Comparative risk-adjusted mortality outcomes following primary surgery, radiation therapy, or androgen deprivation therapy for localized prostate cancer

Matthew R. Cooperberg, MD, MPH[(1),†], **Andrew J. Vickers, PhD**[(2)], **Jeanette M. Broering, RN, MS, MPH**[(1)], **Peter R. Carroll, MD, MPH**[(1)], and **the CaPSURE Investigators**

[(1)] Department of Urology, UCSF Helen Diller Family Comprehensive Cancer Center, San Francisco, CA

[(2)] Department of Epidemiology and Biostatistics, Memorial-Sloan Kettering Cancer Center, New York, NY

Introduction

For the more than 192,000 men expected to be diagnosed with prostate cancer annually,1 decision-making with respect to type and timing of treatment is complex: prostate cancer is surpassed only by lung cancer in its mortality burden among men in the United States1 yet the natural history of the disease is frequently indolent even among those untreated,2 and all available active treatments can be associated with significant adverse effects.3 No contemporary studies randomizing patients across primary treatments have been reported. Indeed, a systematic review recently commissioned by the Agency for Healthcare Research and Quality concluded that insufficient high-quality evidence exists to support one modality over another.4

The American Urological Association's clinical practice guideline for localized prostate cancer states that alternatives offered to patients should include active surveillance, radical prostatectomy, external-beam radiation therapy, and brachytherapy, but draws no conclusions regarding the relative efficacy of these alternatives.5 Primary androgen deprivation monotherapy for localized disease is not endorsed by the guideline, given inadequate evidence regarding outcomes; nonetheless, it is commonly used in practice.6, 7

Given prostate cancer's often prolonged course even among most cases which are ultimately lethal,8 studies with short- to intermediate-term followup may report outcomes only in terms of recurrence-free survival based on prostate specific antigen (PSA)-based definitions.
Because many disparate definitions of biochemical recurrence have been proposed,9 comparing outcomes across modalities using PSA endpoints is problematic. Clinical endpoints—in particular prostate cancer-specific mortality (CSM) and all-cause mortality (ACM)—do not vary across treatments and are ultimately more relevant to patients. However, analyses at these endpoints require long-term followup.

In order to ascertain risk-adjusted comparative effectiveness of primary treatment approaches for prostate cancer, we conducted an analysis comparing CSM and ACM outcomes following prostatectomy, external-beam radiation, or primary androgen deprivation in a well-defined, multi-centre, prospective cohort of prostate cancer patients.

Comment [JEC1]: What makes this a good paper?

Comment [JEC2]: First sentence clearly describes the issue of treatment effectiveness.

Comment [JEC3]: Justifies the need for this study.

Comment [JEC4]: Describes the shortcomings of clinical outcomes and why CSM is the most solid.

Comment [JEC5]: Clearly and concisely describes the study question and method to address it.

Patients and methods

Patient cohort

Data were abstracted from the Cancer of the Prostate Strategic Urologic Research Endeavour (CaPSURE™), a national disease registry accruing men with biopsy-proven prostate adeno-carcinoma managed at one of 40 urology practices, primarily community- based, across the United States. Participating urologists recruit men consecutively at diagnosis, and report initial and followup clinical data including staging tests and treatments. Comorbidities are recorded at baseline and in followup, comorbidity scoring is based on the Charlson index.10 The registry was initiated in 1995. Between 1995 and 1998 accrual was both prospective and retrospective; since 1998 all accrual has been prospective. Patients provide written, informed consent under local and central institutional review board supervision.

Patients are treated per their clinicians' usual practices, and are followed until death or withdrawal from the study. Clinicians report mortality events, and copies of state death certificates are obtained. CSM is determined if prostate cancer is listed as a primary, secondary, or tertiary cause of death on the certificate and no other malignancy is listed as a higher order cause. Perioperative mortality and death due to complications of radiation and/or androgen deprivation counted toward all-cause but not cancer-specific mortality. If the patient has been lost to followup or the certificate is not available, the National Death Index is queried to identify date and cause of death. Previous details regarding CaPSURE's methodology have been reported previously.11, 12

As of July 2008, 13,805 men had enrolled in CaPSURE. Of these, 8982 had localized disease (clinical stage ≤T3aN0), were treated with prostatectomy, external-beam radiation, or primary androgen deprivation, and had at least six months of followup recorded. 1444 with missing data needed to calculate both risk instruments described below were excluded. Thus 7538 men comprised the analytic dataset. Years of treatment ranged from 1987 to 2007; 26% of the patients were treated before 1997, 10% before 1995, and 1% before 1991.

