PROSTATE CANCER-SPECIFIC SURVIVAL DIFFERENCES BETWEEN RADICAL PROSTATECTOMY AND CURATIVE RADIOTherapy

by

Julie Marie DeGroot

A thesis submitted to the Department of Community Health and Epidemiology in conformity with the requirements for the degree of Master of Science

Queen’s University

Kingston, Ontario, Canada

(September, 2009)

Copyright © Julie Marie DeGroot, 2009
Abstract

Background: The relative treatment effectiveness of surgery versus radiotherapy for early-stage prostate cancer is uncertain and randomized clinical trials are unlikely to be performed. This study describes the difference in cause-specific survival between patients treated with radiotherapy versus surgery, using a number of design and analytic steps to mitigate confounding by indication within an observational study.

Methods: We conducted a population-based case-cohort study, sampling patients from the Ontario Cancer Registry who were treated or were candidates for cure by radiotherapy or surgery. Cases were those who died of prostate cancer within 10 years. Cause-specific survival was analyzed using Cox-proportional hazard regression, with variance adjustment for the case-cohort sampling. Analysis using intent to treat was compared to that using treatment received. Propensity scores were also calculated and Cox-proportional hazard regression was conducted within each propensity score quintile. We formed instrumental variable groups based on radiotherapy rates in Cancer Care Ontario Regions (CCORs) using the study population sampling frame and checked the instrumental variable assumption of equal distribution of covariates by comparing those covariates across these groups using data from the subcohort.

Results: The adjusted hazard ratios for risk of prostate cancer death for radiotherapy compared to surgery were 1.44 (95% CI 0.86-2.40) and 1.84 (1.06-3.17) using intent to treat and treatment received respectively. Stratified hazard ratios comparing radiotherapy to surgery for death from prostate cancer from the lowest propensity quintile to the
highest propensity quintile were 0.30 (0.04-2.28), 1.54 (0.35-6.77), 0.90 (0.29-2.82), 2.71 (1.01-7.31) and 1.08 (0.41-2.81). Differences among these hazard ratios were not statistically significant (p=0.13). The distributions of all prognostic indicators were statistically significantly different between instrumental variable groups.

**Conclusion:** Analysis by intent to treat produced a hazard ratio closer to the null than analysis by treatment received, indicating that uncontrolled confounding toward more serious cases getting radiotherapy was present in the analysis by treatment received. Future studies should focus on obtaining enough numbers for subgroup analysis such as the stratification by risk groups. Caution should be used when using the instrumental variable approach in this population, as prognostic indicators were not as equally distributed as expected.
Co-Authorship

This thesis is the research work of Julie DeGroot in collaboration with her supervisors Dr. Patti A. Groome, Dr. Michael Brundage and Dr. Miu Lam.

Manuscript 1: Prostate cancer-specific survival differences in patients treated by radical prostatectomy versus curative radiotherapy. The study design and methods were the work of Dr. Michael Brundage, Ms. Susan Rohland, Dr. Jeremy Heaton, Dr. William Mackillop, Dr. Robert Siemens and Dr. Patti Groome. Julie DeGroot was involved in the statistical analysis, writing the manuscript, and interpreting the results, with supervision and feedback provided by Dr. Patti Groome, Dr. Michael Brundage and Dr. Miu Lam. Ms. Susan Rohland and Dr. Jeremy Heaton have also provided feedback on the manuscript. Dr. Miu Lam provided statistical advice.

Manuscript 2: Assessing the instrumental variable method assumption of equal distribution of unmeasured covariates in a prostate cancer population. The idea of using instrumental variables to compare differences in treatment effectiveness between radiotherapy and surgery was Dr. Patti Groome’s. Statistical analysis, writing the manuscript and interpreting the results was done by Julie DeGroot, with supervision and feedback provided by Dr. Patti Groome, Dr. Michael Brundage and Dr. Miu Lam. Dr. Miu Lam provided statistical advice.
Acknowledgements

First, I would like to thank my primary supervisor Dr. Patti Groome for all of her patience, words of encouragement, and detailed suggestions for my thesis. I cannot imagine a better supervisor and I am very grateful for this positive experience. I would also like to thank Dr. Mike Brundage and Dr. Miu Lam for all of their helpful suggestions and guidance-Dr. Brundage for his clinical insight and Dr. Miu Lam for his statistical assistance.

