

Predicting the characteristics and results of the PRONOUNCE trial: study protocol

Joshua D. Wallach, MS, PhD; Yihong Deng, PhD; Rozalina McCoy, MD, MS; Jeph Herrin, PhD; Eric C. Polley, PhD; Alyssa Berkowitz, MPH, Bill Crown, PhD; Sanket S. Dhruva, MD, MHS; Lindsey R. Sangaralingham, MPH; Xiaoxi Yao, PhD, MPH; Peter A. Noseworthy, MD; Nilay Shah, PhD; Joseph S. Ross, MD, MHS; Timothy D. Lyon, MD

Protocol draft: August 2020

Yale University-May Clinical Center of Excellence in Regulatory Science and Innovation (CERSI)

project area: Understanding the use of existing real-world data for medical product evaluation

PROTOCOL

The purpose of this analysis plan is to provide a detailed description of our methods and decisions, in order to guide our analyst(s). To facilitate replications of our methodology, the information in this document will also be shared in the supplementary materials of the final manuscript.

Protocol changes made after accessing the data are in bold.

1. ABBREVIATIONS

ADT	Androgen deprivation therapy
ASCVD	Atherosclerotic cardiovascular disease
CABG	Coronary artery bypass grafting
CI	Confidence interval
CT	Computed tomography
DBP	Diastolic blood pressure
EHR	Electronic health record
E&M	Evaluation & Management
eGFR	Estimated Glomerular Filtration Rate
FDA	Food and Drug Administration
GnRH	Gonadotropin-releasing hormone
HbA1c	Hemoglobin A1c
HR	Hazard ratio
MACE	Major Adverse Cardiovascular Events
NA	Not applicable
PCI	Percutaneous coronary intervention
PDC	Proportion of days covered
PSA	Prostate-specific antigen
RCT	Randomized controlled trials
RWD	Real world data
SBP	Systolic blood pressure
SSA	Social Security Administration

2. KEY DEFINITIONS

Item	Variable name	Description
Index Date	index_date	The date patients received degarelix (Firmagon) or leuprolide (Lupron depot)
Index Medication	index_med	Patient's first prescription of degarelix (Firmagon) or leuprolide (Lupron depot)
Baseline Period	baseline	Any time before and including the index date used to establish a patient's medical history
Study Period		12/24/2008 (degarelix (Firmagon) FDA approval date) – 6/30/2019
Data Source		OptumLabs Data Warehouse

3. BACKGROUND AND OBJECTIVES

Rigorous double-blind randomized controlled trials (RCTs) are considered the gold standard for generating evidence for the safety, efficacy, and effectiveness of any clinical intervention. Although RCTs are used to inform the decisions made by the Food and Drug Administration (FDA), health insurance payers, professional societies, patients, and clinicians,¹ these studies often have strict inclusion and exclusion criteria, small sample sizes, and short follow-up durations.² These limitations can undermine the generalizability of RCTs to real-world clinical practice, and have stimulated interest in research focused on generating evidence on medical products using observational research methods applied to real-world data sources.

Advances in the quantity, quality, and granularity of real-world data, and improvements in statistical methods used to account for unmeasured confounding, have provided opportunities to use real-world data to inform our understanding of drug safety and efficacy. While many studies have estimated the results of completed RCTs using real-world data and observational methods,³⁻⁹ the majority have focused on comparisons with large prospective cohort studies, as opposed to retrospective analysis of routinely-collected electronic health record (EHR) or

administrative claims data. Furthermore, less is known about the ability to simulate RCT eligibility criteria for ongoing trials to predict the characteristics of real-world populations and determine whether the characteristics of the RCT population and the real-world populations are similar. Moreover, it would be valuable to predict the results of ongoing trials, thereby avoiding the potential biases that could be introduced by trying to replicate the results of RCTs that have already been completed and disseminated among the scientific community.

Using ClinicalTrials.gov, we have identified an illustrative test case to assess the potential opportunities, advantages, and limitations of using retrospective research methods to pursue evaluations of drug safety using real-world data, including as an alternative to RCTs. The PRONOUNCE trial is an ongoing Phase IIIb comparative safety trial focused on comparing the cardiovascular safety of degarelix (Firmagon), a gonadotropin-releasing hormone (GnRH) antagonist, and leuprolide (Lupron Depot), a GnRH agonist, among patients with prostate cancer and cardiovascular disease. Prostate cancer is the second most commonly occurring cancer in men, and among patients with advanced prostate cancer, androgen deprivation therapy (ADT) is a preferred treatment option.¹⁰ However, studies have suggested that ADT in the form of GnRH agonists, such as leuprolide, increase the risk for cardiovascular morbidity and mortality, especially among prostate cancer patients with a history of cardiovascular disease.¹¹ Although GnRH antagonists are believed to be associated with improved cardiovascular safety compared to GnRH agonists,¹² the PRONOUNCE trial was designed to clarify uncertainties and compare the cardiovascular safety profile of the GnRH antagonist degarelix and the GnRH agonist leuprolide.

Using real-world data from OptumLabs, a large database of insurance claims linked with EHR data, we sought to predict the results of the PRONOUNCE trial by applying observational research methods to the real-world patients.

4. METHODS

4.1 Study design

This study will be a retrospective cohort analysis using OptumLabs Data Warehouse.

