Comparing Methods of Applying the Propensity Score to Control Selection Bias in Studies of Prostate Cancer

Xiao Fang¹, Vahakn Bedig Shahinian², Kristin Marie Sheffield³, Alai Tan⁴, Daniel Chaim Jupiter¹ and Yong-Fang Kuo¹

1. Department of Preventive Medicine and Community Health, the University of Texas Medical Branch, Galveston 77555, USA

2. University of Michigan, Internal Medicine-Nephrology, Ann Arbor, USA

3. Global Patient Outcomes and Real World Evidence, Eli Lilly, Indianapolis 46285, USA

4. Sealy Center on Aging, the Ohio State University, Columbus 43210, USA

Abstract: To compare the performance of four common methods of applying propensity scores: covariate adjustment, stratification, matching, and inverse probability of treatment weighting, in addressing issues of selection bias. The linked Surveillance, Epidemiology, and End Results Medicare database was used in this study. We compared mortality of patients with localized prostate cancer, in two different scenarios. In Scenario 1, treatment effect on non-prostate cancer mortality were compared between patients with localized prostate cancer receiving active treatment (radical prostatectomy or radiation) and those only being observed. In Scenario 2, prostate cancer specific mortality was compared between localized prostate cancer patients with and without primary androgen deprivation therapy (ADT). The known confounding factors of comorbidity (Scenario 1) or tumor grade (Scenario 2) were removed to evaluate the ability of each propensity score method to control for the selection bias caused by the presence of the factors. Matching yields slightly better control than the three other methods. None of the propensity score methods can eliminate bias caused by removing a known confounder.

Key words: Selection bias, propensity score, prostate cancer.

1. Introduction

In the United States, prostate cancer is the second most common cancer among men [1]. How to select the best treatment for prostate cancer is one of the most controversial areas for consideration by both physicians and patients [2]. Because of the limitations of randomized controlled trials, such as lack of external validity [3, 4], high cost [5] and the fact that they are time consuming [3, 5], interest is growing for investigators to use observational data to compare the outcomes of different prostate cancer treatments [6-8]. However, a serious concern with observational studies is the existence of selection bias [9]. When different therapies are compared, patients with poorer cancer prognoses may be expected to receive more aggressive

therapy, making therapy appear to be associated negative outcomes. Alternately, patients with better underlying health may be selected for more aggressive therapy, making therapy appear to be associated with better outcomes. Mitigating treatment selection bias is therefore the major challenge for observational studies of treatment effectiveness.

Propensity score methods have increasingly been used to control for the influence of effects of confounding, including selection bias, in observational studies [10-12]. Four of the most common methods of applying the propensity score are covariate adjustment, stratification, matching, and inverse probability of treatment weighting (IPTW) [13, 14]. In this paper, we assessed the performance of these methods in controlling for selection bias in the analyses of treatment effects on outcomes among prostate cancer patients. The novelty of this study was that

Corresponding author: Xiao Fang, M.S., research fields: survival analysis, longitudinal analysis and propensity score.

the performance of each method was evaluated based on its ability to account for a selection bias that was induced by the removal of known confounders from a full multivariate model. We hypothesized that propensity matching would provide results with less bias than those using the other propensity methods.

2. Methods

2.1 Data Sources

The linked Surveillance, Epidemiology and End Results (SEER)-Medicare database 1992-2008 was used. The SEER database includes information on patient and tumor characteristics, treatments, and survival. Medicare is a primary health insurer that covers approximately 97% of US individuals aged 65 years and older. Medicare Part A covers inpatient claims and part B covers outpatient and physician claims. SEER patients who were older than 65 years have been linked to Medicare records by the collaborative effort of the National Cancer Institute and Medicaid Services [15], and this Medicare data is provided within the SEER data set. United States 2000 Census data regarding income and education at the zip code level was used as a surrogate measure of patients' socio-economic status.

2.2 Study Design

We generated two scenarios in which there are unmeasured confounding effects, as well as selection bias mediated by two confounders: general health or comorbidities and cancer severity. In other words, the presence of these confounders in the full multivariate model removes the effect of selection bias from these two confounders, while removing them from the models creates a selection bias. Unmeasured confounding may still remain.

In Scenario 1, patients with localized prostate cancer receiving active treatment (radical prostatectomy or radiation) were compared to those only being observed. The average age at the time of diagnosis of prostate

cancer is about 66 [1], and comorbidity has a large impact on the non-prostate cancer mortality. More importantly to us, comorbidity has an impact on selection of treatment: those with greater burden of comorbidity are more likely to be in the watchful waiting group rather than treated by active treatment.