I would also like to thank Katherine Cook and Lee Watkins for all of their assistance. They have both definitely made completing my thesis much easier.

I would also like to acknowledge Dr. Groome’s other students- Dylan, John, John, and MK for all of their useful suggestions and everyone at CCE for their support. Thanks especially to Sue Rohland for all of her detailed explanations on patient selection and variable formation. Thanks to all of my friends and peers here in Kingston for making my masters such a great experience.

Finally, I would like to thank my parents for their unwavering support throughout this entire process.
Table of Contents

Abstract ...................................................................................................................................................... ii
Co-Authorship ........................................................................................................................................ iv
Acknowledgements .................................................................................................................................... v
Table of Contents .................................................................................................................................... vi
List of Figures .......................................................................................................................................... ix
List of Tables ........................................................................................................................................ x
List of Acronyms .................................................................................................................................. xii

Chapter 1: Introduction ........................................................................................................................ 1
  1.1 Background and Rationale ............................................................................................................. 1
  1.2 Objectives ...................................................................................................................................... 1
  1.3 Thesis Outline ............................................................................................................................... 2
  1.4 References ................................................................................................................................... 2

Chapter 2: Literature Review ................................................................................................................. 4
  2.1 Prostate Cancer ............................................................................................................................ 5
  2.2 Prostate Cancer Treatments ....................................................................................................... 5
  2.3 Treatment Decision Making ....................................................................................................... 6
  2.4 Observational Studies of Prostate Cancer .................................................................................. 11
  2.5 Methodological Strategies for Managing the Impact of Confounding by Indication ............. 13
     2.5.1 Overview .............................................................................................................................. 13
     2.5.2 Intent to Treat ....................................................................................................................... 14
     2.5.3 Restriction ........................................................................................................................... 15
     2.5.4 Multivariate Analysis .......................................................................................................... 16
     2.5.5 Instrumental Variables ....................................................................................................... 16
     2.5.6 Propensity Scores ............................................................................................................... 18
  2.6 Summary ...................................................................................................................................... 20
  2.7 References ................................................................................................................................... 21

Chapter 3: Methods ............................................................................................................................... 28
  3.1 Introduction .................................................................................................................................. 28
  3.2 Study Objectives ........................................................................................................................ 28
     3.2.1 Primary Objectives .............................................................................................................. 28
     3.2.2 Secondary Objectives ........................................................................................................ 29
  3.3 Study Design ............................................................................................................................... 29
6.2 Diagnostic Tests ................................................................................................................... 89
6.3 High Risk Patients ................................................................................................................ 90
6.4 The Instrumental Variable Method ...................................................................................... 92
6.5 References ............................................................................................................................ 98

Chapter 7: General Discussion ......................................................................................................... 99
7.1 Discussion of Key Findings .................................................................................................... 99
  7.1.1 Treatment Effectiveness .............................................................................................. 99
  7.1.2 The Instrumental Variable Method ............................................................................. 102
7.2 Discussion of Supplemental Results .................................................................................. 103
  7.2.1 High Risk Patients ....................................................................................................... 103
  7.2.2 Diagnostic Tests .......................................................................................................... 104
7.3 Statistical Power ................................................................................................................. 105
7.4 Limitations ......................................................................................................................... 106
7.5 Strengths .......................................................................................................................... 109
7.6 Future Directions .............................................................................................................. 110
7.7 References .......................................................................................................................... 111

Appendices ....................................................................................................................................... 113
  Appendix I: Continuous versus Categorical Variables for Age and CIRS-G ...................... 113
  Appendix II: Relevant Excerpts from Chart Abstraction Form ........................................... 114
List of Figures

Chapter 3
Figure 3.1: Sampling of the study population in the original study. The addition of the prostate cancer death cases, other cause death cases and the subcohort does not total 2740 due to overlap between the case groups and the subcohort..............................................................32
Figure 3.2: Exclusion criteria and study population numbers......................................................33
Figure 3.3: Radiotherapy and surgery candidacy designation .....................................................36

