The Thesis Committee for Xiao Fang Certifies that this is the approved version of the following dissertation:

Comparing Different Methods of Applying Propensity Score in Controlling Selection Bias for Studies of Prostate Cancer

Committee:

Yong-Fang Kuo, PhD

Alai Tan, MD, PhD

Kristin M. Sheffield, PhD

Dean, Graduate School

Comparing Different Methods of Applying Propensity Score in Controlling Selection Bias for Studies of Prostate Cancer

by

Xiao Fang

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas Medical Branch

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science

The University of Texas Medical Branch

March, 2014

Acknowledgements

I would especially like to thank Dr. Yong-Fang Kuo, my advisor and committee chair, for her guidance, rigorous standards and boundless support during my MS years and the completion of this project.

Besides my mentor, I would like to give gratefulness to my entire committee as well: Dr. Alai Tan and Dr. Kristin M. Sheffield, for their insightful comments and hard questions.

My sincere thanks also go to the editor, Dr. Marinel M Ammenheuser, for her countless help editing my thesis.

I would like to extend my sincere appreciation to the faculty and staff of the department of Preventive Medicine and Community Health for all their teaching and assistance during my study at the University of Texas Medical Branch.

Deepest gratitude to my dear husband, Yucheng Zhao.

Finally, to my parents, Qing Fang and Hong Wang.

Comparing Different Methods of Applying Propensity Score in Controlling Selection Bias for Studies of Prostate Cancer

Publication No._____

Xiao Fang, B.S.

The University of Texas Medical Branch, 2014

Supervisor: Yong-Fang Kuo

Background: Prostate cancer is the most common cancer among men. There is growing interest in population-based comparative effectiveness research on different therapies for prostate cancer because of the limitations on generalizability of randomized controlled trials. The major challenge for population-based studies using observational data is how to control for selection bias. The objective of this study was to compare the performance of four common methods of applying propensity scores—covariate adjustment, stratification, matching, and inverse probability of treatment weighting—to adjust for confounding in analyses of treatment effects among prostate cancer patients.

Methods: The linked Surveillance, Epidemiology, and End Results (SEER) Medicare database 1992-2007 was used for this research in which we studied two scenarios. In scenario 1, the overall mortality and cause-specific mortality for patients with local prostate cancer receiving active treatment (radical prostatectomy or radiation) vs. observation were compared. The known confounding factor, comorbidity, was removed in the analyses to evaluate the selection bias for general health. In scenario 2, we compared the overall mortality and cause-specific mortality between patients with and without primary androgen deprivation therapy. The known confounding factor, cancer grade, was removed in the analyses to evaluate the selection bias for cancer severity.

Results: Among four propensity methods, matching (greedy matching and caliper matching) produced the estimates that were closest to the estimates when the removed confounder was included in both scenarios. Covariate adjustment and stratification yielded similar controlling effects in both scenarios. Inverse probability of treatment weighting showed a better performance in scenario 1 compared to scenario 2.

Conclusions: Propensity score matching outperformed covariates adjustment, stratification, and inverse probability of treatment weighting. None of the propensity methods produced estimates that were identical to the estimates when the removed confounders, comorbidity or cancer grade, were included.

TABLE OF CONTENTS

List of Tables

List of Figures

List of Abbreviations

IPTW	Inverse probability of treatment weighting
SEER	Surveillance, epidemiology, and end results
PSA	Prostate-specific antigen
ADT	Androgen deprivation therapy
CI	Confidence interval
HR	Hazard ratio
t-PA	Tissue plasminogen activator
HIPPA	Health insurance portability and accountability act
NCI	National cancer institute
SCoA	Sealy center on aging

Chapter 1 Introduction

In the United States, prostate cancer is the most common cancer among men, excluding basal cell and squamous cancer of the skin. There were an estimated 241,740 new cases and 28,170 deaths from prostate cancer in 2012 (American Cancer Society, 2012). How to select the best treatment for prostate cancer is one of the most controversial areas for consideration by both physicians and patients. Because of the limitations of randomized controlled trials, such as lack of external validity, high cost, and the fact that they are time consuming, there is a growing interest for investigators to use observational data to compare the outcomes of different cancer therapies. However, unlike randomized controlled trials, the most serious problem with observational studies is the existence of selection bias. For example, when comparing different therapies, patients with poorer cancer prognoses may be expected to receive more aggressive therapy, which can make therapy appear to cause negative outcomes. How to remove treatment selection bias is the major challenge for observational studies.

In the past, regression adjustment was used by statisticians to adjust for the difference in baseline covariates between treatment and no treatment groups. Stratification was also recommended by Cochran (Cochran, 1968) to account for bias in the study. Lately, there has been growing interest in employing statistical methods (e.g. propensity score methods) to control the influence of confounding in observation studies (Becerril & Abdulai, 2010; Monahan, Lee, & Steinberg, 2011; Schneeweiss et al., 2009).

Propensity score methods are often used in observational studies to control for selection bias. Peter C. Austin published analyses of two clinical studies, including both Monte Carlo simulations and case studies, which showed that propensity score methods can reduce the bias in estimating marginal hazard ratios and relative risks (Austin, 2008, 2013). Propensity score refers to the probability of a patient receiving a treatment based on observed baseline covariates. There are four common methods of applying the

propensity score. These are: covariate adjustment, stratification, matching, and inverse probability of treatment weighting (IPTW) (Rosenbaum & Rubin, 1983, 1984). The present study was aimed to assess the performance of these four methods in adjusting for confounding in the analyses of treatment effects on outcomes among prostate cancer patients using the Surveillance, Epidemiology, and End Results (SEER) - Medicare data. The performance of each method was evaluated based on its ability to account for the removal of known confounder(s) from the model. Based on previous work on simulation (Austin, 2009, 2011), we hypothesized that propensity score matching will yield results with the lowest bias among all the methods.

Chapter 2 Literature Review

PROSTATE CANCER

Prostate cancer is a disease in which malignant cells form in the tissues of the prostate. The causes of prostate cancer have been investigated in several case-control and prospective cohort studies (Clinton & Giovannucci, 1998; Chan, 1998). It is usually found among elderly men (Boyle, Severi, & Giles, 2003). The incidence rates of prostate cancer vary greatly by race/ethnicity. In Asian countries, about 2 to 10 prostate cancer cases exist in 10,000 persons, but the number in the United States was 186 in 10,000 persons in 2013. About 6 cases in 10 are diagnosed in men aged 65 or older in the United State (American Cancer Society, 2013). African Americans have the highest incidence rate, 60% higher than whites (SEER, 2013). Some studies suggest that this difference is mostly due to genetic factors (Haas & Sakr, 1997). During the 1990s, some prospective studies also suggested that specific fatty acids, antioxidant vitamins, and vitamin D alter prostate cancer risk (Boyle, Severi, & Giles, 2003). In the United States, prostate cancer was the most common cancer in 2012, excluding basal and squamous cell cancer of the skin. The estimated new cases were 241,740, and the deaths from prostate cancer were 28,170 (American Cancer Society, 2012).

With the widespread adoption of prostate-specific antigen (PSA) screening, there are growing numbers of early stage and low- or intermediate-grade prostate cancer cases (Albertsen, Hanley, & Fine, 2005; Albertsen et al., 1998; Chodak et al., 1994). Treatments for early stage disease usually are either watchful waiting or active treatments, which include radical prostatectomy with or without radiation therapy after surgery (Wong et al., 2006). Many factors influence the choice of an optimal management strategy for early stage prostate cancer. On the one hand, the speed of tumor growth is significantly different between cases. Patients need to consider both survival

consequences and acute morbidity. On the other hand, the median age at diagnosis is 72, and many patients will die of other comorbidities rather than prostate cancer. The relative 15-year survival rate is about 78% (Thompson et al., 2013). We need also to think about the treatment effects on quality of life (Sanda & Kaplan, 2009). Controversy exists regarding how to select the best treatment (Kollmeier & Zelefsky, 2012). Beside this, another procedure, androgen deprivation therapy (ADT), is also well-established in treating patients with metastatic prostate cancer. ADT is used to treat patients with increasing PSA levels after local treatment (Sharifi, Gulley, & Dahut, 2005). The goal of ADT is to reduce androgen hormone levels, which limits the growth of prostate cancer cells (Perlmutter & Lepor, 2007).

Nonetheless, the adverse effects of treatments need to be considered. First, there are acute adverse events associated with radical prostatectomy, which include bleeding, infection, and urinary retention (Sanda et al., 2008). Patients also need inpatient hospitalization because radical prostatectomy is major surgery (Sanda & Kaplan, 2009). Second, toxic reactions may occur, either right after radiation therapy or after many years. Acute toxic effects include urinary frequency, urgency or dysuria (Potosky et al., 2000; Sanda et al., 2008). Third, several controlled trials have shown that ADT is associated with increased risk of cardiovascular disease, fractures, and low bone density (Shahinian et al., 2005; Levine et al., 2010; Sharifi et al., 2005).

RANDOMIZED CONTROLLED TRIALS

In the face of the burden of localized prostate cancer and the challenges of the cost and implementation of treatment, limited numbers of randomized controlled trials have been conducted to give adequate evidence for the selection of treatment for early stage prostate cancer. Two major randomized controlled trials have been published,

which addressed whether radical prostatectomy can provide a mortality benefit for prostate cancer patients.

The first published trial was a randomized study, done in Sweden, of radical prostatectomy versus watchful waiting in 695 men with a median of 12.8 years follow-up (Bill-Axelson et al., 2005; 2011). Three hundred forty-seven men were assigned to a radical-prostatectomy group, and 55 of them died due to prostate cancer. Three hundred forty-eight men were assigned to a watchful-waiting group, and 81 of them died from prostate cancer. The cumulative incidence of death from prostate cancer at 15 years was 14.6% and 20.7% in the two groups. The difference was 6.1% with a 95% confidence interval (CI) of 0.2% - 12.0%, and a relative risk with surgery of 0.62 (95% CI, 0.44 to 0.87; P=0.01). This study showed that radical prostatectomy was associated with a reduction in the rate of death from prostate cancer.

However, the Veterans Affairs Prostate Cancer Intervention verses Observation trial also compared radical prostatectomy and observation in 731 men (Wilt et al., 2012). During the median follow-up of 10.0 years, 171 of 364 men (47%) assigned to radical prostatectomy died, as compared with 183 of 367 (49.9%) assigned to observation. The hazard ratio (HR) was 0.88 with a 95% CI of 0.71- 1.08 (p=0.22). This trial showed that, compared with observation, radical prostatectomy did not significantly reduce prostate-cancer mortality.

A European study group has conducted a randomized controlled trial to assess the overall survival in patients with localized prostate cancer with immediate or deferred primary ADT therapy (Studer et al., 2006). This trial randomly assigned 985 patients with newly diagnosed prostate cancer to receive ADT therapy immediately (N=493) or on symptomatic disease progression or occurrence of serious complications (N=492) with a median follow-up of 7.8 years. The results showed that the overall survival hazard ratio was 1.21 (95% CI, 1.05 to 1.39; p=.0085) favoring immediate treatment. It demonstrated

that immediate ADT resulted in a modest but statistically significant increase in overall survival but no significant difference in prostate cancer mortality.

Warde et al. conducted a randomized controlled trial to compare the overall survival of patients with locally advanced prostate cancer with adjuvant ADT alone or ADT with radiation therapy (Warde et al., 2011). After a median 6.0 years of follow-up, 175 of 602 patients in the ADT only group and 145 of 603 patients in the ADT and radiation therapy group died during the trial. The addition of radiation therapy to ADT improved overall survival at 7 years (74%, 95% CI, 70% to 78% vs. 66%, 95% CI, 60% to 70%; hazard ratio, 0·77, 95% CI, 0.61 to 0.98; p=0.033). This result indicated that the benefits of combined modality treatment (ADT and radiation therapy) should be discussed by patients with their physicians.

RANDOMIZED CONTROLLED TRIALS VS. OBSERVATIONAL STUDIES

In the evaluation of treatments, randomized controlled trials are the gold standard for determining the efficacy of outcomes, but randomized controlled trials have some disadvantages. First of all, they lack external validity because of highly selected subjects (Rothwell, 2005). All subjects need to meet strict eligibility criteria. These trials also can be expensive and time consuming (Johnston et al., 2006). In the evaluation of prostate cancer treatment, it is difficult to perform randomized controlled trials because of the large sample size required and the longer study duration (Silverman, 2009). Considering the limitations of randomized controlled trials, there is a growing interest among investigators to use observational data to study the comparative outcomes of different cancer therapies.

