from [redacted] in aggregate form.

7.3 Reporting of adverse drug events

Reporting will be carried out according to [redacted] Safety Reporting from Epidemiology Studies and Analyses of Epidemiology Databases.

7.4 Study closure/uninterpretability of results

N/A

7.5 Study milestones

Date	Objective and Tasks to Complete	Deliverables
Start Date: December 2012		
Months 0-3	Protocol development, data analysis plan, obtain IRB approval, obtain IRB ceding or approvals from other regions and get data use agreements in place.	Protocol and data analysis plan; IRB documentation
Months 3-7	Data abstraction for NDI match and primary objectives in all regions, build analytic datasets and transfer data to [redacted] for analyses.	Interim Progress Report (including mortality counts); data to [redacted]
Months 4-7	NDI Match (CDC)	
4-16	Develop and validate natural language processing algorithms for cause of death and metastatic disease (SCAL). Other sites send report and text data for NLP to SCAL. Validate algorithm with each region's data (chart review as needed)	Interim NLP Progress Report; NLP data sent to SCAL.
Months 7-10	Primary analysis of mortality objectives (based on crude coded death categorization)	Report #1
Month 10-12	Apply cause of death algorithm; Random sample validation of prostate cancer deaths (chart review in each region)	Validation report
Months 12-16	Analysis and reporting of Primary objective (using validated death data)	Report #2
Months 16-22	Analysis and reporting of Secondary Objectives	Report #3
Months 22-28	Analysis and reporting of additional exploratory objectives	Final report
End Date: April 2015		

7.6 Study advisory committee

N/A

7.7 Study reporting and publications

The results from this study will be submitted to a peer reviewed journal within 18 months of the completion of data analysis per GSK policy [redacted].

7.8 Resourcing needs

N/A

8. REFERENCES

Abdollah F, Sun M Schmitges J, Thuret R, Tian Z, Shariat SF, Briganti A, Jeldres C, Perrotte P, Montorsi F, Karakiewicz PI. Competing-risk mortality after radiotherapy vs. observation for localized prostate cancer: a population-based study. International Journal of Radiation Oncology Biology Physics 2012;84(1):95-103.

Albertsen PC, Walters S, Hanley JA. A comparison of cause of death determination in men previously diagnosed with prostate cancer who died in 1985 or 1995. The Journal of Urology 2000;163:519-523.

Andriole GL, Bostwick D, Brawley OW, Gomella L, Marberger M, Montorsi F, Pettaway C, Tammela TL, et al. The effect of dutasteride on the usefulness of prostate specific antigen for the diagnosis of high grade and clinically relevant prostate cancer in men with a previous negative biopsy: results from the REDUCE study. Journal of Urology 2011; 185(1):126-131.

Brawley OW. Prostate cancer epidemiology in the United States. World Journal of Urology 2012; 30:195–200.

Danforth KM, Early MI, Ngan S, Kosco AE, Zheng C, Gould MK. Automated identification of patients with pulmonary nodules in an integrated health system using administrative health plan data, radiology reports, and natural language processing. Journal of Thoracic Oncology. 2012;7(8):1257-1262.

De Koning HJ, Blom J Merkelbach JW, Raaijmakers R, Verhaegen H, Van Vliet P, Nelen V, Coebergh JWW, Hermans A, Ciatto S, Makinen T. Determining the cause of death in randomized screening trial(s) for prostate cancer. BJU International 2003;92(2):71-78.Denny JC, Choma NN, Peterson JF, Miller RA, Batarache L, Li M, Peterson NB. Natural language processing improves identification of colorectal cancer testing in the electronic medical record. Medical Decision Making 2012;32(1):188-97.

Eifler JB, Humphreys EB, Agro M, Partin AW, Trock BJ, Han M. Causes of death after radical prostatectomy at a large tertiary center. The Journal of Urology 2012;188:798-802.

Fleshner NE, Lucia MS, Egerdie B, Aaron L, Eure G, Nandy I, Black L, Rittmaster RS. Dutasteride in localised prostate cancer management: the REDEEM randomised, double-blind, placebo-controlled trial. Lancet 2012;379(9821):1103-1111.

Hernan MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survivial of HIV-positive men. Epidemiology 2000;11:561-570.

Jorgensen TL, Hallas J, Friis S, Herrstedt J. Comorbidity in elderly cancer patients in relation to overall and cancer-specific mortality. British Journal of Cancer 2012;106: 1353–1360.