4.2 Data sources

OptumLabs Data Warehouse

OptumLabs Data Warehouse contains longitudinal health information on over 160 million predominantly privately insured individuals, and includes a large number of Medicare Part D and Medicare Advanced beneficiaries.¹³ Patient-level information includes socio-demographic characteristics (i.e., age, sex, and race) and geographic region. Data also include health insurance coverage (dates of enrollment, primary/secondary coverage, Medicaid/Medicare status), administrative claims (diagnosis and procedure codes), and pharmacy claims (prescribing physician, pharmacy, days of supply, drug/quality/strength/data dispensed). These data also include physician & facility claims (physician type and specialty). For roughly 50 million individuals, additional information is available from electronic medical record encounters, including vital signs and laboratory measurements.

ClinicalTrials.gov and the PRONOUNCE trial Protocol

The PRONOUNCE trial was identified using ClinicalTrials.gov (**Box 1**). The trial inclusion and exclusion criteria were established using ClinicalTrials.gov and the descriptions outlined in the PRONOUNCE trial protocol (**Table 1**).¹⁴

Box 1. PRONOUNCE trial Characteristics
Title: A Trial Comparing Cardiovascular Safety of Degarelix Versus Leuprolide in Patients With Advanced Prostate Cancer and Cardiovascular Disease (PRONOUNCE)
Aim(s): To test whether degarelix (Firmagon), a marketed drug for advanced prostate cancer, can reduce the risk of cardiovascular complications compared to leuprolide (Lupron Depot), another drug for advanced prostate cancer, among patients with prostate cancer and cardiovascular disease.
Primary endpoint: Major Adverse Cardiovascular Events (MACE), including death due to any cause, non-fatal myocardial infarction, or non-fatal stroke.
Secondary endpoints: Time from randomization to occurrence of: <ol style="list-style-type: none"> 1. myocardial infarction (fatal, non-fatal) 2. stroke (fatal, non-fatal) 3. unstable angina requiring hospitalization (fatal, non-fatal) 4. cardiovascular-related death
Source: https://clinicaltrials.gov/ct2/show/NCT02663908?term=PRONOUNCE&draw=2&rank=1

Table 1. The PRONOUNCE Trial Eligibility Criteria	
PRONOUNCE Eligibility Criteria	Operational Definition in OLDW
<u>Inclusion Criteria</u>	
Histologically confirmed adenocarcinoma of the prostate Tumor, node, metastasis staging available prior to treatment start (bone scan and/or CT scan and/or MRI) <12 weeks prior to study start. If no radiographic image is available at the time of screening, a bone scan should be performed	NA: Already defined in the initial cohort as: <i>Patients must have at least one “Evaluation and Management” visit with a diagnosis of prostate cancer within 6 months before index date, at least one “Evaluation and Management” visit with a diagnosis of prostate cancer at any time after the index date</i> <i>Sensitivity analysis: prostate biopsy required</i>
Indication to initiate androgen deprivation therapy (ADT)	NA: ‘Index date’ is the first fill of degarelix or leuprolide* <i>*For the leuprolide cohort, we will allow up to one month of bicalutamide prior to leuprolide initiation</i>
Predefined cardiovascular disease inclusion criteria	

<p>Pre-existing ASCVD (confirmed diagnosis, documented) according to at least one of the following criteria:</p> <p>Prior myocardial infarction ≥ 30 days before randomization; prior revascularization procedure ≥ 30 days before randomization); Coronary artery: stent placement/balloon angioplasty or coronary artery bypass graft surgery; Coronary artery: stent placement/balloon angioplasty or endarterectomy surgery; Iliac, femoral, popliteal arteries: stent placement/balloon angioplasty or vascular bypass surgery</p> <p>At least one vascular stenosis $\geq 50\%$ at any time point before randomization by angiogram or CT angiogram Coronary artery Carotid artery Iliac femoral, or popliteal arteries Carotid ultrasound results that documented a vascular stenosis $\geq 50\%$ at any time point before randomization</p> <p>Ankle-brachial pressure index < 0.9 at any time point before randomization.</p>	<p>To establish history of cardiovascular disease using claims data, it is common to use both primary and secondary discharge diagnosis and procedure codes. In particular, we will identify the following using discharge diagnosis and procedure codes as indicators of a history of cardiovascular disease, ≥ 30 days before the index date: Myocardial infarction, percutaneous coronary intervention (PCI), coronary artery bypass grafting (CABG), peripheral artery revascularization, carotid revascularization any position</p> <p><i>We do not have angiogram, CT angiogram, or ankle-brachial pressure data.</i></p>
<p><u>Exclusion Criteria</u></p>	
<p>Previous or current hormonal management of prostate cancer</p> <ul style="list-style-type: none"> - Surgical castration - Any hormonal manipulation <p>Except prior neoadjuvant/adjuvant hormonal therapy, in this case treatment must be terminated > 12 months prior to study start.</p>	<p>NA: 'Index date' is the first fill of degarelix or leuprolide</p> <p><i>*For the leuprolide cohort, we will allow up to one month of bicalutamide prior to leuprolide initiation</i></p>
<p>Main cardiovascular exclusion criteria</p>	

