Optimizing the Effectiveness of Routine Post-Treatment Surveillance in Prostate Cancer Survivors

National Clinical Trial (NCT) Identification Number: NCT02957357

Principal Investigator: Ronald C. Chen

20 Dec 2016

Study Protocol and Statistical Analysis Plan

Specific Aim 1: Compare patient-centered outcomes (including survival, procedures/tests, treatments) in prostate cancer survivors followed with alternative surveillance frequencies.

Hypothesis: More frequent surveillance increases the receipt of additional procedures and tests in all groups of patients, but improves survival only in patients with high-risk prostate cancer.

Synopsis: We will analyze data from the National Cancer Data Base (NCDB), which includes real-world prostate cancer treatment, surveillance and outcomes data on ~70% of US patients. We will compare the effectiveness of the three most common surveillance frequencies (PSA testing every 3 months vs. 6 months vs. 12 months) stratified by risk of recurrence and initial treatment (surgery or radiation).

Introduction: The primary goal for PSA surveillance after prostate cancer treatment is to detect recurrent cancer, thus allowing provision of additional treatments which can improve survival.[1,2] Intuitively, patients who are most likely to recur after initial treatment ("high risk" cancer) may require more frequent surveillance, while those least likely to recur ("low risk") may require less frequent surveillance. Further, patients treated by surgery vs. radiation may require different surveillance frequencies because the definitions of recurrence differs by treatment, and false positives from surveillance PSA are much more common after radiation.[3,4] These considerations underlie the rationale to examine outcomes separately in 6 individual groups based on *standard definitions* of recurrence risk[5] and initial treatment:

- Low-risk prostate cancer (defined by clinical stage ≤T2a, PSA <10, and Gleason score ≤6): Recurrence risk is <10%.[6] Two cohorts will be examined separately, 1) those treated with **initial prostatectomy**, and 2) those treated with **initial radiation** (including external radiation and/or brachytherapy).
- Intermediate-risk prostate cancer (defined by clinical stage T2b-c, and/or PSA 10-20, and/or Gleason 7): Recurrence risk is ~30%.[7] Two cohorts will be examined separately, 3) those treated with **initial prostatectomy**, and 4) those treated with **initial radiation**.
- High-risk prostate cancer (defined by clinical stage ≥T3, PSA >20, or Gleason 8-10): most aggressive type of prostate cancer. Recurrence risk is ~50%.[7] Two cohorts will be examined separately, 5) those treated with **initial prostatectomy**, and 6) those treated with **initial radiation**.

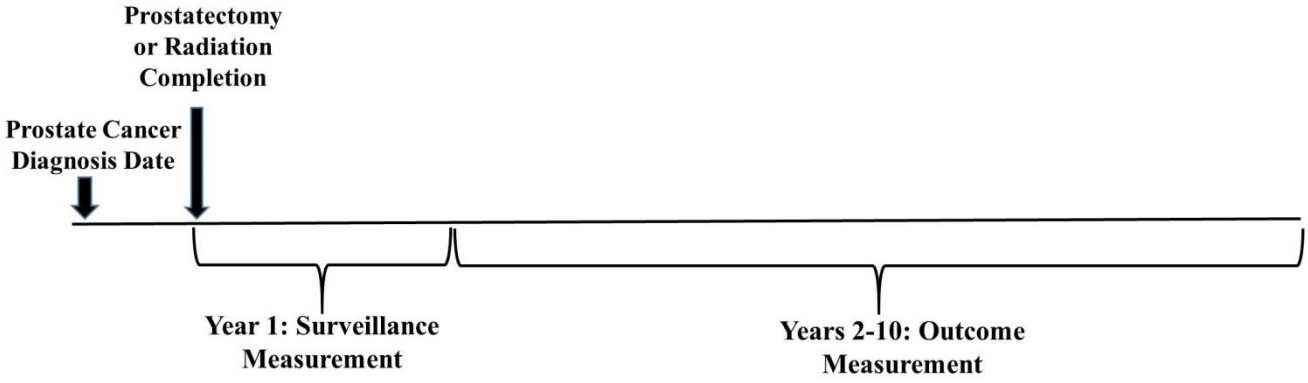
The objective of this aim is to quantify the potential benefits (survival) and harms (procedures/tests, treatments) of different PSA surveillance frequencies in the 6 patient groups, which are currently unknown. It is intuitive that more frequent surveillance will lead to more findings of elevated PSA, thus leading to more downstream procedures and tests, and treatments; with unclear survival benefit. The *rationale* for this aim is that successful completion of the proposed research will contribute novel, currently unavailable, data to inform patient decision-making.