In this scenario, the known confounding factor—comorbidity as assessed using Charlson comorbidity index—was removed from the model, and estimates of treatment effect on non-prostate cancer mortality were compared between the full and reduced models in which the propensity score methods were applied.

In Scenario 2, localized prostate cancer patients with and without primary androgen deprivation therapy (ADT) were compared. Tumor grade has an impact on prostate cancer [16, 17] specific mortality, and also on the selection of treatment: those with higher tumor grade are more likely to receive ADT. This confounding factor was removed from the model. Estimates of treatment effect on prostate cancer mortality were compared between the full and reduced models in which the propensity score methods were applied.

2.3 Study Cohort

The study population included patients aged 66 years or old, diagnosed with prostate cancer from 1992 to 2007. Patients were followed until death or the end of study (December 31, 2008).

In order to better compare with previous studies [18], additional criteria were applied. In Scenario 1, men who (1) were older than 80 years, (2) had poorly differentiated or unknown tumor grade or (3) had unknown income or education information were excluded. In Scenario 2, men treated actively with radical prostatectomy or radiation therapy were excluded.

2.4 Variables

Patient demographic and tumor characteristics were

derived from SEER records and used to categorize patients by age (categorical: Scenario 1 66-67, 68-70, 71-73, 74-77, 78-80; Scenario 2 66-69, 70-74, 75-80, 80+), ethnicity (white, black, Hispania, other), SEER region of residence, year of diagnosis, marital status, clinical stage (T1 and T2), and grade (low: Gleason 2-4; moderate: Gleason 5-7; high: Gleason 8-10; unknowns), education, income, and comorbidity index (0, 1, 2, ≥ 3) [9]. The socio-economic characteristics of each patient were based on the percentage of adults with less than 12 years of education and median household income within the zip code of residence, as derived from U.S. Census 2000 data. Comorbidity was assessed using an adaptation of the Charlson Comorbidity Index for use with Medicare inpatient, outpatient and Part B claims during the 12 months before cancer diagnosis [15]. Use of ADT was defined as the receipt of at least one dose of a gonadotropin-releasing hormone (GnRH) agonist or orchiectomy occurring within 6 months after prostate cancer diagnosis, and was assessed using previously published and validated methods [16]. These data are based on information from either SEER or Medicare claims, using previous published methods [17]. CPT (Current Procedural Terminology), ICD-9-CM (International Classification of Diseases, Ninth Revision, clinical modification), and HCPCS (Healthcare Common Procedure Coding System) used to identify treatment variables are listed in Appendix 1.

The outcome variable was mortality (non-prostate cancer mortality in Scenario 1 and prostate cancer-specific mortality in Scenario 2), obtained from

SEER records.

2.5 Statistical Analyses

Demographic and tumor characteristics were compared between patients receiving and not receiving active treatment for Scenario 1, and between patients receiving and not receiving primary ADT for Scenario 2 (Table 1). Baseline characteristics were also compared between the different levels of comorbidity in Scenario 1, and between tumor grades in Scenario 2, to assess whether these two factors impacted the distribution of other covariates (Table 2). Since all variables were treated as categorical variables, the chi-squared test was used for these tests.

In model 1, Cox proportional hazards models were built to evaluate the non-prostate cancer mortality in patients who received active treatment compared to those managed with observation (Scenario 1), and to evaluate the prostate cancer-specific mortality in patients who received primary ADT therapy compared to those who did not (Scenario 2). All variables between comparison treatment groups, as assessed on the first bivariate test above and listed in Table 2, were adjusted for in this model. The results from model 1 (full models) were used as reference. We built a second model where non-prostate specific and prostate specific mortality were regressed on all variables, excluding comorbidity in Scenario 1 and tumor grade in Scenario 2. Finally, four propensity score adjustment methods were then applied in models where the known confounder (comorbidity in Scenario 1 and tumor grade in Scenario 2) was removed.

Table 1 Treatment and associated CPT, ICD-9, and HCPCS codes.