Uncontrolled type 1 or type 2 diabetes mellitus (defined as HbA1c > 10%) at time of randomization	NA: Restricting patients with laboratory data would limit the sample
Uncontrolled hypertension (SBP >180 mmHg or DBP >110 mmHg) at time of randomization	NA: Restricting patients with laboratory data would limit the sample
A history of congenital long QT syndrome or risk factors for torsade de pointes ventricular arrhythmias (e.g., heart failure, hypokalemia, concomitant medication known to cause QT prolongation)	NA: This exclusion criteria only applies to patients at risk for ventricular arrhythmias
<p>Within 30 days prior to randomization</p> <ul style="list-style-type: none"> - Myocardial infarction - Stroke (hemorrhagic/ischemic) - Coronary, carotid, or peripheral artery revascularization 	<p>To identify recent/active cardiovascular events, we will use primary diagnosis of myocardial infarction and stroke (emergency department or inpatient visits) within 30 days before index date, as patients hospitalized for an acute cardiovascular event would be expected to have these diagnoses listed as the primary discharge diagnosis, while patients with a history of cardiovascular disease would be expected to have these diagnoses listed as secondary discharge diagnoses.</p> <p>Percutaneous coronary intervention (PCI), coronary artery bypass grafting (CABG), peripheral artery revascularization (limb events), carotid revascularization within 30 days before index date</p>
Planned or scheduled cardiac surgery or PCI procedure that is known at the time of randomization	NA: We cannot determine planned or scheduled PCI procedures.
Ankle-brachial pressure index <0.9 at any point before randomization	NA: Restricting patients with laboratory data would limit the sample.
ASCVD = atherosclerotic cardiovascular disease; CT = computed tomography; DBP = diastolic blood pressure; HbA1c = hemoglobin A1c; NA = not applicable; SBP = systolic blood pressure;	

4.3 Study inclusion and exclusion criteria

We will apply the PRONOUNCE trial inclusion and exclusion criteria listed on ClinicalTrials.gov, which we will update with information made available by the PRONOUNCE trial authors,¹⁴ to patients represented in OptumLabs data (**Table 1**). We will not restrict the real-world sample to the planned sample size for the trial, but rather include all patients who

otherwise meet the eligibility criteria. However, we will apply successive inclusion, and then exclusion criteria, determining which criteria have the biggest impact on the size of the population of patients observed in real-world data (**Table 2**).

The PRONOUNCE trial includes male patients, without any age restrictions, with advanced prostate cancer and cardiovascular disease, who were treated with degarelix (Firmagon) or leuprolide (Lupron depot) (**Table 3**). We will first identify all patients who initiated degarelix and leuprolide between 12/24/2008 and 6/30/2019. The start date was selected because degarelix received FDA approval on 12/24/2008. The date of an individual’s first treatment (first fill date) with degarelix or leuprolide will be defined as the index date.

Table 2. Impact of inclusion/exclusion criteria				
PRONOUNCE Eligibility Criteria	Operational Definition in OLDW	Overall Cohort (N=)	Degarelix (Firmagon) (N =)	Leuproli de (Lupron Depot) (N =)
Total % of patients eligible for PRONOUNCE				
Total % of patients ineligible for PRONOUNCE				
Patients who met one of the following exclusion criteria (%)				
Previous or current hormonal management of prostate cancer	A prescription fill of ADT medications within 6 months before the index date <i>OR</i> Procedure codes for bilateral orchiectomy within 6 months before the index date			
Within 30 days prior to randomization: -Myocardial infarction -Stroke (hemorrhagic)	Primary diagnosis of myocardial infarction and stroke (emergency department or inpatient visits) within 30 days before index date; percutaneous coronary intervention, coronary artery			

-Coronary, carotid, or peripheral artery revascularization	bypass grafting, peripheral artery revascularization (limb events), carotid revascularization within 30 days before index date			
--	--	--	--	--

Table 3. Generic Names of Medical Therapy		
Brand name	Drug class	Generic Names
Firmagon	Gonadotropin-releasing hormone receptor antagonists	Degarelix
Lupron Depot	Gonadotropin-releasing hormone receptor agonist	Leuprolide, leuprolide acetate

We will then identify all male enrollees, without any age restrictions, with valid demographic (age and race/ethnicity) and residence data. All enrollees will be required to have at least 6 months of continuous enrollment with medical and pharmacy coverage (i.e. no more than 45 days gap in coverage) before the index date, in order to capture an adequate prior medical history.

We developed an algorithm to identify enrollees with prostate cancer based on clinical expertise and similar methodology outlined in previous studies, which reported positive predictive values between 70% and 82%.¹⁵⁻¹⁸ Specifically, we will require patients to have at least one Evaluation and Management (E&M) visit with a diagnosis of prostate cancer within 6 months before the index date and at least one E&M visit with a diagnosis of prostate cancer any time after the index date. We are not able to ascertain prostate cancer severity, so no categorization by prostate cancer grade will be possible. For our primary analysis, a prostate biopsy will not be required, since patients may receive a diagnosis from a biopsy of metastatic site (e.g. bone or lymph node). As a secondary analysis, we will restrict to a subcohort of patients with at least one prostate biopsy.

To identify patients with a history of cardiovascular disease, it is common to use both primary and secondary discharge diagnosis and procedure codes. This is an established method for administrative claims data research and is used extensively for cohort creation and quality measurement by the Centers for Medicare & Medicaid Services (CMS).¹⁹ In particular, we will identify the following using discharge diagnosis and procedure codes as indicators of a history of cardiovascular disease, at least 30 days before the index date: myocardial infarction, percutaneous coronary intervention (PCI), coronary artery bypass grafting (CABG), peripheral artery revascularization, carotid revascularization any position.

Among patients with prostate cancer and pre-defined cardiovascular disease, we will use pharmacy claims data to exclude patients with a prescription fill of ADT medications within 6 months before the index date. However, for the leuprolide cohort, we will allow patients to remain eligible for inclusion even if they received bicalutamide within one month prior to leuprolide initiation. Leuprolide, a GnRH agonist, can paradoxically lead to a transient increase in testosterone during the first 1 to 3 weeks of treatment. Therefore, bicalutamide is often given for a few weeks before the initial leuprolide injection in order to block any potential adverse effects from the testosterone flare. Degarelix, a GnRH antagonist, does not produce a testosterone flare.