Comparators [RQ-5]: We selected our comparators of surveillance frequency based on published recommendations and by engaging diverse stakeholders (including clinicians who specialize in prostate cancer care). The three most common surveillance frequencies in clinical

practice are every 3, 6, or 12 months. In each of the 6 patient groups, patient outcomes of every 3 vs. 6 vs. 12 month surveillance will be directly compared. We will define our surveillance groups based on the number of PSA tests patients receive during the first year following completion of prostatectomy or radiation therapy (Figure) – using an "intention to treat" approach – and analyze patient outcomes from Years 2-10. Operationally, we will compare the outcomes of patients who have 1 vs. 2-3 vs. ≥4 PSA tests within the first year. We selected this timeframe as we believe that these patterns would be reflective of the physician's intended surveillance schedule as opposed to later years which may be more reflective of patient adherence to surveillance recommendations.

In sensitivity analysis, we will vary this timeframe to define surveillance frequency based on the timing of the first PSA test following completion of prostatectomy or radiation therapy (first PSA within 3 months vs. >3 but within 6 months vs. >6 months); or define comparison groups empirically (most frequent 1/3 of patients vs. middle vs. least frequent 1/3 of patients) – to assess if our findings are robust to these varying exposure groupings [IR-5]. We will also explore using Years 1-2 to define surveillance frequency, and examining patient outcomes from Years 3-10.

Figure 1. Sample Identification and Definition of Timeframes for Measurement



Data Source [IR-1][DR-2][PC-2]: The National Cancer Data Base (NCDB) is a joint program between the Commission on Cancer (CoC) of the American College of Surgeons and the American Cancer Society.[8] NCDB, started in 1989, is a nationwide cancer outcomes database that includes >1,400 CoC-accredited cancer programs and ~70% of all newly diagnosed cancer cases in the US – making its data broadly generalizable and ideal for CER.[8-10] Systematic, longitudinal follow-up data are updated annually, with multiple data-quality assurance steps taken. **The high quality and representative NCDB data are used for the annual US "cancer statistics" report**.[11,12] Thus, our results using the NCDB optimizes external validity and generalizability. Data elements are recorded by certified cancer registrars using nationally standardized data item and coding definitions specified by the Facility Oncology Registry Data Standards. Detailed information about data elements collected by the NCDB, standardized data collection protocol, and quality control methods used to assess reported case information is available at: *www.facs.org/cancer/ncdb/datasubmission.html*.

NCDB Special Study mechanism for primary data collection will be used to obtain the detailed level of information required for this study on data elements which do not currently exist in the NCDB registry – including PSA frequency – and to fill in missing data [MD-1].[13-15] This will be done using chart abstraction by each of the CoC-accredited centers. For this Aim, the NCDB will randomly select 10 prostate cancer patients diagnosed in 2005-07 from each CoC-

accredited center, with purposeful sampling in order to achieve approximately the same overall sample size in each of the 6 risk group x treatment cohorts. By randomly selecting a small number of patients from a large number of institutions, we can ensure a broadly representative sample of hospitals and patients, ideal for CER.

Outcomes [RQ-6]: The NCDB annually collects vital status (survival), cancer status (disease free or recurrence), date of recurrence, and treatments for recurrent cancer. This will allow for evaluation of our primary outcome of overall survival, and secondary outcomes of time to recurrence and treatment for recurrent cancer. Additional secondary outcomes will be collected by the special studies mechanism, and include: rates and types of procedures (e.g. biopsy) and tests (e.g. advanced imaging), results of these tests, and treatment for recurrent cancer.

The special studies mechanism will also confirm existing NCDB information regarding cancer status and date of recurrence. Specifically, the PSA dates and results will be abstracted from medical records. Therefore, the investigators are able to define recurrence per standard definitions. These citations provide official definitions for biochemical recurrence after radiation treatment[16] and radical prostatectomy.[17]

Cancer registry staff will review hospital and provider records to collect these data, using the same rigorous procedures as primary registry data with specific instructions and training for abstraction by the investigative team and personnel at CoC. These data will be combined with the existing NCDB data to provide the analysis dataset. This Table summarizes the data elements (outcomes) to be collected using the special studies mechanism: (Table 1)

Data Point	**Format**
Date of imaging	mm/dd/yy
Type of imaging performed	1) bone scan; 2) CT scan; 3) MRI; 4) X-ray; 5) PET scan; 6) other
Date of biopsy/procedure	mm/dd/yy
Biopsy/procedure results	1) benign; 2) malignant; 3) indeterminate
Date of treatment for recurrence	
Type of treatment	1) radiation; 2) surgery; 3) hormone therapy; 4) chemotherapy; 5) other therapy; 6) no treatment

The full description of all the variables which will be collected is provided in a separate document. Of note, we will first undergo a pilot study with a limited number of participating sites to test the feasibility of collecting each data element retrospectively. Lessons from the pilot study will inform necessary modifications of data elements collected before embarking on the full with sites across all CoC-accredited centers.