Statistical analysis

Demographic and clinical characteristics of patients in each treatment group were compared using analysis of variance or chi-squared, as appropriate for continuous and categorical variables. To ensure that the analysis was not dependent on a specific risk adjustment approach, prostate cancer risk was assessed using two well-validated pre-treatment instruments. The first was the original nomogram published by Kattan et al., which yields a 0 to 100 score estimating likelihood of recurrence-free survival following radical prostatectomy from the PSA, biopsy Gleason grade, and clinical T stage.13–16 For this analysis, risk was expressed as 100-Kattan score, such that higher numbers indicate greater disease risk.

The second instrument was the UCSF Cancer of the Prostate Risk Assessment (CAPRA) score, a 0 to 10 score calculated from the PSA, biopsy Gleason grade, clinical T stage, age at diagnosis, and percent of biopsy cores positive.16–19 The CAPRA score predicts pathologic stage and biochemical recurrence-free survival, with each two-point increase in score indicating roughly a doubling of recurrence risk. Most recently, the score has been also shown to predict metastasis, CSM, and ACM across multiple primary treatments.20

Kaplan-Meier time to event curves were generated21 and outcomes compared via the log-rank test.

Comment [JEC6]: Thorough description of the patient cohort, including type of data collection and informed consent.

Comment [JEC7]: Clearly states that the study is observational.

Comment [JEC8]: Thorough explanation of how the CSM outcome is defined and the data is obtained.

Comment [JEC9]: Detailed account for the analytic population, including attrition due to missing data and follow up.

Comment [JEC10]: TIP: This information most often is presented in the first paragraph of the Results section.

Comment [JEC11]: Patient and clinical descriptions should be reported first. These are the most basic parts of the analysis.

Comment [JEC12]: Detailed and cited descriptions of validated clinical risk measures

Comment [JEC13]: Next, unadjusted time-to-event is described, followed by adjusted time-to-event model.

Weibull parametric survival models were then constructed to compare outcomes, adjusting for case mix using either Kattan or CAPRA score and age. The primary endpoint was CSM; ACM was assessed as a secondary endpoint. In each case the hazard ratio (HR) with 95% confidence intervals (CI) was calculated for radiation and androgen deprivation compared to prostatectomy. The model was used to predict CSM at 10 years at various levels of risk. For the CSM analyses, patients dying of other causes were censored at the time of death. As a sensitivity analysis, the CSM analyses were also conducting using competing risks regression.22 Tests for interaction between risk and treatment were also performed.

Adjustment for neoadjuvant androgen deprivation did not alter the statistical significance of any variable in the model, and had minimal impact on the parameter estimates; therefore this variable was not included in the final model. The model was also tested excluding the 136 men who received adjuvant radiation therapy after prostatectomy, both with and without inclusion of adjuvant radiation as an additional predictor variable in the model. To limit the analysis to those receiving radiation treatment under relatively contemporary standards we performed a subset analysis limited to those treated since 1998. Finally, although the models have been shown to be accurate in predicting CSM across multiple treatments,20 it is possible that neither the Kattan nor CAPRA scores adequately reflect differences in risk across patients. Therefore, as an additional test we reassessed the model with Kattan scores for radical prostatectomy patients artificially increased, progressively by 5-point increments, to estimate the degree of unmeasured confounding beyond measured risk which would need to be assumed in order to nullify the results. All statistical tests were two-sided, and analyses were performed using Stata version 11 (Stata Corp., College Station, TX).

Results

In total, 1293 (17.2%) men died, 226 (3.0%) of prostate cancer. Sociodemographic and clinical factors among patients in each primary treatment group are summarized in table 1. All comparisons among treatment groups for clinical and sociodemographic factors were statistically significant ($p<0.001$). Prostatectomy patients were younger, more frequently Caucasian, and had less comorbidity and lower risk disease features than those in other groups. There were 3 peri-operative deaths. Overall, 49.7% of the radiation patients received neoadjuvant and/or adjuvant hormonal therapy: 33.7%, 50.6%, and 67.6%, respectively, of those with CAPRA scores 0–2, 3–5, and 6–10. 6.7% of the prostatectomy patients received neoadjuvant therapy: 4.5%, 7.7%, and 19.3% respectively of those in each CAPRA score group. Mean ± SD duration of therapy was 7.9 ± 3.1 months.