TYPES OF SELECTION BIAS

Unlike randomized controlled trials, the most serious problem with observational data analysis is the existence of selection bias. Selection bias is caused when there is a lack of comparability between groups that are studied, and there are two main sources that can lead to selection bias: general health and tumor prognosis (Giordano et al., 2008). For example, a common type of selection bias arises from physicians' propensity to prescribe medication and perform procedures. When comparing different therapies, patients with poorer cancer prognoses may be prescribed to receive more aggressive therapy, which can make therapy appear to cause negative outcomes (Walker, 1996). Moreover, when studying the comparative toxicity of two treatments, patients with better general health may be able to choose more aggressive, but also more toxic, treatments. Also, patients who exercise on a regular basis or avoid unhealthy life style may associate with a reduced risk of some health outcomes, all of which may have some direct or indirect effects on treatment effect estimates (Brookhart et al., 2007, 2010).

PROPENSITY SCORE

Propensity score methods are often used to control for selection bias in observational studies. A description of this method was first published by Rosenbaum and Rubin in 1983 (Rosenbaum & Rubin, 1983). The propensity score is the probability of a patient receiving treatment based on observed baseline covariates. Treated and untreated subjects with the same propensity score will have the same multivariate distributions of measured baseline covariates. The derived propensity score, including overlap, balance, and ability to pull across deciles, needs to be examined first before applying the propensity score in controlling for selection bias (Austin, 2011) .

The propensity score methods assume that (a) there are no hidden confounders that will affect treatment assignment and (b) probability of receiving either treatment is

between 0 and 1. The first assumption addresses that the type of treatment or the mechanism of assignment is conditional independent for the outcome giving measured covariates. Usually, there are four different methods for applying propensity scores: covariate adjustment, stratification, matching, and inverse probability of treatment weighting (Rosenbaum & Rubin, 1983, 1984).

The first standard method for applying propensity scores in observational studies is covariate adjustment. In this model, the propensity score is used as a covariate. The outcome variable is regressed on an indicator variable representing treatment status, covariates, and the estimated propensity score. The second method is stratification on the propensity score. For this method, the subjects are stratified into subgroups based on their estimated propensity score (Rosenbaum & Rubin, 1984). Based on the research of Cochran (Cochran, 1968), the use of five subclasses is sufficient to remove at least 90% of bias for many continuous distributions. Usually, the stratum begins at the highest propensity scores. The third method is matching on the propensity score. One-to-one matching appears to be the most common method. This may eliminate approximately 99% of the selection bias due to measured confounders (Ming & Rosenbaum, 2000). We also need to sample from the potential controls to produce a control group to compare with the treated group. Each pair of treated and control subjects has similar propensity score values. The best technique to determine how close the two propensity scores are is to define a specified caliper distance. It is suggested that a caliper of width equal to 0.2 standard deviations of the logit of the propensity score will minimize the mean square error of the estimated treatment effect in several scenarios (Austin, 2011). Besides one-to-one matching, many-to-one (M: 1) matching can also be used because it can increase the power, due to a bigger sample size. During this process, multiple (M) untreated subjects are selected to match to each treated subject. This approach has shown that it can improve the bias reduction (Ming & Rosenbaum, 2000). The best way to select the matched group is using the same database because, in this way, the control group will

have a similar distribution of covariates. The last method is inverse probability of treatment weighting (IPTW). Here, the inverse of the propensity score is used to weight each subject in the treated group, and one minus the inverse of the propensity score is used for the controls. However, a weakness of the IPTW method is the possibility of extreme propensity scores that can result in very large or very small weights, which will bias the estimates. Harder et al. proposed a stabilization technique to adjust the extreme weights (Harder, Stuart, & Anthony, 2010). The big advantage of weighting is that this will include all the data and it does not depend on the randomization of sampling, resulting in greater reliability (Austin, 2011).

Several studies have been done to evaluate the effect of controlling selection bias in observational studies using different methods for applying propensity scores. One study evaluated the performance of propensity score methods for estimating relative risks. Monte Carlo simulation was applied, and the results showed that covariate adjustment could eliminate between 97.9% and 99.3% of the bias created by baseline covariates. Stratification on the quintiles eliminated between 87.5% and 99.4% of bias. However, approximately 99% of the bias can be eliminated by applying propensity score matching (Austin, 2008). The other empirical study was designed to evaluate the effect of tissue plasminogen activator (t-PA) on death following a stroke (Kurth et al., 2006). The crude odds ratio between t-PA treatment and death after ischemic stroke was 3.35 (95% CI, 0.67 to 1.84). After applying propensity score matching, the odds ratio dropped down to 1.17 (95% CI, 0.68 to 2.00). And the odds ratio was 1.53 (95% CI, 0.95 to 2.48) after adjusting for propensity score as a continuous variable. The matching results were close to the estimated odds ratio from a cumulative meta-analysis of several randomized trials, which was 1.16 (95% CI, 0.95 to 1.43) (Wardlaw, Sandercock, & Berge, 2003).

The aim of the present study is to assess the performance of the four methods in adjusting for confounding in analyses of treatment effects among prostate cancer patients. Two questions will be answered: (1) how much do results vary in use of these four

propensity score methods in controlling for selection bias and (2) what are the strengths and limitations of the four methods in controlling for selection bias related to general health versus selection bias related to tumor prognosis? This research will provide information on using propensity score methods to control for selection bias in observational studies.

Chapter 3 Methods

STUDY DESIGN

A retrospective cohort study was conducted. Four different methods for applying propensity scores were used to study two scenarios for prostate cancer. In scenario 1, the study was done in males aged between 66 and 80, diagnosed with well or moderately differentiated localized prostate cancer. It involved determining how receipt of active treatment compared to how observation alone affects the overall and cause-specific mortality in patients diagnosed with prostate cancer in 1992-2007. These patients were followed until the end of 2008. In scenario 2, the study was done in males aged 66 or above diagnosed with localized prostate cancer that did not have active treatment. It determined how receipt of primary ADT compared to no ADT treatment, affected the overall and cause-specific mortality in patients within the same study period. In each scenario the known confounders, comorbidity and grade, were removed. The performance, after using four methods for applying the propensity score, was evaluated based on its ability to account for the removal of known confounder from the model.

DATA SOURCE

The linked Surveillance, Epidemiology, and End Results (SEER) Medicare database was used for this research. The SEER database is supervised by the National Cancer Institute (NCI). This database contains routinely collect information on all newly diagnosed cancer cases in individuals residing in SEER registry areas. It also includes information on tumor histology, size, and grade. The SEER program began to collect data in January 1, 1973, and the number of registries has expanded over time. This population-based data represented approximately 14% of the US population until 2001 and 26% thereafter. All data are highly valid because all SEER registries maintain the highest level

of certification of data quality. Medicare is a primary health insurer that covers approximately 97% of US individuals aged 65 years and older. Part A of the Medicare file contains inpatient claims and part B includes outpatient and physician claims The linkage of the SEER data with Medicare data by matching a person's social security number, name, sex, and date of birth using an algorithm is accomplished by the collaborative effort of the NCI and Medicaid Services (Warren et al., 2002).

Also, the United States Census data were used to extract income and education information at zip codes as surrogate measures of patients' socio-economic status.

INCLUSION AND EXCLUSION CRITERIA OF STUDY POPULATION

In scenario 1, the study population was patients aged between 66 and 80 who were diagnosed with prostate cancer from 1992 to 2007, as recorded in the linked SEER-Medicare data. Patients who met the following conditions were excluded from the study population:

- Men diagnosed at death or after autopsy or who had Medicare entitlement based on end-stage renal disease
- Men enrolled in a managed care plan from 12 months before diagnosis to 6 months after diagnosis
- Men not enrolled in both Medicare part A and part B from 12 months before diagnosis to 6 months after diagnosis
- Men diagnosed at stage T3 to T4 and those who had poorly differentiated or unknown tumor grade
- Men treated with hormonal therapy alone
- Men with unknown income and education information

In scenario 2, the study population was patients aged 66 or above, diagnosed with prostate cancer from 1992 to 2007 as listed in the SEER- Medicare data. Patients who met the following conditions were excluded from the study population:

- Men diagnosed at death or after autopsy or who had Medicare entitlement based on end-stage renal disease
- Men enrolled in a managed care plan from 12 months before diagnosis to 6 months after diagnosis
- Men not enrolled in both Medicare part A and part B from 12 months before diagnosis to 6 months after diagnosis
- Men diagnosed at stage T3 to T4
- Men treated actively with radical prostatectomy or radiation

DATA MANAGEMENT

During this research project, all data management and analyses were conducted on computers provided by the Office of Biostatistics and the Sealy Center on Aging (SCoA) at the University of Texas Medical Branch. The SCoA maintains a complete database of SEER-Medicare files for the study cohort. Data were linked across datasets using a common identifier. These data were routinely cleaned and examined for outliers, and have served as the basis for numerous peer-reviewed publications. All data abstraction, data management, and data analyses were conducted in full compliance with HIPPA regulations and SEER-Medicare data user agreement.

MEASUREMENT

Table 1: Independent variables

Category	***Data Source***	***Definition***
Age	SEER	Age at diagnosis in years

Year	SEER	The year of diagnosis
Race	SEER	White, Black, Hispanic, Other
Marital status	SEER	Married at time of diagnosis or not
SEER site	SEER	San Francisco (CA), Connecticut, Detroit (MI), Hawaii, Iowa, New Mexico, Seattle (CA), Utah, Rural Georgia, San Jose (CA), Los Angeles (CA), Atlanta (GA), Kentucky, Louisiana, New Jersey, Greater California
Socio-economic status: income[a]	1990 and 2000 US Census	Available only at patient zip code level: % of adults living below poverty line For 1992-1995: Applying 1990 census data For 1996-2007: Applying 2000 census data
Socio-economic status: education[a]	1990 and 2000 US Census	Available only at patient zip code level: % of adults with less than 12 years of education For 1992-1995: Applying 1990 census data For 1996-2007: Applying 2000 census data
Number of comorbidities[b]	Medicare	Based on Klabunde's development of the Charlson index The Charlson index is derived from a counting system based on a count of certain comorbid diagnoses. Klabunde's modification allows identification of comorbid diagnoses from Medicare inpatient, outpatient and physician claims.
Tumor stage	SEER	For 1992-2003: Using clinical T stage from SEER Extent of Disease – Clinical Extension coding; categorized as T1, T2 For 2004-2007: Using Collaborative Staging System extension – Clinical Extension codes; categorized as T1, T2

Tumor grade[c]	SEER	For 1992-2002, 2004-2007: Grading is categorized as well differentiated (Gleason 2-4), moderately differentiated (Gleason 5-7), poorly differentiated (Gleason 8-10) or unknown. For 2003: Grading is categorized as well differentiated (Gleason 2-4), moderately differentiated (Gleason 5-6), poorly differentiated (Gleason 7-10) or unknown.
Radical Prostatectomy	SEER and Medicare	From SEER coding on site-specific surgery or any of the following from Medicare claims: Current Procedural Terminology (CPT) codes 55810, 55812, 55815, 55801, 55821, 55831, 55842, 55845; International Classification of Diseases, 9th revision (ICD-9) procedure code 60.5.
Radiation Therapy	SEER and Medicare	From SEER coding on radiation treatment or any of the following from Medicare claims: 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[d]	Medicare	Orchiectomy: Any claim with CPT codes 54520, 54521, 54522, 54530, or 54535; or ICD-9 procedure code 62.4. Gonadotropin-releasing hormone (GnRH) agonist: Any outpatient or carrier claim with the Healthcare Common Procedure Coding System (HCPCS) codes J9202 (goserelin), J1950 (leuprolide), J9217 (leuprolide), J9218 (leuprolide), J9219 (leuprolide implant) or J3315 (triptorelin).

a. Income and education information were obtained from United States Census as described above.
b. Comorbidities were collected by Medicare inpatient, outpatient and Part B claims during the 12 months before cancer diagnosis.

c. Tumor grade was recoded according to the Gleason score from 2004. Gleason code 7 has been moved from moderately differentiated to poorly differentiated since 2003. Cases diagnosed in 2003 stayed in the new grading system because of lacking the detailed Gleason score information.
d. Use of ADT was defined as the receipt of at least 1 dose of a GnRH or orchiectomy, based on previously published and validated methods (Shahinian, Kuo, & Gilbert, 2010).

Table 2: Dependent variables

Category	***Source***
Overall and cause-specific mortality (divided as prostate cancer-specific and non-prostate cancer mortality) by Dec. 31, 2008[a]	SEER

a. The outcomes of this study were the overall and cause-specific mortality obtained from the SEER database.