Covariates [IR-3][CI-4]: Data collected at time of diagnosis by the NCDB include, but are not limited to, the following:

- *Demographics*: age, race/ethnicity, marital status, county of residence, census tract data, occupation, tobacco history, rural/urban continuum, insurance

- *Cancer specific variables*: histology/behavior (ICD-0-2, ICD-0-3), diagnostic date. Starting in 2004: PSA level, Gleason score and clinical stage – used to classify patients into low-, intermediate-, and high-risk categories.[7]
- *Comorbidities:* Up to 10 per patient are abstracted and scored per the Charlson-Deyo comorbidity score.[18]
- *Hospital characteristics*: unique facility identifier, type of cancer center
- *First course of treatment:* surgery, radiation, hormonal/endocrine, treatment dates.

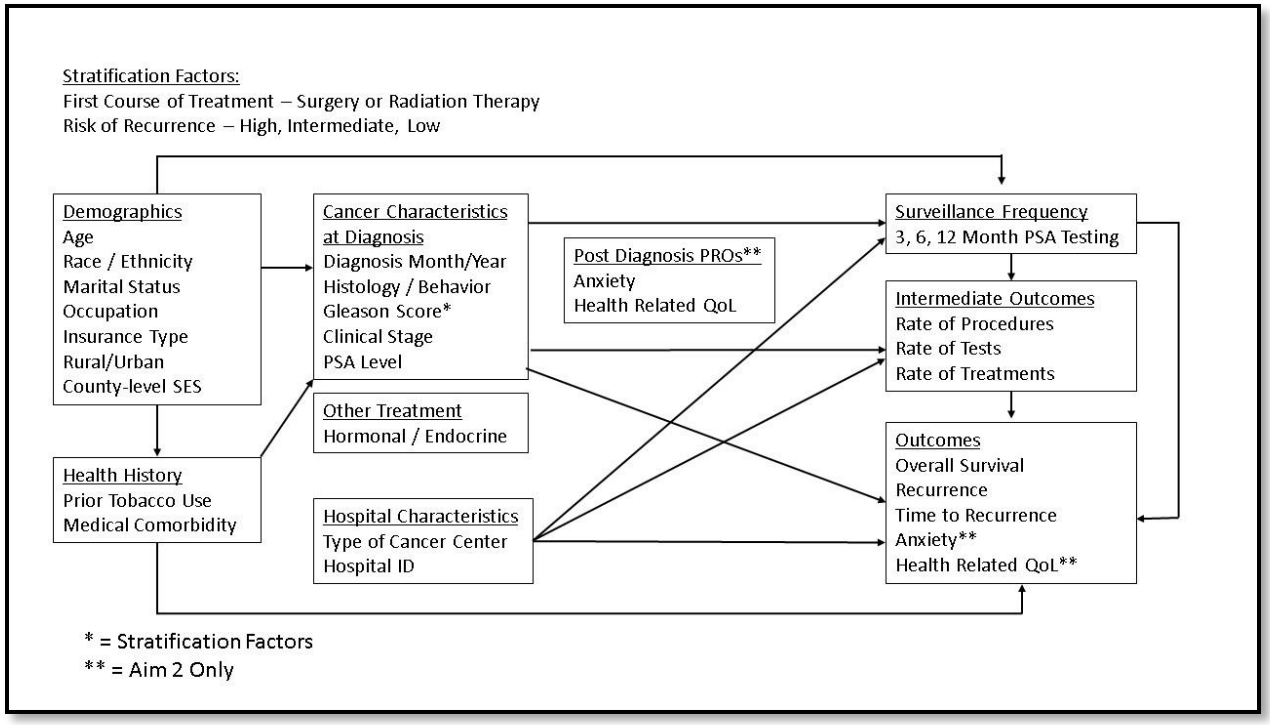
Data Analysis [IR-3]: Patients diagnosed in 2005-2007 will be included. By the time of data analysis, the NCDB will have survival data through 2015, providing 8-10 years of follow-up data for each patient.