Chapter 6
Figure 6.1: Proportion of patients who were radiotherapy (RT) candidates based on (a) county and (b) CCOR with 95% confidence intervals...............................................................97
List of Tables

Chapter 2
Table 2.1: Brief description of observational studies comparing survival or biochemical relapse-free survival (bRFS) in prostate cancer patients between radiotherapy (RT) and radical prostatectomy (RP) ................................................................................................................................................................................. 8

Chapter 3
Table 3.1: Sampling distribution based on Cancer Care Ontario Region (CCOR) for the subcohort, prostate cancer (PC) death cases and other cause (OC) death cases ............... 31

Chapter 4: Manuscript 1
Table 1: Baseline characteristics of the subcohort and cases based on radiotherapy and surgical candidacy ................................................................................................................................................................................................................................................................................................................. 68
Table 2: Hazard ratios for death from prostate cancer using intent to treat and actual treatment received for radiotherapy compared to surgery and for study covariates (with 95% confidence intervals). ................................................................................................................................................................................................................................................................................................................................................................................. 69
Table 3: Distribution of covariates between radiotherapy and surgical candidates in the subcohort within each propensity score quintile (percentage) ................................................................................................................................................................................................................................................................................................................................................................................. 70

Chapter 5: Manuscript 2
Table 1: Comparison of prognostic indicator distribution between radiotherapy and surgical patients in the chart-reviewed cohort ................................................................................................................................................................................................................................................................................................................................................................................. 86
Table 2: Radiotherapy rates by Cancer Care Ontario Region (CCOR) using the registry-based cohort and the chart-reviewed cohort ................................................................................................................................................................................................................................................................................................................................................................................. 86
Table 3: Comparison of prognostic indicators between instrumental variable groups .......... 87

Chapter 6
Table 6.1: Number and proportion of patients with unknown covariates in the subcohort and case groups by candidacy ................................................................................................................................................................................................................................................................................................................................................................................. 88
Table 6.2: Comparison of included patients and patients excluded in the subcohort due to unknown PSA, Gleason score, and/or T category ................................................................................................................................................................................................................................................................................................................................................................................. 89
Table 6.3: Proportion (%) of patients in the subcohort who had specific diagnostic tests done, not done, unknown if done, or who had information missing ................................................................................................................................................................................................................................................................................................................................................................................. 90
Table 6.4: Proportion of patients in the subcohort who had diagnostic tests or had tests recommended grouped by radiotherapy and surgical candidates ................................................................. 90

Table 6.5: Characteristics of radiotherapy and surgical patients in the subcohort when high risk patients are included, and the proportion of surgical patients for each variable .................. 91

Table 6.6: Hazard ratios for death from prostate cancer for radiotherapy compared to surgery and for covariates ........................................................................................................ 92

Table 6.7: Percentage of patients in the subcohort who were radiotherapy candidates based on their Cancer Care Ontario Region (CCOR) of residence and county. .............................. 94

Table 6.8: Comparison of characteristics across instrumental variable groups, using county as the instrument ................................................................................................................ 95

Table 6.9: Comparison of characteristics across instrumental variable groups, using CCOR as the instrument ................................................................................................................ 96
List of Acronyms

bRFS  biochemical relapse-free survival  
CCE  Cancer Care & Epidemiology  
CCO  Cancer Care Ontario  
CCOR  Cancer Care Ontario Region  
CI  confidence interval  
CIRS-G  Cumulative Illness Rating Scale  
HR  hazard ratio  
OCR  Ontario Cancer Registry  
PSA  prostate specific antigen  
RCT  randomized clinical trial  
TMN  Tumor, node, and metastases (staging system)  
TURP  Transurethral resection of the prostate  
TRUS  Transrectal ultrasound
Chapter 1: Introduction

1.1 Background and Rationale

Prostate cancer is the leading non-skin cancer diagnosed in Canadian men. The relative treatment effectiveness of radiotherapy compared to surgery is uncertain due to the lack of randomized clinical trial (RCT) evidence, as only one RCT has been completed and it has been heavily criticized for methodological shortcomings. Other attempts to conduct RCTs comparing radiotherapy to surgery in this study population have failed due to lack of patient recruitment. Due to the lack of RCTs, observational studies have been relied on to provide evidence on treatment effectiveness. Observational studies have had mixed results: several claim that there is no difference in treatment effectiveness between radiotherapy and surgery and some claim that surgery provides a survival advantage.