STATISTICAL ANALYSIS

First, descriptive statistics of age, year of diagnosis, race, marital status, SEER site, income, education, number of comorbidities and tumor stage and grade was conducted. Means and standard deviations were used to describe continuous variables. Categorical variables were examined by frequency and proportion. Data were plotted and examined for errors and outliers.

Second, the propensity score was estimated by generating a logistic regression model. The binary treatment status was regressed on all covariates in this model. The expected probability P that an event occurs for a set of given covariates x is:

$$P = \frac{\exp(\beta_1 x_{i1} + \cdots + \beta_k x_{ik})}{\exp(\beta_1 x_{i1} + \cdots + \beta_k x_{ik}) + 1}$$

The estimated propensity score was the predicted probability of receiving treatment as derived from the fitted logistic regression.

Third, survival analysis was conducted comparing the overall mortality and cause-specific mortality in patients who received active treatment compared to those who

were managed with observation and patients who received primary ADT therapy and those who didn't. A Cox proportional hazard model, in which survival time was regressed on all covariates, was built in this step. A Cox proportional hazard model is usually written as

$$h_i(t)=\lambda_0(t)\exp(\beta_1 x_{i1}+\cdots+\beta_k x_{ik})$$

The hazard for individual i at time t is a function of two factors: the baseline hazard function $\lambda_0(t)$ and a linear function of a set of k fixed covariates. The hazard for any individual is a fixed proportion of the hazard for any other. For two individuals i and j, the ratio of the hazard ratio is:

$$\frac{h_i(t)}{h_j(t)}=\exp\left\{\beta_1(x_{i1}-x_{j1})+\cdots+\beta_k(x_{ik}-x_{jk})\right\}$$

Now the function $\lambda_0(t)$ has been canceled. As the result, the ratio of the hazards is constant over time. Still, the assumption of proportionality was tested first using PROC PHREG in SAS. This assumption indicates that the survival curves for two strata must have hazard functions that are proportional over time in the regression model. The outcome of this model will be used as the reference.

Fourth, bivariate analysis was used to compare study characteristics between treatments for two scenarios. The Cochran Mantel-Haenszel chi-square test was used to analyze the balance of covariates after adjusting the propensity score.

Fifth, the known confounders, comorbidity in scenario 1 and tumor grade in scenario 2, were removed from the model. The Cox proportional hazards methods that incorporated estimated propensity scores with four methods is described, as shown in Table 3.

Table 3: Building models for applying propensity scores

Method for applying propensity score	***Building this model***
Covariate adjustment	Use the Cox proportional hazard model in which survival outcome was regressed on the estimated propensity score and the covariates that were not balanced after controlling for the propensity score.
Stratification	Rank the estimated propensity score first, and then divide individuals into five equal strata. The stratified Cox proportional hazard model was built with adjustment of unbalanced covariates.
Matching	Both greedy and caliper one-to-one matching were applied. One untreated subject was matched to each treated subject. The criterion for selecting subjects with similar propensity score was a caliper of width equal to 0.2 of standard deviation of the logit of the propensity score. The Cox proportional hazard model was built with only treatment status.
Inverse probability of treatment weighting (IPTW)	The inverse of the propensity score was used to weight each subject in the treated group, and one minus the inverse of the propensity score was used for the controls. The mean weight was calculated to normalize the IPTW weights to adjust for the extreme values. The weights were then used in a Cox proportional hazard model with other unbalanced covariates.

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

Propensity scores calculations and survival analyses were performed using SAS 9.3. Statistical significance was set at .05, and all tests were 2-tailed.

Chapter 4 RESULTS

DESCRIPTION OF SCENARIO 1: PROSTATE CANCER PATIENTS WITH ACTIVE TREATMENT (RADICAL PROSTATECTOMY OR RADIATION) VS. OBSERVATION

The cohort selection for studying the effect of active treatment on overall and cause-specific mortality for patients with localized prostate cancer is summarized in Figure 1. Of patients who met selection criteria, 86,516 (78.07%) were in the active treatment group and 24,305 (21.93%) were in the observation group.

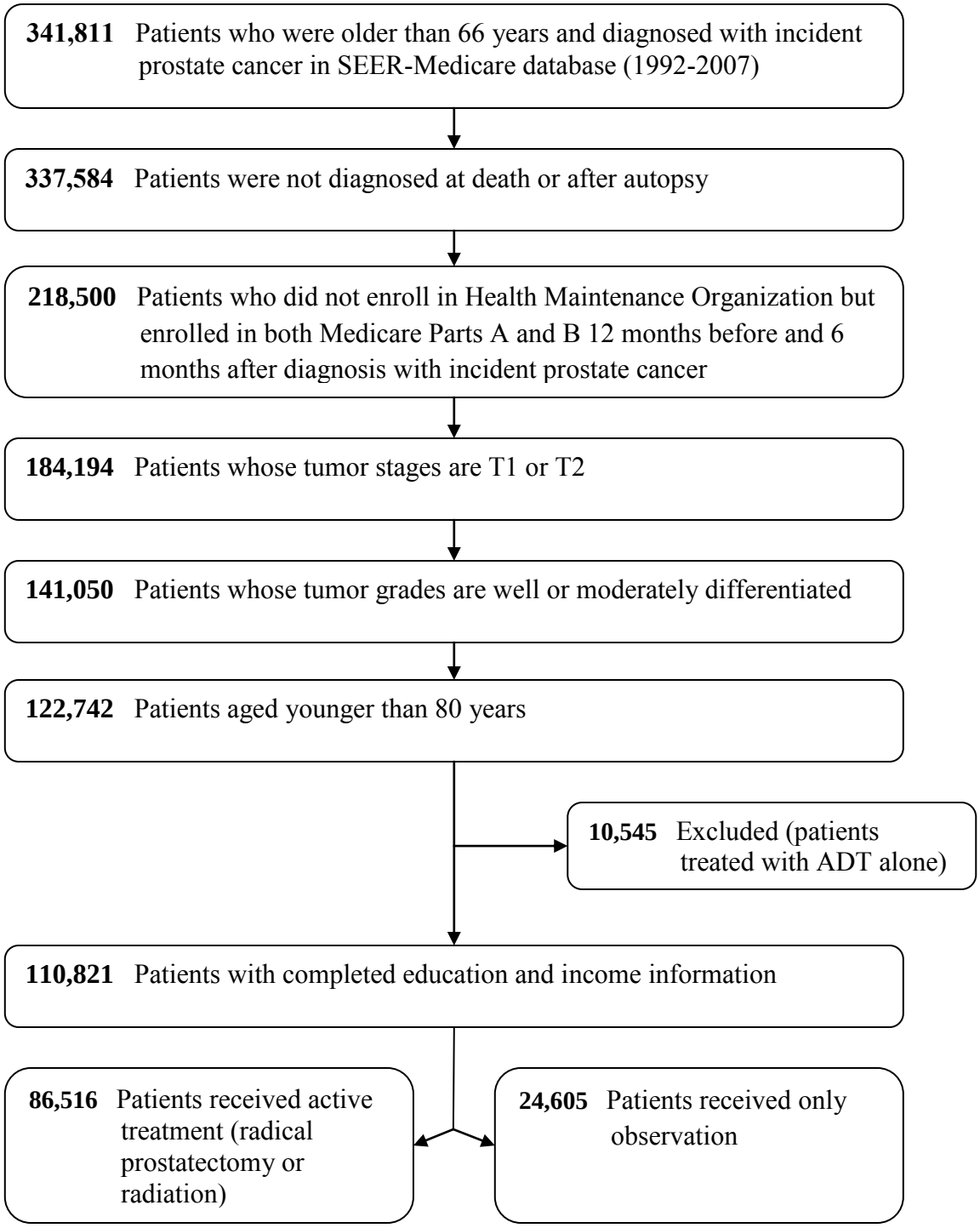

Figure 1: Study-selection summary for the overall cohort of patients with active treatment and observation

Table 4: Characteristics of patients receiving active treatment and observation

	Overall Cohort			Adjusted by Propensity Score	Applied IPTW
Characteristic	Observation N=24305	Active Treatment N=86516	P	P	P
Age(years)			<.0001	<.0001	0.7566
66-67	9.10%	17.19%			
68-70	17.06%	26.35%			
71-73	20.22%	24.51%			
74-77	31.51%	23.65%			
78-80	22.11%	8.30%			
Mean±SD	73.6±4.1	71.6±3.8			
Race			<.0001	0.2874	0.0133
White	76.42%	81.68%			
Black	11.11%	8.74%			
Hispanic	5.01%	4.69%			
Others	7.46%	4.89%			
Year of diagnosis			<.0001	0.0016	0.4964
1992-1995	22.55%	15.71%			
1996-1999	17.02%	15.20%			
2000-2003	27.93%	32.53%			
2004-2007	32.50%	36.56%			
SEER Regions			<.0001	0.4349	0.0015
San Francisco	5.67%	4.26%			
Connecticut	8.34%	8.32%			
Detroit	12.35%	12.47%			
Hawaii	1.27%	1.64%			
Iowa	7.72%	7.83%			
New Mexico	5.00%	3.31%			
Seattle	8.55%	7.83%			
Utah	5.76%	4.22%			
Atlanta / Rural Georgia	4.16%	4.48%			
San Jose	2.98%	2.56%			
Los Angeles	10.16%	8.65%			

Greater California	11.48%	12.44%			
Kentucky	4.42%	5.22%			
Louisiana	4.26%	4.68%			
New Jersey	7.86%	12.11%			
Marital Status			<.0001	0.0004	0.1150
Not Married	37.47%	23.98%			
Married	62.53%	76.02%			
Education, % of adults with < 12 years of education			<.0001	0.1627	0.3877
<8.00	22.25%	25.75%			
8.00 to 14.26	23.22%	25.47%			
14.26 to 23.49	25.36%	24.92%			
≥23.49	29.17%	23.86%			
Income, % of adults living below poverty line			<.0001	0.1222	0.4644
<3.80	21.35%	25.89%			
3.80 to 7.02	23.43%	25.50%			
7.02 to 13.40	25.54%	24.91%			
≥13.40	29.68%	23.70%			
Comorbidity index			<.0001		
0	67.53%	74.29%			
1	20.33%	18.46%			
2	6.99%	4.90%			
≥3	5.15%	2.34%			
Tumor stage			<.0001	0.0190	0.3694
T1	53.78%	44.44%			
T2	46.22%	55.26%			
Tumor Grade			<.0001	<.0001	0.9503
Well differentiated	18.28%	5.49%			
Moderately differentiated	81.72%	94.51%			

The characteristics of all covariates are described in above Table 4. Patients who were younger, white, or married were more likely to receive active treatment. In addition, patients with larger tumors or moderately differentiated tumors were more likely to

receive active treatment. Patients with three or more comorbidities, or who had higher education or higher income were less likely to received active treatment.

Kaplan-Meier survival curves comparing the overall mortality and prostate cancer-specific mortality in patients who received active treatment, compared to those who were managed with observation, were generated and shown in Figure 2.