Radical Prostatectomy	CPT codes 55810, 55812, 55815, 55801, 55821, 55831, 55842, 55845; ICD-9 procedure code 60.5.
Radiation Therapy	CPT codes 77401-77499 and 77750-77799; ICD-9 procedure codes 92.21-92.29, and ICD-9 diagnosis codes V58.0, V66.1 and V67.1.
ADT	Orchiectomy: Any claim with CPT codes 54520, 54521, 54522, 54530, or 54535; or ICD-9 procedure code 62.4. GnRH agonist: Any outpatient or carrier claim with the HCPCS codes J9202 (goserelin), J1950 (leuprolide), J9217 (leuprolide), J9218 (leuprolide), J9219 (leuprolide implant) or J3315 (triptorelin).

Table 2 Patient and tumor characteristics stratified by treatments.

Characteristics	Scenario 1 Overall cohort					Scenario 2 Overall cohort				
	Observation	Active Treatment	<i>P</i> -value	Adjusted Propensity Score	Applied IPTW	ADT	No ADT	<i>P</i> -value	Adjusted Propensity Score	Applied IPTW
	N = 24,305	N = 86,516		<i>P</i> -value	<i>P</i> -value	N = 27,344	N = 40,413		<i>P</i> -value	<i>P</i> -value
Age(years)			< 0.0001	< 0.0001	0.7566			< 0.0001	0.0027	0.9984
Mean ± SD	73.6 ± 4.1	71.6 ± 3.8				78.7 ± 6.2	76.6 ± 6.2			
Race			< 0.0001	0.2874	0.0133			< 0.0001	0.9900	0.9242
White	76.42%	81.68%				75.87%	76.47%			
Black	11.11%	8.74%				10.17%	11.06%			
Hispanic	5.01%	4.69%				5.20%	4.81%			
Others	7.46%	4.89%				8.76%	7.66%			
Year of diagnosis			< 0.0001	0.0016	0.4964			< 0.0001	0.7504	1.0000
1992-1995	22.55%	15.71%				13.86%	22.66%			
1996-1999	17.02%	15.20%				15.22%	17.46%			
2000-2003	27.93%	32.53%				38.89%	29.44%			
2004-2007	32.50%	36.56%				32.03%	30.44%			
SEER Regions			< 0.0001	0.4349	0.0015			< 0.0001	0.1097	0.9914
San Francisco	5.67%	4.26%				3.46%	5.81%			
Connecticut	8.34%	8.32%				8.85%	8.66%			
Detroit	12.35%	12.47%				10.48%	12.88%			
Hawaii	1.27%	1.64%				1.90%	1.44%			
Iowa	7.72%	7.83%				10.08%	8.11%			
New Mexico	5.00%	3.31%				2.74%	4.94%			
Seattle	8.55%	7.83%				5.36%	8.37%			
Utah	5.76%	4.22%				4.00%	5.37%			
Atlanta/Rural Georgia	4.16%	4.48%				2.86%	3.79%			
San Jose	2.98%	2.56%				3.57%	2.98%			
Los Angeles	10.16%	8.65%				9.89%	9.71%			
Greater California	11.48%	12.44%				11.58%	11.40%			
Kentucky	4.42%	5.22%				5.65%	4.13%			
Louisiana	4.26%	4.68%				8.52%	4.38%			
New Jersey	7.86%	12.11%				11.06%	8.04%			

Table 2 continued

Characteristics	Scenario 1					Scenario 2				
	Overall cohort		Adjusted Propensity Score	Applied IPTW		Overall cohort		Adjusted Propensity Score	Applied IPTW	
	Observation	Active Treatment				ADT	No ADT			
	N = 24,305	N = 86,516	P-value	P-value		N = 27,344	N = 40,413	P-value	P-value	P-value
Marital Status			< 0.0001	0.0004	0.1150			< 0.0001	0.0824	0.2542
Not Married	37.47%	23.98%				43.78%	40.29%			
Married	62.53%	76.02%				56.22%	59.71%			
Education, % of adults with < 12 years of education			< 0.0001	0.1627	0.3877			< 0.0001	0.8243	0.9946
	22.25%	25.75%				22.25%	25.75%			
	23.22%	25.47%				23.22%	25.47%			
	25.36%	24.92%				25.36%	24.92%			
	29.17%	23.86%				29.17%	23.86%			
Unknown						1.80%	1.59%			
Income, % of adults living below poverty line			< 0.0001	0.1222	0.4644			< 0.0001	0.2435	0.0833
	21.35%	25.89%				23.77%	25.11%			
	23.43%	25.50%				24.03%	24.93%			
	25.54%	24.91%				24.98%	24.35%			
	29.68%	23.70%				25.42%	24.02%			
Unknown						1.80%	1.59%			
Charlson comorbidity index			< 0.0001					< 0.0001	0.7301	0.9963
0	67.53%	74.29%				61.03%	66.31%			
1	20.33%	18.46%				23.21%	20.47%			
2	6.99%	4.90%				9.04%	7.62%			
≥ 3	5.15%	2.34%				6.72%	5.60%			
Tumor stage			< 0.0001	0.0190	0.3694			< 0.0001	0.0037	0.3210
T1	53.78%	44.44%				33.87%	52.77%			
T2	46.22%	55.26%				66.13%	47.23%			
Tumor Grade			< 0.0001	< 0.0001	0.9503					
Well differentiated	18.28%	5.49%				3.76%	14.50%	< 0.0001		
Moderately differentiated	81.72%	94.51%				57.46%	66.59%			
Poorly differentiated						34.61%	11.58%			
Unknown						4.17%	7.33%			