An RCT is unlikely to ever be performed in this study population to provide a more definitive answer to the treatment effectiveness question. Therefore, we attempted to provide information using data from an observational case-cohort study using a number of approaches to minimize the effect of confounding by indication; we aimed to describe the difference in treatment effectiveness using traditional multivariate analysis, propensity scores and instrumental variables while restricting the study population to those who are candidates for both treatments and using an intent to treat approach.

1.2 Objectives

The primary objective of this study was to examine the difference in treatment effectiveness between surgery and radiotherapy for localized prostate cancer analyzing by intent to treat with a restricted population based on treatment eligibility guidelines. We were also interested in determining the effects of analyzing by intent to treat and using a restricted population on the point estimates. Therefore, analysis was repeated analyzing by actual treatment received and was also repeated using an unrestricted population. We also used propensity scores
to assess the consistency of treatment effect across subgroups whose indication for one treatment versus the other varies. Finally, we explored the instrumental variable assumption of equal distribution of unobservable covariates after finding that it was violated in this study population.

1.3 Thesis Outline

The second chapter provides a literature review summarizing the studies done to date comparing treatment effectiveness between radiotherapy and surgery for prostate cancer, factors contributing to the controversy in prostate cancer treatment decision-making, and methods for limiting bias in observational studies. The third chapter provides a description of the study design, population and analyses used in this study. The fourth chapter is a draft of a manuscript comparing the treatment effectiveness of radiotherapy to surgery in localized prostate cancer. The fifth chapter is a draft of a manuscript that describes our exploration of the instrumental variable assumption of equal distribution of unobservable covariates. The sixth chapter provides supplemental results that were not included in the manuscripts. The seventh chapter discusses the overall results, conclusions, strengths and limitations of this thesis as well as future directions for treatment effectiveness research in prostate cancer. The appendices provide results to support decisions that we made, and excerpts from the chart abstraction form.

1.4 References


Chapter 2: Literature Review

2.1 Prostate Cancer

Excluding non-melanoma skin cancer, prostate cancer is the leading cancer diagnosed in Canadian men, accounting for over a quarter of new cases. In 2009, it is expected that 25,500 men in Canada will be diagnosed with prostate cancer and 4,400 men will die from it. The age-standardized incidence of prostate cancer has not been stable over the last 30 years, due to an increased uptake in prostate specific antigen (PSA) screening which peaked in 1993 and again to a smaller extent in 2001. The first peak followed the introduction of PSA screening while the second is thought to be due to the publicity around the prostate cancer diagnosis of the Canadian Minister of Health following multiple PSA tests. Increased screening during these periods detected prevalent cancers that otherwise would not be clinically apparent until more advanced stages and may never have been apparent to due competing risks of death, thereby temporarily increasing the incidence of prostate cancer.

Most prostate cancers are diagnosed before they have metastasized, with over 90% of men diagnosed with localized or regional disease. Because prostate cancer is a slow-growing disease and is usually diagnosed in older men, survival is quite high with a 5-year relative survival of almost 100% and 10-year relative survival of close to 95% for localized disease. The 10-year disease specific survival rates vary from 43% to 94% depending on a patient’s prognostic indicators. Although these statistics are from data collected in the United States, we expect that patterns in Canada would be similar. The Canadian Cancer Statistics reports that the 5-year relative survival rate for all prostate cancer patients is 95%, however this is the only comparable statistic known in Canada.
2.2 Prostate Cancer Treatments

Prostate cancer has several treatment options including surgery, external beam radiotherapy, brachytherapy, active surveillance, and hormone therapy. According to the United States National Cancer Data Base, in 1995 34.1% of men diagnosed with prostate cancer had a radical prostatectomy, 26.3% received external beam radiation, 2.2% received radiation implants, 11.7% had hormone therapy, 21.6% had no treatment and 4.1% received a combination of these treatments, chemotherapy, biologic response modifiers or experimental treatment. For localized prostate cancer, external beam radiotherapy and surgery remain the two most popular options. In British Columbia in 2000, 37% of curative patients received external beam radiotherapy, 50% received surgery, and the remaining 13% received brachytherapy.