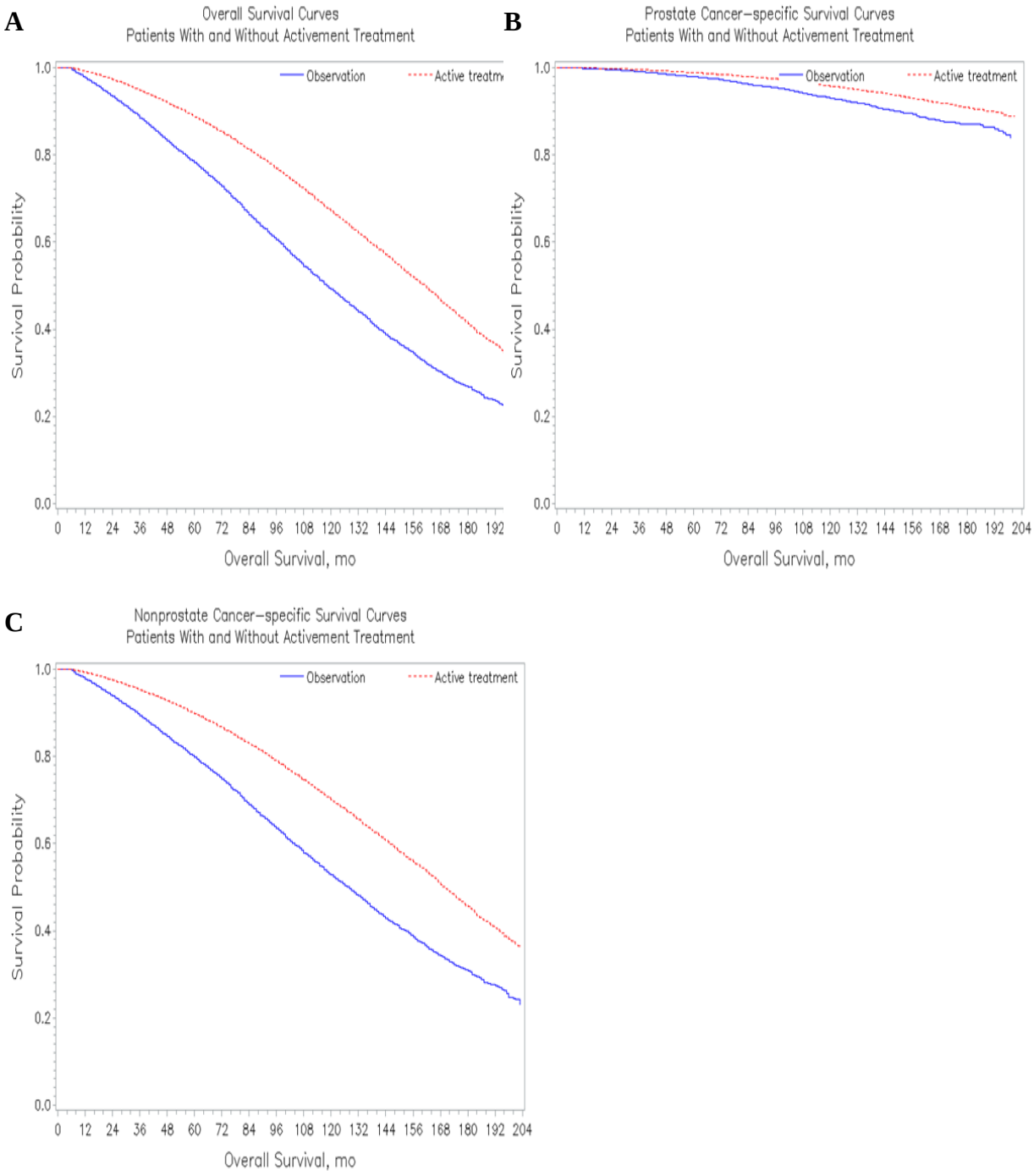

Figure 2: Kaplan-Meier analyses for patients with active treatment compared to observation

In Figure 2A and 2C, both overall and non-prostate cancer-specific survival were higher among patients received radical prostatectomy or radiation (32.18% versus 19.35% at 17 years of follow-up, P<.0001; 36.28% versus 23.09% at 17 years of follow-up, P<.0001). However, patients who received active treatment had slightly higher prostate cancer-specific survival compared to patients who received only observation (88.77% versus 83.93%, P<.0001). The effect of active treatment on mortality, estimated from the Cox proportional hazard model adjusted for all covariates (see Table 7), was used as a reference in the final analysis.

Propensity scores of patients were estimated by generating a logistic regression model with treatment as the outcome of interest. Age, year of diagnosis, race, marital status, SEER site, income, education, number of comorbidities, tumor stage and tumor grade were the independent variables. The distributions of the propensity score across two groups, patients receiving observation and patients receiving active treatment, were compared to evaluate the quality of the propensity scores. Ideally, propensity score distribution would overlap fully across the entire range. In this process, 36 subjects (9 in the observation group and 27 in the active treatment group) were detected and removed from the cohort and final comparison was described in Figure 3. 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. Since the propensity score estimates the probability of receiving treatment, the active treatment group had a higher propensity score.

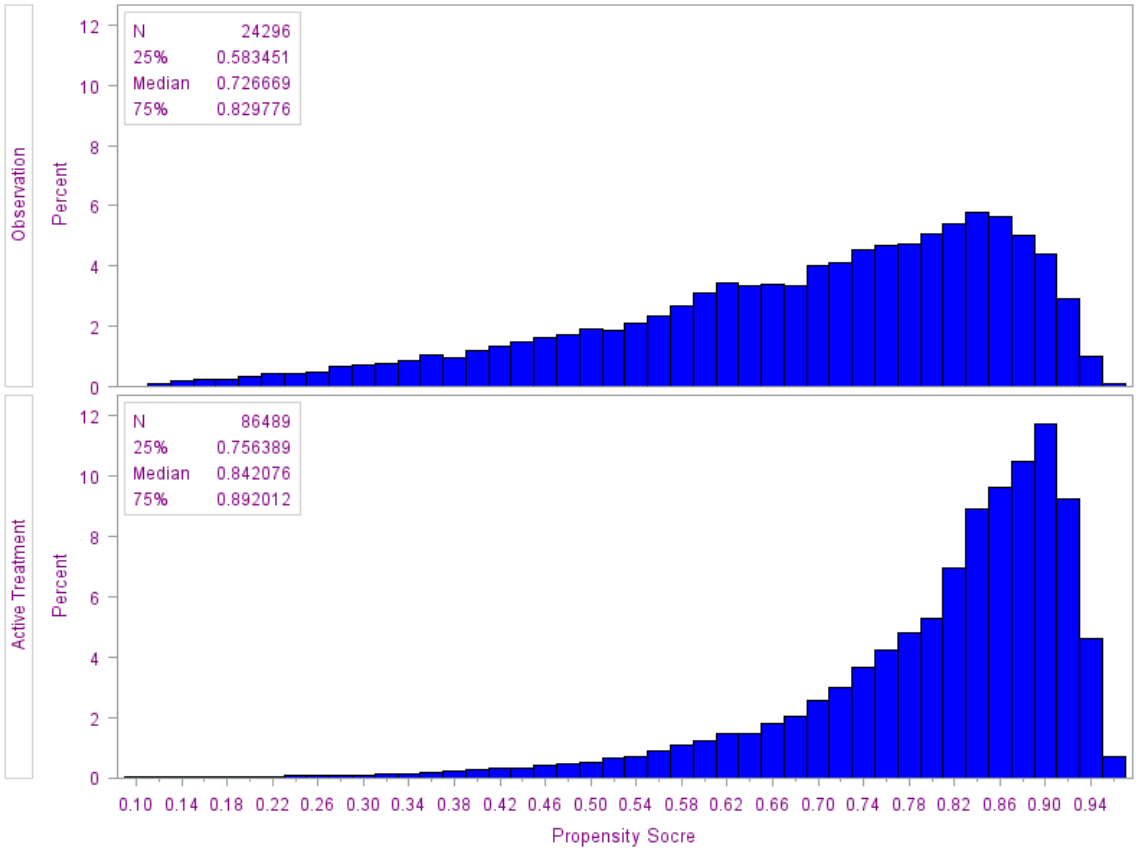

Figure 3: Distribution of the propensity scores for patients with active treatment and with observation

After generating the propensity score, the known confounder, comorbidity, was taken out of the independent variables. The Cox proportional hazards method that incorporated estimated propensity score with four methods is described below.

COVARIATE ADJUSTMENT

First, assessment of the balance of each covariate took place with adjustment for the propensity score. The Cochran Mantel-Haenszel chi-square test was used for this purpose. The results are listed in Table 4. Age, year of diagnosis, marital status, tumor stage, and tumor grade were not balanced after adjusting the propensity score. Then, the Cox proportional hazard model was used to compare survival in patients between two

groups, adjusting the propensity score as a continuous variable together with those five unbalanced variables.

STRATIFICATION

In this method, the propensity score was divided into five equal strata. The distribution of each group by propensity score quintiles is shown in Table 5.

Table 5: Distribution of patients by propensity score quintiles

	Quintile (range) of propensity score				
	0.06-0.68	0.68-0.79	0.79-0.85	0.85-0.90	0.90-0.96
Observation	10,121	5656	3907	2795	1808
Active treatment	12,033	16,493	18,252	19,360	20,351

The assessment of balance was tested by applying the same procedure used in covariate adjustment, and the result was identical to that of as in covariate adjustment. Propensity score stratification was then applied by building strata Cox proportional hazard model in which the propensity score quintile was adjusted as a strata variable. Age, year of diagnosis, marital status, tumor stage and tumor grade were not balanced, even after adjustment for propensity scores.

Inverse probability of treatment weighting (IPTW)

The active treatment group and observation group were weighted, as described in the Table 3. Table 4 describes the balance of covariates before and after propensity score weighting. Race and SEER regions were not balanced in this method. The Cox proportional hazard model was built by adjusting these two unbalanced variables.

Matching

In greedy one-to-one matching, an observation patient was matched to a treated patient based on their propensity score. First, patients in the observation group were matched to the ones in the treatment group on 8 digits of the propensity score. For those that did not match, patients in the observation group were matched to the ones in the treatment group on 7 digits of the propensity score. The algorithm proceeded sequentially to the lowest digit match on the propensity score (1 digit). This was referred to as the 8→1 Digit Match. In this study, 8→1, 5→1 and 3→1 Digit Matches were all applied to provide more options for comparison.

The second matching mechanism was caliper one-to-one matching. The criteria of selecting subjects with a similar propensity score was a caliper of width equal to 0.2 of standard deviation of the logit of the propensity score.

Figure 4 shows the standardized difference after using these two matching tools.

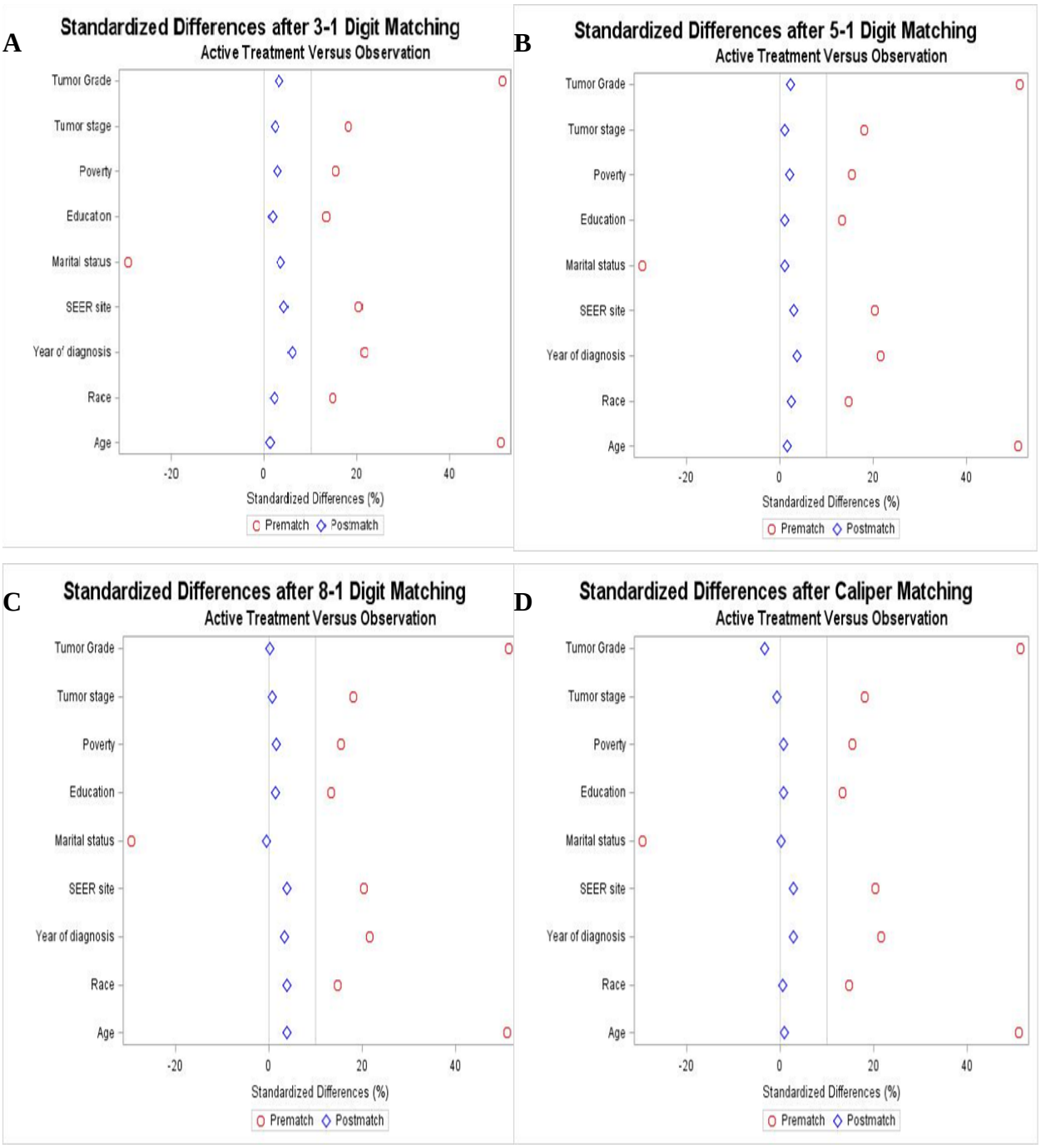

Figure 4: Standardized differences of covariates after matching

Before matching, absolute standardized differences between the two groups were greater than 10% for all covariates. However, differences dropped down to less than 5% after matching. It appears that matched samples were well balanced.