Mean ± SD and median times to death were 6.8 ± 4.0 and 6.4 years, respectively, and mean and median followup times among those surviving were 4.2 ± 3.3 and 3.9 years. Median followup times were similar across treatments: 3.9, 4.5, and 3.6 years, respectively, for prostatectomy, radiation, and androgen deprivation patients; and across risk groups: 3.6, 4.1, and 4.0 years, respectively, for CAPRA 0–2, CAPRA 3–5, and CAPRA 6–10 patients.
Unadjusted time-to-event curves for CSM are presented in figure 1. The differences in outcomes across treatments were statistically significant by log-rank test ($p<0.001$). Relative to prostatectomy, the unadjusted HRs for CSM were 2.46 (1.8–3.43) for radiation and 4.36 (3.21–5.93) for androgen deprivation.

The results of the primary risk-adjusted analysis are presented in table 2a. Adjusting for age and case mix using the Kattan score, the HRs for CSM relative to prostatectomy for radiation and androgen deprivation were 2.21 (1.50–3.24) and 3.22 (2.16–4.81), respectively. The HR for CSM for androgen deprivation relative to radiation was 1.45 (1.02–2.07). Adjusting for CAPRA rather than Kattan score

Comment [JEC14]: Multivariate regression models are described, including independent covariates. Risk of mortality events over time analyzed with time-to-event models, such as Weibull parametric and Cox proportional hazards (semi-parametric) regression. Events/endpoints must be clearly defined.

Comment [JEC15]: Sub/sensitivity analyses were performed because clinical risk and treatment are interrelated exposures and may have an interaction effect, in addition to the separate effect of each individual variable.

Comment [JEC16]: (I ASSUME) this issue is addressed here because such treatment would complicate the treatment groups. If preliminary analysis showed that neoadj treatment was not statistically significant, it can be eliminated from multivariate models, making them easier to interpret and less prone to collinearity.

Comment [JEC17]: This subset analysis is valuable because change in practice patterns over time is always of concern when analyzing diagnosis and treatment trends. This helps make the data less prone to historical bias.

Comment [JEC18]: Additional subanalysis to address unmeasured confounding is not very common in a paper like this and strengthens overall findings.

Comment [JEC19]: It is good/common practice to specify that two-sided tests were used, as well as significance p-value and type of statistical software.

Comment [JEC20]: In the Results section, present each finding in the same order as in the Methods section. Patient description is first.

Comment [JEC21]: AREN'T SUCH DEATHS RARE? SEEMS ALARMING BUT I DON'T THINK THIS IS ADDRESSED IN THE DISCUSSION.

Comment [JEC22]: TIP: Avoid using "respectively" in this way because it makes the reader work harder to understand the findings.

Also, decimal places are not necessary here.

Comment [JEC23]: TIP: Present the median and interquartile range, instead of reporting both the Mean(SD) and median. Two measures of central tendency are not needed and median is better than mean for this variable because the values are skewed.

Comment [JEC24]: TIP: If follow up times do not differ significantly be group, then just report the overall time once and note that was similar across treatment groups and across risk groups.

Comment [JEC25]: Confidence intervals should always be reported for hazard ratios and odds ratios.

Comment [JEC26]: It is informative to add this additional pairwise comparison to demonstrate that all the groups are different from each other, not just relative to RP.

yielded somewhat lower but similar HRs relative to prostatectomy: 1.63 (1.09–2.45) for radiation and 2.65 (1.75–4.01) for androgen deprivation, and 1.62 (1.11–2.36) for androgen deprivation relative to radiation. Use of competing risks regression likewise yielded similar results: relative to prostatectomy, the HRs were 2.00 (1.33–3.01) and 2.56 (1.62–4.03) for radiation and androgen deprivation, respectively; relative to radiation, the HR was 1.27 (0.88–1.84) for androgen deprivation.