Lastly, we will exclude patients with recent/active cardiovascular events. We will use primary diagnosis of myocardial infarction and stroke (emergency department or inpatient visits) within 30 days before index date, as patients hospitalized for an acute cardiovascular event would be expected to have these diagnoses listed as the primary discharge diagnosis, while patients with a history of cardiovascular disease would be expected to have these diagnoses listed as secondary discharge diagnoses.

5. Measurements

5.1 Baseline characteristics

We will record and summarize key baseline characteristics, including socio-demographic characteristics, comorbidities, and prior and concurrent medication use (**Table 4**). Socio-demographic characteristics include age (mean, median, categories (≤ 54 , 55-64, 65-74, 75+), and race/ethnicity (Asian, Black, Hispanic, White, Unknown). The race/ethnicity data provided by OptumLabs is primarily self-reported, with imputation by the data provider based on other available administrative data when it is missing.²⁰

Update August 2020: After accessing the data and running preliminary analyses, we updated our protocol to account for additional baseline comorbidities (*italicized and underlined* in Table 4). These comorbidities were selected to account for residual confounding by severity of disease. In particular, we observed that degarelix was paradoxically associated with increased mortality. To further account for observed imbalances between the degarelix and leuprolide patients, we also matched on state.

Medical history will be determined using patients' physician, facility, and pharmacy claims before or on the index date. We will use all data available to us to establish patients' medical history (**Table 4**). Data from the baseline 6-months period will be used for all covariates, unless otherwise specified. In OptumLabs, approximately one third of patients with insurance claims data have linked laboratory results, and the availability depends on the contracts between labs testing facilities and the OptumLabs Data Warehouse. For the patients with laboratory data, we will determine serum prostate-specific antigen (PSA) levels and estimated Glomerular Filtration Rate (eGFR). We will also determine whether patients had a prostate

biopsy or received radiotherapy within 6 months before the index date. Previous treatment with bicalutamide and other baseline medications will be determined 6 months prior to index date.

Table 4. Baseline characteristics before and after propensity score matching^a								
	Before PS Matching				After PS Matching			
	Degarelix (N=)	Leuprolide (N=)	Total (N=)	SMD	Degarelix (N=)	Leuprolide (N=)	Total (N=)	SMD
Age								
Mean (SD)								
Median (IQR)								
Age group								
<=54								
55-64								
65-74								
75+								
Race/Ethnicity								
Asian								
Black								
Hispanic								
White								
Unknown								
Geographic Region								
Midwest								
Northeast								
South								
West								
Unknown								
Serum PSA level								
N								
Mean (SD)								
Median (IQR)								
Estimated Glomerular Filtration Rate (eGFR)								
Number								
Mean (SD)								
Median (IQR)								
Prior prostate biopsy (within 6 months of index date)								
Baseline comorbidities								
Coronary artery disease								
Chronic kidney disease								
Congestive heart failure								
Cerebrovascular disease								
Peripheral vascular disease								
Obesity								
Atrial Fibrillation								
Sleep apnea								

Table 4. Baseline characteristics before and after propensity score matching^a

	Before PS Matching				After PS Matching			
	Degarelix (N=)	Leuprolide (N=)	Total (N=)	SMD	Degarelix (N=)	Leuprolide (N=)	Total (N=)	SMD
Hypertension								
<u>MI</u>								
<u>Stroke</u>								
<u>PCI</u>								
<u>CABG</u>								
<u>PAD</u>								
<u>Dementia</u>								
<u>COPD</u>								
<u>Peptic ulcer disease</u>								
<u>Mild liver disease</u>								
<u>Diabetes without chronic complication</u>								
<u>Diabetes with chronic complication</u>								
<u>Metastatic solid tumor</u>								
<u>Rheumatic disease</u>								
<u>Charlson comorbidity score</u>								
<u>Mean (SD)</u>								
<u>Median (IQR)</u>								
Prior radiotherapy (within 6 months before index date)								
Prior use of bicalutamide (within 6 months before index date)								
Other baseline medications (within 6 months before index date)								
Statin								
Non-statin lipid lower medications								
ACEi								
ARB								
ACEi/ARB								
Sacubitril / Valsartan								
Warfarin								
DOAC								
Beta-blockers								
Loop diuretics								
Aldosterone antagonist								
Digoxin								
Calcium channel blocker								
Antiplatelet								
<u>Number of hospitalizations</u>								
0								

Table 4. Baseline characteristics before and after propensity score matching^a

	Before PS Matching				After PS Matching			
	Degarelix (N=)	Leuprolide (N=)	Total (N=)	SMD	Degarelix (N=)	Leuprolide (N=)	Total (N=)	SMD
<u>1</u>								
<u>2+</u>								
Number of ER visits								
<u>0</u>								
<u>1</u>								
<u>2+</u>								
Year of Cohort Entry								
2008								
2009								
2010								
2011								
2012								
2013								
2014								
2015								
2016								
2017								
2018								
2019								
<u>State</u>								
<u>AL</u>								
<u>AZ</u>								
<u>CA</u>								
<u>CT</u>								
<u>FL</u>								
<u>GA</u>								
<u>IL</u>								
<u>IN</u>								
<u>MA</u>								
<u>MN</u>								
<u>MO</u>								
<u>NC</u>								
<u>NJ</u>								
<u>NY</u>								
<u>OH</u>								
<u>RI</u>								
<u>SC</u>								
<u>TN</u>								
<u>TX</u>								
<u>UT</u>								
<u>VA</u>								
<u>WI</u>								
<u>OTHER</u>								