Table 6: Distribution of patients by Matching

Patients	3→1 Digit Matching	5→1 Digit Matching	8→1 Digit Matching	Caliper Matching
Observation	23,088	23,091	23,091	23,122
Active Treatment	23,088	23,091	23,091	23,122

In greedy matching, 95.00% of patients in the observation group were captured, while 95.13% of patients were captured in caliper matching.

Table 7 describes the final outcome models after applying four different propensity score methods.

Table 7: Outcome of patients with active treatment vs. observation for localized prostate cancer

Active treatment versus observation (Referent category)		All-cause mortality HR (95% CI)	Mortality from prostate cancer HR (95% CI)	Mortality from other causes HR (95% CI)
All variables adjusted		0.695 (0.667-0.714)	0.654 (0.601-0.711)	0.701 (0.681-0.721)
Adjusted without comorbidity		0.666 (0.649-0.684)	0.648 (0.596-0.705)	0.668 (0.650-0.687)
Covariate adjustment		0.667 (0.650-0.685)	0.65 (0.598-0.707)	0.67 (0.651-0.689)
Stratification		0.668 (0.650-0.686)	0.646 (0.595-0.702)	0.671 (0.652-0.690)
IPTW		0.674 (0.659-0.689)	0.657 (0.613-0.704)	0.676 (0.661-0.693)
Matching	3→1 digit	0.689 (0.667-0.711)	0.704 (0.637-0.778)	0.687 (0.664-0.711)
	5→1 digit	0.681 (0.659-0.703)	0.714 (0.646-0.789)	0.677 (0.654-0.700)
	8→1 digit	0.689 (0.668-0.712)	0.714 (0.646-0.789)	0.686 (0.664-0.710)
	Caliper	0.68 (0.659-0.703)	0.687 (0.622-0.760)	0.68 (0.657-0.703)

The performance of propensity score methods used to control selection bias between treatment and overall and cause-specific mortality risk was examined. Patients

who underwent active treatment had significantly lower all-cause mortality (HR, 0.695; 95% CI, 0.667-0.714) compared with patients on observation. In the model that removed comorbidity and was combined with propensity score matching, the HR for all-cause mortality was 0.689 (95% CI, 0.667-0.711). For general health, men who were treated with active treatment had a lower risk of death from other causes (HR, 0.701; 95% CI, 0.681-0.721). After applying propensity score matching, the HR was 0.687 (95% CI, 0.664-0.711).

Comparing all four propensity score methods, matching outperformed covariates adjustment, stratification, and IPTW. Greedy matching improved estimates better than caliper matching. The implementation of 3→1 and 8→1 digit matching was best among the other matching tools in calculating all causes mortality, while 3→1 digit matching was best in controlling non-prostate cancer-specific mortality.

DESCRIPTION OF SCENARIO 2: PROSTATE CANCER PATIENTS TREATED WITH VS. WITHOUT ADT

The cohort selection for studying the effect of ADT on overall and cause-specific mortality for patients with localized prostate cancer is summarized in Figure 5. Of patients who met selection criteria, 40,413 (59.64%) did not receive ADT and 27,344 (40.36%) received ADT treatment.

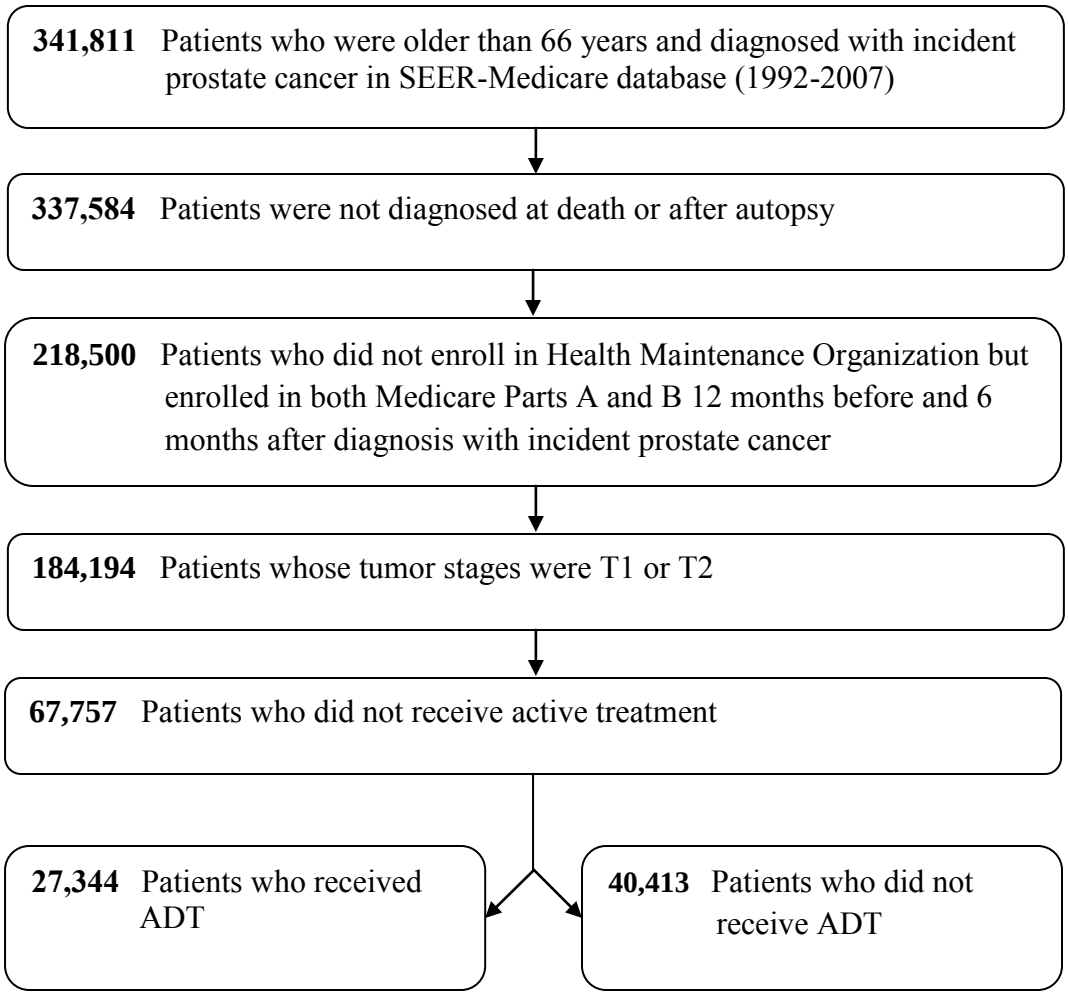

Figure 5: Study-selection summary for the overall cohort of patients treated with and without ADT

Table 8: Characteristics of patients treated with and without ADT

	Overall Cohort			Adjusted by Propensity Score	Applied IPTW
Characteristic	ADT N=27344	No ADT N=40413	P	P	P
Age(years)			<.0001	0.0027	0.9984
66-69	8.70%	14.85%			
70-74	17.32%	24.70%			
75-79	27.63%	28.89%			
≥80	46.35%	31.55%			
Mean±SD	78.7±6.2	76.6±6.2			
Race			<.0001	0.9900	0.9242
White	75.87%	76.47%			
Black	10.17%	11.06%			
Hispanic	5.20%	4.81%			
Others	8.76%	7.66%			
Year of diagnosis			<.0001	0.7504	1.0000
1992-1995	13.86%	22.66%			
1996-1999	15.22%	17.46%			
2000-2003	38.89%	29.44%			
2004-2007	32.03%	30.44%			
SEER Regions			<.0001	0.1097	0.9914
San Francisco	3.46%	5.81%			
Connecticut	8.85%	8.66%			
Detroit	10.48%	12.88%			
Hawaii	1.90%	1.44%			

Iowa	10.08%	8.11%			
New Mexico	2.74%	4.94%			
Seattle	5.36%	8.37%			
Utah	4.00%	5.37%			
Atlanta / Rural Georgia	2.86%	3.79%			
San Jose	3.57%	2.98%			
Los Angeles	9.89%	9.71%			
Greater California	11.58%	11.40%			
Kentucky	5.65%	4.13%			
Louisiana	8.52%	4.38%			
New Jersey	11.06%	8.04%			
Marital Status			<.0001	0.0824	0.2542
Not Married	43.78%	40.29%			
Married	56.22%	59.71%			
Education, % of adults with					
< 12 years of education			<.0001	0.8243	0.9946
<9.32	22.60%	25.91%			
9.32 to 16.05	25.04%	24.23%			
16.05 to 26.59	24.55%	24.64%			
≥26.59	26.01%	23.63%			
Unknown	1.80%	1.59%			
Income, % of adults living					
below poverty line			<.0001	0.2435	0.0883
<4.25	23.77%	25.11%			
4.25 to 8.22	24.03%	24.93%			
8.22 to 15.99	24.98%	24.35%			

≥15.99	25.42%	24.02%			
Unknown	1.80%	1.59%			
Comorbidity index			<.0001	0.7301	0.9963
0	61.03%	66.31%			
1	23.21%	20.47%			
2	9.04%	7.62%			
≥3	6.72%	5.60%			
Tumor stage			<.0001	0.0037	0.3210
T1	33.87%	52.77%			
T2	66.13%	47.23%			
Tumor Grade			<.0001		
Well differentiated	3.76%	14.50%			
Moderately differentiated	57.46%	66.59%			
Poorly differentiated	34.61%	11.58%			
Unknown	4.17%	7.33%			

Patients who were older, better educated, and had higher income were more likely to receive ADT. In addition, patients with a larger tumor or moderately or poorly differentiated tumors were more likely to receive ADT. Patients with 1 or 2 comorbidities and those who were married were less likely to received ADT treatment. Patients who diagnosed between 2000 and 2003 were more likely to receive ADT treatments than patients who were diagnosed between 2004 and 2007. This indicates the decreasing trend of use of ADT in the treatment of localized prostate cancer patients.

Kaplan-Meier survival curves comparing the overall mortality and prostate cancer-specific mortality in patients who were treated with ADT compared to those who were not treated with ADT were shown in Figure 6.

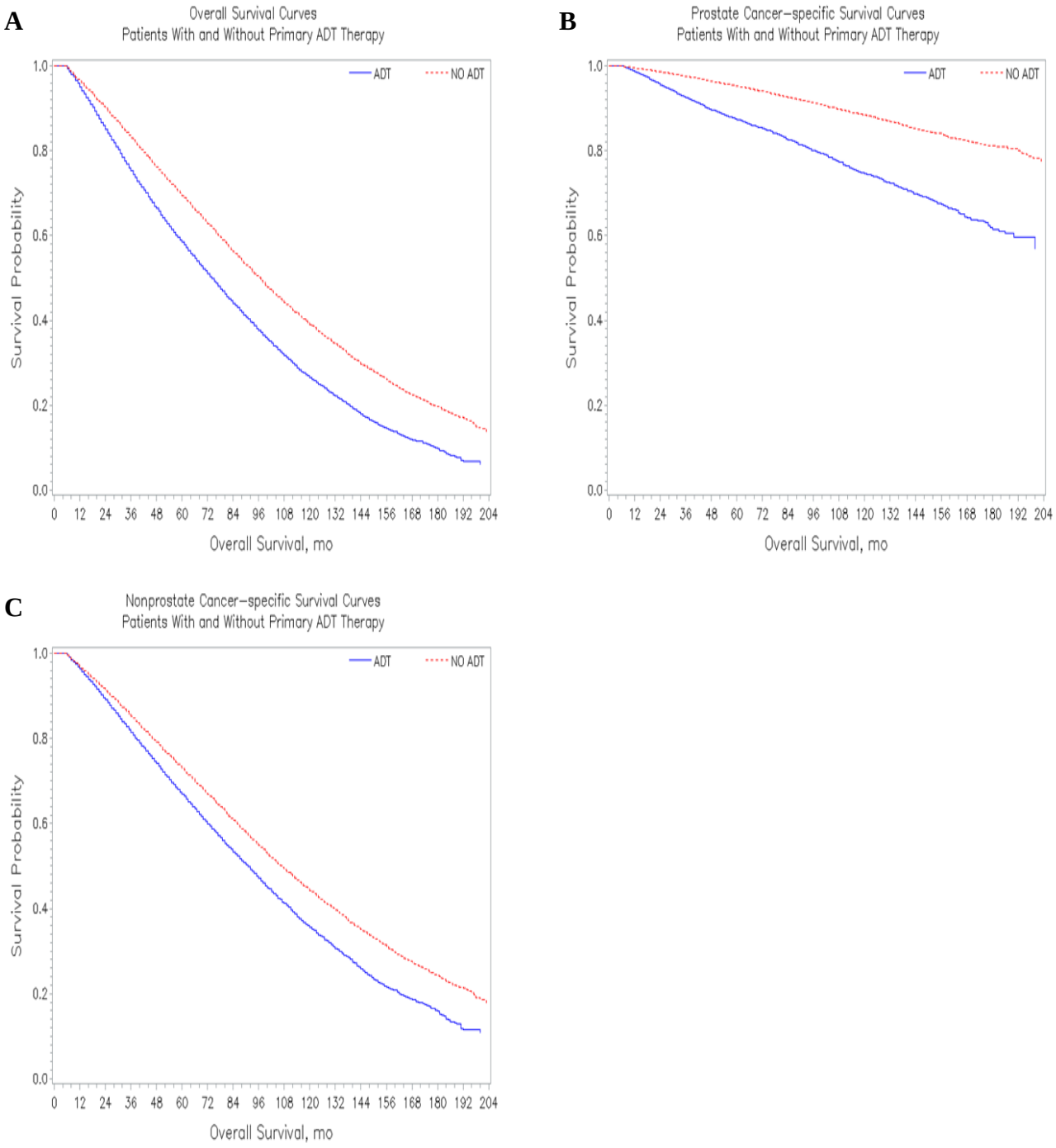

Figure 6: Kaplan-Meier analyses for patients treated with and without ADT

The ADT group had worse overall and cause-specific survival. Patients who were not treated with ADT had improved overall survival (13.71% versus 6.18% at 17 years of