Excluding 136 men receiving adjuvant radiation therapy after prostatectomy had no effect on the results of the model, whether or not radiation was included as a predictor in the model. In interaction analyses, there was no evidence that the difference between prostatectomy and radiation depended on baseline risk (p=0.20). There was suggestion that improved outcome with radiation compared to androgen deprivation increased for patients with higher risk disease (p=0.07), but as this did not meet statistical significance, survival differences were modeled assuming constant relative risk among treatments across different levels of risk.

Table 2b presents the results for ACM: adjusting for age, Kattan score, and comorbidity, the HR relative to prostatectomy for radiation was 1.58 (1.32–1.89) and for androgen deprivation was 2.25 (1.86–2.72). Relative to radiation, the HR for ACM for androgen deprivation was 1.43 (1.21–1.69). Virtually identical results were produced with adjustment for CAPRA rather than Kattan score. Figure 2 and table 3 present predicted 10-year CSM by 100-Kattan and CAPRA score, respectively, for each treatment. Predicted CSM increases consistently with rising CAPRA score, from 1.5% to 32.8% for prostatectomy, 2.5 to 48.7% for radiation, and 4.0 to 66.3% for androgen deprivation.

In restricting the analysis to those treated since 1998, the number of CSM events fell to 67 among 5143 at risk. The HRs for CSM relative to prostatectomy in this subset were 2.7 (1.2–6.2) for radiation therapy and 6.5 (3.1–13.5) for androgen deprivation. In our sensitivity analysis for unmeasured confounding, in calculating the model with Kattan scores artificially increased for prostatectomy patients, the difference between prostatectomy and radiation patients remained statistically significant until the Kattan scores were increased by 20 points for prostatectomy patients, and did not change direction until the scores were increased by over 30 points (table 4).

Discussion

Uncertainty regarding optimal management of localized prostate cancer has produced wide and excessive local and regional variation in the utilization of various interventions.23–25 In general, with increasing risk men are less likely to receive prostatectomy, more likely to receive radiation, and much more likely to receive androgen deprivation monotherapy.26 Over time, utilization of androgen deprivation in particular has increased for high-risk men. 6, 26 Although several large centers have recently reported outcomes of prostatectomy in high risk patients which compare favorably to those from older series,27 there are no indications that these findings have yet impacted community practice.

These trends have not been evidence-driven; indeed, given the existing dearth of high- quality comparative data, the Institute of Medicine recently included treatment for localized prostate cancer among the 25 most important topics for comparative effectiveness research. 28 Only three randomized trials have been published comparing major primary management approaches. Bill-Axelson et al found a survival benefit for prostatectomy over watchful waiting, with a 35% relative reduction in risk of CSM at 10 years.29 An older, smaller randomized study likewise reported longer overall survival for prostatectomy patients compared to watchful waiting patients.30 Another recent trial randomized patients with cT3N0M0 disease to flutamide with or without radiation therapy. The study found a strong

Comment [JEC27]: Although findings were similar for both risk measures, it is informative to report them both because the HRs were higher for Kattan/lower for CAPRA.

Comment [JEC28]: Same here – findings were significant for all 3 models but the HRs varied. If pressed for space, though, it is ok to report simply that the findings were similar for the competing risk model, since it was not the primary model.

Comment [JEC29]: Good to report this specifically because use/non-use of adjuvant treatment could create variations within the primary treatment groups and affect duration of follow up time.

Comment [JEC30]: Well reported. While the p-value was close to <0.05, it was NOT significant and should not be treated as if it were.

Comment [JEC31]: Again, each finding in this section should follow the order laid out in the Methods.

Comment [JEC32]: Both the subanalysis and sensitivity results reinforce the findings of the main analysis.

Comment [JEC33]: The first paragraph of the Discussion elaborates on paragraph 2 of the Introduction and cites additional supporting studies.

benefit for the combination treatment arm,31 though flutamide monotherapy would generally be considered inadequate therapy by contemporary standards, particularly for locally advanced disease.

