DOI: 10.1093/jnci/djq393
Advance Access publication on October 13, 2010

ARTICLE

Comparative Effectiveness of Prostate Cancer Treatments: Evaluating Statistical Adjustments for Confounding in Observational Data

Jack Hadley, K. Robin Yabroff, Michael J. Barrett, David F. Penson, Christopher S. Saigal, Arnold L. Potosky

Manuscript received April 7, 2010; revised September 2, 2010; accepted September 10, 2010.

Correspondence to: Jack Hadley, PhD, Office of the Dean and Department of Health Administration and Policy, College of Health and Human Services, George Mason University, 4400 University Dr, MS 2G7, Fairfax, VA 22030 (e-mail: jhadley1@gmu.edu).

Background Using observational data to assess the relative effectiveness of alternative cancer treatments is limited by patient selection into treatment, which often biases interpretation of outcomes. We evaluated methods for addressing confounding in treatment and survival of patients with early-stage prostate cancer in observational data and compared findings with those from a benchmark randomized clinical trial.

Methods We selected 14302 early-stage prostate cancer patients who were aged 66–74 years and had been treated with radical prostatectomy or conservative management from linked Surveillance, Epidemiology, and End Results–Medicare data from January 1, 1995, through December 31, 2003. Eligibility criteria were similar to those from a clinical trial used to benchmark our analyses. Survival was measured through December 31, 2007, by use of Cox proportional hazards models. We compared results from the benchmark trial with results from models with observational data by use of traditional multivariable survival analysis, propensity score adjustment, and instrumental variable analysis.

Results Prostate cancer patients receiving conservative management were more likely to be older, nonwhite, and single and to have more advanced disease than patients receiving radical prostatectomy. In a multivariable survival analysis, conservative management was associated with greater risk of prostate cancer–specific mortality (hazard ratio [HR] = 1.59, 95% confidence interval [CI] = 1.27 to 2.00) and all-cause mortality (HR = 1.47, 95% CI = 1.35 to 1.59) than radical prostatectomy. Propensity score adjustments resulted in similar patient characteristics across treatment groups, although survival results were similar to traditional multivariable survival analyses. Results for the same comparison from the instrumental variable approach, which theoretically equalizes both observed and unobserved patient characteristics across treatment groups, differed from the traditional multivariable and propensity score results but were consistent with findings from the subset of elderly patient with early-stage disease in the trial (ie, conservative management vs radical prostatectomy: for prostate cancer–specific mortality, HR = 0.73, 95% CI = 0.08 to 6.73; for all-cause mortality, HR = 1.09, 95% CI = 0.46 to 2.59).

Conclusion Instrumental variable analysis may be a useful technique in comparative effectiveness studies of cancer treatments if an acceptable instrument can be identified.

J Natl Cancer Inst 2010;102:1780–1793

Comparing the effectiveness of alternative cancer treatments is a critical objective for medical and health services researchers. Practicing physicians need to have valid information about the risks and benefits of alternative treatments to discuss options with their patients and make treatment recommendations. Moreover, in a climate of rapidly growing health-care costs and constrained national resources, health-care policymakers need comparative effectiveness information to make decisions about reimbursement rates and insurance coverage.

The randomized controlled trial is considered the most valid methodology for assessing treatments' efficacy. However, randomized controlled trials are costly, time consuming, and frequently not feasible because of ethical constraints. Moreover, some randomized controlled trial results have limited generalizability because of differences between randomized controlled trial study populations, who may be screened for eligibility on the basis of age and comorbidities, and community populations, who are likely to be much more heterogeneous with regard to health conditions and socioeconomic characteristics.

Given the need for comparative effectiveness information and the limitations of randomized controlled trials, investigating the feasibility of using observational data from actual medical practice in comparative effectiveness studies as a complement to randomized controlled trials is important. However, observational studies are subject to bias that are caused by selection of patients into treatments for reasons related to expected survival (eg, patients

with a better prognosis may be more likely to receive one treatment over another) and the inability to observe all relevant information (1–4). Patient selection into specific treatments is an important consideration in all observational studies, but particularly for those in prostate cancer, because incidence is highest in the elderly who are also most likely to have multiple comorbidities.

The number of published studies that used observational data to assess the effectiveness of cancer treatment has increased dramatically in the past decade (5–9) and is likely to increase even more rapidly with the growing emphasis on comparative effectiveness research (10). In addition to traditional multivariable regression analyses, researchers have used propensity score analysis to adjust for differences in observed patient and physician characteristics. Observational studies (1,11–13) have previously used traditional regression and propensity score methods to evaluate associations between specific prostate cancer treatments with survival. In these studies, the propensity score methods did not completely balance (ie, equalize) important patient characteristics such as tumor grade, size, and comorbidities across treatment groups. Furthermore, patients who received active treatment had better survival for noncancer causes of death than patients who received conservative management, indicating that unobserved differences between groups affected both treatment choice and survival.

Instrumental variable analysis is a statistical technique that uses an exogenous variable (or variables), referred to as an "instrument," that is hypothesized to affect treatment choice but not to be related to the health outcome (14–17). Variations in treatment that result from variations in the value of the instrument are considered to be analogous to variations that result from randomization and so address both observed and unobserved confounding. Instrumental variable analysis has been used with observational data to investigate clinical treatment effects among patients with breast cancer (18–20), lung cancer (21), or prostate cancer (5,22).

Early-stage prostate cancer is an ideal disease for comparing analytic techniques for addressing observed and unobserved confounding because it is a common cancer among men with uncertainty as to which treatment is optimal (23). Geographic variation in surgical treatment has been consistently reported in the United States (7,24,25), and patient selection into different treatments on the basis of their ages and/or comorbid conditions is particularly a concern in evaluations of prostate cancer treatments (24–27). In this study, we evaluated traditional multivariable regression, propensity score, and instrumental variable analyses for addressing observed and unobserved confounding among patients with early-stage prostate cancer who were treated with radical prostatectomy or conservative management. We also compared findings from these analyses with those from a benchmark randomized clinical trial that evaluated the same two treatments (28,29).

CONTEXT AND CAVEATS

Prior knowledge

Although randomized controlled trials provide the best assessment of alternative treatments, such trials are costly, time consuming, and may have limited generalizability. Observational data might be an alternative, but these data can be limited by confounding.

Study design

Data from early-stage, elderly prostate cancer patients who had been treated with radical prostatectomy or conservative management were from the linked Surveillance, Epidemiology, and End Results–Medicare database. Observational data were examined by use of traditional multivariable survival analysis, propensity score adjustment, and instrumental variable analysis. Results were compared with those from a benchmark randomized trial.

Contribution

Propensity score adjustments resulted in similar patient characteristics across treatment groups, and survival was similar to that of traditional multivariable survival analyses. The instrumental variable approach, which theoretically equalizes both observed and unobserved patient characteristics across treatment groups, differed from multivariable and propensity score results but were consistent with findings from a subset of elderly patient with early-stage disease in the randomized trial.

Implications

Instrumental variable analysis may be a useful technique in comparative effectiveness studies of cancer treatments.

Limitations

The study population was restricted to Medicare enrollees in fee-for-service plans. Data from only one randomized trial were used, and its study sub-population of elderly prostate patients may have been under-powered.

From the Editors

Patients and Methods

This study evaluated three statistical techniques (ie, traditional multivariable regression analysis, propensity score analysis, and instrumental variable analysis) for assessing survival among early-stage prostate cancer patients who received either radical prostatectomy or conservative management within 6 months of diagnosis. We used the linked Surveillance, Epidemiology, and End Results (SEER)–Medicare data to identify patients with early-stage prostate cancer, to measure treatment, and to assess prostate cancer–specific and all-cause mortality (30). Findings were compared with those from a benchmark randomized controlled trial that was conducted in Scandinavia (28,29). This Scandinavian trial compared radical prostatectomy with conservative management among newly diagnosed patients with early-stage prostate cancer. To the extent possible, we selected patients by use of the same eligibility criteria as the clinical trial.

Patient Data

We used data from the SEER program maintained by the National Cancer Institute that was linked to Medicare claims data. The SEER registries (which include the metropolitan areas San Francisco–Oakland, Detroit, Seattle, Atlanta, San Jose–Monterey, and Los Angeles county, and the states of Connecticut, Hawaii, Iowa, Kentucky, Louisiana, New Mexico, New Jersey, Utah, and the remainder of California) represent approximately 26% of the US population (1). For each person diagnosed within these defined geographic catchment areas, the SEER registries collect information on every occurrence of a primary incident cancer, the month

and year of diagnosis, cancer site, stage, histology, initial treatment, and vital status including cause of death for patients who died.

Cancer patients reported to SEER from January 1, 1973, through December 31, 2005, have been matched against Medicare's master enrollment file, and Medicare claims have been extracted for those with fee-for-service coverage. Among patients aged 65 years or older with a cancer diagnosis recorded in the SEER data, 94% have been linked with Medicare enrollment data (2). A more detailed description of the linked SEER-Medicare data is available at http://healthservices.cancer.gov/seermedicare/.