Kim MM, Hoffman KE, Levy LB, Frank SJ, Pugh TJ, Choi S, Nguyen ON, McGuire SE, Lee AK, Kuban DA. Prostate cancer-specific mortality after definitive radiation therapy: who dies of disease? European Journal of Cancer 2012;48:1664–1671

Miller AB, Yurgalevitch S, Weissfeld JL. Death review process in the prostate, lung, colorectal, and ovarian (PLCO) cancer screening trial. Controlled Clinical Trials 2000;21:400S-406S.

Penson DF, Albertsen PC, Nelson PS, Barry M, Stanford JL. Determining cause of death in prostate cancer: are death certificates valid. Journal of the national cancer institute 2001;93(23):1822-1823.

Pinsky PF, Black A, Grubb R, Crawford ED, Andriole G, Thompson I, Parnes H. Projecting prostate cancer mortality in the PCPT and REDUCE chemoprevention trials. Cancer 2012; 1-9.

Saville BR, Kim YS, Koch GG. Graphical displays for clarifying how allocation ratio affects total sample size for the two sample logrank test. Pharmaceutical Statistics 2011;10:40-44.

Scosyrev E, Messing EM, Mohile S, Golijanin D, Wu G. Prostate cancer in the elderly. Cancer 2012;118:3062-70

Thompson IM, Tangen C, Gooman P. The prostate cancer prevention trial: design, status, and promise. World Journal of Urology 2003;21(1):28-30.

9. TABLES

A selection of key tables for this study is presented below. The inclusion of additional tables will be considered as the analysis plan is developed for the study.

Table 2: Pre-treatment characteristics by exposure group after matching (5ARI and alpha-blocker)

	Overall (n= _)	5ARI users (n= _)	Alpha-blocker users (n= _)	p-value
Characteristic				
Age at treatment initiation (mean (sd), median)				
Age at treatment initiation (%)				
<60				
60-69				
70+				
Race (%)				
Non-Hispanic White				
African American				
Hispanic				
Asian/Pacific Islander				
Other				
Missing				
SES				
PSA level (mean (sd), median				
Prostate volume				
BMI (kg/m2) (%)				
<25 mg/kg2				

<=25 mg/kg2 <30 mg/kg2				
>=30 mg/kg2				
Smoking status				
Charlson Comorbidity index (%)				
0				
1				
2+				
History of cardiovascular disease				
History of high blood pressure				
History of hyperlipidemia				
History of diabetes				
History of cancer				
Use of other medications				

Table 3: Post-treatment characteristics by exposure group

	Overall (n= _)	5ARI users (n= _)	Alpha-blocker users (n= _)	p-value
Characteristic				
Cumulative exposure time, years (mean (sd), median)				
Cumulative dose				
Duration of follow-up time, years (mean (sd), median)				
Patients with >=1 PSA test (%)				
Number of PSA tests (mean (sd), median)*				
Patients with >=1 biopsy (%)				
Number of Biopsies (mean (sd), median)				
Number of cores/biopsy** (mean(sd), median)				
Number of positive cores/biopsy (mean(sd), median)				
Prostate cancer diagnosis (%)				
Prostate cancer incidence rate, /1,000 person-years				

***Patients without PSA tests not included in percentages.**

****Patients without biopsies not included in percentages.**

Table 4: Characteristics of men developing prostate cancer over the follow-up period by exposure group.

	Overall (n= _)	5ARI users (n= _)	Alpha-blocker users (n= _)
Characteristic			
Number of PSA tests (mean (sd), median)			
Number of Biopsies (mean (sd), median)			
Number of cores/biopsy (mean(sd), median)			
Number of positive cores/biopsy (mean(sd), median)			
Age at diagnosis (mean (sd), median)			
Age at diagnosis (%)			
<60			
60-69			
70+			
Gleason score at diagnosis(n (%))			
<=6			

7			
7-10			
8-10			
Radical prostatectomy (n(%))			
Reclassification of Gleason score at prostatectomy (%same, % upgrade, % downgrade))			
<=6			
7			
7-10			
8-10			
Stage at diagnosis			
I			
II			
III			
IV			
Bone scan at diagnosis (%)			
Positive bone scans (%)			
Other tests for metastatic disease at diagnosis (%)			
Positive tests (%)			
Prostate cancer treatment within 6 months of diagnosis (%)			