During the period of the current study, conventional external beam radiotherapy was most often used. It involves a total dose of 64 to 70 Gy in fractions of 180 to 200 cGy for seven to eight weeks. During this time, a more advanced type of radiotherapy known as 3-dimensional conformal radiotherapy started to become available. Its increased accuracy of delivering radiotherapy to the prostate enabled the dose of radiotherapy to be increased with limited toxicity to the surrounding tissues. It involves a total dose of 66 to 76 Gy in daily treatments for 33-38 days. Radiation may also be directed at the surrounding lymph nodes, as this has been shown to improve progression free survival in patients with high risk for lymph node metastases and has shown a trend for decreased biochemical relapse.

Radical prostatectomy involves the removal of the entire prostate usually including the seminal vesicles and may be accompanied by a lymph node dissection, in which the lymph nodes surrounding the prostate are removed first in order to determine if the cancer has spread regionally. If there is evidence of nodal spread, the removal of the prostate is abandoned. The two most popular methods of radical prostatectomy are the perineal method, in which the surgical incision is made between the scrotum and the anus, and the retropubic method, in which the
incision is made in the abdominal wall. Laparoscopic surgery is currently being integrated into surgical practice, with promising benefits including shorter hospital stay and recovery, however the technique is difficult and takes a relatively long time to become proficient.

2.3 Treatment Decision Making

Radiotherapy and prostatectomy are the two main options to treat localized prostate cancer. However, the treatment decision process is quite complicated for many reasons. Prostate cancer is a slow-growing disease, and treatment effects on survival may not be noticeable until 10 to 15 years post-diagnosis. Also, the indications for surgery vary based on prognostic indicators. There is a lack of randomized clinical trial evidence demonstrating the superiority of one treatment over the other, and observational studies have provided conflicting results. Both treatments have adverse side effects, and patients may weigh the possibility of each side effect differently. Also, there are disagreements between urologists and radiation oncologists regarding which treatment is best overall and which treatment is best in different situations. Decision aids have been proposed in this population to assist a patient in determining what values he considers important to help make this difficult treatment decision.

Competing risks for death play an important role in determining eligibility for treatment. Prostate cancer is a relatively slow-growing disease and over 75% of prostate cancers are diagnosed over the age of 60. Therefore, men diagnosed with prostate cancer may die of causes other than prostate cancer even if they remain untreated. Many men have other illnesses that put them at a much higher risk for death within 10 to 15 years than prostate cancer does, and providing curative treatment may not be beneficial. A study by Albertsen et al showed that men 70 to 74 with a Gleason score between 2 to 4 who were managed conservatively (received no treatment other than hormone therapy) were over ten times more likely to die of other causes than prostate cancer within 15 years. Because of the potential for inconvenient and bothersome side
effects, providing any curative treatment may not be appropriate for men who do not have a life expectancy greater than 10 years due to their age and/or comorbid illnesses.\textsuperscript{21}

Patients who are indicated only for radiotherapy tend to be different than patients who are indicated for either treatment. Men who have radiotherapy tend to have worse prognostic indicators: higher PSA levels, higher Gleason scores and/or higher T categories.\textsuperscript{22} These men also tend to be older\textsuperscript{23} and have more comorbid illnesses.\textsuperscript{24} When cancers penetrate the prostate capsule, surgery often does not remove the entire tumour and the suitability of these patients for surgery remains unclear.\textsuperscript{2} Partin and Walsh developed a nomogram, commonly called “Partin Tables”, which predicts the probability of spread of prostate cancer using PSA, Gleason score and T category.\textsuperscript{25} For example, a patient with a Gleason score between 2 to 4, a PSA $\leq 2.5$ and a T category of T1c has a 91\% chance of having organ confined disease and a 0\% chance of positive lymph nodes compared to a person with a Gleason score between 8 to 10, a PSA $>10$, and a T category of T2c, who has a 6\% chance of having organ confined disease and a 38\% chance of positive lymph nodes.\textsuperscript{26} Physicians can use these nomograms to predict the probability of organ-confined disease and spread to lymph nodes in order to counsel patients on the best treatment action.