Patient demographic characteristics and vital status were obtained from Medicare enrollment data. Information about inpatient and outpatient care, specifically surgery, radiation therapy, injectable hormone treatment, and chemotherapy was obtained from SEER data and Medicare claims. The inpatient (Medicare Provider Analysis and Review), Hospital Outpatient, and Carrier Medicare claims files were used in this study.

We used the eligibility criteria in the randomized controlled trial that compared radical prostatectomy and conservative management (28,29) to select the study population, including newly diagnosed and previously untreated patients with prostate cancer who were younger than 75 years and whose tumor stage was T1 or T2 (28,29). We selected newly diagnosed early-stage prostate cancer (*International Classification of Diseases for Oncology* [*ICD-O* code C61.9]) patients aged 66–74 years in linked SEER-Medicare data from January 1, 1995, through December 31, 2003. Survival was observed for up to 12 years, through December 31, 2007. The median survival time from date of diagnosis to December 31, 2007, was 78 months (interquartile range = 48 months). We used diagnosis and procedure codes from the inpatient and outpatient Medicare claims in the year before diagnosis to classify patient comorbid status that was based on a series of condition indicators and condition-specific weights by use of the NCI combined comorbidity index (26).

We defined radical prostatectomy within 6 months of diagnosis from SEER surgery codes and *International Classification of Diseases, Ninth Edition* (*ICD-9*), and *Current Procedural Terminology, Fourth Edition* (*CPT-4*), codes from the Medicare claims (Appendix Table 1). Conservative management was defined as no radiation, surgery, hormonal treatment, or chemotherapy within 6 months of diagnosis. Among the 110 857 newly diagnosed elderly patients with prostate cancer who had fee-for-service coverage, patients were excluded for the following reasons: unusual histology (n = 2149), identified as having cancer through a death certificate or autopsy (n = 291), not from a SEER registry (n = 283), month of diagnosis or date of death unknown (n = 977), aged 65 years and no data for previous year (n = 10 806), incomplete Medicare Part A and Part B data because of managed care enrollment or only Part A enrollment for 1 year before or after diagnosis (n = 39 417), distant stage or not clinical stage T1 or T2 disease (n = 21 512), and treatment with chemotherapy, radiation therapy, or hormone therapy but without surgery (n = 17 607). The remaining 17 815 patients were used to construct the propensity scores and the primary instrumental variable (ie, the lagged [previous year's], local area, adjusted probability of receiving conservative management). The final sample of 14 302 patients for the estimation of survival models resulted from eliminating patients in geographic areas with fewer than 50 patients over the entire observation period (n = 561) and using a lagged value of the primary instrumental variable (n = 2952).

The primary health outcome measures are the number of months of survival from diagnosis to death or the end of the observation period. We measured both death from prostate cancer from SEER and death from any cause from Medicare claims.

Alternative Statistical Methodologies

The key assumption of a randomized controlled trial is that random assignment to treatment groups effectively holds constant the effects of all observed and unobserved patient characteristics on health. In this analysis, we investigated three analytic alternatives to a randomized controlled trial design that used observational data and statistical methods, rather than randomization, to hold constant the effects of factors other than the treatments that might affect health at the end of the observation period.

Multivariable Regression Analysis. Multivariable regression analysis is the conventional analytic approach for comparing groups. It holds constant the effects of observable factors by including them as covariates in the regression model, but its key assumption is that unobserved factors that affect health are not associated with the treatment received. If this assumption is violated, then the magnitude and possibly the direction of the estimated treatment effect will be biased. The clearest examples of this type of bias occur when there are systematic differences in unobserved health between patients receiving different treatments (1).

Propensity Score Analysis. Propensity score analysis addresses the potential problem that some patient characteristics may vary systematically and substantially across treatment groups so that the other independent variables cannot adequately control for the effects of nonoverlapping characteristics and thus leads to a biased estimate of the effect of the treatment on health. Propensity score analysis also implicitly makes the fundamental assumption that balancing (ie, equalizing) the observable patient characteristics across treatment groups minimizes the potential bias from unobservable factors (31–33).

In practice, there are several approaches to balancing characteristics across treatment groups that typically begin by estimating a logistic regression model to calculate a propensity score [ie, the probability of receiving the particular treatment as a function of all measured factors (ie, confounders) that might affect the treatment outcome (34)]. Factors that arguably influence only treatment choice, but not outcome, should not be included in estimating the propensity score.

After each patient is assigned a propensity score, there are three general strategies for "balancing" patients' characteristics across treatment groups (35): 1) grouping, which subdivides patients into homogeneous subgroups on the basis of the propensity score; 2) matching, which essentially pairs patients with identical or nearly identical propensity scores across treatment arms; and 3) weighting, which assigns patients differential weights on the basis of their propensity score. In each of these approaches, the primary goal is to develop samples of treated and untreated patients who are as similar as possible and arguably mimic samples that would be created by randomization.

Previous studies (34,35) indicate that the weighting approach is the most general and most efficient because it uses all available data and does not require any arbitrary decisions with regard to grouping or matching. Therefore, in this analysis, we proceeded by estimating the propensity score (*ps*) from a logistic regression model of the probability of conservative management relative to aggressive (surgical) treatment as a function of clinical and demographic characteristics. We then compared two weighting approaches that make different assumptions about the distribution of propensities between the treated and untreated (or control) populations (34). 1) Specifically, the inverse probability of treatment weighting assigns weights of $1/ps$ for patients receiving conservative management and $1/(1 - ps)$ for patients receiving radical prostatectomy. This approach assumes that the two patient populations are reasonably similar and that the treatment could be applied to the entire study population. 2) Standardized mortality ratio weighting assigns a weight of 1 for treated patients (conservative management) and a weight of $[ps/(1 - ps)]$ for untreated patients (radical prostatectomy). Under this assumption, the estimated treatment effect applies to the subpopulation that has characteristics similar to the treated population. This approach is more appropriate when the study populations in the two treatment groups are very different.

Instrumental Variable Analysis. Instrumental variable analysis addresses a potential limitation of traditional regression analysis and propensity score analysis by seeking to adjust for the effects of both observable and unobservable characteristics. The key challenge of instrumental variable analysis is identifying at least one factor that statistically significantly affects treatment choice but is not related to the health outcome. Variations in treatment that result from this exogenous factor, referred to as an instrument, can be regarded as similar to randomization because random assignment is itself essentially an exogenous instrument in that it affects treatment but is unrelated to the subsequent effect on health. Like treatment groups created by randomization, patients who receive different treatments because of variation in the value of the instrument should have similar observable and unobservable characteristics.

In general, there are two assumptions and conditions for using instrumental variable analysis (36,37). First, the instrument should have a statistically significant impact on the treatment (ie, should be a statistically significant cause of treatment variation). In practice, this condition has been translated into a rule of thumb that the test for the statistical significance of the instrument(s) should have an F statistic value of at least 10 and that the instrument should account for a meaningful share of the observed variation in the treatment, because, if the instrument does not explain very much of the variation in the treatment, then predicted treatment will not differ very much across patient populations.

Second, the instrument should not be associated with the health outcome or with unobserved factors that might influence the health outcome. However, there is no definitive statistical test that proves an instrument's validity. In effect, any potential instrument also needs to satisfy a plausibility condition (ie, a logical and convincing argument that justifies the instrument as a factor that statistically significantly influences the treatment received but is not associated with either the patient's current health or the treatment outcome).

Applying instrumental variable analysis requires a two-stage estimation approach. The first-stage equation predicts the likelihood of receiving the treatment as a function of the instrument and other exogenous factors, and the second-stage equation estimates the effect of the treatment on the health outcome incorporating the "instrumental variable" generated from the first-stage equation. If the health outcome model is nonlinear (eg, a logistic or hazard model), as in this analysis, then the appropriate instrumental variable procedure is the two-stage residual inclusion method (38), which adds the residual (ie, the difference between the actual value of the treatment choice variable and the predicted value) from the first-stage equation as an additional variable in the second-stage equation. In principle, only one instrument is needed to implement instrumental variable analysis for a single comparison of two treatment choices. However, there may be several potential instruments, and investigators should assess the relative strengths and robustness of estimates generated from alternative individual or combinations of instruments.

Although not a definitive classification scheme, past instrumental variable analyses (5,18,19,21,39–42) have tended to select instruments of the following types: 1) the frequency of a particular treatment in a geographic area, sometimes referred to as a local area treatment pattern or treatment signature; 2) the treatment pattern of the patient's provider (ie, physician, hospital, clinic) that is based on the pattern of care received by other patients with the same condition treated by that provider; 3) the distance to or availability of a key type of medical resource that is strongly associated with the treatment of interest; 4) the economic cost to the provider and/or the patient of alternative treatments; or 5) natural "experiments" that occur because of changes in policy or institutional structure that are independent of individual patients' health.