The analysis is stratified by two important factors, risk of recurrence (low, intermediate, high) and type of initial treatment (surgery, radiation); thus, 6 separate analyses will be performed. All outcomes will be defined from years 2-10 from end of prostatectomy or radiation treatment (Figure 1). We will use two modeling strategies to evaluate our outcomes of interest: 1) modified Poisson/binary regression[19,20] with sandwich standard errors to estimate risks of the events of interest during a set follow-up period; and 2) Cox Proportional Hazards regression [CI-3]. Using the modified Poisson/binary regression models, we will estimate the risk of experiencing the outcome of interest (death, recurrence, use of procedures and tests, treatment for recurrence) during a defined interval beginning in the second year following treatment. Additionally, we will estimate overall survival using adjusted Cox Proportional Hazards regression models. We will evaluate the proportional hazards assumption by generating Kaplan-Meier survival curves and through testing interactions between relevant covariates and time in the Cox model. Covariates that violate the proportional hazards assumption will be used as stratification variables as appropriate. In sensitivity analyses, we will generate a propensity score (where the patient, provider and hospital characteristics measured at the time of prostatectomy or radiation completion are included in the propensity score) and create a matched cohort of patients to evaluate using Kaplan-Meier survival models [described in detail in C3]. For secondary outcomes, time to recurrence will be modeled for each surveillance frequency using a competing risk analysis as described by Fine and Gray.[21] Similar methods will be used to calculate the cumulative incidence of procedures and tests and treatment for recurrence.

Several outside/environmental factors can affect a patient's surveillance frequency. In our preliminary analysis using the US population-based Surveillance, Epidemiology and End Results (SEER)-Medicare dataset, we found race, marital status, and regional socioeconomic determinants (high school graduation rate) to be associated with different PSA surveillance frequencies. The NCDB captures a comprehensive list of covariates (see above). We will examine if surveillance frequency is associated with each individual factor and account for these factors using modeling as above. Finally, we recognize that there may still be important unmeasured confounders in our study; therefore, we plan to use sensitivity analyses to estimate the impact of a hypothetical unmeasured/unobserved confounder, using similar methodology used in a prior study by the co-I (Dusetzina).[22] This analysis will allow us to determine the extent of confounding that would substantively change our study findings.

A conceptual diagram (directed acyclic graph) demonstrating the relationship between measured covariates and our selected study outcomes for Aim 1 is provided in Figure 2 below.

Figure 2. Directed Acyclic Graph for the Hypothesized Relationship between Selected Study Variables and Outcomes for Aim 1



Footnote: Analyses will be stratified by the type of initial treatment (surgery, radiation) and recurrence risk (low, intermediate, high) and the hypothesized relationships may vary somewhat by each clinical scenario. Relationship between variables identified as pertaining to Aim 2 are noted in Figure 3.

Statistical power: There are >1,400 CoC-accredited hospitals. With 10 patients sampled at each hospital, we expect 14,000 patients will be available for analysis. We will intentionally sample approximately the same number of patients in each of the 6 cohorts (risk group x primary treatment), providing 2,333 patients in each cohort. Within each cohort, patients receiving different surveillance frequencies will be compared. Based on preliminary analysis of SEER-Medicare data, the national patterns of PSA surveillance in the first year are: (Table 2)

Number of PSAs	Low Risk	Intermediate Risk	High Risk
1	20%	18%	17%
2-3	61%	59%	53%
4	19%	23%	30%

Power calculations used a range of 10-year overall survival rates from 70-80%, α=.05, and two-sided test. Using distributions in Table 2, there is >80% power to detect a 5% survival difference among groups of different surveillance frequencies. Our stakeholders have indicated this level of survival difference to be meaningful.

Potential problems and alternative strategies:
Although the analysis will stratify by prostate cancer risk status, it is possible that within each

risk strata, patients with clinical and pathological features that are associated with a worse prognosis may be followed more frequently than patients with a better prognosis. One way we have proposed to address this issue is with instrumental variable analysis (Page 11). We will test two commonly-used, potential instrumental variables: region and the provider's most commonly used surveillance regimen (commonly called "provider preference"). Another way to address this in our primary analysis is to account for risk variables (clinical and pathological features) within each risk strata. We will plan to incorporate this in the primary analysis as control variables within the outcome models or within the propensity score model, as appropriate to the research question.