Due to the lack of well-designed RCTs, observational studies have been relied on to provide information on treatment effectiveness. Since the widespread use of prostate-specific antigen (PSA) testing, many studies have used endpoints of biochemical relapse-free survival (bRFS) when comparing treatment regimens. Studies comparing radiotherapy to surgery have generally shown that there is no difference in bRFS, however some studies have shown improved survival and/or bRFS following radical prostatectomy. A brief description of studies that compare radiotherapy to surgery, which either controlled for confounders or present stratified analyses, is provided in Table 2.1. Because these studies have not reached the same consensus regarding treatment effectiveness, whether or not surgery provides a survival advantage over radiotherapy for localized prostate cancer remains uncertain.
Table 2.1 Brief description of observational studies comparing survival or biochemical relapse-free survival (bRFS) in prostate cancer patients between radiotherapy (RT) and radical prostatectomy (RP)

<table>
<thead>
<tr>
<th>Authors</th>
<th>Population</th>
<th>Year of Diagnosis</th>
<th>Outcome and Control for Confounding</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albertsen et al (2007) 27</td>
<td>642 RT, 596 RP PSA&lt;50</td>
<td>1990-1992</td>
<td>13-year prostate cancer-specific survival, adjusted for PSA, Gleason score, age, clinical stage, and Charlson comorbidity score</td>
<td>Hazard ratio RT vs. RP 2.2 (95% CI 1.6-3.1)</td>
</tr>
<tr>
<td>D’Amico et al (2002) 28</td>
<td>381 RT, 2635 RP T1c, T2</td>
<td>1988-2000</td>
<td>8-year bRFS, stratified by prognostic group (based on PSA, T category and Gleason score)</td>
<td>RP vs RT: Low risk: 88% (95% CI 85-90) vs.78% (72-83) Intermediate, low volume: 79% (73-85) vs. 65% (58-72) Intermediate, high volume: 36% (27-44) vs 35% (12-55) High risk: 33% (27-39) vs. 40% (28-52)</td>
</tr>
<tr>
<td>Kupelian et al (2002) 22</td>
<td>628 RT, 1054 RP T1-T2</td>
<td>1990-1998</td>
<td>8-year bRFS adjusted for prior TURP*, age, race, T category, PSA, Gleason, neoadjuvant AD*, and year of therapy</td>
<td>Hazard ratio RT vs. RP 1.01 (95% CI 0.77-1.32)</td>
</tr>
<tr>
<td>Kupelian et al (2004) 29</td>
<td>1034 RP, 484 RT &lt;72, 301 RT ≥72</td>
<td>1990-1998</td>
<td>8-year bRFS adjusted for T category, PSA, Gleason, AD*, and year of therapy</td>
<td>Hazard ratio RT &lt;72 vs. RP 1.83 (95% CI 1.83-2.73) Hazard ratio RT ≥72 vs. RP 1.08 (95% CI 0.78-1.50)</td>
</tr>
<tr>
<td>Martinez et al (2000) 30</td>
<td>225 RT, 157 RP PSA ≤10, Gleason ≤6</td>
<td>1987-1994</td>
<td>7-year bRFS, adjusted for PSA, Gleason score and age</td>
<td>Relative risk RT vs. RP 0.98 (95% CI 0.55-1.74)</td>
</tr>
<tr>
<td>Study Authors (Year)</td>
<td>Patients</td>
<td>Diagnosis Period</td>
<td>Follow-up</td>
<td>Outcome Description</td>
</tr>
<tr>
<td>----------------------</td>
<td>----------</td>
<td>------------------</td>
<td>-----------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Merglen et al (2007)</td>
<td>205 RT, 158 RP</td>
<td>1989-1998</td>
<td>10 year prostate cancer-specific survival</td>
<td>adjusted for age, period of diagnosis, method of detection, lymph node status, clinical stage, PSA and differentiation</td>
</tr>
<tr>
<td>Potters et al (2004)</td>
<td>340 RT, 746 RP</td>
<td>1992-1998</td>
<td>7 year bRFS</td>
<td>adjusted for clinical stage, Gleason score, PSA, race, and age</td>
</tr>
<tr>
<td>Tewari et al (2006)</td>
<td>934 RT, 1014 RP</td>
<td>1980-1987</td>
<td>15 year prostate cancer-specific survival</td>
<td>adjusted for age, race, comorbid illnesses, SES*, grade, year of diagnosis</td>
</tr>
<tr>
<td>Tewari et al (2007)</td>
<td>137 RT, 119 RP</td>
<td>1980-1997</td>
<td>15 year prostate cancer-specific survival</td>
<td>adjusted for comorbid illnesses, diagnosis year and propensity score (included age, race comorbid illnesses, grade, year of diagnosis and estimated income)</td>
</tr>
<tr>
<td>Vicini et al (2002)</td>
<td>1500 RT, 1635 RP</td>
<td>1989-1998</td>
<td>5 year bRFS</td>
<td>stratified by prognostic group (based on PSA and Gleason score)</td>
</tr>
</tbody>
</table>