Propensity scores were generated using logistic regression. In Scenario 1, having active treatment or being in the observation group was the dependent variable, and age, race, year of diagnosis, SEER region, marital status, education, income, tumor stage and tumor grade were independent variables. In Scenario 2, having ADT or being in the no ADT group was the dependent variable, and age, race, year of diagnosis, SEER region, marital status, comorbidity, education, income, and tumor stage were independent variables. Since all covariates were categorical variables, the chi-square test was used to determine whether the covariates were balanced before applying propensity score.

2.5.1 Propensity Score-Covariate Adjustment

Survival outcome was regressed on the estimated propensity score as a continuous variable, and on those covariates that were not balanced, as assessed by Cochran Mantel Haenszel Chi-square test after controlling for the quintile of propensity score. Age, year of diagnosis, marital status and tumor grade were not balanced in Scenario 1 and age and tumor stage were not balanced in Scenario 2.

2.5.2 Propensity Score-Stratification

Propensity scores were used to group patients into five equal strata. A stratified Cox proportional hazard model was built with adjustment for unbalanced covariates described earlier.

2.5.3 Propensity Score-Matching

Both greedy and caliper one-to-one matching were applied. In greedy matching, 8→1, 5→1 and 3→1 digit matching were all applied to provide more options for comparison. The criteria in caliper matching for selecting subjects with a similar propensity score was a caliper of width equal to 0.2 of the standard deviation of the logit of the propensity score [10]. The test for matching using chi-square test was repeated, and all variables were found to be balanced after matching. The Cox proportional hazard model was built using treatment status only, in all four matching methods.

2.5.4 Propensity Score-Inverse Probability of Treatment Weighting (IPTW)

The inverse of the propensity score was used to weight each subject in the treated group, and the inverse of one minus the propensity score was used for the controls. The mean weight was calculated and then used to normalize the IPTW weights to adjust for the extreme values [19]. Balance of covariates was assessed by chi-square test weighted by the created weights. The weights were then used in a Cox proportional hazard model with unbalanced covariates. SEER regions and race were not balanced in Scenario 1 and all variables were balanced in Scenario 2.

The results from the four methods were compared subjectively to the reference model to analyze which method produces an estimator of treatment effect similar to the reference model.

3. Results

Table 1 presents the patient and tumor characteristics stratified by the relevant treatment group for each scenario. Patients who received active treatment were younger, White, married, and had less comorbidity, more advanced stage and higher grade prostate tumors. Patients who received primary ADT were older, and had more advanced stage and higher grade prostate tumors.

Table 2 displays the comparison between the different levels of comorbidity in Scenario 1, and between tumor grades in Scenario 2, to assess whether these two factors impacted the distribution of other covariates. In Scenario 1, patients with higher Charlson comorbidity index tended to be older, minorities, diagnosed later, not married, educated, have higher income, in T1 tumor stage, and well differentiated tumor grade. In Scenario 2, patients with poorly differentiated tumor grade were older, minorities, diagnosed later, not married, and with T2 tumor stage. Differences are more pronounced in Scenario 1 than in Scenario 2.

The overlapping of propensity scores was checked in both scenarios and is described in Figure 1. In Scenario

Comparing Methods of Applying the Propensity Score to Control Selection
Bias in Studies of Prostate Cancer

1, 36 patients in the non-overlap region were excluded. For patients who received observation or active treatment, the ranges of the propensity scores were 0.102 to 0.962 and 0.010 to 0.962, respectively. In

Scenario 2, 95 patients in the non-overlap region were excluded. For patients treated with and without primary ADT, the ranges of the propensity scores were 0.092 to 0.790 and 0.090 to 0.790, respectively.