Our definition of comparison groups is based on published recommendations and stakeholder guidance. However, schedules for surveillance PSA tests in this observational cohort may not precisely match the anticipated 3, 6, and 12 month intervals. Multiple classification schemes will be assessed for characterizing the surveillance intervals, as described on page 2. The analysis will be repeated using alternative classification schemes, and sensitivity analyses will be used to determine if there are meaningful differences in the results based on classification of surveillance intervals, and to optimize our classification of surveillance frequency [IR-5].

Patients treated within a site may be more similar with regard to surveillance frequency than patients treated within other sites. Given this, we will evaluate the need for clustering patients within sites by estimating the variance in surveillance frequency and outcomes within versus between sites. If clustering is indicated we will incorporate a site level term into our regression models as 1) a grouping variable; 2) as a fixed effect within the regression model (to control for time-invariant omitted variables and methodological artifacts of temporal correlation), and 3) as a random effect (to control for unobserved heterogeneity which may be constant over time and correlated with independent variables). These analyses will be in addition to the analytic techniques previously proposed (propensity score and instrumental variables methods), which provide alternative approaches to examining for these issues, and are discussed in C3. If the variation between clusters is small, we will present regression results without clustering (focused on patient-level analyses) for our primary analytic strategy.

We will assess the prevalence of patients receiving treatment before meeting official definition of recurrence, and whether this prevalence is similar or different in the comparison groups. In our primary analysis, we will assess outcomes (procedures, treatments) regardless of whether patients received treatments before meeting official definition of recurrence – because this is likely the result of patient/physician anxiety caused by frequent PSA surveillance testing. However, in sensitivity analysis, we will exclude patients who received treatment before meeting official definition of recurrence.

C2. Specific Aim 2: Compare patient-reported outcomes in prostate cancer survivors followed with alternative surveillance frequencies.

Hypothesis: More frequent surveillance is associated with more patient-reported anxiety and worse health-related quality of life.

Introduction: Patient stakeholders have indicated that increased surveillance testing increases anxiety, and that the quality of life impact from frequent surveillance is important to inform decision-making. *The objective of this aim* is to quantify the impact of different surveillance frequencies from the patient's perspective. Aim 2 meets PCORI methodology standards [PC-3] by capturing patient-reported outcomes when the patient is the best source of information. In addition to anxiety, we will measure health-related quality of life in a population-based cohort of prostate cancer survivors undergoing routine surveillance. The *rationale* for this aim is that successful completion of the proposed research will provide further patient-centered information to inform decision-making. Our use of a population-based cohort facilitates generalizability of study findings to the overall prostate cancer survivor population.

Comparators: Comparison groups are the same as Aim 1. PSA frequency will be similarly defined using medical record abstraction.

Data Source [PC-2]: The North Carolina Prostate cancer Comparative Effectiveness & Survivorship Study (NC ProCESS), PI Ronald Chen, is a prospective population-based cohort of >1,000 patients with newly diagnosed prostate cancer, enrolled from January 2011 through June 2013.[23] Patients were identified in collaboration with the North Carolina Cancer Registry's Rapid Case Ascertainment system, and all patients were enrolled prior to treatment and actively followed prospectively with repeated collection of patient-reported outcomes data and medical records. The primary goal of NC ProCESS is to compare the effectiveness of different surgery and radiation treatments for prostate cancer; these participants are now receiving routine post-treatment surveillance. As a population-based cohort, NC ProCESS participants have similar characteristics (age, prostate cancer diagnosis, treatment patterns) as prostate cancer patients across the US, and are diverse, with 31% non-Caucasian (including 27% African American). These characteristics make NC ProCESS an ideal cohort to examine patient-reported outcomes.

The data available in NC ProCESS include:

- Already available: information regarding prostate cancer diagnosis, primary treatment received, and comorbidities.
- Will be abstracted from medical records: PSA dates and results, procedures, tests, and treatments.

Outcomes [RQ-6][IR-4][PC-3]:

Primary outcome: Our stakeholders identified anxiety and HRQOL (see below) as potential harms of frequent surveillance. The *validated* Memorial Anxiety Scale for Prostate Cancer (MAX-PC)[24] will be used as primary outcome for Aim 2. This instrument contains 18 questions assessing anxiety related to prostate cancer (11 items, total score 0 to 33), PSA testing (3 items, total score 0 to 9), and fear of recurrence (4 items, total score 0 to 12). A higher score in each subscale indicates more anxiety.

Secondary outcomes: Prostate cancer treatments can have detrimental HRQOL effects, including effects on urinary, bowel and sexual function symptoms;[25] and global quality of life.[26] However, HRQOL outcomes in patients who receive different post-treatment surveillance frequencies have not been previously studied. It is likely (we hypothesize) that more frequent surveillance will detect more recurrences, leading to more procedures and treatments resulting in worse HRQOL.