ACEi, angiotensin-converting-enzyme inhibitors; ARB, angiotensin II receptor blocker; CABG, coronary artery bypass grafting; COPD, chronic obstructive pulmonary disease; DOAC, direct-acting oral anticoagulants; IQR, interquartile range; MI = myocardial infarction; PAD, peripheral artery disease; PCI, percutaneous coronary intervention

^a **Underlined values were added after data were accessed**

5.2 Follow-up and outcome ascertainment

OptumLabs Data Warehouse is continuously updated on a monthly basis and the data are complete within 6 months of the service being provided. The analyses of this study will be performed from May to September 2020, implying that the most recent data available to us will be up to October, 2019. Therefore, patients will be followed until the end of the study period (07/31/2019), the end of enrollment in health insurance plans, or death, whichever is first.

5.3 Study outcomes

We will use similar primary and secondary endpoints as the PRONOUNCE trial (**Box 1**). The primary endpoint in the PRONOUNCE trial is the time to first occurrence of the composite Major Adverse Cardiovascular Event (MACE) endpoint, defined as death due to any cause, non-fatal myocardial infarction, or non-fatal stroke. The secondary endpoints in the PRONOUNCE trial include: time from randomization to occurrence of fatal and non-fatal myocardial infarction, fatal and non-fatal stroke, fatal and non-fatal unstable angina requiring hospitalization, and cardiovascular-related death as separate outcomes. Using OptumLabs data, we are able to determine stroke, myocardial infarction, and angina, but are unable to distinguish between fatal and non-fatal events (**Table 5**). However, we will use commonly used, published, and previously validated diagnosis and procedure codes for MACE. For instance, previous evaluations suggest that the performance of similar MACE outcome codes are relatively good, with positive predictive values between 88.4% and 94% for myocardial infarction, 85% for ischemic stroke, and 80%-98% for hemorrhagic stroke.²¹⁻²⁵

Table 5. PRONOUNCE trial Endpoint Definition			
	PRONOUNCE trial Definition	Operational Definition in OLDW	ICD codes
Primary endpoint	Time from randomization to the first confirmed (adjudicated) occurrence of the composite Major Adverse Cardiovascular Event (MACE) endpoint, defined as death due to any cause, non-fatal myocardial infarction or non-fatal stroke [up to 336 days]	Time from randomization to the first confirmed (adjudicated) occurrence of the composite Major Adverse Cardiovascular Event (MACE) endpoint, defined as death due to any cause, myocardial infarction or stroke [up to 336 days]	Myocardial infarction: ICD-9: 41001, 41011, 41021, 41031, 41041, 41051, 41061, 41071, 41081, 41091 ICD-10: I2101, I2102, I2109, I2111, I2119, I2121, I2129, I213, I214, I220, I221, I222, I228, I229
Secondary endpoints	Time from randomization to occurrence of myocardial infarction (fatal, non-fatal) [up to 336 days]	Time from randomization to occurrence of myocardial infarction [up to 336 days]	Stroke: ICD-9: 43301, 43311, 43321, 43331, 43381, 43391, 43401, 43411, 43491, 436, 430, 431, 444X, 435X
	Time from randomization to occurrence of stroke (fatal, non-fatal) [up to 336 days]	Time from randomization to occurrence of stroke [up to 336 days]	ICD-10: I693, I63X
	Time from randomization to occurrence of unstable angina requiring hospitalization (fatal, non-fatal) [up to 336 days]	Time from randomization to occurrence of angina [up to 336 days]	Angina: ICD-9: 413X ICD-10: I20X
	Time from randomization to occurrence of cardiovascular-related death [up to 336 days]	Time from randomization to occurrence of all-cause mortality [up to 336 days]	We cannot determine cardiovascular-related death
ICD = International Classification of Diseases			

Mortality will be identified based on the Social Security Death Master File and discharge status. Before November 2011, the Social Security Death Master File has complete mortality data. However, effective on November 1st, 2011, Section 205(r) of the Social Security Act prohibits the Social Security Administration (SSA) from disclosing state death records that SSA receives through its contracts with the states, except in limited circumstances. Thus, if the SSA knows of a death only from the states and not from any of its other sources of death information, which happens roughly one-third of the time, those death data will not appear on the Death Master File.²⁶ Using discharge status (i.e. in-hospital death), we typically capture an additional 30% of deaths beyond what has been captured by Death Master File; we anticipate therefore that most of the deaths missing from Death Master File should be captured by discharge status, particularly since most deaths occur in an institutional setting. We acknowledge that a small proportion of patients who died out of hospital and were not captured by Death Master File could be missing, however, this should be non-differential between treatment groups and should not influence our comparison.

5.4 Study follow-up

For each patient, we will also determine the follow-up time, which will start the day after initiation of degarelix or leuprolide. Follow-up will continue until the date when the patient experiences any of the following events:

- (a) An outcome of interest
- (b) End of insurance coverage (end of patient enrollment)
- (c) Death
- (d) Reaches the maximum anticipated follow-up of the trial (336 days)

5.5 Missing data

Patients will be considered to have a condition, comorbidity, outcome, or drug exposure if they have a corresponding claim, and will be considered not having a comorbidity, outcome or drug exposure if they do not have a corresponding claim. Although we will therefore not have missing comorbidities, drug use, or outcomes data, misclassification may exist. While this is a limitation of using claims data, the algorithms used to define our inclusion/exclusion criteria, outcomes of interest, and important covariates are commonly used and have demonstrated good performance in previous studies. We suspect that any existing misclassification will be unrelated to treatment group and should not meaningfully impact our findings.