Fig. 1 Distribution of propensity score stratified by active treatment (a) and primary androgen deprivation therapy (b).

Figure 2 (a) shows the hazard ratio (HR) of non-prostate cancer mortality estimated from the models with (model 1) and without comorbidity (model 2), compared to those obtained by applying the four different propensity score methods. In the full model men who received active treatment had a lower risk of death from other causes (HR, 0.70; 95% Confidence Interval CI, 0.68-0.72); the hazard ratio without comorbidity was 0.67 (95% CI, 0.65-0.69). After applying propensity score 3 to 1 digit matching, the HR was 0.69 (95% CI, 0.66-0.71). Figure 2 (b) shows the hazard ratio of prostate cancer mortality estimated

from the model with (model 1) and without (model 2) tumor grade compared to those obtained from the four propensity score methods. Patients who underwent ADT had significantly higher prostate cancer-specific mortality (adjusted HR, 1.65; 95% CI, 1.56-1.74) compared with patients not treated with ADT. After tumor grade was removed from analysis, the hazard ratio was 2.25 (95% CI, 2.13-2.37). In the model with propensity score caliper matching, the HR for prostate cancer-specific mortality was 2.13 (95% CI, 2.00-2.26). For both scenarios, matching yielded a slight less biased result than covariate adjustment, stratification, or IPTW.

Fig. 2 Kaplan-Meier analyses for patients with active treatment compared to observation (a) and treated with and without ADT (b).

Unadjusted HRs for non-prostate cancer mortality (Scenario 1) and prostate cancer-specific mortality (Scenario 2) were calculated in Table 3. Tumor grade has a very large effect on prostate cancer mortality (HR: 8.687, 95% CI, 7.728-9.765) for patients with poorly differentiated tumors in comparison to patients with well differentiated tumors. While comorbidity impacts less (HR: 4.612, 95% CI, 4.375-4.864).

The proportional hazards assumption was violated in scenario 1, as seen by checking the interaction between treatment groups and logarithm of follow-up time, and by visual inspection of the graph of hazard function, comparing treatment groups. We repeated our analyses using Cox models with time-dependent effect of treatment (up to 6 years and after 6 years). Results of comparisons across different methods were similar to those obtained without this correction (results not shown).

4. Discussion

In this study, we evaluated the ability of four common methods of applying propensity scores to control for selection bias induced by the removal of known confounders from a full multivariate model. Matching was shown to be the method that eliminates the greater proportion of bias, as compared to other methods of applying propensity scores. Our results are consistent with those of several prior studies [10, 20, 21].

Propensity score matching produced an estimate closer to that of the full model in Scenario 1 than in Scenario 2. The impact on mortality of tumor grade is greater than that of comorbidity (Table 3), which suggests that tumor grade is a very important confounder to assess prostate cancer specific mortality when comparing ADT patients and no ADT patients. Given the importance of grade, and its large impact, none of the propensity scoring methods could account for its influence. Further, it may be confounded by

unmeasured confounders, for which we cannot control.

In Scenario 2, IPTW estimated a treatment effect even higher than that calculated from the conventional Cox proportional hazard model not including tumor grade. This is likely due to the fact that, given that extreme weights may not be eliminated in the entire study population even after normalization by mean weight, a high hazard ratio may occur [22].

In our study, none of the four propensity score methods could perfectly remove the bias due to selection bias induced by removing a variable. This conclusion is similar to those of previous reports [9, 23]. The inability to balance unmeasured confounders is a major limitation of observational studies. In our empirical study, we can only try to balance the measured confounders and assume that this balance reduces the overall bias [24].

Instrumental variable analysis has been proposed to reduce the variance and bias caused by unmeasured confounding [25, 26]. It requires identifying variables strongly related to the exposure but unrelated to the outcome. Several studies have used instrumental variable analysis to control selection bias [27, 28]. Correct estimation by using this method relies largely on the thoughtful consideration and selection of an appropriate instrumental variable(s).

One important assumption of applying propensity score methods is that there are no hidden confounders that will affect treatment assignment [14]. The selection bias caused by unmeasured confounders exists in the dataset and is not able to be evaluated, as we saw reflected in scenario 2. The main advice for selecting the best propensity score method is to select the one that yields the most balanced study sample. In our study of prostate cancer patients, propensity score matching outperformed covariates adjustment, stratification, and IPTW. However, none of the propensity methods produced estimates identical to the estimates when the removed confounders, comorbidity or cancer grade, were included.