If indeed more frequent surveillance is associated with worse HRQOL, this is critically important data to inform patients about the trade-offs (potential benefit: improved survival; harm: worse HRQOL) of this decision. We will measure 1) Cancer-specific HRQOL using the *validated*, and well-published instrument Prostate Cancer Symptom Indices (PCSI).[27-34] PCSI measures treatment-related morbidity,[34] and is comprised of scales which measure sexual, bowel and urinary symptoms. In each scale, patient answers are converted to a score from 0 (no symptom) to 100 (maximum symptoms). 2) Global HRQOL will be measured by the *validated* Short-Form 12 (SF-12), one of the most frequently used HRQOL instruments in medicine, and has been used frequently in prostate cancer.[35] SF-12 provides the Mental (MCS) and Physical Component Summary scores (PCS), calculated using norm-based scoring, with the mean score 50 and standard deviation 10.[36,37] A lower score indicates lower HRQOL.

NC ProCESS contains complete PCSI and SF12 data at baseline (pre-treatment), and 12 months after treatment; and MAX-PC data at 12 months. NC ProCESS continues to collect these data at 12-month interval time points. By the end of this proposed study, all participants will complete 60-month (5-year) data collection; these data will be available for Aim 2 analysis.

Covariates [IR-3][CI-4]: Data collected at the time of diagnosis by the NC ProCESS include

From Cancer Registry: county of residence;

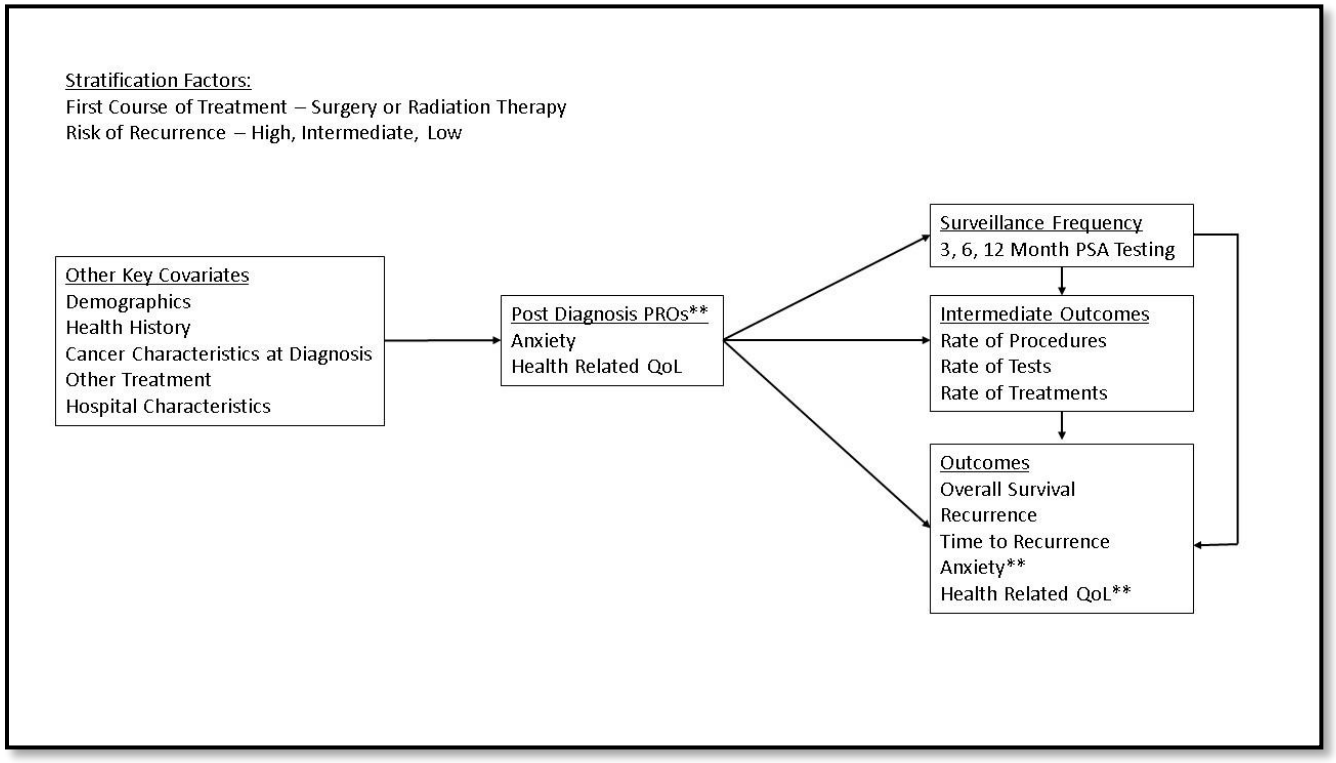
From Medical records: age at diagnosis, comorbidities (scored using the validated Charlson index), prostate cancer diagnosis (low, intermediate, high risk), treatments;

From patient report: education, household income, race/ethnicity, insurance, marital status, employment.