We will exclude patients with invalid demographic data during the cohort creation process (e.g., missing residence region or inconsistent birth year). However, we anticipate fewer than 1% of patients being excluded during the cohort creation. For race/ethnicity, the categories in the database are non-Hispanic white, non-Hispanic black, Hispanic, Asian, other and unknown. The other and unknown will be used as a separate category in the propensity score model.

6. STATISTICAL METHODS

6.1 Main analysis using OptumLabs cohort

For our primary analyses, we will focus on OptumLabs patients who would be eligible for PRONOUNCE based on the operational definitions of the inclusion and exclusion criteria in **Table 1** (base population).

Propensity score matching will be used to balance the difference in baseline characteristics between patients who received degarelix versus those who received leuprolide. A

propensity score, the probability of receiving degarelix, will be estimated using a logistic regression model which includes patient characteristics presented in **Table 4**. No interaction terms will be used. One-to-one nearest neighborhood caliper matching will be used to match patients based on the logit of the propensity score using a caliper equal to 0.2 of the standard deviation of the logit of the propensity score.²⁷ Standardized differences will be used to assess the balance of covariates after matching and a standardized difference within 0.1 will be considered acceptable.²⁸ Covariates with standardized differences above 0.1 will be adjusted for in the regression models.

Cox proportional hazards regression will be used to compare patients receiving degarelix versus those who received leuprolide for the primary and secondary outcomes in the propensity matched cohort, with robust sandwich estimates to account for the clustering within matched sets.²⁹ The proportional hazard assumption will be tested on the basis of Schoenfeld residuals.³⁰ If the proportional hazard assumption is not met, we will assess alternative time to event models, including parametric models, using Akaike information criterion and Bayesian information criterion to determine the final model specification. The Fine and Gray method will be used to consider death as a competing risk when assessing non-fatal outcomes.³¹ All primary analyses will compare the assigned treatment groups under the intention-to-treat principle.

All analyses will be conducted using SAS 9.4 (SAS Institute Inc., Cary, North Carolina) and Stata 16 (Stata Corp, College Station, TX).

6.2 Subgroup Analyses

First, we will repeat our analyses restricted to a subcohort of patients with at least one prostate biopsy. Next, we will perform subgroup analyses for the primary outcome stratified by age, race, diabetes mellitus, and renal function, using receipt of hemodialysis to identify patients

with end-stage renal disease. In addition, for the patients with laboratory data, we will generate subgroups of patients with eGFR <45 and >45. The subgroup analyses will be performed separately in patients who were eligible for the trial (primary analysis) and patients who failed to meet the inclusion criterion/exclusion criteria (secondary analyses). Within each subgroup, we will re-examine the standardized differences to assess the balance of the covariates. If the majority of the standardized differences are above 0.01, we will rematch the patients within each subgroup. Since an increasing number of subgroup analyses could increase the chance of false positive results, we pre-specified the above subgroups since they are either key demographic characteristics or risk factors strongly associated with the primary outcome. However, we will not perform any adjustment for multiple testing.

6.3 Sensitivity Analyses

We will conduct the following sensitivity analyses to assess the robustness of the findings:

1. We will repeat our analyses across two subgroups: (1) patients who failed to meet any one of the cardiovascular inclusion criteria for PRONOUNCE; and (2) patients who met at least one of the cardiovascular exclusion criteria. These subgroups are of particular interest because they represent patient populations that may have been eligible for degarelix or leuprolide treatment, outside of the stringent inclusion/exclusion criteria used by clinical trials. In particular, patients in subgroup 1 are likely to have lower event rates and patients in subgroup 2 are likely to be at higher risk and bad candidates for new drugs. Some patients may have both failed to meet the inclusion criteria and met the exclusion criteria. In the stratified analyses for clinical outcomes, such patients will be classified as those who met the exclusion criteria. Second

2. **Update August 2020: After accessing the data and running preliminary analyses, we discovered that there was cross-over between the two treatments. Therefore, we added three sensitivity analyses to account for the cross-over: Leuprolide, a GnRH agonist, can paradoxically lead to a transient increase in testosterone during the first 1 to 3 weeks of treatment. Therefore, degarelix, a GnRH antagonist that does not produce a testosterone flare and requires fewer office visits, can be given prior to leuprolide. To account for the possibility of patients crossing-over between the two treatment groups, we will repeat our analyses: (1) excluding all patients that cross-over between the two treatments and (2) censoring patients at the point at which they switch. We used inverse probability of treatment weighting (IPTW) instead of propensity score matching to minimize confounding.**
3. We will conduct a stratified analysis based on the adherence to degarelix and leuprolide, i.e., patients with proportion of days covered (PDC) $<80\%$ and those with PDC $\geq 80\%$, since the adherence to medical therapy in practice is often lower than that in clinical trials. The adherence will consider all drugs that a patient used during follow up, even if they were different from the initial treatment. For degarelix, which is administered monthly, we will use 30 days supply. For leuprolide, there are multiple dosing intervals, which makes timing of the next dose dependent on the dose given at the last injection. For leuprolide patients without dose information, we will assume that fills were for 30 days. After matching the degarelix patients with PDC $\geq 80\%$ with the leuprolide patients with PDC $<80\%$, we will conduct the cox regression analyses to compare the outcomes between the two groups. Analyses will be repeated among patients with PDC $<80\%$.