Exogenous Instrumental Variables. We selected the lagged (ie, previous year's) local area treatment pattern for conservative management as the primary instrumental variable for three reasons. First, it varies substantially across geographic areas (Supplementary Table 1, available online) and has a highly statistically significant impact on the actual treatment received (Supplementary Table 2, available online). Second, it satisfies the key plausibility criterion of being independent of a current patient's health and other characteristics because it reflects provider treatment decisions from a previous time period (ie, the year before the patient was diagnosed). Third, it can be constructed from readily available data.

The instrument was created by grouping eligible patients from the SEER-Medicare database into hospital referral regions as developed by the Dartmouth Atlas of Health Care (43). A hospital referral region is the set of contiguous zip codes around a major hospital (defined as a hospital that performs cardiovascular surgery and neurosurgery) from which the hospital draws substantial proportions of patients admitted for major cardiovascular surgery or neurosurgery. We defined geographic areas as hospital referral regions because of probable intra-area heterogeneity in treatment patterns within the several geographically large SEER registry areas and small sample concerns that are associated with geographically smaller health-care service areas (ie, 561 patients were deleted after applying the constraint that a hospital referral region have at least 50 patients during the entire study period and 19

adjoining or nearby hospital referral regions were combined to satisfy this condition) (see Supplementary Table 1, available online for a list of the hospital referral regions and their sample sizes).

We constructed the primary instrumental variable for treatment received by use of a two-step process. First, we used the entire dataset (n = 17 815) to estimate the probability of receiving conservative management as a function of patients' clinical characteristics (tumor stage and grade, NCI comorbidity index, and Medicare reimbursements for medical care in the previous year), demographics (age, race, ethnicity, and marital status), year of diagnosis, and all possible interactions among these variables. Second, we calculated the difference between the actual proportion of patients receiving conservative management and the average predicted probability of receiving conservative management (generated from the logistic regression model) in each hospital referral region by year. Areas with relatively large positive differences between the actual and predicted proportions of patients receiving conservative management favor a conservative management treatment pattern, and areas with large negative differences between the actual and predicted proportions of patients receiving conservative management favor a radical prostatectomy treatment pattern. We then lagged this measure of the local area treatment pattern by 1 year and linked it to each patient in the analysis to enhance the instrument's independence from patients' current health and unobserved characteristics.

Additional potential instruments were used to construct a second instrumental variable for sensitivity analysis. These variables measure the availability of medical resources in patients' counties of residence in 2000: total number of patient care physicians, urologists, radiation oncologists, and hospital beds per 100 000 population [from the Area Resource File (44)]. Controlling for the overall availability of physicians, we hypothesized that conservative management will be less likely in areas with more hospital beds and more specialists who tend to concentrate on care of prostate cancer patients. Although these measures are likely not to be associated with patients' underlying health, their direct association with prostate cancer treatment may be relatively weak because they are not specific to Medicare patients with prostate cancer. Moreover, they are measured at only a single point in time, 2000, during the 14-year observation period in this study.

Statistical Estimation

All statistical models included the following control variables: age (66–69 or 70–74 years), race or ethnicity (white non-Hispanic, white Hispanic, African American, or all other races), marital status (single or married), tumor characteristics (stage and grade), previous health problems [as measured by the NCI combined comorbidity index (26) and Medicare reimbursements in the 12 months before diagnosis], and year of diagnosis. Year of diagnosis captured the combined effects of several factors that were changing over the study's time period, including the increase in prostate-specific antigen testing, movements by Medicare beneficiaries into and out of managed care organizations, and changes in Medicare physician reimbursement. These trends potentially change the nature of the underlying population of elderly fee-for-service Medicare patients who are diagnosed with prostate cancer.

Treatment propensity (ie, the predicted probability of receiving conservative management) for the propensity score analysis and for constructing the lagged area treatment pattern for the instrumental variable analysis was estimated with logistic regression. The survival models were estimated with Cox proportional hazard models. Visual inspection of the parallelism of the Kaplan–Meier plots of the logarithms of the estimated cumulative survival models by treatment supported the proportionality assumption. The instrumental variable version of the Cox hazard model was estimating with the two-stage residual inclusion method (38), which has been shown to be appropriate for nonlinear outcome models. This approach adds a separate variable that measures the residual (ie, the difference between the actual value of the [0,1] dependent variable and the predicted probability generated by the logistic model from the first-stage model for predicting treatment choice as a function of the instrumental variable. All models were estimated with the SAS statistical software (Cary, NC). All statistical tests were two-sided.

Results

Prostate cancer patients receiving conservative management were statistically significantly older, much more likely to be African American, more likely to have more advanced disease and much less likely to be married than patients receiving radical prostatectomy (Table 1). The conservative management group was also much less likely to have any Medicare claims in the year before diagnosis and, as a result, has a much higher proportion with unknown comorbidities. Characteristics of the two treatment groups changed after applying the statistical adjustments designed to correct for potential observational data biases. As expected after propensity score reweighting that used the inverse probability of treatment weights, the weighted characteristics of the two treatment groups were virtually identical. The one characteristic that was not equalized was the lagged difference between the actual and predicted proportions of patients who received conservative management. This difference between treatment groups in actual and predicted proportions of patients who received conservative management indicates that unobserved differences, which influenced the treatment received between the reweighted populations, may persist even after reweighting. (In data not shown, using the standardized mortality ratio weight also balances the characteristics of the two treatment groups by adjusting the radical prostatectomy group to have the same reweighted characteristics as the conservative management group.)

Characteristics of the sample were also grouped by the value of the primary instrument for the instrumental variable approach (ie, the lagged area-wide difference between the actual and predicted proportions of patients who were treated by conservative management). Splitting the sample at the median value of the instrument resulted in a 12.8% difference in the instrument's value between the two groups, which meant that by holding clinical characteristics constant, patients in the above-median group were, on average, 6.4% more likely to receive conservative management, and those in the below-median group were 6.4% less likely to receive conservative management. (The range across individual hospital referral regions was from −13.9% in the region including

Table 1. Comparison of unweighted and propensity score (PS) reweighted samples, by treatment, and for samples grouped by median value of instrumental variable*

Variable	Unweighted				PS reweighted (IPTW)			Below- or above-median value of instrumental variable		
	All	RP	CM†	*P*	RP	CM‡	*P*	Below	Above	*P*
Lagged difference between actual and predicted proportions receiving CM, % (instrumental variable)	−0.001	−0.014	0.007	<.001†	−0.016	.004	<.001†	−0.064	0.064	<.001†
Actual treatment, % CM	0.330	0.000	1.000	—	—	—	—	0.271	0.354	<.001
Age group, y										
66–69	0.504	0.532	0.441	<.001†	0.504	0.506	0.79†	0.510	0.498	0.18†
70–74	0.496	0.468	0.559		0.496	0.494		0.490	0.502	
Race or ethnicity				<.001‡			0.96‡			<.001‡
White non-Hispanic	0.775	0.807	0.700		0.773	0.776		0.782	0.767	
White Hispanic	0.067	0.069	0.063		0.066	0.066		0.076	0.057	
African American	0.106	0.078	0.169		0.108	0.106		0.084	0.128	
All other	0.052	0.046	0.068		0.052	0.052		0.057	0.048	
Marital status										
Single	0.234	0.185	0.342	<.001†	0.230	0.236	0.44†	0.219	0.251	<.001†
Married	0.766	0.815	0.658		0.770	0.764		0.781	0.749	
Stage										
T1	0.63	0.649	0.610	<.001†	0.637	0.634	0.71†	0.636	.637*	0.97†
T2	0.364	0.351	0.390		0.363	0.366		0.364	.363*	
Grade				<.001‡			0.30‡			<.001‡
Well differentiated	0.079	0.071	0.096		0.081	0.078		0.084	0.074	
Moderately differentiated	0.704	0.708	0.696		0.701	0.709		0.699	0.711	
Poorly differentiated	0.188	0.210	0.140		0.188	0.184		0.193	0.181	
Unknown	0.029	0.011	0.069		0.030	0.029		0.024	0.035	
NCI comorbidity index				.002‡			0.32‡			0.61‡
0	0.698	0.754	0.578		0.701	0.699		0.706	.691	
1	0.096	0.100	0.084		0.097	0.096		0.094	.098	
≥2	0.092	0.093	0.089		0.094	0.089		0.093	.090	
Unknown	0.114	0.054	0.234		0.108	0.116		0.107	.121	
Any Medicare claims in year before diagnosis	0.899	0.957	0.771	<.001†	0.904	0.892	0.27†	0.903	.895	.002†
Adjusted reimbursement in year before diagnosis	$2162	$2290	$1876	<.001†	$2178	$2183	0.05†	$2184	$2139	0.68†

* Data are expressed as the mean value. CM = conservative management; IPTW = inverse probability of treatment weights; NCI = National Cancer Institute; RP = radical prostatectomy.

† A two-sided *t* test was used.