follow-up, P<.0001), better prostate cancer-specific survival (76.93% versus 56.92% at end of follow-up, P<.0001), and better non-prostate cancer-specific survival (17.90% versus 10.95% at end of follow-up, P<.0001).

After checking the overlapping of propensity scores, 95 subjects (3 in the ADT group and 92 in the without ADT group) were detected and removed from the cohort and final comparisons are described in Figure 7. For patients treated with and without ADT, the ranges of the propensity scores were 0.092 to 0.790 and 0.090 to 0.790, respectively.

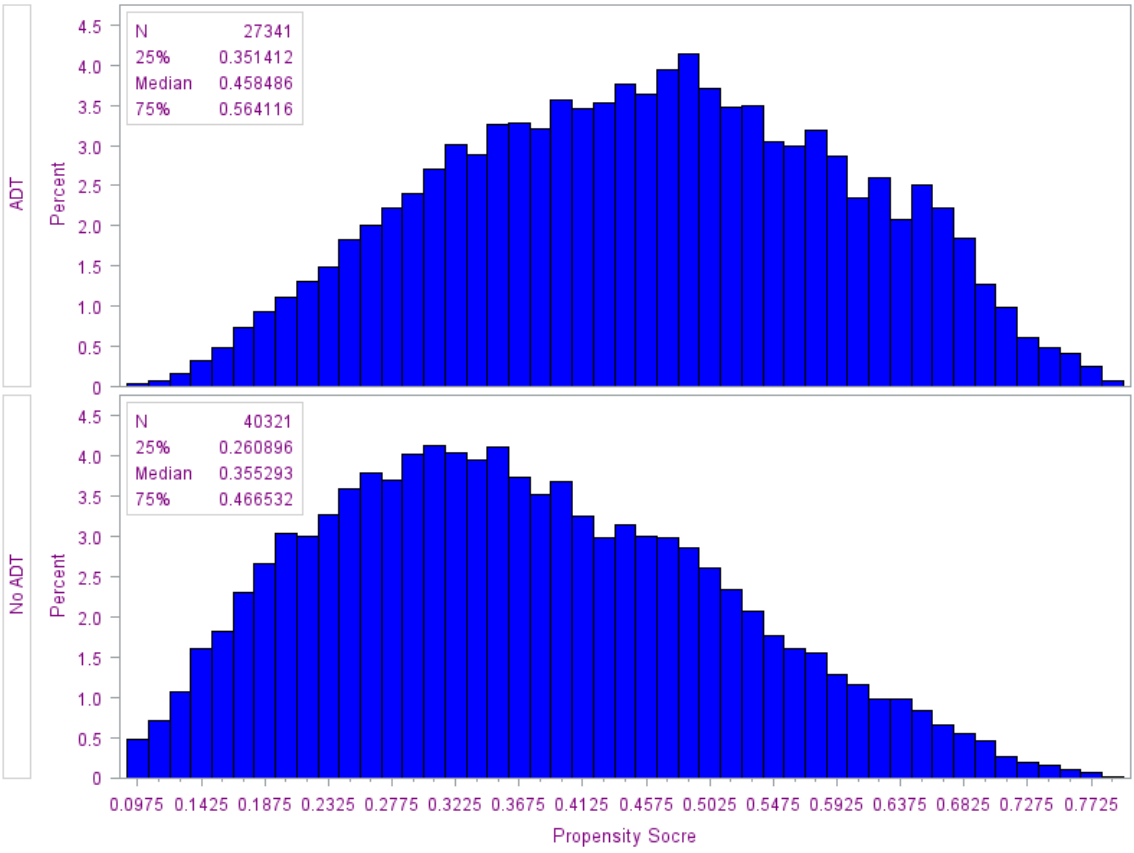

Figure 7: Distribution of propensity scores for patients treated with and without ADT

After generating the propensity score, the known confounder, tumor grade, was taken out of the independent variables. The cox proportional hazards method that incorporates the estimated propensity score with four methods is described below.

COVARIATE ADJUSTMENT

First, assessment of balance of each covariate took place with adjusting propensity score. The Cochran Mantel-Haenszel chi-square test was used for this purpose. The result is listed in Table 8.

According to our result, age and tumor stage were not balanced after adjusting the propensity score. Then the Cox proportional hazard model was used to compare survival in patients between two groups, adjusting the propensity score as a continuous variable together with two unbalanced variables.

STRATIFICATION

In this method, the propensity score was divided into five equal strata. The distribution of each group by propensity score quintiles is shown in Table 9.

After generating the propensity score, the known confounder, tumor grade, was taken out of the independent variables. The Cox proportional hazards method that incorporated the estimated propensity score with four methods is described below.

Table 9: Distribution of patients by propensity score quintiles

	Quintile (range) of propensity score				
	0.09-0.27	0.27-0.35	0.35-0.44	0.44-0.54	0.54-0.79
ADT	2760	4269	5436	6549	8327
No ADT	10,776	9261	8095	6993	5196

The assessment of balance was identical as in covariate adjustment because they were tested by applying the same method. Propensity score stratification was then applied by building strata Cox proportional hazard model in which propensity score quintile was adjusted as a strata variable together with age, and tumor stage, which were not balanced after adjustment for the propensity scores.

INVERSE PROBABILITY OF TREATMENT WEIGHTING (IPTW)

Table 8 describes the balance of covariates before and after propensity score weighting. All variables were balanced in this method.

MATCHING

Figure 8 shows the standardized difference after use of these two matching tools.

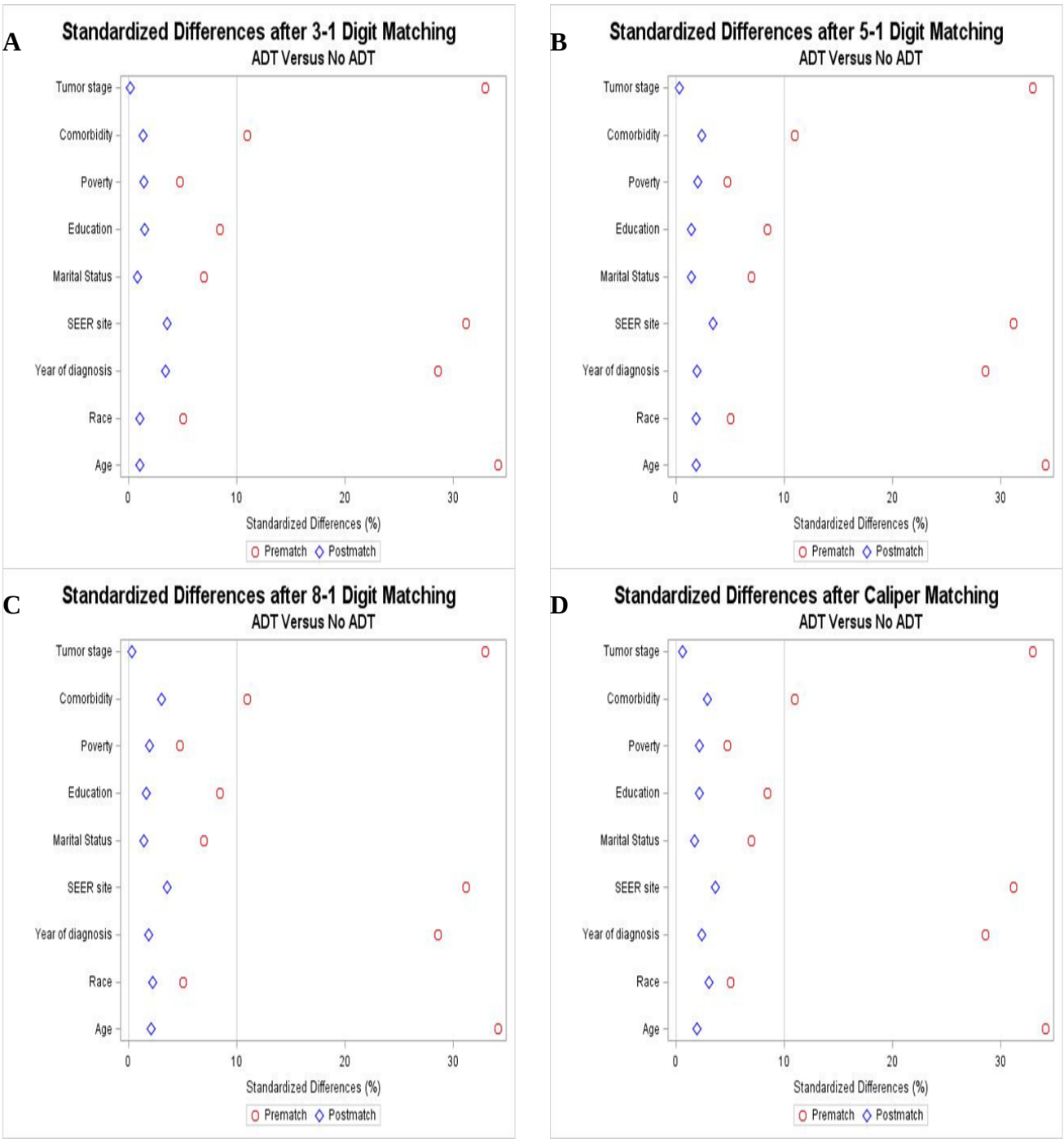

Figure 8: Standardized differences of covariates after matching

Before matching, absolute standardized differences of some variables between the two groups were greater than 10%. This included age, year of diagnosis, SEER site,

comorbidity, and tumor stage. However, differences dropped down to less than 5% after matching. It appeared that matched samples were well balanced.

Table 10: Distribution of patients in matching

Patients	3→1 Digit Matching	5→1 Digit Matching	8→1 Digit Matching	Caliper Matching
ADT	24,174	24,152	24,152	24,134
No ADT	24,174	24,152	24,152	23,122

Almost 88.9% of patients in the no ADT group were captured by propensity score matching methods. Table 10 describes the final outcome models obtained by applying four different propensity score methods.

Table 11: Outcome of patients treated with and without ADT for localized prostate cancer

ADT versus No ADT (Referent category)		All cause mortality HR (95% CI)	Mortality from prostate cancer HR (95% CI)	Mortality from other causes HR (95% CI)
All variables adjusted		1.151 (1.124-1.178)	1.648 (1.558-1.743)	1.064 (1.036-1.092)
Adjusted without comorbidity		1.253 (1.225-1.282)	2.246 (2.129-2.371)	1.101 (1.073-1.129)
Covariate adjustment		1.24 (1.212-1.269)	2.238 (2.121-2.362)	1.086 (1.059-1.114)
Stratification		1.245 (1.217-1.274)	2.222 (2.107-2.345)	1.093 (1.066-1.121)
IPTW		1.251 (1.225-1.278)	2.264 (2.151-2.383)	1.09 (1.065-1.116)
Matching	3→1 digit	1.234 (1.203-1.265)	2.128 (2.004-2.260)	1.087 (1.057-1.118)
	5→1 digit	1.25 (1.219-1.282)	2.138 (2.014-2.269)	1.101 (1.071-1.132)
	8→1 digit	1.25 (1.219-1.282)	2.138 (2.014-2.269)	1.101 (1.071-1.132)
	Caliper	1.237 (1.206-1.268)	2.125 (2.002-2.256)	1.09 (1.060-1.120)

The performance of propensity score methods to control selection bias between patients who received ADT or not, and overall and cause-specific mortality risk were examined.