Surgery			
Hormonal therapy			
Chemotherapy			
Immunotherapy			
Radiation therapy			
Other therapy			
Follow-up time after cancer diagnosis, years (mean (sd, median)			
Time since last exposure to BPH medication			

Table 5: Progression to metastatic disease by stage and grade at diagnosis

	Incidence rate of metastatic disease		
	Overall (n= _)	5ARI users (n= _)	Alpha-blocker users (n= _)
Patient Stratum			
Gleason score at diagnosis (%)			
<=6			
7			
8			
9			
10			
Stage at diagnosis			
I			
II			
III			

Table 6: Pre-treatment characteristics after matching, prostate cancer incidence, and prostate cancer mortality by 5ARI user type

	Overall (n= _)	**5ARI monotherapy users (n= _)**	**5ARI+Alpha-blocker users (n= _)**
Characteristic			
Age at treatment initiation (mean (sd), median)			
Age at treatment initiation (%)			
<60			
60-69			
70+			
Race (%)			
Non-Hispanic White			
African American			
Hispanic			
Asian/Pacific Islander			
Other			
Missing			
SES			
PSA level (mean (sd), median			
Prostate volume			
BMI (kg/m2) (%)			
<25 mg/kg2			

<=25 mg/kg2 <30 mg/kg2			
>=30 mg/kg2			
Smoking status			
Charlson Comorbidity index (%)			
0			
1			
2+			
History of cardiovascular disease			
History of high blood pressure			
History of hyperlipidemia			
History of diabetes			
History of cancer			
Use of other medications			
Prostate cancer diagnosis (%)			
Prostate cancer incidence rate, /1,000 person-years			
Prostate cancer deaths (%)			
Prostate cancer mortality rate, /1,000 person-years			

Table 7: Crude and adjusted prostate cancer related mortality rates and hazard ratios stratified by characteristics of interest.

	No of deaths/ No at risk		Mortality rate/ 1,000 p-y		Hazard ratio (95% CI)	Adjusted hazard ratio (95% CI)
	5ARI	Alpha-blocker	5ARI	Alpha-blocker		
Overall						
Cumulative exposure						
➢ 3 months						
< 1year						
1 year-<2 years						
2+ years						
Cumulative dose						
1st quartile						
2nd quartile						
3rd quartile						
4th quartile						

Table 8: **Crude and adjusted prostate cancer mortality or metastatic cancer rates and hazard ratios stratified by characteristics of interest.**

	No of events/ No at risk		Event rate/ 1,000 p-y		Hazard ratio (95% CI)	Adjusted hazard ratio (95% CI)
	5ARI	Alpha-blocker	5ARI	Alpha-blocker		
Overall						
Cumulative exposure						
➢ 3 months						
< 1year						
1 year-<2 years						
2+ years						
Cumulative dose						
1st quartile						
2nd quartile						
3rd quartile						
4th quartile						

Table 9: **Crude and adjusted all cause mortality rates and hazard ratios stratified by characteristics of interest.**

	No of events/ No at risk		Mortality rate/ 1,000 p-y		Hazard ratio (95% CI)	Adjusted hazard ratio (95% CI)
	5ARI	Alpha-blocker	5ARI	Alpha-blocker		
Overall						
Cumulative exposure						
➢ 3 months						
< 1year						
1 year-<2 years						
2+ years						
Cumulative dose						
1st quartile						
2nd quartile						
3rd quartile						
4th quartile						

Table 10: Crude and adjusted prostate cancer mortality rates and hazard ratios stratified by pre-treatment characteristics of interest.

	No of deaths/ No at risk		Mortality rate/ 1,000 p-y		Hazard ratio (95% CI)	Adjusted hazard ratio (95% CI)
	5ARI	Alpha-blocker	5ARI	Alpha-blocker		
Overall						
Baseline characteristics						
Race (%)						
Non-Hispanic White						
African American						
Hispanic						
Asian/Pacific Islander						
Other						
Missing						
SES						
BMI (kg/m2) (%)						
<25 mg/kg2						
<=25 mg/kg2 <30 mg/kg2						
>=30 mg/kg2						
Smoking status						
Charlson Comorbidity index (%)						
0						

1						
2+						
History of cardiovascular disease						
History of high blood pressure						
History of hyperlipidemia						
History of diabetes						
History of cancer						

Table 11: **Crude and adjusted prostate cancer mortality or metastatic cancer rates and hazard ratios stratified by pre-treatment characteristics of interest.**