‡ A two-sided χ^2 test was used.

Owensboro, Paduca, and Nashville, TN, to 19.4% in New Haven, CT.) Similarly, 35% of the patients in the above-median group actually received conservative management compared with 27.1% in the below-median group. Thus, the instrument successfully distinguishes between patients that are more or less likely to receive conservative management for reasons that were independent of their observed health characteristics. Grouping patients by the value of the instruments narrowed, although did not eliminate, several of the differences in the observed characteristics. It equalized the values of characteristics measuring disease stage, the number of comorbidities, and the presence and amount of Medicare claims in the year before diagnosis.

The unadjusted observational data and the propensity score reweighted data clearly indicate a statistically significant survival advantage for both prostate cancer–specific death and death from all causes associated with radical prostatectomy (Table 2). (The longer survival time shown for all-cause mortality reflected a lag in reporting cause-specific mortality in the SEER-Medicare data.) The same comparisons for patients grouped by the value of the instrument narrowed the differences in both months of prostate-specific and all-cause survival and mortality percentages, which are not statistically significantly different from each other. The absence of statistically significant differences in the mortality rates was consistent with the findings from the randomized controlled trial after 12 years of follow-up (29).

As in the comparison of means in Table 2, the hazard rates from the models estimated with the unweighted observational sample and the two propensity score reweighted samples indicated that a large and statistically significant survival advantage was associated with radical prostatectomy (Tables 3–5). In unweighted multivariable survival analysis, conservative management was associated with greater risk of prostate cancer–specific mortality (Tables 3 and 4; hazard ratio [HR] = 1.59, 95% confidence interval [CI] = 1.27 to 2.00) and all-cause mortality (Tables 3 and 5; HR = 1.47, 95% CI = 1.35 to 1.59) than radical prostatectomy. Hazard rates for conservative management compared with radical prostatectomy by using both propensity score reweighting approaches were similar for prostate cancer–specific and all-cause mortality. In contrast, the hazard rates estimated by instrumental variable analysis did not show a statistically significant survival advantage for radical prostatectomy (Tables 3 and 6; for prostate cancer–specific mortality, HR = 0.73, 95% CI = 0.08 to 6.73; for all-cause mortality, HR = 1.09, 95% CI = 0.46 to 2.59). Moreover, the instrumental variable estimates were also very similar to the relative risk rates calculated by the benchmark randomized controlled trial. (Tables 4–6 report the complete models underlying the results summarized in Table 3.)

The strength of the primary instrumental variable was indicated by its statistical significance in the first-stage equation that predicts treatment choice for each case and its lack of statistical significance in the second-stage survival model. It was highly statistically significant in the first-stage model (F = 109.5 and $P < .001$) and accounted for 4.2% of the explained variation, which, although not as large as one would like, was partially attributed to the fact that the first stage–dependent variable was dichotomous rather than continuous. Its independence of the survival outcomes

Table 2. Comparisons of mean values of outcome variables by estimation method and treatment*

Estimation method and treatment	Prostate cancer–specific death		All-cause death	
	% died (95% CI)	Survival, mo (95% CI)	% died (95% CI)	Survival, mo (95% CI)
All patients (unweighted)	0.028	73.2	0.200	83.0
Observational (unweighted)				
RP	0.025 (0.022 to 0.028)	75.0 (74.46 to 75.62)	0.177 (0.170 to 0.185)	85.1 (84.49 to 85.70)
CM	0.036 (0.030 to 0.041)	69.2 (68.38 to 70.05)	0.249 (0.237 to 0.263)	78.4 (77.57 to 79.33)
P for difference in means	<.001	<.001	<.001	<.001
PS reweighted (IPTW)				
RP	0.026 (0.023 to 0.030)	75.2 (74.6 to 75.7)	0.185 (0.177 to 0.193)	85.1 (84.5 to 85.7)
CM	0.035 (0.029 to 0.040)	68.4 (67.6 to 69.2)	0.236 (0.223 to 0.248)	77.8 (76.9 to 78.7)
P for difference in means	<.001	<.001	<.001	<.001
PS reweighted (SMRW)				
RP	0.030 (0.026 to 0.033)	75.4 (74.8 to 76.0)	0.203 (0.195 to 0.211)	85.2 (84.6 to 85.8)
CM	0.036 (0.030 to 0.041)	69.2 (68.4 to 70.0)	0.250 (0.237 to 0.263)	78.4 (77.6 to 79.3)
P for difference in means	<.001	<.001	<.001	<.001
Instrumental variable†				
Less than median value of instrument	0.027 (0.023 to 0.031)	73.7 (73.05 to 74.41)	0.192 (0.183 to 0.201)	83.6 (82.94 to 84.35)
Median value of instrument or higher	0.030 (0.026 to 0.034)	72.7 (72.05 to 73.39)	0.208 (0.199 to 0.218)	82.4 (81.71 to 83.12)
P for difference in means	.351	.038	.012	.016
RCT (12 y of follow-up)‡				
RP	0.131 (0.088 to 0.195)	NR	0.420 (0.350 to 0.505)	NR
CM	0.132 (0.089 to 0.196)	NR	0.393 (0.325 to 0.477)	NR
P for difference in means	NR		NR	

* A two-sided *t* test was used. CM = conservative management; IPTW = inverse probability of treatment weights; NR = not reported; PS = propensity score; RCT = randomized controlled trial; RP = radical prostatectomy; SMRW = standardized mortality ratio weights.

† Lagged area treatment was the only instrument.

‡ Bill-Axelson et al. (29).

Table 3. Adjusted hazard ratios (HR) comparing conservative management vs radical prostatectomy by the estimation method*

	Death from prostate cancer		Death from all causes	
Estimation method	**Adjusted HR (95% CI)**	***P***	**Adjusted HR (95% CI)**	***P***
Observational (unweighted)	1.59 (1.27 to 2.00)	<.001	1.47 (1.35 to 1.59)	<.001
PS reweighted (IPTW)	1.60 (1.40 to 1.83)	<.001	1.54 (1.46 to 1.62)	<.001
PS reweighted (SMRW)	1.39 (1.10 to 1.76)	<.001	1.46 (1.33 to 1.59)	<.001
Instrumental variable†	0.73 (0.08 to 6.73)	.78	1.09 (0.46 to 2.59)	.84

* Data are from the randomized controlled trial (29) with a follow-up of 12 years that reported the following relative risks (RR): for death from prostate cancer for men aged 65–74 years at diagnosis, RR = 0.87, 95% CI = 0.51 to 1.49, $P = .55$; and for death from all causes, RR = 1.04, 95% CI = 0.77 to 1.40, $P = .81$. CI = confidence interval; IPTW = inverse probability of treatment weights; PS = propensity score; SMRW = standardized mortality ratio weights.

† Lagged area treatment was the only instrument.

was confirmed by its lack of statistical significance as an independent variable in an alternative version (data not shown) of the Cox survival models ($P = .68$ in the all-cause survival model and $P = .34$ in the prostate-specific survival model). A second set of instrumental variable estimates that used additional area variables to construct the treatment instrument resulted in even smaller hazard rates than reported in Table 3. However, these estimates were not as reliable because of the relatively weaker association between general area variables and prostate cancer treatment choices in older men.

Table 4. Prostate-specific Cox proportional hazard models, unadjusted and propensity score (PS) adjusted*