Data Analysis [IR-3]: Consistent with prostate cancer HRQOL literature, we will treat each patient-reported outcome measure as a continuous variable. We will model changes in anxiety (primary outcome) and HRQOL (secondary outcomes) from the first time point through 60 months using repeated measures methodology such as generalized estimating equations to examine change over time while controlling for patient related covariates. Models will be estimated using PROC GENMOD (SAS 9.4) with an identity link and normal distribution for continuous variables. Models will account for the covariates described above. For the primary outcome of anxiety, we will compare the overall anxiety score among comparison groups, as well as scores for each of the 3 subscales to determine if there is a particular aspect of anxiety which differs by surveillance frequency. For PCSI, scale scores for sexual, urinary, and bowel function will be compared across groups. For the SF-12, mental and physical component scores will be calculated and compared.

A conceptual diagram (directed acyclic graph) demonstrating the relationship between measured covariates and our selected study outcomes for Aim 2 is provided in Figure 3 below. Of note, we anticipate similar relationships between the categories of variables identified in the "other key covariates" section of the directed acyclic graph below to mirror those in Aim 1 / Figure 2. For ease of interpretation we focus primarily on the relationship between variables unique to Aim 2 in the graphic below.

Figure 3. Directed Acyclic Graph for the Hypothesized Relationship between Selected Study Variables and Outcomes for Aim 2



Footnote: Analyses will be stratified by the type of initial treatment (surgery, radiation) and recurrence risk (low, intermediate, high) and the hypothesized relationships may vary somewhat by each clinical scenario.

Statistical power: A total of 1,079 participants have completed MAX-PC (primary outcome), PCSI and SF12 (secondary outcomes) at 12 months. If there is attrition of 25% from 12 to 60-month time points, that leaves N=809 with complete data. Using the distributions in Table 2, this sample size provides >90% power to detect a ½-standard deviation (clinically significant) difference in anxiety among the different surveillance frequency groups. More specifically, a sample size as small as 154 in each group will provide 90% power to detect ½-standard deviation difference among groups. For secondary outcomes PCSI and SF-12, assuming a standard deviation of 10,[36,37] we have >90% power to detect clinically meaningful differences (1/2 standard deviation = 5 points) among groups. All α=.05; two-sided tests.

Potential problems and alternative strategies: Patients with different baseline HRQOL and anxiety may receive different surveillance frequencies. The longitudinal design of Aim 2 and availability of pre-treatment data allow us to analytically account for these potential differences among comparison groups. We do not have MAX-PC data pre-treatment, but will examine if pre-treatment anxiety (an item on SF-12) differs by surveillance frequency groups. If so, we will adjust for this baseline anxiety measure in data analysis. We will address potential clustering as described on Page 6.

We will evaluate the extent to which attrition impacts our analysis and subsequent interpretation. We have planned multiple sensitivity analyses to ensure the robustness of our findings, including

evaluations of the impact of a hypothetical unmeasured confounder, instrumental variables analysis, stratification and restriction.

C3. Avoidance of bias [IR-3][GM-3][CI-5][CI-6]: The large population cohorts used in Aim 1 and Aim 2 will provide a high level of evidence from real-world settings. The following analytic methods to minimize bias and confounding are consistent with the PCORI Methods Report ("Standards for Causal Inference Methods," dated 3/15/12) and the recent AHRQ publication, "Developing a Protocol for Observational Comparative Effectiveness Research, A User's Guide."