	No of events/ No at risk		Event rate/ 1,000 p-y		Hazard ratio (95% CI)	Adjusted hazard ratio (95% CI)
	5ARI	Alpha-blocker	5ARI	Alpha-blocker		
Overall						
Baseline characteristics						
Race (%)						
Non-Hispanic White						
African American						
Hispanic						
Asian/Pacific Islander						
Other						
Missing						
SES						
BMI (kg/m2) (%)						
<25 mg/kg2						
<=25 mg/kg2 <30 mg/kg2						
>=30 mg/kg2						
Smoking status						
Charlson Comorbidity index (%)						
0						
1						

2+						
History of cardiovascular disease						
History of high blood pressure						
History of hyperlipidemia						
History of diabetes						
History of cancer						

Table 12: Crude and adjusted all cause mortality rates and hazard ratios stratified by characteristics of interest.

	No of deaths/ No at risk		Mortality rate/ 1,000 p-y		Hazard ratio (95% CI)	Adjusted hazard ratio (95% CI)
	5ARI	Alpha-blocker	5ARI	Alpha-blocker		
Overall						
Baseline characteristics						
Race (%)						
Non-Hispanic White						
African American						
Hispanic						
Asian/Pacific Islander						
Other						
Missing						
SES						
BMI (kg/m2) (%)						
<25 mg/kg2						
<=25 mg/kg2 <30 mg/kg2						
>=30 mg/kg2						
Smoking status						
Charlson Comorbidity index (%)						
0						
1						

2+						
History of cardiovascular disease						
History of high blood pressure						
History of hyperlipidemia						
History of diabetes						
History of cancer						

10. FIGURES

Kaplan Meier curves for prostate cancer related mortality, prostate cancer related mortality or metastatic prostate cancer, all cause mortality, metastatic prostate cancer, and prostate cancer comparing 5ARI and alpha-blocker users will be constructed.

11. APPENDICES

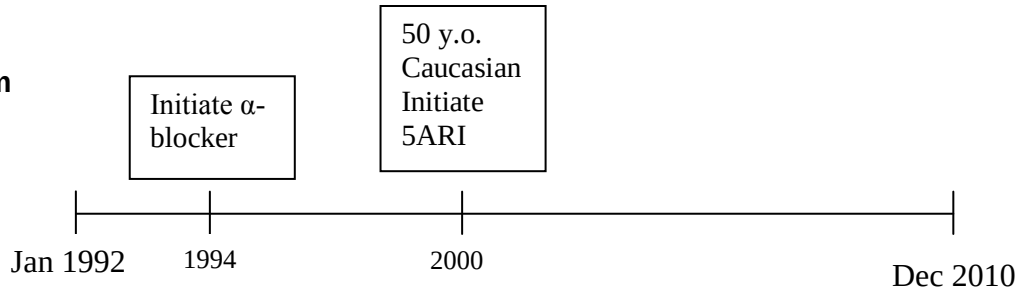
Appendix 1: Main table from feasibility study