4. We will assess falsification endpoints to test for residual confounding. Treatment effects estimated in observational studies are prone to unmeasured confounding. In recent years, a falsification end point, also called a control outcome, has become a popular method to assess for unmeasured confounding.³²⁻³⁴ A falsification endpoint is a health outcome that researchers believe is highly unlikely to be casually related to the treatment in question. If a significant relationship is found between the treatment and a falsification endpoint, it may indicate the treatment groups are different in some unmeasured ways, i.e. the existence of unmeasured confounding. This method is similar to a negative control, a routine precaution taken in the design of biologic laboratory experiments, and is recommended to be used to detect confounding and bias in observational studies.^{33,35,36} This method is particularly useful in observational studies comparing different treatment options, because the unmeasured confounding in these studies tend to make one group systematically healthier or less susceptible to adverse outcomes than the other group.

We selected two endpoints that that are unlikely to be associated with use of either degarelix or leuprolide – chronic obstructive pulmonary disease and appendicitis/cholecystitis. If a significant relationship were to be found between degarelix and any of these endpoints, it would indicate the existence of residual confounding.

6.4 Comparison of cohort and trial population characteristics and results

Once the ongoing trials have been completed, we will compare the trial population to the population of patients identified in the claims component of OptumLabs data after application of the pre-specified eligibility criteria, as described above, to determine how accurately the characteristics of trial populations can be predicted. For each individual characteristic included in

Table 4 and reported in the PRONOUNCE trial publication, we will make pairwise comparisons between the trial population and the population of patients identified in the claims component of OptumLabs, stratified by treatment arm. In particular, within the degarelix and leuprolide arms, we will take the paired differences between the standardized mean differences (Cohen's d) from the trial population and the real-world population. Differences between standardized mean differences within 0.2 will be considered acceptable.

We will also compare rates of missing data and loss to follow-up across arms. If we observe significant differences in the characteristics of the real-world and trial populations, we will estimate a real-world population reweighted to mirror the characteristics of the PRONOUNCE trial population.

Once the PRONOUNCE trial have been completed and published, we will compare the final primary and secondary endpoint results to the results estimated from A) the population of real-world patients meeting pre-specified trial eligibility criteria and (when needed as explained above) B) the population of real-world patients reweighted to mirror the characteristics of the final enrolled population of patients in the trial, for each analytic approach employed as described above. As needed, we will also compare each of the observational approaches used above to the RCT results, providing a better understanding of the tradeoffs inherent to each of the proposed methods.

Two approaches for comparing results will be used. First, as a simple method, results from both the real-world data and the trial will be characterized as positive (i.e., degarelix statistically significantly reduces the risk of cardiovascular complications as compared to leuprolide), neutral (no statistically significant difference between degarelix and leuprolide), or negative (degarelix statistically significantly increases the risk of cardiovascular complications as

compared to leuprolide) and a percent agreement will be estimated. Statistical tests will be 2-sided and significance will be set at $P < 0.05$.

Second, we will pursue a more sophisticated method. The hazard ratios calculated for the primary and secondary outcomes using the real-world data will be converted to natural logarithm hazard ratios (lnHR). For each outcome, we will then take the difference between the lnHR calculated using the real-world data and the lnHR reported by the PRONOUNCE trial. After exponentiating each difference, a ratio of hazard ratios greater than 1.0 will imply greater (more beneficial) treatment effects in the real-world population than in the PRONOUNCE trial population. We will calculate 95% confidence intervals for the ratios of the hazard ratios by taking the square root of the sum of the variance for the hazard ratio derived from the real-world data and the variance for the hazard ratio from the PRONOUNCE trial population. Our variance calculations will be based on assumption of independence (i.e. correlation coefficients of zero, indicating that the outcomes from the real-world data and trial data are independent).³⁷