	Unadjusted		PS adjusted (IPTW)		PS adjusted (SMRW)	
Variable label	**HR (95% CI)**	***P*†**	**HR (95% CI)**	***P*†**	**HR (95% CI)**	***P*†**
Conservative management vs radical prostatectomy	1.593 (1.267 to 2.003)	<.001	1.599 (1.396 to 1.831)	<.001	1.389 (1.098 to 1.757)	.006
Age at diagnosis 70–74 vs 66–69 y	1.546 (1.265 to 1.89)	<.001	1.578 (1.378 to 1.806)	<.001	1.574 (1.232 to 2.01)	<.001
Black non-Hispanic vs white non-Hispanic	1.023 (0.749 to 1.397)	.887	1.225 (0.998 to 1.504)	.052	1.336 (0.999 to 1.788)	.051
Hispanic vs white non-Hispanic	1.057 (0.712 to 1.57)	.783	0.993 (0.75 to 1.314)	.960	1.237 (0.77 to 1.99)	.379
Other race vs white non-Hispanic	0.505 (0.29 to 0.881)	.016	0.435 (0.286 to 0.661)	<.001	0.472 (0.247 to 0.902)	.023
Married vs single	0.829 (0.662 to 1.038)	.102	0.85 (0.729 to 0.991)	.038	0.844 (0.663 to 1.074)	.168
Stage T2 vs T1	1.265 (1.037 to 1.543)	.021	1.38 (1.206 to 1.581)	<.001	1.181 (0.932 to 1.498)	.169
Grade moderately differentiated vs well differentiated	1.929 (1.116 to 3.335)	.019	2.184 (1.463 to 3.262)	.001	2.469 (1.327 to 4.593)	.004
Grade poorly differentiated vs well differentiated	8.953 (5.167 to 15.516)	<.001	10.376 (6.942 to 15.51)	<.001	10.017 (5.317 to 18.873)	<.001
Grade unknown vs well differentiated	4.441 (2.26 to 8.728)	<.001	5.487 (3.385 to 8.895)	<.001	5.984 (3.037 to 11.791)	<.001
NCI comorbidity index of 0 vs unknown	0.417 (0.227 to 0.766)	.005	0.342 (0.234 to 0.501)	<.001	0.344 (0.186 to 0.634)	<.001
NCI comorbidity index of 1 vs unknown	0.447 (0.225 to 0.89)	.022	0.292 (0.186 to 0.457)	<.001	0.285 (0.133 to 0.611)	.001
NCI comorbidity index of ≥2 vs unknown	0.681 (0.347 to 1.335)	.263	0.451 (0.292 to 0.696)	<.001	0.47 (0.229 to 0.966)	.040
Inflation-adjusted reimbursement in the year before diagnosis (2003 prices)	1 (1 to 1)	.148	1 (1 to 1)	<.001	1 (1 to 1)	.210
Any inpatient, outpatient, or carrier claims in year before diagnosis vs no claims	1.572 (0.823 to 2.999)	.170	1.995 (1.32 to 3.014)	.001	1.956 (1.043 to 3.666)	.036
Diagnosed in 1997 vs 1996	1.256 (0.938 to 1.682)	.126	1.475 (1.202 to 1.81)	<.001	1.492 (1.048 to 2.125)	.026
Diagnosed in 1998 vs 1996	0.739 (0.512 to 1.068)	.108	0.763 (0.59 to 0.987)	.039	1.056 (0.698 to 1.599)	.797
Diagnosed in 1999 vs 1996	0.777 (0.524 to 1.15)	.207	0.918 (0.706 to 1.194)	.523	0.709 (0.422 to 1.191)	.194
Diagnosed in 2000 vs 1996	0.999 (0.666 to 1.498)	.996	1.051 (0.798 to 1.383)	.723	1.165 (0.717 to 1.894)	.538
Diagnosed in 2001 vs 1996	0.645 (0.429 to 0.971)	.036	0.661 (0.502 to 0.871)	.003	0.742 (0.454 to 1.212)	.234
Diagnosed in 2002 vs 1996	0.964 (0.643 to 1.445)	.858	0.983 (0.745 to 1.298)	.905	1.069 (0.648 to 1.763)	.795
Diagnosed in 2003 vs 1996	0.586 (0.348 to 0.987)	.045	0.67 (0.477 to 0.942)	.021	0.712 (0.379 to 1.335)	.289

* CI = confidence interval; HR = hazard ratio; IPTW = Inverse probability of treatment weights; NCI = National Cancer Institute; SMRW = standardized mortality ratio weights.

† A two-sided Wald χ^2 test was used.

Table 5. All-cause Cox proportional hazard survival models, unadjusted and propensity score (PS) adjusted*

	Unadjusted		PS adjusted (IPTW)		PS adjusted (SMRW)	
Variable label	**HR (95% CI)**	***P*†**	**HR (95% CI)**	***P*†**	**HR (95% CI)**	***P*†**
Conservative management vs radical prostatectomy	1.467 (1.351 to 1.594)	<.001	1.537 (1.46 to 1.618)	<.001	1.457 (1.332 to 1.593)	<.001
Age at diagnosis 70–74 vs 66–69 y	1.627 (1.508 to 1.756)	<.001	1.573 (1.494 to 1.657)	<.001	1.573 (1.432 to 1.727)	<.001
Black non-Hispanic vs white non-Hispanic	1.293 (1.163 to 1.438)	<.001	1.365 (1.269 to 1.468)	<.001	1.37 (1.227 to 1.53)	<.001
Hispanic vs white non-Hispanic	0.946 (0.81 to 1.104)	0.479	0.924 (0.829 to 1.031)	.158	0.978 (0.802 to 1.193)	.828
Other race vs white non-Hispanic	0.667 (0.553 to 0.805)	<.001	0.708 (0.623 to 0.805)	<.001	0.615 (0.497 to 0.761)	<.001
Married vs single	0.701 (0.646 to 0.761)	<.001	0.743 (0.702 to 0.786)	<.001	0.73 (0.667 to 0.8)	<.001
Stage T2 vs T1	0.955 (0.884 to 1.032)	.242	0.995 (0.944 to 1.05)	.865	0.947 (0.865 to 1.037)	.242
Grade moderately differentiated vs well differentiated	0.888 (0.786 to 1.003)	.057	0.885 (0.813 to 0.963)	.005	0.893 (0.778 to 1.024)	.105
Grade poorly differentiated vs well differentiated	1.348 (1.172 to 1.55)	<.001	1.46 (1.326 to 1.607)	<.001	1.458 (1.234 to 1.722)	<.001
Grade unknown vs well differentiated	0.94 (0.75 to 1.179)	.594	0.911 (0.775 to 1.072)	.261	0.921 (0.749 to 1.131)	.431
NCI comorbidity index of 0 vs unknown	0.701 (0.507 to 0.968)	.031	0.626 (0.505 to 0.776)	<.001	0.661 (0.474 to 0.922)	.015
NCI comorbidity index of 1 vs unknown	1.144 (0.817 to 1.601)	.435	1.046 (0.836 to 1.309)	.695	1.034 (0.725 to 1.473)	.855
NCI comorbidity index of ≥2 vs unknown	1.897 (1.36 to 2.646)	<.001	1.593 (1.275 to 1.989)	<.001	1.695 (1.198 to 2.398)	.003
Inflation-adjusted reimbursement in the year before diagnosis (2003 prices)	1 (1 to 1)	<.001	1 (1 to 1)	<.001	1 (1 to 1)	<.001
Any inpatient, outpatient, or carrier claims in year before diagnosis vs no claims	1.118 (0.798 to 1.567)	.515	1.295 (1.033 to 1.625)	.025	1.213 (0.864 to 1.704)	.265
Diagnosed in 1997 vs 1996	1.118 (0.993 to 1.259)	.065	1.147 (1.055 to 1.246)	.001	1.164 (1.009 to 1.342)	.037
Diagnosed in 1998 vs 1996	0.921 (0.807 to 1.052)	.226	0.87 (0.794 to 0.953)	.003	0.921 (0.787 to 1.077)	.302
Diagnosed in 1999 vs 1996	0.923 (0.799 to 1.066)	.277	0.889 (0.806 to 0.982)	.020	0.862 (0.722 to 1.03)	.102
Diagnosed in 2000 vs 1996	0.971 (0.829 to 1.138)	.719	0.902 (0.809 to 1.006)	.063	0.932 (0.771 to 1.127)	.468
Diagnosed in 2001 vs 1996	0.839 (0.727 to 0.967)	.016	0.783 (0.711 to 0.863)	<.001	0.824 (0.696 to 0.976)	.025
Diagnosed in 2002 vs 1996	0.799 (0.683 to 0.936)	.005	0.755 (0.679 to 0.839)	<.001	0.769 (0.637 to 0.929)	.007
Diagnosed in 2003 vs 1996	0.676 (0.562 to 0.814)	<.001	0.633 (0.56 to 0.717)	<.001	0.656 (0.524 to 0.82)	<.001

* CI = confidence interval; HR = hazard ratio; IPTW = inverse probability of treatment weights; NCI = National Cancer Institute; SMRW = standardized mortality ratio weights.

† A two-sided Wald χ^2 test was used.

Discussion

In this study, we evaluated statistical methods for addressing observed and unobserved confounding in treatment of early-stage prostate cancer patients and prostate cancer–specific and all-cause mortality in observational data. We compared our findings with those from a clinical trial that provided the benchmark results that patients aged 65 years or older receiving radical prostatectomy or conservative management had similar prostate cancer–specific mortality (relative risk [RR] = 0.87, 95% CI = 0.51 to 1.49) and all-cause mortality (RR = 1.04, 95% CI = 0.77 to 1.40) (28,29). Contrary to these benchmark results, both multivariable regression analysis and the propensity score reweighting methods produced very similar implications (ie, that aggressive treatment by radical prostatectomy was associated with statistically significant better survival than conservative management). Consistency between propensity score reweighting and traditional multivariable analysis is not uncommon (45,46).

The instrumental variable results, which accounted for unobserved confounding, were more similar to results from the benchmark randomized controlled trial than to results from the unadjusted multivariable regression and propensity score reweighted analyses. Alternative specifications of the instrumental variable produced different point estimates of the hazard rate, although all were found to be non-statistically significant. Findings from this study suggest that the instrumental variable approach may be useful in comparative effectiveness studies of observational databases of other treatments for other diseases. In particular, the lagged area treatment variable that we constructed from the SEER-Medicare data may be a good instrument in studies of other cancer treatments.