*transurethral resection of the prostate (TURP); androgen deprivation (AD); socioeconomic status (SES)
Because of the location of the prostate, there is substantial potential for acute and long-term damage to surrounding organs following prostatectomy or radical radiotherapy. The rate of complications varies in the literature quite considerably.\textsuperscript{36-39} Part of this variation is likely due to the time frame studied following treatment and the differing definitions of the study population. One of the more reliable studies was the Prostate Cancer Outcomes Study, as it was prospective, had a large sample size, surveyed patients rather than performed chart abstractions, and surveyed patients at multiple time points.\textsuperscript{39} They studied 1591 men who underwent prostatectomy or radiotherapy and showed that two years post-treatment, men who receive radiotherapy are more likely than men who undergo prostatectomy to have diarrhea (37\% versus 21\%), bowel urgency (36\% versus 15\%), painful hemorrhoids (16 \% versus 10\%) and perianal wetness (22\% versus 14\%).\textsuperscript{39} However, men who undergo prostatectomy are more likely than those who receive radiotherapy to have no urinary control or frequent leaks (10\% versus 3\%) and to have erectile dysfunction (80\% versus 62\%).\textsuperscript{39} Acute complications within the first two months also differ,\textsuperscript{39} and patients may weigh the possibilities of each acute and long-term complication differently when making their treatment decisions.\textsuperscript{40}

Further complicating the decision making process is that urologists and radiation oncologists appear to have conflicting views on which treatment is best to treat prostate cancer, and may influence patient decisions based on these opinions. A study completed in the Canadian population, published in 2008, showed that 60\% of urologists compared to 21\% of radiation oncologists felt that surgery definitely offered a survival advantage over radiotherapy (p<0.0001).\textsuperscript{19} Also, more radiation oncologists than urologists recommended radiation for low-risk patients (p<0.01) and urologists were more likely than radiation oncologists to recommend prostatectomy for intermediate risk patients (p<0.001).\textsuperscript{19} In a survey completed in the United States in 1998, 72\% of radiation oncologists reported that they believed radiotherapy and surgery offer equivalent survival benefits compared to 6\% of urologists when asked about patients with a
life expectancy of at least 10 years with prostate-confined, moderate grade prostate cancer. The vast majority of urologists (93%) reported that prostatectomy provided a survival benefit. Although hearing conflicting views on the best course of treatment may be confusing for a patient, meeting with both a radiation oncologist and urologist’s, i.e. taking a multidisciplinary approach to patient care, may be a good approach in order for the patient to be well-informed for his decision-making process.

Along with disease and doctor related factors, the treatment decision-making process also, of course, involves the patient. The type of information needed to make a treatment decision varies by patient. Decision aids have been proposed in this population to assist a patient in determining what values he considers important and how they relate to each treatment, in order to integrate this information to form a treatment decision. These decision aids can be used to help a patient decide which treatment he prefers, or in the case of a patient who strongly prefers one treatment option, to ensure that his treatment decision is based on fact.

### 2.4 Observational Studies of Prostate Cancer

Prostate cancer patients who are surgical candidates are nearly always also radiotherapy candidates, and those who are candidates for both treatments tend to be different than those who are only radiotherapy candidates. Therefore, confounding by indication may hinder the question of treatment effectiveness in observational studies of prostate cancer. Confounding by indication is defined as:

“A distortion of the effect of a treatment on the outcome that is caused by the presence of a sign or symptom that is associated with the treatment and with the outcome; or a distortion of the effect of a treatment that is caused by the presence of an indication, or a contraindication, for the treatment, that is also associated with the outcome.”