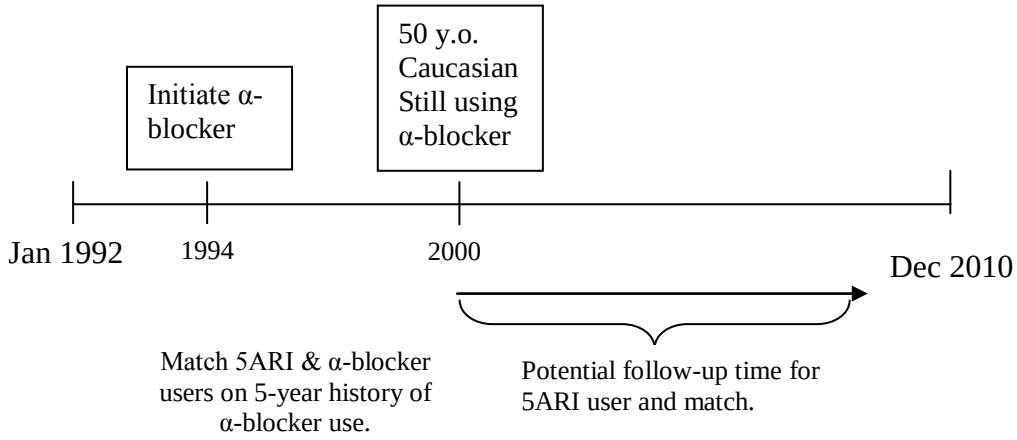
	α-blocker Only (N=104,272)	5ARI Only (N=1,619)	Combination Therapy (N=17,612)
BPH diagnoses			
Number of patients with BPH diagnosis	75,133	917	17,199
Median age at BPH diagnosis	67.7	68.3	68.1
Time from BPH diagnosis to first Rx (in months, % were based on BPH diagnosed)			
0 (first Rx before diagnosis)	37673 (50.1)	233 (25.4)	9547 (55.5)
0-3	24653 (32.8)	258 (28.1)	4509 (26.2)
3-6	923 (1.2)	26 (2.8)	207 (1.2)
6-12	1517 (2.0)	42 (4.6)	388 (2.3)
12+	10367 (13.8)	358 (39.0)	2548 (14.8)
Median (range)	0 (0,319)	1 (0,324)	0 (0,307)
BPH Treatment			
Median age at first medication Rx	65.7	64.2	66.3
Number of patients with follow-up time[2] (years)	N=104187	N=1619	N=17611
< 1	8158 (7.8)	156 (9.6)	235 (1.3)
1-2	6786 (6.5)	145 (9.0)	246 (1.4)
2-4	17712 (17.0)	436 (26.9)	1822 (10.4)
4-6	20339 (19.5)	442 (27.3)	3226 (18.3)
6-10	25686 (24.7)	248 (15.3)	4692 (26.6)
10+	25506 (24.5)	192 (11.9)	7390 (42.0)
Median (range)	5.9 (0, 20.0)	4.2 (0, 18.8)	8.7 (0, 20.0)
Number of patients with cumulative duration of medication supply, (%)[3]			
< 6 months	14543 (14.0)	269 (16.6)	92 (0.5)
6-12 months	9856 (9.5)	184 (11.4)	251 (1.4)
1-2 years	11662 (11.2)	219 (13.5)	483 (2.7)
2-4 years	12156 (11.7)	220 (13.6)	918 (5.2)
4+ years	56055 (53.8)	727 (44.9)	15868 (90.1)
Median (in years, range)	4.9 (0, 2382.5)	3.2 (0, 355.6)	38.8 (0.1, 2627.6)
PSA			
Total number of PSA tests	571671	8346	141740
Total number of PSAs with value	570760	8327	141494
Number of Patients with at least 1 PSA test, (%)[3]	95791 (91.9)	1413 (87.3)	17002 (96.5)
Median Number of PSAs	5	4	7
Median number (min, p25, p75, max) of PSA tests by year of initiation of BPH medication			
1st year post initiation	1 (1, 1, 2, 20)	1 (1, 1, 1, 5)	1 (1, 1, 2, 16)
2nd year post initiation	1 (1, 1, 1, 18)	1 (1, 1, 1, 6)	1 (1, 1, 2, 13)
3rd year post initiation	1 (1, 1, 1, 21)	1 (1, 1, 1, 8)	1 (1, 1, 2, 15)
4th year post initiation	1 (1, 1, 1, 40)	1 (1, 1, 1, 10)	1 (1, 1, 2, 15)
5th year post initiation	1 (1, 1, 1, 14)	1 (1, 1, 1, 5)	1 (1, 1, 2, 16)