Table 3 Patient and tumor characteristics stratified by comorbidity (Scenario 1) and tumor grade (Scenario 2).

Characteristics	Comorbidity (Scenario 1)				<i>P</i> -value	Tumor Grade (Scenario 2)				<i>P</i> -value
	0	1	2	≥ 3		Well differentiated	Moderately differentiated	Poorly differentiated	Unknown	
	N = 80,686	N = 20,916	N = 5,942	N = 3,277		N = 6,889	N = 42,624	N = 14,139	N = 4,105	
Age(years)					< 0.0001					< 0.0001
Mean ± SD	71.9 ± 3.9	72.4 ± 4.0	72.8 ± 4.0	72.8 ± 4.1		76.2 ± 6.1	77.0 ± 6.1	79.1 ± 6.5	77.7 ± 7.0	
Race					< 0.0001					< 0.0001
White	82.34%	77.27%	74.40%	67.90%		80.77%	76.02%	74.54%	76.61%	
Black	8.08%	10.97%	13.67%	19.22%		8.00%	10.53%	12.14%	12.11%	
Hispanic	4.58%	5.56%	6.19%	6.84%		5.53%	4.88%	4.94%	5.04%	
Others	5.00%	6.21%	5.74%	6.04%		5.70%	8.57%	8.38%	6.24%	
Year of diagnosis					< 0.0001					< 0.0001
1992-1995	18.07%	15.45%	13.64%	13.79%		49.49%	14.74%	15.59%	25.73%	
1996-1999	16.26%	14.13%	13.32%	13.01%		25.58%	15.25%	15.07%	20.28%	
2000-2003	31.32%	32.39%	32.70%	28.62%		19.38%	33.33%	39.74%	33.39%	
2004-2007	34.35%	38.03%	40.34%	44.58%		5.55%	36.68%	29.60%	20.60%	
SEER Regions					< 0.0001					< 0.0001
San Francisco	4.75%	4.20%	3.85%	3.75%		5.43%	4.74%	4.59%	5.99%	
Connecticut	8.54%	7.91%	7.61%	6.90%		10.54%	8.25%	9.53%	7.99%	
Detroit	11.41%	14.99%	14.98%	17.06%		12.29%	11.71%	10.14%	19.49%	
Hawaii	1.52%	1.63%	1.87%	1.53%		2.25%	1.28%	2.49%	1.64%	
Iowa	8.24%	7.05%	6.28%	4.82%		10.94%	8.50%	9.52%	7.50%	
New Mexico	3.94%	3.17%	2.81%	2.08%		6.94%	3.65%	3.47%	5.38%	
Seattle	8.74%	6.54%	4.85%	4.52%		8.39%	7.05%	6.74%	7.58%	
Utah	4.96%	3.60%	3.60%	2.50%		8.11%	4.44%	4.31%	4.95%	
Atlanta/Rural Georgia	4.46%	4.37%	4.17%	3.84%		4.18%	3.29%	3.19%	4.17%	
San Jose	2.86%	2.25%	1.65%	1.89%		2.95%	3.13%	3.52%	3.65%	
Los Angeles	8.88%	8.70%	9.79%	11.69%		14.28%	9.87%	8.15%	6.92%	
Greater California	12.30%	12.29%	11.48%	11.63%		5.04%	12.81%	11.63%	7.87%	
Kentucky	4.74%	5.89%	6.29%	4.88%		2.74%	5.03%	5.38%	2.90%	
Louisiana	4.35%	4.87%	5.60%	6.62%		2.45%	6.18%	7.60%	5.38%	
New Jersey	10.32%	12.55%	15.16%	16.30%		3.45%	10.06%	9.74%	9.09%	
Marital Status					< 0.0001					< 0.0001
Not Married	25.57%	29.30%	31.94%	36.37%		34.14%	41.43%	45.32%	44.65%	
Married	74.43%	70.70%	68.06%	63.63%		65.86%	58.57%	54.68%	55.35%	

Table 4 Outcome characteristics stratified by comorbidity (Scenario 1) and tumor grade (Scenario 2).