Patients who underwent ADT had significantly higher prostate cancer-specific mortality (adjusted HR, 1.648; 95% CI, 1.558-1.743) compared with patients who were not treated with ADT. In the model that removed tumor grade and was combined with propensity score caliper matching, the HR for prostate cancer-specific mortality was 2.125 (95% CI, 2.002-2.256).

Generally, matching outperformed covariates adjustment, stratification, and IPTW. With all-cause mortality and prostate cancer-specific mortality, matching produced the smallest bias.

3→1 digit matching implemented best among the other matching tools in calculating all causes mortality and other cause-specific mortality.

Chapter 5 SUMMARY AND DISCUSSION

SUMMARY

Among four propensity score methods, matching (greedy and caliper matching) produced the closest estimates when the removed confounder was included. In scenario 1, 3→1 digit matching appeared to control most of the bias when calculating the non-prostate cancer mortality between patients with or without active treatment.

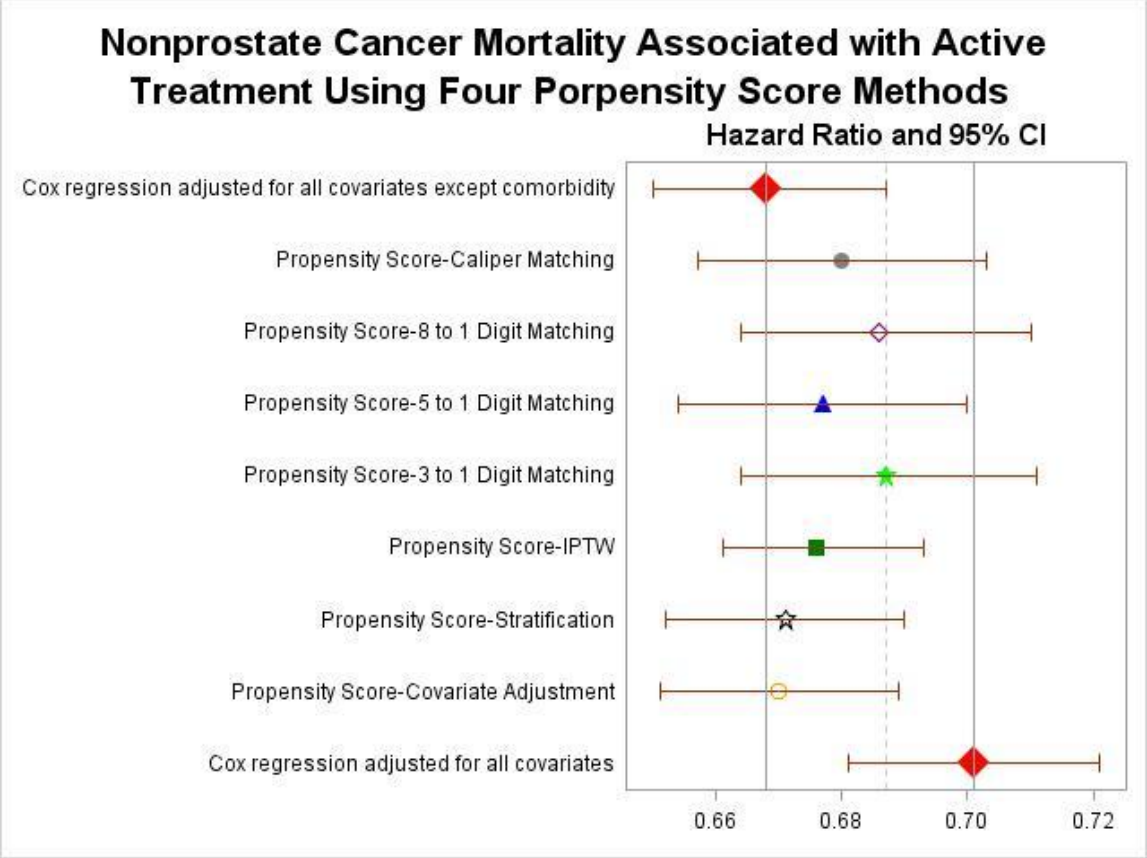

Figure 9: Results of four methods applying the propensity score for the non-prostate cancer mortality associated with active treatment

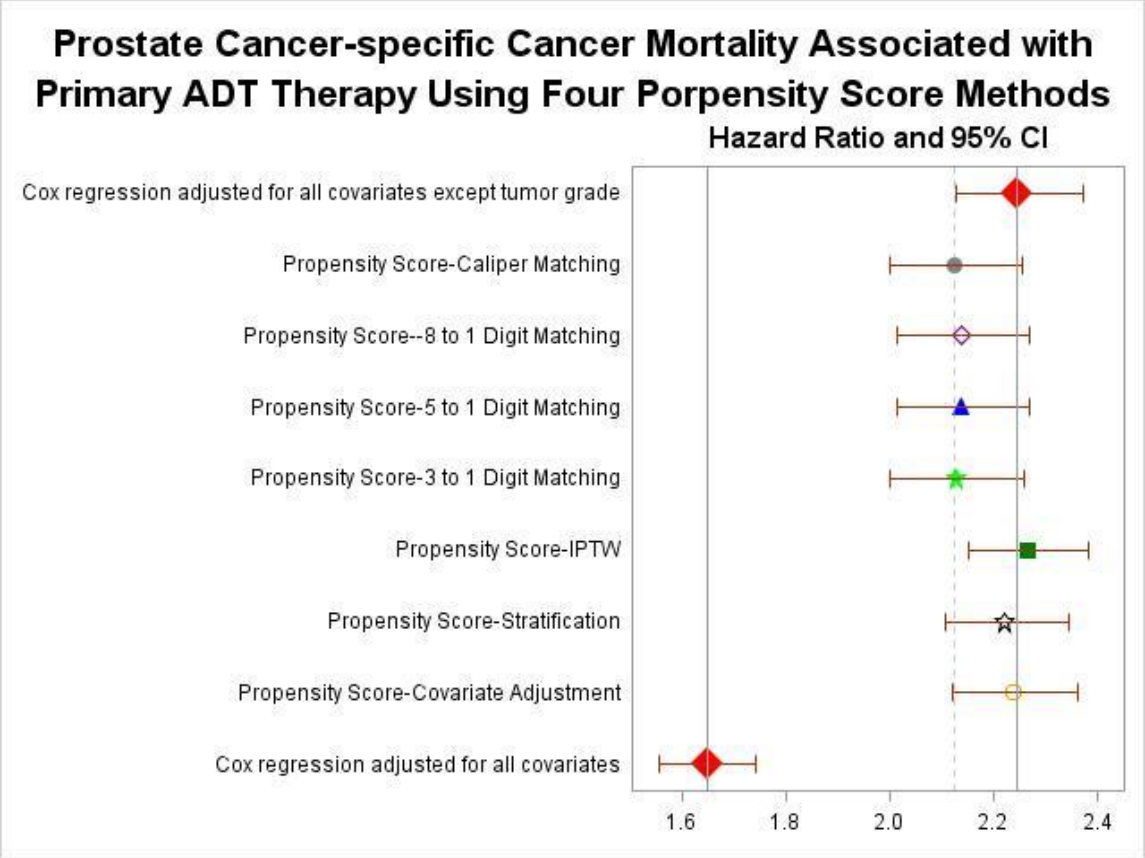

Figure 10: Results of four methods applying the propensity score for the prostate cancer mortality associated with ADT

In scenario 2, caliper matching outperformed other propensity score methods when computing prostate cancer mortality between patients with or without receiving ADT.

Covariate adjustment and stratification produced similar weak controlling effects in both scenarios. IPTW showed a performance that was not as good as matching in scenario 1. IPTW yielded a poorer estimate after including tumor grade in scenario 2. The hazard ratio was not in the range between the models adjusted for all covariates and the one adjusted for all covariates except tumor grade.

However, none of the propensity score methods produced the identical estimates when the removed confounder was included.

DISCUSSION

In scenario 1, propensity score matching produced a closer estimate when the removed confounder was included than in scenario 2. One possibility is that tumor grade has a bigger effect on the mortality from prostate cancer. Patients with poorly differentiated tumor grade had a much higher hazard than patients with well differentiated tumor grade (HR, 6.232). The hazard ratio of mortality from other causes associated with the largest Charlson comorbidity index in scenario 1 as described in Table 12 was 3.968. In this way, propensity score matching can show a better controlling ability.

Table 12: Hazard ratio of comorbidity associated with active treatment

	All-cause mortality	Mortality from other causes
Variable	HR (95% CI)	HR (95% CI)
Comorbidity index		
0	1.000	1.000
1	1.688 (1.641-1.736)	1.766 (2.258-2.479)
2	2.231 (2.133-2.334)	2.366 (2.258-2.479)
≥3	3.699 (3.510-3.898)	3.968 (3.759-4.188)

Table 13: Hazard ratio of tumor grade associated with ADT

Variable	All-cause mortality HR (95% CI)	Mortality from prostate cancer HR (95% CI)
Tumor grade		
Well differentiated	1.000	1.000
Moderately differentiated	1.071 (1.034-1.109)	1.752 (1.553-1.976)
Poorly differentiated	1.649 (1.584-1.716)	6.232 (5.512-7.047)
Unknown	1.290 (1.226-1.358)	3.236 (2.800-3.740)

In scenario 2, there was greater difference between the hazard ratio calculated from the propensity score and the one calculated from the model with the removed confounder. Given that extreme weights may not be eliminated in the entire study population, a high hazard ratio could occur (Kurth et al., 2006).

In our study, none of four propensity score methods could modify the bias after removing a confounder from the models. This conclusion was similar to some previously suggested reports (Giordano et al., 2008b; Shah et al., 2005). As discussed in the first chapter, there are two assumptions for propensity score methods. The first one indicates that there are no unmeasured confounders for the association between treatment and outcome, conditional on the measured covariates. In a randomized controlled trial, measured and unmeasured covariates are balanced between the treatment and the control group. The inability to balance unmeasured confounders is a major limitation of observational studies. Propensity score methods are not able to openly correct the hidden bias. Whereas, in our empirical study, we can only try to balance the measured confounders and assume that this balance reduces the overt bias (Harder et al., 2010). The Cochran Mantel-Haenszel chi-square test was established to assess the balance of all

covariates in covariate adjustment, stratification, and IPTW. Standardized differences for both continuous and categorical variables were checked in matching, and all covariates after matching were balanced in both scenarios.

Instrumental variable analysis has been proposed to reduce the variance and bias caused by unmeasured confounding (Greenland, 2000; Brookhart et al., 2006). It requires identifying variables that are strongly related to the exposure but unrelated to the outcome. Several studies have demonstrated instrumental variable analysis to control selection bias (Sheffield et al., 2013; Thérèse et al., 2007). Correct estimation by using this method largely relies on the thoughtful consideration of selection an appropriate instrumental variable(s). However, it is not clear that the healthcare utilization data (e.g. SEER-Medicare or private insurance companies) have such variables recorded or not. However, statistical methods can not eliminate all bias and confounding in observational study.

After generating the propensity score model, the overlap of the propensity score between two groups was checked first, because of the second assumption of propensity score techniques that the probability of receiving treatment is positive between 0 and 1. Those individuals with the propensity score outside this range were excluded from the original study population.

The first propensity score method in this study was covariate adjustment using the propensity score. This approach differs from stratification, matching, and IPTW in that covariate adjustment does not change the design of the study. Because with stratification, matching, and IPTW, once the specified models are built, the effect of treatment can be directly estimated in the stratified, matched, or weighted population sample. In contrast, for covariate adjustment, a new fitted regression model, in which outcome is regressed on an indicator variable denoting treatment status and propensity score, is built. By using this method, we assume that the relationship between the propensity score and the

outcome is correctly demonstrated. This is a major limitation of applying covariate adjustment.

Stratification on the propensity score creates subgroups so that treated and untreated will fall into the same range of the propensity score stratum. Therefore, the distribution of measured covariates will be approximately similar. In our study, the strata Cox proportional hazard model was adapted. However, the balance for all baseline covariates was assessed by whole study cohorts. In our study, covariate adjustment and stratification yielded similar controlling effects.