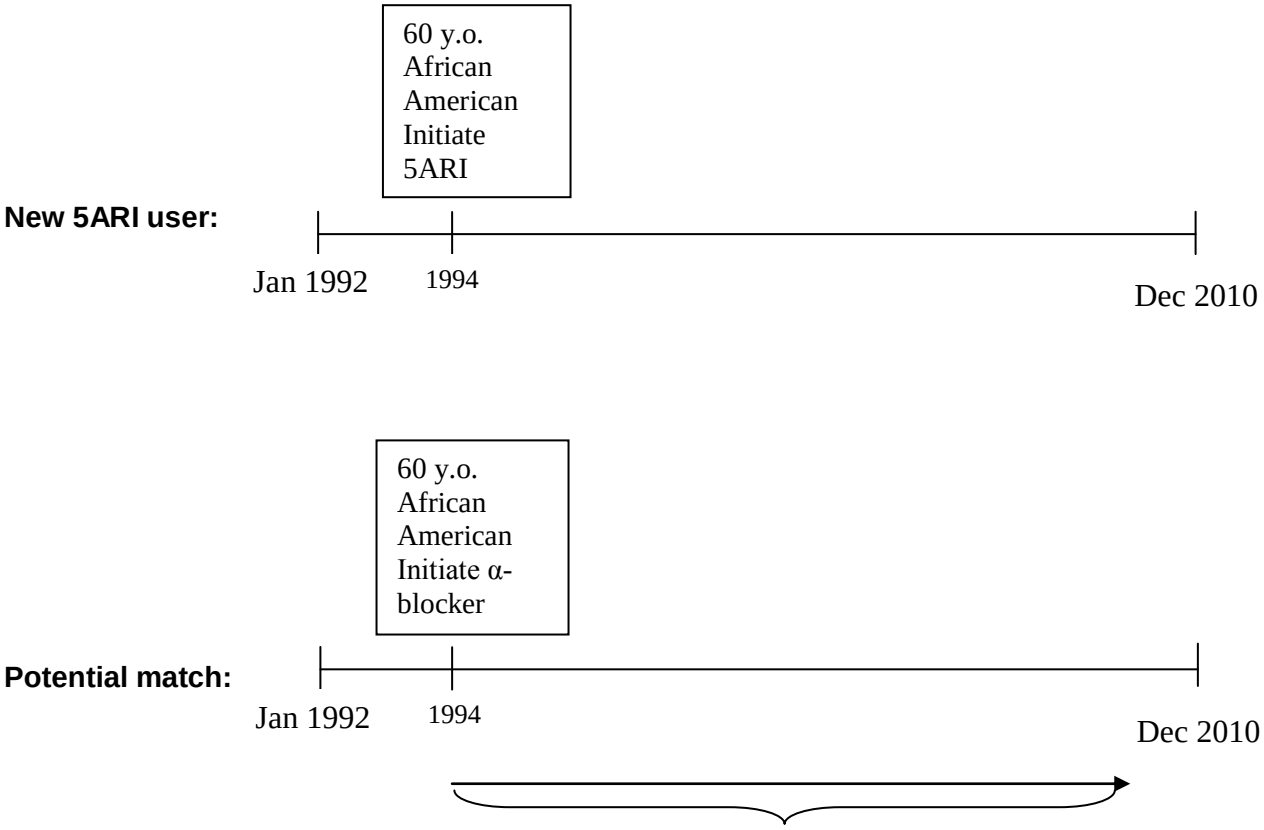
Prostate Cancer	N (%) or median (SE)	N (%) or median (SE)	N (%) or median (SE)
Number of men with prostate cancer, (%)[3]	7856 (7.5)	50 (3.1)	1185 (6.7)
Age at diagnosis	69.7 (0.1)	70.4 (1.4)	72.6 (0.2)
Time since first medication prescription (in years)	2.4 (0)	1.0 (0.4)	4.3 (0.1)
Time since BPH diagnosis (in years)	1.2 (0.1)	2.0 (1.1)	2.7 (0.1)
Gleason Score[4] (sum of both primary and secondary) at diagnosis	N=3267	N=28	N=665
≤6	1621 (49.6)	13 (46.4)	386 (58.0)
7	1064 (32.6)	12 (42.9)	158 (23.8)
8	255 (7.8)	0 (0)	52 (7.8)
9	296 (9.1)	3 (10.7)	61 (9.2)
10	31 (0.9)	0 (0)	8 (1.2)
Cancer stage at diagnosis[5]	N=5579	N=37	N=972
I	46 (0.8)	0 (0)	17 (1.8)
II	4840 (86.8)	29 (78.4)	856 (88.1)
III	301 (5.4)	2 (5.4)	30 (3.1)
IV	392 (7.0)	6 (16.2)	69 (7.1)
Prostate cancer treatment within 6 months of diagnosis, (%)[3]			
Surgery	2078 (26.5)	18 (36.0)	261 (22.0)
Hormonal Therapy	1484 (18.9)	9 (18.0)	246 (20.8)
Chemotherapy	15 (0.2)	1 (2.0)	4 (0.3)
Immunotherapy	2 (0.03)	0 (0)	0 (0)
Radiation Therapy	869 (11.1)	6 (12.0)	116 (9.8)
Other Therapy	5 (0.06)	0 ()	1 (0.1)
Prostate Cancer Mortality	**N (%) or median (range)**	**N (%) or median (range)**	**N (%) or median (range)**
Total number of deaths, (%)[3]	749 (0.7)	10 (0.6)	87 (0.5)
Age at death, median (range)	80.7 (51.1, 97.6)	81.2 (57.4, 93.8)	80.9 (59.5, 100.5)
Time since first medication prescription (in years), median (range)	5.6 (0.1, 17.1)	3 (0.5, 14.9)	5.8 (0.7, 17.3)
Time since BPH diagnosis (in years)	3.5 (0.0,28.7)	1.9 (0.5,26.4)	4.3 (0.6,27.1)
Time since prostate cancer diagnosis (in years)[6]	N=608 2.9 (0, 16.5)	N=6 1.1 (0.4, 6.0)	N=71 1.8 (0, 15.1)
Lost to Follow-Up[7]	**N (%)**	**N (%)**	**N (%)**
Number of patients lost to follow-up	26689 (25.6)	298 (18.4)	2393 (13.6)