We have proposed using multivariable modified Poisson/binomial regression models and adjusted Cox Proportional Hazards models as our primary analytic approach. Although estimates from multivariate models have been shown to be similar to those using propensity score methods,[38] we plan to conduct propensity score matched and weighted sensitivity analyses for all outcomes in an attempt to address potential residual confounding. For propensity score analyses, we will use logistic regression to estimate the probability of being in each of the surveillance groups as a function of measured patient and hospital characteristics in an attempt to achieve balance on measured characteristics and to obtain unbiased estimates of the treatment effect under the assumption of no unmeasured confounding.[39,40] Variables that will be included in the propensity score are those that are predictors of the outcome of interest or confounders of the treatment (surveillance frequency) and the outcome.[41] Separate propensity score models will be used for each outcome of interest and modified based on theoretical relationships between the measured variables and the specified model outcome. We will use both propensity score weighting and matching as the application method may provide insight into treatment effect heterogeneity.[42]

1) Propensity score weights: We will examine the distribution of propensity scores across groups and trim observations from the non-overlapping ends of the distribution if necessary.[43] We will create inverse probability of treatment weights (IPTW) and stabilize these weights to reflect the sample size of each group. The IPTW will be used in Aims 1 and 2 to balance the measured characteristics between the comparison groups.

2) Propensity score matching is another approach to balance sample characteristics,[44] and will be explored as an additional sensitivity analysis [IR-5]. Matching provides an intuitive way to evaluate the extent to which the surveillance groups overlap with one another and automatically excludes individuals who do not have a similar (matching) individual available. This method provides an estimate of the effect among individuals who would have been candidates for any of the three surveillance methods. Individuals who are unmatched can then be evaluated to determine if specific patient characteristics drive surveillance frequency. After we obtain propensity scores for each individual using the methods described above, matching can be conducted using a "greedy algorithm"[44] across comparison groups to ensure good overlap among groups.

3) Assessment of sample balance: To ensure balance of sample characteristics, we will test for differences among comparison groups in the mean (for continuous variables) or distribution (for categorical variables) of each covariate using t-tests and chi-squared/Fisher's exact tests, respectively, both before and after applying the generated propensity score weights. In addition, for the propensity score matched sample, we will assess the balance of study covariates using standardized differences across comparison groups. If there is residual

imbalance after propensity score weighting or matching, we will include unbalanced variables in the outcome model to ensure proper control of these confounding factors.

4) Instrumental variables (IV): We propose this method as an exploratory analysis, [IR-5] and acknowledge potential challenges with identifying an appropriate instrument. IV methodology requires finding an appropriate "instrument" related to the exposure (surveillance frequency) but not directly related to outcomes, and meeting all underlying IV assumptions.[45] We will test two commonly-used, potential instrumental variables to evaluate the robustness of our findings generated using multivariable regression and propensity score matched and weighted analyses. These two proposed instruments are region and the provider's most commonly used surveillance regimen (commonly called "provider preference"). Region was selected as it has been shown in some of our preliminary work in SEER-Medicare data to be independently related to post-treatment surveillance behaviors. For example, patients living in the Pacific region receive a mean of 2.95 PSA tests during the first year post-treatment, while those living in New England receive a mean of 2.26. Provider preference for surveillance frequency will also be tested since this instrument has been hypothesized to be a strong predictor of physician behavior and to have no direct impact on outcomes.[46] We have experience using IV methodology in prostate cancer comparative effectiveness research using large data sources.[47]

5) Handling of missing data [IR-3][IR-5][MD-1-5]: We will follow the PCORI standards ("Minimal standard for prevention and handling of missing data," dated 3/15/12). This study has a unique and important mechanism to address missing data – using the NCDB special studies mechanism. NCDB registry data on all study cases for Aim 1 will be reviewed for missing data [MD-1], and addressed as part of the primary data collection protocol. For missing data elements which cannot be identified even through medical records abstraction in Aim 1, and for missing data in Aim 2, we will undertake 3-steps to minimize impact of missing data on the study results including: 1) examine missing data mechanism, 2) multiple imputation to fill in missing data if appropriate, and 3) sensitivity analysis to estimate the impact of the missingness. First, we will carefully examine the pattern of missing data by individual items. We will determine whether the missing data is informative, missing not at random (MNAR), missing completely at random (MCAR), or missing at random (MAR) by looking at patient characteristics. We will also compare those with missing data to patients with complete data. Next, we will impute the missing data using a widely-used multiple imputation approach.[48] Instead of filling a single value for each missing value, multiple imputation method replaces the missing data with a set of *m* possible values by creating *m* imputed datasets. Each imputed dataset will then be analyzed using analytic methods that are used for complete datasets. Finally, these results are pooled together to provide valid statistical inferences.[49] These methods will enable us to use data from the entire study sample, potentially increasing statistical power to detect differences among comparison groups. Finally, we will conduct sensitivity analysis to examine the impact of the missing data on results.[50-53] In data reporting, we will describe reasons for missing data and account for all patients.