Randomized trials in localized prostate cancer face challenges related to high costs associated with long followup and patient and/or clinician biases a priori in favour of one approach or another. The Surgical Prostatectomy Versus Interstitial Radiation Intervention Trial (SPIRIT) trial intended to randomize men to radical prostatectomy vs brachytherapy. Despite a 90 minute patient educational session intended to facilitate accrual, only 56 patients accrued at 31 centers over two years, and the study was closed early.32 The Prostate cancer Intervention Versus Observation Trial (PIVOT) screened 13,022 men at 52 sites over 7 years to identify 5023 eligible men, of whom 731 (14.5%) agreed to be randomized between surgery and observation. Initial results are expected later this year.33 The Prostate testing for cancer and Treatment (ProtecT) study is the only ongoing randomized trial including more than one active treatment arm—prostatectomy, external-beam radiation, and watchful waiting. It has had greater success attributed to a complex intervention aimed to increase acceptance of randomization.34 Results will require years, however, to reach maturity.

Meanwhile, important insights into outcomes have been gained from research based on large data sources such as Surveillance, Epidemiology and End Results (SEER) and Medicare.35,36 However, these analyses are limited by relatively scant clinical information in the datasets—for example, absent PSA, Gleason, and treatment details. Therefore, prospective disease registries provide an important source of evidence for comparative effectiveness research analyses.37 We performed such an analysis in CaPSURE, a large, national, community-based registry of men followed prospectively and uniformly from diagnosis regardless of treatment selection.

The present analysis finds evidence for significant CSM and ACM differences across primary treatments, controlling for age, disease risk, and comorbidity. Especially striking is the progressive increase in differences across treatments with increasing risk (figure 2 and table 3). Mortality at 10 years is uncommon among men with low-risk disease regardless of treatment, whereas among those with higher risk disease—in contrast to observed treatment trends26—men receiving prostatectomy are much less likely to die than those receiving external-beam radiation, and men in both local therapy groups have better survival than those receiving androgen deprivation alone.

Several caveats should be considered. CaPSURE practice sites are not a random sample of the U.S. population. However, they represent a range of practice locations, sizes, and treatment patterns, and do approximate the community prostate cancer patient experience in the U.S..12 CaPSURE patients reaching mortality endpoints are more likely to have been diagnosed earlier, with a sextant biopsy; their likelihood of clinical understaging is thus greater than would be expected for contemporary patients undergoing extended-template biopsy. Therefore the mortality predictions from this analysis may be higher than might be expected in contemporary practice. It is possible that improvements in technique and outcomes among radiation patients have been more pronounced over the past decade than those for surgery patients; however, we found that the survival differences were if anything greater when restricting the analysis to a more contemporary cohort.

CaPSURE does not include consistent data on radiation dose and technique, nor on tertiary Gleason scores. There were insufficient events to control adequately for type and timing of salvage therapies, which are quite variable—reflecting inconsistent community practices in the face of little evidence-based guidance—and have been discussed in detail previously.38 In a recent report from a large academic cohort comparing prostatectomy with radiation under relatively uniform protocols,

Comment [JEC34]: The second paragraph of the Discussion reiterates paragraph 1 of the introduction and cites past studies within the context of the CSM outcome.

Comment [JEC35]: This paragraph provides an overview of current comparative effectiveness trials (as opposed to Capsure which is a prospective, observational registry), describing findings and shortcomings that warrant an additional study with Capsure data.

Comment [JEC36]: Restate and interpret this study's findings.

Comment [JEC37]: Study limitations must be specific and address considerations about the study design of a registry and historical bias due to changing practice patterns.

adjustment for salvage therapy had no impact on the outcomes of the analysis.39

Higher doses of radiation have been associated with a 12% improvement in recurrence-free survival,40 but have not been demonstrated to improve likelihood of CSM or ACM,4 nor have variations in technique such as intensity-modulation. Variation in radiation practice seems unlikely to explain more than a fraction of the results of this analysis. Indeed, the academic series noted above included only radiation patients receiving 81Gy or more. The results were concordant with the present study, with approximately 3-fold reduction in case mix-adjusted rates of metastasis and prostate cancer-specific mortality in the surgery group. 39 CaPSURE does include a large cohort of patients treated with brachytherapy and active surveillance/watchful waiting. However, they generally were diagnosed in the more recent years of the registry, and their followup is not yet sufficiently mature to assess mortality.