Appendix 2

Matching diagram

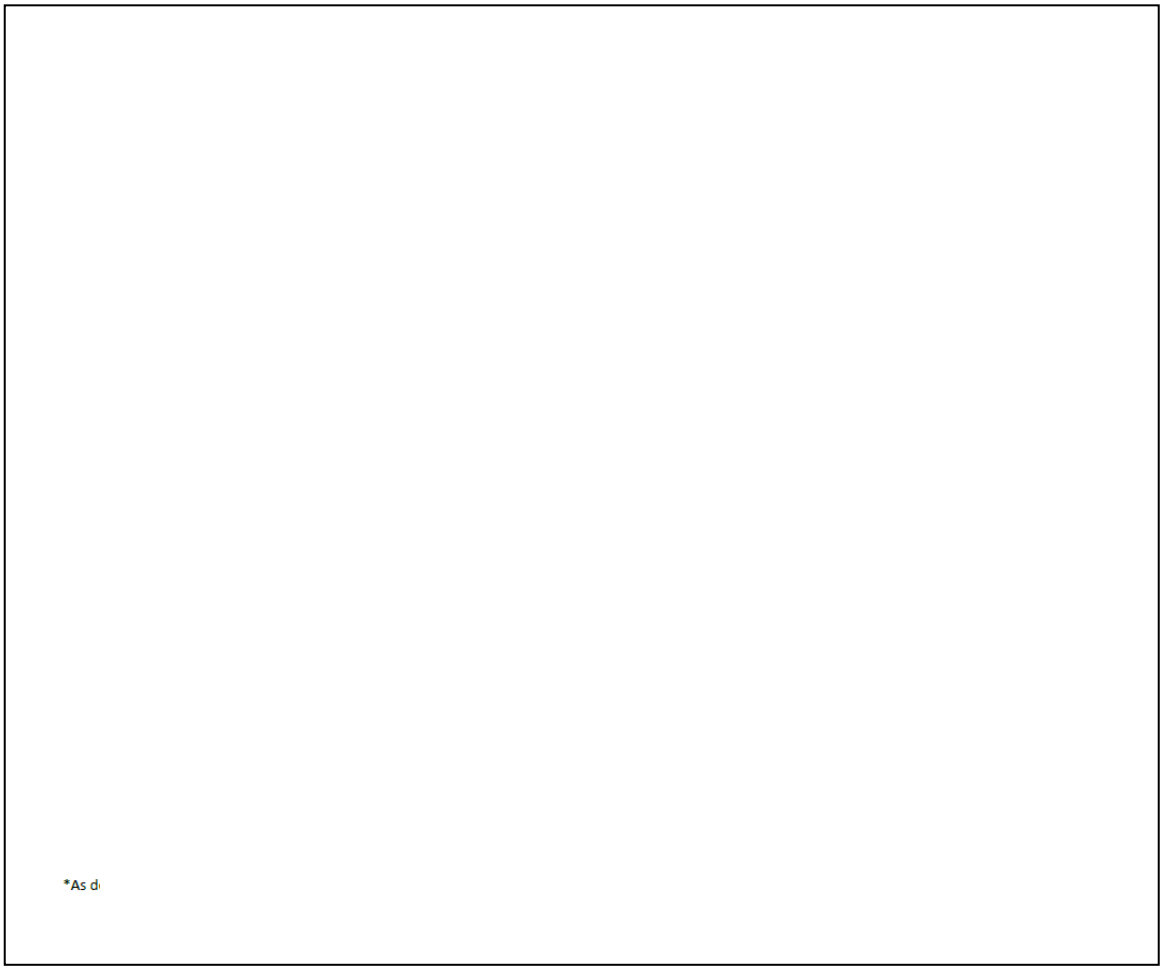





60 y.o. African American Initiate 5ARI
New 5ARI user:
Jan 1992
1994
Dec 2010
60 y.o. African American Initiate α-blocker
Potential match:
Jan 1992
1994
Dec 2010
Potential follow-up time for 5ARI user and match.