IPTW addressed a study population in which covariates for subjects were more balanced in both scenarios. That is because the difference between two groups was measured using weighted samples. A treated subject's weight is the inverse of the propensity score, and one minus the inverse of the propensity score is used for controlled subjects. The use of IPTW is similar to the use of survey sampling weights (Austin, 2011). However, one thing that we need to concern about was that sometimes there are extreme weights due to a very low probability of receiving treatment. In our study, the weights were stabilized, which gave less influence to extreme weights.

The fourth propensity score method, matching, is considered to be the method that can eliminate a greater proportion of bias than stratification on the propensity score or covariate adjustment using the propensity score. Several studies have demonstrated this conclusion and our results agreed with them (Austin, 2009, 2011; Austin, Grootendorst, & Anderson, 2007; Austin & Mamdani, 2006).

In some situations, matching and IPTW removed systematic differences between treated and untreated groups at a comparable level; however, in some situations, IPTW removed fewer imbalances than did matching (Austin, 2009). In our study, IPTW performed less controlling ability than matching, but more than covariate adjustment and stratification.

The main advice for selecting the best propensity score method is to select the one that yields the most balanced study sample. Matching appears to be the method that usually produces samples that have the smallest standardized difference.

Bibliography

Albertsen, P. C., Hanley, J. A., & Fine, J. (2005). 20-Year outcomes following conservative management of clinically localized prostate cancer. The Journal of the American Medical Association, 293(17), 2095-2101.

Albertsen, P. C., Hanley, J. A., Gleason, D. F., & Barry, M. J. (1998). Competing risk analysis of men aged 55 to 74 years at diagnosis managed conservatively for clinically localized prostate cancer. The Journal of the American Medical Association, 280(11), 975-980.

American Cancer Society (2012). Cancer Facts & Figures 2012. From http://www.cancer.org/research/cancerfactsfigures/cancerfactsfigures/cancer-facts-figures-2012.

American Cancer Society (2013). From http://www.cancer.org.

Austin, P. C. (2008). The performance of different propensity-score methods for estimating relative risks. Journal of Clinical Epidemiology, 61(6), 537-545.

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(1), Article 13.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behavioral Research, 46(3), 399-424.

Austin, P. C. (2013). The performance of different propensity score methods for estimating marginal hazard ratios. Statistics in Medicine, 32(16), 2837-2849.

Austin, P. C., Grootendorst, P., & Anderson, G. M. (2007). A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. Statistics in Medicine, 26(4), 734-753.

Austin, P. C., & Mamdani, M. M. (2006). A comparison of propensity score methods: A case-study estimating the effectiveness of post-AMI statin use. Statistics in Medicine, 25(12), 2084-2106.

Becerril, J., & Abdulai, A. (2010). The impact of improved maize varieties on poverty in Mexico: a propensity score-matching approach. World Development, 38(7), 1024-1035.

Bill-Axelson, A., Holmberg, L., Ruutu, M., Garmo, H., Stark, J. R., Busch, C., et al. (2011). Radical prostatectomy versus Watchful Waiting in Early Prostate Cancer. The New England Journal of Medicine, 364(18), 1708-1717.

Bill-Axelson, A., Holmberg, L., Ruutu, M., Haggman, M., Andersson, S., Bratell, S., et al. (2005). Radical Prostatectomy versus Watchful Waiting in Early Prostate Cancer. The New England Journal of Medicine, 352(19), 1977-1984.

Boyle, P., Severi, G., Giles, G. G. (2003). The epidemiology of prostate cancer. Urologic Clinics of North America, 30(2), 209-217.

Brookhart, M. A., Patrick A. R., Dornuth C., et al. (2007). Adherence to Lipid-lowering Therapy and the Use of Preventive Health Services: An Investigation of the Healthy User Effect. American Journal of Epidemiology, 166(3), 348-354.

Brookhart, M. A., Schneeweiss, S., Rothman, K. J., Glynn, R. J., Avorn, J., Sturmer, T. (2006). Variable Selection for Propensity Score Models. American Journal of Epidemiology, 163(12), 1149-1156.

Brookhart, M. A., Sturmer, T., Glynn, R. J., Rassen, J., Schneeweiss, S. (2010). Confunding Control in Healthcare Database Research, Challenges and Potential Approaches. Medical Care, 48(6), S114-S120.

Chodak, G. W., Thisted, R. A., Gerber, G. S., Johansson, J., Adolfsson, J., Jones, G. W., et al. (1994). Results of conservative management of clinically localized prostate cancer. The New England Journal of Medicine, 330(4), 242-248.

Clinton, S. K., & Giovannucci, E. (1998). Diet, Nutrition, and prostate cancer. Annual Review of Nutrition, 18(1), 413-440.

Cochran, W. G. (1968). The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies. Biometrics, 24(2), 295-313.

Giordano, S. H., Kuo, Y. F., Duan, Z., Hortobagyi, G. N., Freeman, J., & Goodwin, J. S. (2008). Limits of observational data in determining outcomes from cancer therapy. Cancer, 112(11), 2456-2466.

Greenland, S. (2000). An Introduction to Instrumental Variables for Epidemiologists. International Journal of Epidemiology, 29(4), 722-729.

Chan, J. M., Stampfer, M. J., Giovannucci, E. L. (1998). What causes prostate prostate cancer? A brief summary of the epidemiology. Seminars in Cancer Biology, 8(4), 263-273.

Haas, G. P., & Sakr, W. A. (1997). Epidemiology of prostate cancer. CA: A Cancer Journal for Clinicians, 47(5), 273-287.

Harder, V. S., Stuart, E. A., & Anthony, J. C. (2010). Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. Psychological methods. Psychological Methods, 15(3), 234-249.

Johnston, S. C., Rootenberg, J. D., Katrak, S., Smith, W. S., & Elkins, J. S. (2006). Effect of a US National Institutes of Health programme of clinical trials on public health and costs. The Lancet, 367(9519), 1319-1327.

Kollmeier, M. A., & Zelefsky, M. J. (2012). How to select the optimal therapy for early-stage prostate cancer. Critical Reviews in Oncology/Hematology, 83(2), 225-234.

Kurth, T., Walker, A. M., Glynn, R. J., Chan, K. A., Gaziano, J. M., Berger, K., et al. (2006). Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. American Journal of Epidemiology, 163(3), 262-270.

Levine, G. N., D'Amico, A. V., Berger, P., Clark, P. E., Eckel, R. H., Keating, N. L., et al. (2010). Androgen-Deprivation Therapy in Prostate Cancer and Cardiovascular Risk: A Science Advisory from the American Heart Association, American Cancer Society, and American Urological Association: Endorsed by the American

Society for Radiation Oncology. CA: A Cancer Journal for Clinicians, 60(3), 194-201.

Ming, K., & Rosenbaum, P. R. (2000). Substantial Gains in Bias Reduction from Matching with a Variable Number of Controls. Biometrics, 56(1), 118-124.

Monahan, K. C., Lee, J. M., & Steinberg, L. (2011). Revisiting the impact of part-time work on adolescent adjustment: Distinguishing between selection and socialization using propensity score matching. Child Development, 82(1), 96-112.

Perlmutter, M. A., & Lepor, H. (2007). Androgen Deprivation Therapy in the Treatment of Advanced Prostate Cancer. Reviews in Urology, 9(Suppl 1), S3-S8.

Potosky, A. L., Legler, J., Albertsen, P. C., Stanford, J. L., Gilliland, F. D., Hamilton, A. S., et al. (2000). Health outcomes after prostatectomy or radiotherapy for prostate cancer: results from the prostate cancer outcomes study. Journal of the National Cancer Institute, 92(19), 1582-1592.

Rosenbaum, P. R., & Rubin, D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika, 70(1), 41-55.

Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing Bias in Observational Studies Using Subclassification on the Propensity Score. Journal of the American Statistical Association, 79(387), 516-524.

Rothwell, P. M. (2005). External validity of randomised controlled trials: "To whom do the results of this trial apply?". The Lancet, 365(9453), 82-93.

Sanda, M. G., Dunn, R. L., Michalski, J., Sandler, H. M., Northouse, L., Hembroff, L., et al. (2008). Quality of life and satisfaction with outcome among prostate-cancer survivors. The New England Journal of Medicine, 358(12), 1250-1261.

Sanda, M. G., & Kaplan, I. D. (2009). A 64-Year-Old Man with Low-risk Prostate Cancer: Review of Prostate Cancer Treatment. American Medical Association, 301(20), 2141-2151.

Schneeweiss, S., Rassen, J. A., Glynn, R. J., Avorn, J., Mogun, H., & Brookhart, M. A. (2009). High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. Epidemiology, 20(4), 512-522.

SEER. (2013). Racial/Ethnic Patterns of Cancer in the United State, 1988-1992. From http://seer.cancer.gov/publications/ethnicity/

Shah, B. R., Laupacis, A., Hux, J. E., & 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(6), 550-559.

Shahinian, V. B., Kuo, Y. F., & Freeman, J. L. (2005). Risk of Fracture after Androgen Deprivation for Prostate Cancer. The New England Journal of Medicine, 352(2), 154-164.

Shahinian, V. B., Kuo, Y. F., & Gilbert, S. M. (2010). Reimbursement policy and Androgen-Deprivation Therapy for Prostate Cancer. The New England Journal of Medicine, 363(19), 1822-1832.

Sharifi, N., Gulley, J. L., & Dahut, W. L. (2005). Androgen Deprivation Therapy for Prostate Cancer. The Journal of the American Medical Association, 294(2), 238-244.

Sheffield, K. M., Riall, T. S., Han, Y., Kuo, Y. F., Townsend, C. M. Jr., Goodwin, J. S. (2013). The Journal of the American Medical Association, 310(8), 812-820.

Silverman, S. L. (2009). From randomized controlled trials to observational studies. The American Journal of Medicine, 122(2), 114-120..

Studer, U. E., Whelan, P., Albrecht, W., Casselman, J., Reijke, T. d., Hauri, D., et al. (2006). Immediate or deferred androgen deprivation for patients with prostate cancer not suitable for local treatment with curative intent: European Organisation for Research and Treatment of Cancer (EORTC) Trial 30891. Journal of Clinical Oncology, 24(12), 1868-1876.

Thérèse, A. S., Elliott, S. F., David, E. W., David, A. A., Daniel, J. G., Marian J. V. (2007). 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. The Journal of American Medical Association, 297(3), 278-285.

Thompson, I. M. Jr., Goodman, P. J., Tangen, C. M., Parnes, H. L., Minasian, L. M., Godley, P. A., et al. (2013). Long-term survival of participants in the Prostate Cancer Prevention Trial. New England Journal of Medicine, 369(7), 603-610.

Walker, A. M. (1996). Confounding by Indication. Epidemiology, 7(4), 335-336.

Warde, D. P., Mason, M., Ding, K., Kirkbride, P., Brundage, M., Cowan, R., et al. (2011). Combined androgen deprivation therapy and radiation therapy for locally advanced prostate cancer: a randomised, phase 3 trial. The Lancet, 378(9809), 2104-2111.

Wardlaw, J. M., Sandercock, P. A. G., & Berge, E. (2003). Thrombolytic Therapy With Recombinant Tissue Plasminogen Activator for Acute Ischemic Stroke. Where Do We Go From Here? A Cumulative Meta-Analysis. Stroke, 34(6), 1437-1442.

Warren, J. L., Klabunde, C. N., Schrag, D., Bach, P. B., & Riley, G. F. (2002). Overview of the SEER-Medicare Data: Content, Research Applications, and Generalizability to the United States Elderly Population. Medical Care, 40(8), IV3-IV18.

Wilt, T. J., Brawer, M. K., Jones, K. M., Barry, M. J., Aronson, W. J., Fox, S., et al. (2012). Radical Prostatectomy versus Observation for Localized Prostate Cancer. The New England Journal of Medicine, 367(3), 203-213.

Wong, Y. N., Mitra, N., Hudes, G., Localio, R., Schwartz, J. S., Wan, F., et al. (2006). Survival Associated With Treatment vs Observation of localized prostate cancer in Elderly Men. The Journal of the American Medical Association, 296(22), 2683-2693.

Vita

Xiao Fang was born 07 November 1986 in Bengbu, Anhui, China. She graduated from Nanjing University of Science and Technology with a Bachelors of Science in Optical Information of Science and Technology in 2010. In August 2012, she began her Masters of Science in Preventive Medicine and Community Health with a curriculum in Biostatistics, at the University of Texas Medical Branch at Galveston.

This thesis was typed by Xiao Fang.