Characteristics	Comorbidity (Scenario 1)				P-value	Tumor Grade (Scenario 2)				P-value
	0 N = 80,686	1 N = 20,916	2 N = 5,942	≥ 3 N = 3,277		Well differentiated N = 6,889	Moderately differentiated N = 42,624	Poorly differentiated N = 14,139	Unknown N = 4,105	
Death										
Overall death	23.45%	31.65%	36.57%	48.28%	< 0.0001	64.74%	42.84%	61.81%	58.25%	< 0.0001
Prostate cancer specific death						4.63%	5.64%	20.32%	11.33%	< 0.0001
Non-prostate cancer specific death	20.63%	29.23%	34.33%	45.93%	< 0.0001					

Table 5 Hazard ratio of comorbidity associated with active treatment

	All-cause mortality	Mortality from other causes
Comorbidity	HR (95% CI)	HR (95% CI)
0	1.000	1.000
1	1.765 (1.716-1.815)	1.850 (1.797-1.906)
2	2.445 (2.338-2.557)	2.602 (2.485-2.725)
≥ 3	4.290 (4.074-4.517)	4.612 (4.375-4.864)

Table 6 Hazard ratio of tumor grade associated with ADT.

	All-cause mortality	Mortality from prostate cancer
Tumor grade	HR (95% CI)	HR (95% CI)
Well differentiated	1.000	1.000
Moderately differentiated	1.091 (1.056-1.128)	1.967 (1.749-2.213)
Poorly differentiated	1.950 (1.312-1.379)	8.687 (7.728-9.765)

Funding: This study was not funded by any grant.

Conflict of Interest: Dr. Vahakn B. Shahinian's work was supported by an American Cancer Society Research Scholar Grant (118244-RSGI-10-076-01-CPHPS to VBS). The sponsor had no role in the design or conduct of the study, nor in the interpretation of data, or in the preparation of the manuscript.

Dr. Daniel Jupiter has previously consulted for

Texas Physician Hospitals Advocacy Center, currently consults for State Association of Free Standing Emergency Rooms: Texas, and has received an honorarium from Solana Surgical for presenting a statistical methodology lecture. For the remaining authors there were no disclosures.

This article does not contain any studies with human participants or animals performed by any of the authors.

Comparing Methods of Applying the Propensity Score to Control Selection Bias in Studies of Prostate Cancer

23,088 cases were matched in 3 to 1 matching;
 23,091 cases were matched in 5 to 1 matching;
 23,091 cases were matched in 8 to 1 matching;
 23,122 cases were matched in Caliper matching.

24,174 cases were matched in 3 to 1 matching;
 24,152 cases were matched in 5 to 1 matching;
 24,152 cases were matched in 8 to 1 matching;
 24,134 cases were matched in Caliper matching.

Fig. 3 Comparing hazard ratio of non-prostate cancer mortality (a, Scenario 1) and prostate cancer specific mortality (b, Scenario 2) estimated from the models with and without removed confounding factor, and hazard ratio estimated from for propensity score methods.