Table 6. Instrumental variable estimates of Cox proportional hazard survival models*

Variable label	Prostate-specific survival		All-cause survival	
	HR (95% CI)	P†	HR (95% CI)	P†
Watchful waiting	0.725 (0.078 to 6.729)	.777	1.094 (0.462 to 2.593)	.838
Residual (waiting)—from Supplementary Table 2, available online	2.442 (0.215 to 27.71)	.471	1.491 (0.64 to 3.475)	.354
Age at diagnosis 70–74 vs 66–69 y	1.536 (1.144 to 2.061)	.004	1.678 (1.509 to 1.865)	<.001
Black non-Hispanic vs white non-Hispanic	1.125 (0.725 to 1.745)	.599	1.367 (1.181 to 1.583)	<.001
Hispanic vs white non-Hispanic	0.868 (0.542 to 1.389)	.554	0.933 (0.799 to 1.09)	.381
Other race vs white non-Hispanic	0.504 (0.259 to 0.982)	.044	0.7 (0.573 to 0.855)	.001
Married vs single	0.743 (0.496 to 1.112)	.149	0.662 (0.573 to 0.765)	<.001
Stage T2 vs T1	1.197 (0.946 to 1.515)	.135	0.976 (0.898 to 1.061)	.568
Grade moderately differentiated vs well differentiated	1.969 (1.008 to 3.843)	.047	0.848 (0.728 to 0.988)	.035
Grade poorly differentiated vs well differentiated	8.079 (3.612 to 18.071)	<.001	1.227 (0.975 to 1.546)	.082
Grade unknown vs well differentiated	6.491 (2.312 to 18.224)	<.001	1.067 (0.763 to 1.493)	.705
NCI comorbidity index of 0 vs unknown	0.322 (0.148 to 0.703)	.004	0.668 (0.462 to 0.965)	.031
NCI comorbidity index of 1 vs unknown	0.354 (0.152 to 0.825)	.016	1.095 (0.752 to 1.594)	.637
NCI comorbidity index of ≥2 vs unknown	0.488 (0.213 to 1.117)	.090	1.845 (1.277 to 2.667)	.001
Inflation-adjusted reimbursement in the year before diagnosis (2003 prices)	1 (1 to 1)	.704	1 (1 to 1)	<.001
Any inpatient, outpatient, or carrier claims in year before diagnosis vs no claims	1.349 (0.583 to 3.119)	.484	0.989 (0.677 to 1.444)	.953
Diagnosed in 1997 vs 1996	1.185 (0.857 to 1.637)	.304	1.104 (0.979 to 1.244)	.106
Diagnosed in 1998 vs 1996	0.772 (0.522 to 1.142)	.195	0.929 (0.814 to 1.061)	.279
Diagnosed in 1999 vs 1996	0.987 (0.653 to 1.492)	.952	0.928 (0.803 to 1.073)	.313
Diagnosed in 2000 vs 1996	1.176 (0.741 to 1.866)	.492	0.998 (0.848 to 1.175)	.983
Diagnosed in 2001 vs 1996	0.733 (0.453 to 1.187)	.207	0.851 (0.733 to 0.988)	.034
Diagnosed in 2002 vs 1996	1.061 (0.632 to 1.783)	.822	0.844 (0.711 to 1.002)	.052
Diagnosed in 2003 vs 1996	0.871 (0.441 to 1.72)	.690	0.723 (0.592 to 0.884)	.002

* CI = confidence interval; HR = hazard ratio; NCI = National Cancer Institute.

† A two-sided Wald χ^2 test was used.

Previous studies (21,39,47) have used both propensity score and instrumental variable analyses with the same data to assess treatment outcomes. Earle et al. (21) examined the effect of chemotherapy on survival in elderly patients with stage IV non–small cell lung cancer and compared these results with those of a randomized controlled trial. In that study, both the instrumental variable and propensity score analyses produced results that were similar to those of the randomized controlled trial, although median follow-up was much shorter for this acute disease, and there were stronger associations between observable clinical characteristics and survival among patients with lung cancer than among those with prostate cancer. Other studies have investigated the effects of invasive cardiac management on acute myocardial infarction survival (39) and adherence to two oral antidiabetic drug therapies that differ in patient tolerance, adverse events, and side effects (47). The acute myocardial infarction survival study compared multivariable regression, two propensity score methods, and instrumental variable analysis with randomized controlled trial findings and reported similar findings from the multivariable and two propensity score methods. However, the instrumental variable findings were comparable to the results from randomized controlled trials, indicating that there was selection bias that was caused by unobservable confounders that could not be adjusted by propensity score analysis. In the drug adherence study, multivariable, propensity score, and instrumental variable findings were similar, indicating that any selection bias was caused by observable factors because all three methods of adjusting for confounding produced similar results.

This study and previous studies (21,39,47) indicated that if unobservable factors were not a major source of bias, then instrumental variable and propensity score methods should provide similar results. Whether the instrumental variable and propensity score results support or contradict unadjusted results depends on the extent of selection bias in assigning patients to alternative treatments. These differences across studies also suggest that it may not be possible to generalize about the choice of a statistical method across different clinical conditions. Instrumental variable analysis in principle is the more robust approach because it adjusts for both observable and unobservable potential sources of bias. However, this outcome depends critically on the identification of a valid and plausible instrument, which is controversial because there is no definitive test of the instrument's lack of association with the health outcome and, if the instrument is not strongly associated with the treatment received, the estimate of the treatment effect will be highly imprecise. Thus, it is difficult to distinguish between a true lack of statistical significance between treatment outcomes and an imprecise statistical estimate from a weak instrument. The differences in the estimated hazard rates between the two instrumental variable models that we used illustrate this concern.

One advantage of using SEER-Medicare data for comparative effectiveness studies of alternative cancer treatments is that the lagged treatment pattern in the local geographic area, which we

used in this study, is a potentially readily available choice for an instrumental variable, as long as there is sufficient variation across small geographic areas and there are enough patients in each treatment group to generate reasonably stable local area estimates. Similar treatment propensity measures can be constructed for other cancers. An important innovation in this study was that the instrumental variable was defined as the difference between the actual and predicted treatment proportions in the geographic area because the underlying characteristics of patients are not likely to be similar across geographic areas.

Findings from this study have important ramifications for physicians who rely on the medical literature to counsel newly diagnosed patients with localized prostate cancer regarding treatment and also patients who learn of newly published findings on the comparative effectiveness of various prostate cancer treatments in the popular press. Given the difficulties in successfully conducting randomized trials of prostate cancer treatments, observational data may form the preponderance of evidence that treating physicians will rely on to guide their discussions with patients. Many practicing physicians may not have the time or expertise to evaluate the biases inherent in observational reports published in academic journals. Thus, when observational data analyses are published without the appropriate methodology to account for observed and unobserved sources of bias, treating physicians may ascribe inappropriate validity to their findings when advising patients about treatment choice.

A recent study (13) and accompanying editorial (2) underscore the very real nature of this problem. Using a propensity score methodology, Wong et al. (13) found, as we did in this analysis, that aggressive management (either surgery or radiation) was associated with better survival than conservative management among older patients with prostate cancer. The accompanying editorial (2) noted that the findings of the study seemed counter to clinical intuition and that perhaps there was inadequate risk adjustment. Despite this concern, the article received substantial coverage in the lay press (48,49).

Sensitivity analyses (50,51) can reassure clinicians that the results are robust to alternative assumptions about the presence of a hypothetical confounder. For example, Wong et al. (13) showed that relative to their primary result that active treatment was associated with statistically significant better survival than conservative management (HR for mortality = 0.69, 95% CI = 0.66 to 0.72), the effect of an omitted confounder would have to be large to generate a result of no difference in mortality. However, such a finding does not mean that there were no unobserved confounders or that actual treatment decisions might not be influenced by multiple unobserved factors, which alone might make a small contribution but in combination might influence treatment decisions in a systematic way. Moreover, our analysis indicated that unobserved confounding may in fact be large because we found that the propensity score–adjusted survival (HR range = 1.46–1.56) was higher than the instrumental variable estimate (HR = 1.09, 95% CI = 0.46 to 2.59) and the estimate from the randomized controlled trial (RR = 1.04, 95% CI = 0.77 to 1.40) (Table 3).

For the sake of patients and health-care providers who use study results to make life-changing decisions, researchers need to use multiple methods of risk adjustment, such as propensity score reweighting and instrumental variable analysis, to confirm that the results are not sensitive to the method of risk adjustment. If differences are noted, the clinical plausibility and statistical validity of the various approaches should be reexamined and results should be considered with appropriate caution. Patient selection into specific treatments on the basis of factors related to prognosis is an important consideration in all observational studies, but particularly in studies involving cancer in which the incidence is highest in the elderly who are also most likely to have multiple comorbidities.