Overall, 51% of the external-beam radiation patients in this analysis received neoadjuvant and/or adjuvant androgen deprivation therapy, a proportion similar to the 56% reported in the recent academic series noted above.39 In CaPSURE, likelihood of receiving neoadjuvant therapy together with external-beam radiation for high-risk disease has increased steadily over time.6 Adjustment for neoadjuvant androgen deprivation with radiation therapy did not modify the results, likely because use of neoadjuvant therapy in CaPSURE associates closely with disease risk: higher risk patients are much more likely to receive neoadjuvant therapy,6 so the impact of neoadjuvant therapy is reflected in the risk-adjustment, and in a model adjusting for risk, likelihood of neoadjuvant therapy is not an independent predictor. The mean duration of therapy was longer than in the academic series (7.9 months in the present cohort vs. 3 to 6 months in the academic cohort).39 A recent analysis of duration of neoadjuvant therapy found a relatively small (<5%) difference in cancer-specific survival for those receiving longer-term therapy, and an overall survival difference only among those with high-grade disease.41 Longer duration of therapy among higher-risk radiation patients in CaPSURE might therefore be expected to improve outcomes.

To address the possibility that our results were affected by differences in death rates from causes other than prostate cancer, we used competing risk regression, with minimal changes in findings. The attribution of CSM may not be accurate in all cases, particularly those ascertained from the National Death Index; however, the findings were seen for both CSM and ACM and were robust to different considerations of risk as well as several other sensitivity analyses. Finally, it is possible that other unmeasured confounding might explain some part of results. The Charlson score, for example, may not adequately reflect differences in comorbidity which could drive treatment decision-making. (A subset of CaPSURE patients have completed a more comprehensive comorbidity evaluation,42 but this group was too small for the present analysis.)

To evaluate the possibility that the limitations discussed—or other sources of unmeasured confounding—may explain the results, we artificially raised the Kattan scores for radical prostatectomy patients, finding that that in the risk-adjusted model the benefit for surgery over radiation persisted until the scores for prostatectomy patients were increased by at least 20 points. In other words, the nomogram would need to systematically underestimate radiation patients' risk of progression relative to surgical patients' by 20 absolute percentage points; thus, a patient undergoing radiation, for example, with a Gleason 3+3, PSA 4.0 ng/ ml, stage T1c tumor would have to have the same true risk as a surgical patient with Gleason 3+4, PSA 9.0 ng/ml, stage T2a tumor. Prediction model accuracy in predicting CSM is 80% across multiple treatments,20 and we cannot identify unmeasured confounders which would be expected to have such a large impact on risk-adjusted outcomes. The magnitude of the differences between treatments might be expected to vary with additional adjustment, but a qualitative change in

Comment [JEC38]: Data limitations are describe separately in following paragraph

Comment [JEC39]: THERE IS A LOT OF MENTION OF ACADEMIC VS COMM SITES IN THE DISCUSSION BUT IT WASN'T REPORTED IN THE RESULTS.

Comment [JEC40]: Discussion of variations within treatment groups.

Comment [JEC41]: Clearly interpret/address neoadjuvant treatment issues. NOT SURE WHY SO MUCH IS SAID ABOUT THIS BECAUSE IT ALREADY WAS DISCUSSED IN THE METHODS AND BARELY MENTIONED IN THE RESULTS.

the findings seems very unlikely. An additional strength of this analysis is that the Kattan and CAPRA scoring systems assign different relative weights to the various prognostic factors included, reducing the likelihood that the outcome of the model is dependent on the specific risk stratification system. In the Zelefsky et al study, likewise, different considerations of risk did not substantially modify the outcomes.39

Comment [JEC42]: Citation nicely corroborates the point.

Conclusion

In a multi-institutional, prospective cohort of prostate cancer patients, we found a low overall risk of cancer-specific mortality. After rigorous case-mix adjustment and multiple sensitivity analyses, however, we identified roughly two- and three-fold increases in risk of cancer mortality among those undergoing external-beam radiation or primary androgen deprivation, respectively, compared to those undergoing radical prostatectomy, with the greatest differences seen for higher-risk patients. These findings should be verified with randomized trial data when available, and with longer followup in CaPSURE and other large registries as more men ultimately reach mortality endpoints.

Comment [JEC43]: The Conclusion states specific findings (ie. "we found ..") without broad, sweeping conclusions (ie. "there is a low risk ...") and specifies next steps.