REFERENCES

1. Office of Technology Assessment, U.S. Congress. The Impact of Randomized Clinical Trials on Health Policy and Medical Practice: Back-ground paper (OTA-BP-H-22). August 1983: <https://repository.library.georgetown.edu/bitstream/handle/10822/708483/8310.PDF?sequence=1>. Accessed July 26, 2018. In.
2. Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-World Evidence - What Is It and What Can It Tell Us? *N Engl J Med*. 2016;375(23):2293-2297.
3. Hernán MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health*. 2006;60(7):578-586.
4. Melloni C, Washam JB, Jones WS, et al. Conflicting results between randomized trials and observational studies on the impact of proton pump inhibitors on cardiovascular events when coadministered with dual antiplatelet therapy: systematic review. *Circ Cardiovasc Qual Outcomes*. 2015;8(1):47-55.
5. Fralick M, Kesselheim AS, Avorn J, Schneeweiss S. Use of Health Care Databases to Support Supplemental Indications of Approved Medications. *JAMA Intern Med*. 2018;178(1):55-63.
6. Lonjon G, Boutron I, Trinquart L, et al. Comparison of treatment effect estimates from prospective nonrandomized studies with propensity score analysis and randomized controlled trials of surgical procedures. *Ann Surg*. 2014;259(1):18-25.
7. Fretheim A, Zhang F, Ross-Degnan D, et al. A reanalysis of cluster randomized trials showed interrupted time-series studies were valuable in health system evaluation. *J Clin Epidemiol*. 2015;68(3):324-333.
8. Tannen RL, Weiner MG, Xie D, Barnhart K. A simulation using data from a primary care practice database closely replicated the women's health initiative trial. *J Clin Epidemiol*. 2007;60(7):686-695.
9. Tannen RL, Weiner MG, Xie D. Replicated studies of two randomized trials of angiotensin-converting enzyme inhibitors: further empiric validation of the 'prior event rate ratio' to adjust for unmeasured confounding by indication. *Pharmacoepidemiol Drug Saf*. 2008;17(7):671-685.
10. Dawson NA. Patient education: Treatment for advanced prostate cancer (Beyond the Basics). In: UpToDate, Post TW (Ed), UpToDate, waltham, MA. (Accessed on 3 April, 2020). In.
11. Nanda A, Chen MH, Braccioforte MH, Moran BJ, D'Amico AV. Hormonal therapy use for prostate cancer and mortality in men with coronary artery disease-induced congestive heart failure or myocardial infarction. *JAMA*. 2009;302(8):866-873.
12. Albertsen PC, Klotz L, Tombal B, Grady J, Olesen TK, Nilsson J. Cardiovascular morbidity associated with gonadotropin releasing hormone agonists and an antagonist. *Eur Urol*. 2014;65(3):565-573.
13. Wallace PJ, Shah ND, Dennen T, Bleicher PA, Bleicher PD, Crown WH. Optum Labs: building a novel node in the learning health care system. *Health Aff (Millwood)*. 2014;33(7):1187-1194.
14. Melloni C, Solvin SF, Blemings A, et al. Cardiovascular Safety of Degarelix Versus Leuprolide for Advanced Prostate Cancer: The PRONOUNCE Trial Study Design. 2020;2(1).
15. Unger JM, Hershman DL, Till C, et al. Using Medicare Claims to Examine Long-term Prostate Cancer Risk of Finasteride in the Prostate Cancer Prevention Trial. *J Natl Cancer Inst*. 2018;110(11):1208-1215.
16. McClish DK, Penberthy L, Whittemore M, et al. Ability of Medicare claims data and cancer registries to identify cancer cases and treatment. *Am J Epidemiol*. 1997;145(3):227-233.
17. Nordstrom BL, Whyte JL, Stolar M, Mercaldi C, Kallich JD. Identification of metastatic cancer in claims data. *Pharmacoepidemiol Drug Saf*. 2012;21 Suppl 2:21-28.
18. Cooper GS, Yuan Z, Stange KC, Dennis LK, Amini SB, Rimm AA. The sensitivity of Medicare claims data for case ascertainment of six common cancers. *Med Care*. 1999;37(5):436-444.

19. Centers for Medicare & Medicaid Services. Measures Methodology. <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Measure-Methodology>. Accessed 17 July 2020. In.
20. Hershman DL, Tsui J, Wright JD, Coromilas EJ, Tsai WY, Neugut AI. Household net worth, racial disparities, and hormonal therapy adherence among women with early-stage breast cancer. *Journal of Clinical Oncology*. 2015;33(9):1053-1059.
21. Kiyota Y, Schneeweiss S, Glynn RJ, Cannuscio CC, Avorn J, Solomon DH. Accuracy of Medicare claims-based diagnosis of acute myocardial infarction: estimating positive predictive value on the basis of review of hospital records. *Am Heart J*. 2004;148(1):99-104.
22. Wahl PM, Rodgers K, Schneeweiss S, et al. Validation of claims-based diagnostic and procedure codes for cardiovascular and gastrointestinal serious adverse events in a commercially-insured population. *Pharmacoepidemiol Drug Saf*. 2010;19(6):596-603.
23. Andrade SE, Harrold LR, Tjia J, et al. A systematic review of validated methods for identifying cerebrovascular accident or transient ischemic attack using administrative data. *Pharmacoepidemiol Drug Saf*. 2012;21 Suppl 1:100-128.
24. Tirschwell DL, Longstreth WT. Validating administrative data in stroke research. *Stroke*. 2002;33(10):2465-2470.
25. Roumie CL, Mitchel E, Gideon PS, Varas-Lorenzo C, Castellsague J, Griffin MR. Validation of ICD-9 codes with a high positive predictive value for incident strokes resulting in hospitalization using Medicaid health data. *Pharmacoepidemiol Drug Saf*. 2008;17(1):20-26.
26. da Graca B, Filardo G, Nicewander D. Consequences for healthcare quality and research of the exclusion of records from the death master file. *Circulation: Cardiovascular Quality and Outcomes*. 2013;6(1):124-128.
27. Austin PC. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharm Stat*. 2011;10(2):150-161.
28. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat Med*. 2009;28(25):3083-3107.
29. Gayat E, Resche-Rigon M, Mary JY, Porcher R. Propensity score applied to survival data analysis through proportional hazards models: a Monte Carlo study. *Pharm Stat*. 2012;11(3):222-229.
30. Grambsch PM, Therneau TM. Proportional hazards test and diagnostics base don weighted residuals. *Biometrika*. 1994;81(3):515-526.
31. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American statistical association*. 1999;94(446):496-509.
32. Prasad V, Jena AB. Prespecified falsification end points: can they validate true observational associations? *JAMA*. 2013;309(3):241-242.
33. Dusetzina SB, Brookhart MA, Maciejewski ML. Control Outcomes and Exposures for Improving Internal Validity of Nonrandomized Studies. *Health Serv Res*. 2015;50(5):1432-1451.
34. Ioannidis JP. Are mortality differences detected by administrative data reliable and actionable? *Jama*. 2013;309(13):1410-1411.
35. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology*. 2010;21(3):383-388.
36. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med*. 2015;34(28):3661-3679.
37. Wallach JD, Ciani O, Pease AM, et al. Comparison of treatment effect sizes from pivotal and postapproval trials of novel therapeutics approved by the FDA based on surrogate markers of disease: a meta-epidemiological study. *BMC Med*. 2018;16(1):45.