As we have noted, instrumental variable analysis does not guarantee that all observational data bias is eliminated. The variable(s) selected as the instrument should have a statistically significant association ith treatment choice but not with the health outcome or with unobserved factors that influence the health outcome. Although there are guidelines for assessing whether an instrument is a strong predictor of the treatment received, there is no definitive test for an instrument's validity. If the instrument is weak, the extent of bias in the instrumental variable estimate may be greater than in the unadjusted observational data. For example, our alternative instrumental variable analysis that used both the lagged area treatment and measures of local area medical resources, which were not strongly related to the treatment received, resulted in much better all-cause survival (HR = 0.71, 95% CI = 0.31 to 1.59). Although this estimate had a large confidence interval and the hypothesis of no difference in survival could not be rejected, its low point value could lead some to conclude that conservative management was associated with better survival than radical prostatectomy in this population of men between the ages of 66 and 74 years. Although use of multiple instruments may increase the ability to explain the treatment received, it can also increase the likelihood of an association between the instrumental variable and the health outcome. It is also important to recognize that the result of the instrumental variable analysis is limited primarily to the population on the treatment margin (ie, men who do not have strong indications that would favor one treatment approach over the other). Given these caveats, however, if a conceptually plausible instrument that has a strong and statistically significant association with the treatment can be found, instrumental variable analysis should provide a potentially important alternative and complementary methodology to propensity score methods for assessing treatment outcomes without having to make any a priori assumptions about the potential magnitude of unobserved confounders.

There were several limitations in our study. First, the benchmark Scandinavian randomized controlled trial (28,29) that we used to assess the alternative statistical methods of adjusting for observational data bias is only one study and is not representative of all elderly prostate patients in the United States. It was also limited to the comparison of radical prostatectomy and conservative management, excluding men who were treated by radiation therapy. Although we selected our study sample of patients from the SEER database to be as similar as possible to patients in the benchmark clinical trial, we were not able to include prostate cancer patients who were younger than 66 years, a group that represented approximately 46% of the trial population (28,29). The effectiveness of treatment varied in the two age groups (ie, <65 and ≥65 years), and we compared our findings with the population aged 65 years or older.

Consequently, our analysis should be viewed primarily from a methodological perspective rather than as an analysis with direct implications for clinical treatment. Second, our sample was restricted to the approximately 85% of Medicare enrollees in fee-for-service plans. Prostate cancer stage at diagnosis has been reported to be similar in Medicare managed care and fee-for-service settings; however, among patients with clinically localized disease, treatment varies by setting (8). Patients with early-stage prostate cancer in managed care were less likely to receive radical prostatectomy and more likely to receive radiation or conservative management than similar patients in fee-for-service settings (8). Third, enrollment in Medicare managed care changed during the study period, which may also limit the generalizability of our findings to the managed care population. Fourth, we did not have information about prostate-specific antigen screening before diagnosis. Use of prostate-specific antigen screening increased dramatically over the period of our study (52–54), and the number of men diagnosed with early-stage prostate cancer increased accordingly (55). Although we included year of diagnosis in our models, we could not identify which prostate cancer patients were diagnosed because of elevated prostate-specific antigen levels and which were diagnosed because of symptoms. Lastly, a complete statistical assessment of the Cox hazard model's proportionality assumption indicated that the effects of some covariates may not be time invariant, especially in the analysis of all-cause mortality. Although a sensitivity analysis of the effects of allowing time-varying covariates did not alter the principle findings with regard to treatment effects, further analysis of time-varying effects may be warranted.

In summary, survival after radical prostatectomy or conservative management in elderly patients with early-stage prostate cancer as calculated by instrumental variable estimation of Cox proportional hazard models from observational data was similar to that calculated by Cox proportional hazard models from clinical trial data. Consequently, instrumental variable analysis may be a useful technique in comparative effectiveness studies of prostate cancer and other cancer treatments if an acceptable instrument can be identified. Future research is warranted to evaluate additional methods of addressing confounding in observational data and to compare results from these methods with those from benchmark randomized controlled trials.

Appendix Table 1. Definitions of radical prostatectomy and conservative management from Medicare claims and Surveillance, Epidemiology, and End Results (SEER) registry data*

Treatment		Medicare claims	SEER
Radical prostatectomy†	Exclusion	*ICD-9* procedure codes 60.2-60.6 *CPT-4* codes 55840, 55845, 55810, 55815	Surgery codes 30, 50, 80
Conservative management‡	Radiation	*ICD-9* diagnosis codes V58.0, V66.1, V67.1 *ICD-9* procedure codes 92.21-92.29 *CPT-4* codes 77261-77299, 77300, 77305, 77401-77499, 77750-77799 Revenue center codes 0330 or 0333	Radiation codes 1-5, 7
	Chemotherapy	*ICD-9* diagnosis code V58.1 *ICD-9* procedure code 99.25 *CPT-4* 96400-96549, 95990, 95991, 96530, G0355, G0357-G0359, J0640, J2405, J8520-J8999, J9000-J9164, J9166-J9201, J9203-J9216, J9220-J9999, K0415, K0416, Q0083-Q0085, Q0179, S0177, S0181; Revenue center codes 0331, 0332, or 0335	Chemotherapy codes 01, 02, 03
	Hormonal therapy	J1950, J9217-J9219, J9202, J9165	Hormone therapy 01
	Prostate cancer–directed surgery	*ICD-9* procedure codes 60.2–60.6 *CPT-4* codes 55840, 55845, 55810, 55815	Surgery codes 30, 50, 80

* *ICD-9 = International Classification of Diseases, Ninth Edition*; *CPT-4 = Current Procedural Terminology, Fourth Edition.*

† Radical prostatectomy was defined as receiving radical prostatectomy within 6 months of diagnosis.

‡ Conservative management was defined as not receiving any prostate cancer–directed surgery, radiation therapy, chemotherapy, or hormonal therapy within 6 months of diagnosis.

References

1. Giordano SH, Kuo Y, Duan Z, Hortobagyi GN, Freeman J, Goodwin JS. Limits of observational data in determining outcomes from cancer therapy. *Cancer.* 2008;112(11):2456–2466.
2. Litwin MS, Miller DC. Treating older men with prostate cancer: survival [or selection] of the fittest? *JAMA.* 2006;296(22):2733–2734.
3. Kleinbaum DG. *Epidemiologic Research: Principles and Quantitative Methods.* Belmont, CA: Lifetime Learning Publications; 1982.
4. Gordis L. *Epidemiology* [Internet]. 4th ed. Amsterdam, the Netherlands: Elsevier; 2008. http://www.us.elsevierhealth.com/product.jsp?isbn=9781416040026#description. Accessed January 20, 2010.
5. Lu-Yao GL, Albertsen PC, Moore DF, et al. Survival following primary androgen deprivation therapy among men with localized prostate cancer. *JAMA.* 2008;300(2):173–181.
6. Lu-Yao GL, Albertsen PC, Moore DF, et al. Outcomes of localized prostate cancer following conservative management. *JAMA.* 2009;302(11):1202–1209.