Appendix 3

Prostate cancer death validation process flow chart



Appendix 4

DETERMINING CAUSE OF DEATH ALGORITHM ID_____

1. Does subject have adenocarcinoma of the prostate?			Data Source
a. Is there a pathology report of a prostate biopsy that shows adenocarcinoma?	YES NO	If yes, go to 2.	NLP-pathology reports
b. Is there a PSA level > 10?	YES NO	If yes, go to 2.	Lab data
c. Is there an acid phosphatase level greater than normal?	YES NO	If yes, go to 2.	Lab data
d. Is there a DRE which the examiner states is very suspicious for cancer?	YES NO	If yes AND e yes, go to 2.	NLP- clinic notes
e. Is there a report of an abnormal bone scan that is suspicious for metastatic cancer?	YES NO	If yes AND d yes, go to 2.	NLP- radiology reports
If a through c are no AND either d or e is no, death was not due to prostate cancer. STOP HERE			
2. Is there another clear-cut cause of death? Starting with the date of death and working backwards in the record, review all clinical data including outpatient clinic visits, hospital summaries, hospice notes, telephone calls, etc.			
a. Is there a clear statement that the cause of death is something other than prostate cancer?	YES NO	If yes, go to b. If no, go to 3.	Mortality data
b. Is the cause of death stated to be one of the	YES NO	If no, then the death was not due to prostate	Mortality data

following? Pneumonia Pulmonary emboli Urosepsis Septic shock Anoxic brain damage Renal failure Failure to thrive Multi-organ failure Unknown		cancer. STOP HERE. If yes, go to 3.	
3. Does subject have evidence of metastatic disease?			
a. bone scan that is abnormal and suggestive of metastatic disease	YES NO	If yes, go to 4.	NLP- radiology reports
b. PSA level >100	YES NO	If yes, go to 4.	Lab data
c. elevated acid phosphatase level?	YES NO	If yes, go to 4.	Lab data
d. prostate cancer found in biopsy of a metastatic lesion	YES NO	If yes, go to 4.	NLP-pathology reports
e. evidence of metastatic disease to lymph nodes or distant organs and no other known primary site of cancer besides prostate cancer	YES NO	If yes, go to 4.	NLP- clinic notes
f. blastic lesion on bone x-rays	YES NO	If yes, go to 4.	NLP- radiology

If the answers to a through f are all no, go to 5.			
4. Does subject have evidence of progressive and debilitating disease during the interval immediately preceding death?			
a. rising PSA	YES NO	If yes, PC death.	Lab data
b. worsening bone scan	YES NO	If yes, PC death.	NLP-radiology
c. pathological fractures	YES NO	If yes, PC death.	EMR-Diagnoses
d. requiring blood transfusions	YES NO	If yes, PC death.	EMR- Diagnoses
e. increasing pain	YES NO	If yes, PC death.	NLP-clinic notes
f. increasing narcotic requirements	YES NO	If yes, PC death.	Pharmacy data
h. renal failure due to ureteral obstruction	YES NO	If yes, PC death.	EMR Diagnoses
g. enrolled in hospice	YES NO	If yes, PC death.	EMR
If all of a through g are no, go to 5.			
5. Did subject die of a cause related to prostate cancer? Starting with the date of death and working backwards in the record, review all outpatient clinic notes and hospital summaries.			

a. urosepsis due to outlet obstruction from locally-advance prostate cancer	YES NO	If yes, PC death.	EMR- Diagnoses
b. pulmonary emboli	YES NO	If yes, PC death.	EMR- Diagnoses
c. post-operative complications of prostate surgery (1) infection (2) bleeding (3) pulmonary emboli (4) urosepsis as a post-operative complication of prostate ca. surgery	YES NO	If yes, PC death.	Diagnoses
If all are no, death was not due to prostate cancer. STOP HERE			
Was death due to prostate cancer?	YES NO		