Reference

- [1] What are the Key Statistics about Prostate Cancer? 2015.
- [2] Kollmeier, M. A., and Zelefsky, M. J. 2012. How to Select the Optimal Therapy for Early-stage Prostate Cancer Critical Reviews in *Oncology/Hematology* 83: 225-34.
- [3] Hannan, E. L. 2008. "Randomized Clinical Trials and Observational Studies: Guidelines for Assessing Respective Strengths and Limitations." *JACC: Cardiovascular Interventions* 1: 211-7.
- [4] Rothwell, P. M. 2005. External Validity of Randomised Controlled Trials: "to whom do the Results of this Trial Apply?" *The Lancet* 365: 82-93.
- [5] Johnston, S. C., Rootenberg, J. D., Katrak, S., Smith, W. S., and Elkins, J. S. 2006. "Effect of a US National Institutes of Health Programme of Clinical Trials on Public Health and Costs." *The Lancet* 367: 1319-27.
- [6] Baillargeon, J., Kuo, Y.-F., Fang, X., and Shahinian, V. B. 2015. "Long-term Exposure to Testosterone Therapy and the Risk of High Grade Prostate Cancer." *The Journal of Urology*.
- [7] Lu-Yao, G. L., et al. 2009. "Outcomes of Localized Prostate Cancer following Conservative Management." *Jama* 302: 1202-9.
- [8] Lu-Yao, G. L., Albertsen, P. C., Moore, D. F., Shih, W., Lin, Y., DiPaola, R. S., and Yao, S.-L. 2008. "Survival following Primary Androgen Deprivation Therapy among Men with Localized Prostate Cancer." *Jama* 300: 173-81.
- [9] Giordano, S. H., Kuo, Y. F., Duan, Z., Hortobagyi, G. N., Freeman, J., and Goodwin, J. S. 2008. "Limits of Observational Data in Determining Outcomes from Cancer Therapy." *Cancer* 112: 2456-66.
- [10] Austin, P. C. 2011. "An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies." *Multivariate Behavioral Research* 46: 399-424.
- [11] Ross, M. E., Kreider, A. R., Huang, Y.-S., Matone, M., Rubin, D. M., and Localio, A. R. 2015. "Propensity Score Methods for Analyzing Observational Data like Randomized Experiments: Challenges and Solutions for Rare Outcomes and Exposures." *American Journal of Epidemiology*: 469.
- [12] Schneeweiss, S., Rassen, J. A., Glynn, R. J., Avorn, J., Mogun, H., and Brookhart, M. A. 2009. "High-dimensional Propensity Score Adjustment in Studies of Treatment Effects using Health Care Claims Data." *Epidemiology* (Cambridge, Mass) 20: 512.
- [13] Rosenbaum, P. R., and Rubin, D. B. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70: 41-55.
- [14] Rosenbaum, P. R., and Rubin, D. B. 1984. "Reducing Bias in Observational Studies using Subclassification on the Propensity Score." *Journal of the American Statistical Association* 79: 516-24.
- [15] Klabunde, C. N., Potosky, A. L., Legler, J. M., and Warren, J. L. 2000. "Development of a Comorbidity Index using Physician Claims Data." *Journal of Clinical Epidemiology* 53: 1258-67.
- [16] Shahinian, V. B., Kuo, Y.-F., Freeman, J. L., and Goodwin, J. S. 2005. "Risk of Fracture after Androgen Deprivation for Prostate Cancer." *New England Journal of Medicine* 352: 154-64.
- [17] Shahinian, V. B., Kuo, Y.-F., and Gilbert, S. M. 2010. "Reimbursement Policy and Androgen-deprivation Therapy for Prostate Cancer." *New England Journal of Medicine* 363: 1822-32.
- [18] Wong, Y.-N., et al. 2006. "Survival Associated with Treatment vs Observation of Localized Prostate Cancer in Elderly Men." *Jama* 296: 2683-93.
- [19] Robins, J. M., Hernan, M. A., and Brumback, B. 2000. "Marginal Structural Models and Causal Inference in Epidemiology." *Epidemiology* 11: 550-60.
- [20] Austin, P. C. 2008. "The Performance of Different Propensity-score Methods for Estimating Relative Risks." *Journal of Clinical Epidemiology* 61: 537-45.
- [21] Austin, P. C. 2009. "Type I Error Rates, Coverage of Confidence Intervals, and Variance Estimation in Propensity-score Matched Analyses." *The International Journal of Biostatistics* 5.
- [22] Kurth, T., Walker, A. M., Glynn, R. J., Chan, K. A., Gaziano, J. M., Berger, K., and Robins, J. M. 2006. "Results of Multivariable Logistic Regression, Propensity Matching, Propensity Adjustment, and Propensity-based Weighting under Conditions of Nonuniform Effect." *American Journal of Epidemiology* 163: 262-70.
- [23] Shah, B. R., Laupacis, A., Hux, J. E., and Austin, P. C. 2005. "Propensity Score Methods Gave Similar Results to Traditional Regression Modeling in Observational Studies: a Systematic Review." *Journal of Clinical Epidemiology* 58: 550-9.
- [24] Harder, V. S., Stuart, E. A., and Anthony, J. C. 2010. "Propensity Score Techniques and the Assessment of Measured Covariate Balance to Test Causal Associations in Psychological Research." *Psychological Methods* 15: 234.
- [25] Brookhart, M. A., Stürmer, T., Glynn, R. J., Rassen, J., and Schneeweiss, S. 2010. "Confounding Control in Healthcare Database Research: Challenges and Potential Approaches." *Medical Care* 48: S114.
- [26] Greenland, S. 2000. "An Introduction to Instrumental Variables for Epidemiologists." *International Journal of*

Epidemiology 29: 722-9.

- [27] Kuo, Y.-F., Montie, J. E., and Shahinian, V. B. 2012. "Reducing Bias in the Assessment of Treatment Effectiveness: Androgen Deprivation Therapy for Prostate Cancer." *Medical Care* 50: 374.

- [28] Sheffield, K. M., Riall, T. S., Han, Y., Kuo, Y.-F., Townsend, C. M., and Goodwin, J. S. 2013. "Association between Cholecystectomy with vs without Intraoperative Cholangiography and Risk of Common Duct Injury." *JAMA* 310: 812-20.