7. Lu-Yao G, Stukel TA, Yao S. Changing patterns in competing causes of death in men with prostate cancer: a population based study. *J Urol.* 2004;171(6, pt 1):2285–2290.
8. Riley GF, Warren JL, Potosky AL, Klabunde CN, Harlan LC, Osswald MB. Comparison of cancer diagnosis and treatment in Medicare fee-for-service and managed care plans. *Med Care.* 2008;46(10):1108–1115.
9. Saigal CS, Gore JL, Krupski TL, Hanley J, Schonlau M, Litwin MS. Androgen deprivation therapy increases cardiovascular morbidity in men with prostate cancer. *Cancer.* 2007;110(7):1493–1500.
10. Institute of Medicine. *Initial National Priorities for Comparative Effectiveness Research—Institute of Medicine* [Internet]. Washington, DC: Institute of Medicine; 2009.http://www.iom.edu/en/Reports/2009/ComparativeEffectivenessResearchPriorities.aspx. Accessed November 5, 2009.
11. Albertsen PC, Hanley JA, Penson DF, Barrows G, Fine J. 13-year outcomes following treatment for clinically localized prostate cancer in a population based cohort. *J Urol.* 2007;177(3):932–936.
12. Tewari A, Raman JD, Chang P, Rao S, Divine G, Menon M. Long-term survival probability in men with clinically localized prostate cancer treated either conservatively or with definitive treatment (radiotherapy or radical prostatectomy). *Urology.* 2006;68(6):1268–1274.
13. Wong Y, Mitra N, Hudes G, et al. Survival associated with treatment vs observation of localized prostate cancer in elderly men. *JAMA.* 2006;296(22):2683–2693.
14. Joseph P Newhouse, Mark McClellan. Econometrics in outcomes research: the use of instrumental variables. *Ann Rev Pub Hlth.* 1998;19(1):17–34.
15. McClellan MB, Newhouse JP. Overview of the special supplement issue. *Hlth Srvcs Rsch.* 2000;35(5 pt 2):1061–1069.
16. Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables, I: instrumental variables exploit natural variation in nonexperimental data to estimate causal relationships. *J Clin Epidemiol.* 2009;62(12):1226–1232.
17. Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables, II: instrumental variable application—in 25 variations, the physician prescribing preference generally was strong and reduced covariate imbalance. *J Clin Epidemiol.* 2009;62(12):1233–1241.
18. Brooks JM, Chrischilles EA, Scott SD, Chen-Hardee SS. Was breast conserving surgery underutilized for early stage breast cancer? Instrumental variables evidence for stage II patients from Iowa. *Health Serv Res.* 2003;38(6, pt 1):1385–1402.
19. Hadley J, Polsky D, Mandelblatt JS, et al. An exploratory instrumental variable analysis of the outcomes of localized breast cancer treatments in a Medicare population. *Health Econ.* 2003;12(3):171–186.
20. Basu A, Heckman JJ, Navarro-Lozano S, Urzua S. Use of instrumental variables in the presence of heterogeneity and self-selection: an application to treatments of breast cancer patients. *Health Econ.* 2007;16(11):1133–1157.
21. Earle CC, Tsai JS, Gelber RD, Weinstein MC, Neumann PJ, Weeks JC. Effectiveness of chemotherapy for advanced lung cancer in the elderly: instrumental variable and propensity analysis. *J Clin Oncol.* 2001;19(4):1064–1070.
22. Zeliadt SB, Potosky AL, Penson DF, Etzioni R. Survival benefit associated with adjuvant androgen deprivation therapy combined with radiotherapy for high- and low-risk patients with nonmetastatic prostate cancer. *Int J Radiat Oncol Biol Phys.* 2006;66(2):395–402.
23. Wilt TJ, MacDonald R, Rutks I, Shamliyan TA, Taylor BC, Kane RL. Systematic review: comparative effectiveness and harms of treatments for clinically localized prostate cancer. *Ann Intern Med.* 2008;148(6):435–448.
24. Harlan L, Brawley O, Pommerenke F, Wali P, Kramer B. Geographic, age, and racial variation in the treatment of local/regional carcinoma of the prostate. *J Clin Oncol.* 1995;13(1):93–100.
25. Harlan LC, Potosky A, Gilliland FD, et al. Factors associated with initial therapy for clinically localized prostate cancer: prostate cancer outcomes study. *J Natl Cancer Inst.* 2001;93(24):1864–1871.
26. Klabunde CN, Legler JM, Warren JL, Baldwin L, Schrag D. A refined comorbidity measurement algorithm for claims-based studies of breast, prostate, colorectal, and lung cancer patients. *Ann Epidemiol.* 2007;17(8):584–590.
27. Klabunde CN, Harlan LC, Warren JL. Data sources for measuring comorbidity: a comparison of hospital records and Medicare claims for cancer patients. *Med Care.* 2006;44(10):921–928.
28. Bill-Axelson A, Holmberg L, Ruutu M, et al. Radical prostatectomy versus watchful waiting in early prostate cancer. *N Engl J Med.* 2005;352(19):1977–1984.
29. Bill-Axelson A, Holmberg L, Filen F. for the Scandinavian Prostate Cancer Group Study Number 4. Radical prostatectomy versus watchful waiting in localized prostate cancer: the Scandinavian Prostate Cancer Group-4 Randomized Trial. *J Natl Cancer Inst.* 2008;100(16):1144–1154.
30. Warren JL, Klabunde CN, Schrag D, Bach PB, Riley GF. Overview of the SEER-Medicare data: content, research applications, and generalizability to the United States elderly population. *Med Care.* 2002;40(suppl 8)IV-3–18.
31. Cochran WG. The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics.* 1968;24(2):295–313.
32. Rosenbaum P, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika.* 1983;70(1):41–55.
33. Rosenbaum P, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *J Am Stat Assoc.* 1984;79(387):516–524.
34. Kurth T, Walker AM, Glynn RJ, et al. Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. *Am J Epidemiol.* 2006;163(3):262–270.
35. Curtis LH, Hammill BGM, Eisenstein ELD, Kramer JM, Anstrom KJ. Using inverse probability-weighted estimators in comparative effectiveness analyses with observational databases. *Med Care.* 2007;45(10):S103–S107.
36. Bound J, Jaeger DA, Baker RM. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *J Am Stat Assoc.* 1995;90(430):443.
37. Staiger D, Stock JH. Instrumental variables regression with weak instruments. *Econometrica.* 1997;65(3):557–586.
38. Terza JV, Basu A, Rathouz PJ. Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. *J Health Econ.* 2008;27(3):531–543.
39. Stukel TA, Fisher ES, Wennberg DE, Alter DA, Gottlieb DJ, Vermeulen MJ. Analysis of observational studies in the presence of treatment selection bias: effects of invasive cardiac management on AMI survival using propensity score and instrumental variable methods. *JAMA.* 2007;297(3):278–285.
40. Brookhart MA, Rassen JA, Wang PS, Dormuth C, Mogun H, Schneeweiss S. Evaluating the validity of an instrumental variable study of neuroleptics: can between-physician differences in prescribing patterns be used to estimate treatment effects? *Med Care.* 2007;45(10 suppl 2):S116–S122.
41. Brooks JM, Chrischilles EA. Heterogeneity and the interpretation of treatment effect estimates from risk adjustment and instrumental variable methods. *Med Care.* 2007;45(10 suppl 2):S123–S130.
42. Yoo B, Frick KD. The instrumental variable method to study self-selection mechanism: a case of influenza vaccination. *Value Health.* 2006;9(2):114–122.
43. Dartmouth Medical School. *The Dartmouth Atlas of Health Care, 1998.* Hanover NH: Dartmouth Medical School; 1998.
44. United States. *The Area Resource File (ARF) System: Information for Health Resources Planning and Research.*Rockville, MD: The Bureau; 1989.
45. Shah BR, Laupacis A, Hux JE, Austin PC. Propensity score methods gave similar results to traditional regression modeling in observational studies: a systematic review. *J Clin Epidemiol.* 2005;58(6):550–559.
46. Stürmer T, Joshi M, Glynn RJ, Avorn J, Rothman KJ, Schneeweiss S. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. *J Clin Epidemiol.* 2006;59(5):437.e1–437.e24.

47. Leslie RS, Ghomrawi H. *The Use of Propensity Scores and Instrumental Variable Methods to Adjust for Treatment Selection Bias* [Internet]. SAS Global Forum. Cary, NC: SAS Institute; 2008. http://www2.sas.com/proceedings/forum2008/366-2008.pdf. Accessed December 16, 2009.
48. MedHelp. *Survival Without Treatment—Gleason Score 7. Prostate Cancer* [Internet]. MedHelp. 2008. http://www.medhelp.org/posts/Prostate-Cancer/Survival-without-Treatment--Gleason-Score-7/show/590910. Accessed January 13, 2010.
49. Mallet K, *Treatment For Prostate Cancer Helps Older Men Live Longer, Versus Observation* [Internet]. Medical News Today. 2006. http://www.medicalnewstoday.com/articles/38408.php. Accessed January 13, 2010.
50. Lin DY, Psaty BM, Kronmal RA. Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics*. 1998;54(3):948–963.
51. Mitra N, Heitjan D. Sensitivity of the hazard ratio to non-ignorable treatment assignment in an observational study. *Stat Med*. 2007;26(6):1398–1414.
52. Farwell WR, Linder JA, Jha AK. Trends in prostate-specific antigen testing from 1995 through 2004. *Arch Intern Med*. 2007;167(22):2497–2502.
53. Mariotto AB, Etzioni R, Krapcho M, Feuer EJ. Reconstructing PSA testing patterns between black and white men in the US from Medicare claims and the National Health Interview Survey. *Cancer*. 2007;109(9):1877–1886.
54. Ross LE, Berkowitz Z, Ekwueme DU. Use of the prostate-specific antigen test among U.S. men: findings from the 2005 National Health Interview Survey. *Cancer Epidemiol Biomarkers Prev*. 2008;17(3):636–644.
55. American Cancer Society. *Cancer Facts and Figures 2009* [Internet]. Atlanta, GA: American Cancer Society; 2009. http://www.cancer.org/acs/groups/content/@nho/documents/document/500809webpdf.pdf. Accessed September 27, 2010.

Funding

National Cancer Institute (HHSN2612007003 39P).

Notes

Authors had full responsibility in the design of the study; the collection, the analysis, and interpretation of the data; the decision to submit the article for publication; and the writing of the article.

Affiliations of authors: Office of the Dean and Department of Health Administration and Policy, College of Health and Human Services, George Mason University, Fairfax, VA (JH); Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD (KRY); Information Management Services, Inc, Rockville, MD (MJB); Department of Urologic Surgery, Center for Surgical Quality and Outcomes Research, Vanderbilt University, Nashville, TN (DFP); Department of Urology, David Geffen School of Medicine at UCLA, Los Angeles, CA (CSS); Department of Oncology, Lombardi Cancer Center, Georgetown University, Washington, DC (ALP).