As mentioned in the Section 2.3, radiotherapy patients tend to have worse prognostic indicators: higher PSA levels, higher Gleason scores and/or higher T categories. These men also tend to be
older\textsuperscript{23} and have more comorbid illnesses.\textsuperscript{24} Clearly physicians have a different prognostic and personal profiles for surgical candidates compared to patients who are candidates for both treatments. Therefore, confounding by indication may have influenced observational studies on survival differences between radiotherapy and surgery. As also mentioned previously, surgery often does not remove the entire tumor when cancers penetrate the prostate capsule and the suitability of these patients for surgery remains unclear.\textsuperscript{2} Partin tables can be used to predict the likelihood or organ confined disease and the likelihood that the cancer has spread to the lymph nodes.\textsuperscript{25} Because patients who have a high probability for extracapsular spread or positive lymph nodes may not be offered surgery as a curative treatment, again confounding by indication may bias treatment effectiveness studies of prostate cancer that compare surgery to radiotherapy.

Both observational studies and, more commonly, randomized clinical trials seek to determine treatment effects, however in an observational study the allocation of treatment is not randomly assigned.\textsuperscript{47} Randomization is usually able to produce groups that are comparable in both observable and unobservable baseline characteristics.\textsuperscript{47} Because an observational study does not divide groups in this fashion, the study may be biased as groups may differ in ways that affect an outcome in ways other than the treatment received.\textsuperscript{47} Overt bias is less problematic as it can be controlled for, however some bias may not be controllable in analysis as the confounding information was not recorded or was unobservable.\textsuperscript{47} These systematic differences between treatment groups can lead to results that may not represent the true effect of the exposure on the outcome.\textsuperscript{48}

Although a randomized clinical trial would be the ideal study design to explore the difference in treatment effectiveness, attempts to conduct randomized trials comparing radiotherapy to surgery have been unsuccessful.\textsuperscript{49} The one trial which ran to completion was conducted on 106 men between 1974 and 1978, 59 of whom were assigned to radiation and 47 were assigned to prostatectomy.\textsuperscript{50} This study found that surgery offered a survival advantage over radiation,\textsuperscript{50} however it has been criticized for several reasons. First, patients were analyzed by the
treatment they received and not their intended treatment, and only 88% of patients assigned to the radiation arm received radiation and 81% of patients assigned to the surgical arm received surgery. Second, 56% of patients either died of other causes or were lost to follow-up, which makes the results unreliable. Finally, the rate of metastases in the radiotherapy group was higher than those found in other studies.

Several other randomized clinical trials have been attempted in this population, with little success. In the 1970s, a phase III study conducted in the United States comparing surgery to radiotherapy in men with clinical stage A2 and B, using the Whitmore-Jewitt staging system, with an accrual goal of 1128 patients was closed after only 6 patients were recruited within 24 months (SWOG 8890). Similarly, in the United Kingdom, the Medical Research Council also attempted a randomized clinical trial comparing radiotherapy, surgery and watchful waiting (Medical Research Council PR06) with an accrual goal of 1800 patients which was closed after only 35 patients were recruited within 28 months. The National Prostate Cancer Treatment Group attempted a study comparing radiotherapy to surgery on 200 patients (National Prostate Cancer Treatment Group 2000), which was also closed early. Similarly, the Surgical Prostatectomy Versus Interstitial Radiation Intervention Trial (SPIRIT; ACOSOG Z0070 NCIC PR10), a Canadian randomized trial aiming to accrue 1980 patients to compare surgery to brachytherapy, was closed after accruing only 56 patients in 2 years. Because attempts to conduct randomized clinical trials have been unsuccessful, researchers and clinicians can only use observational data in order to draw conclusions on treatment effectiveness in this population.

2.5 Methodological Strategies for Managing the Impact of Confounding by Indication

2.5.1 Overview

Observational studies used to infer relative treatment efficacy are subject to bias that inevitably accompanies this type of study. The following section describes ways to limit the impact of confounding by indication in the prostate cancer population when examining
differences in treatment effectiveness. Section 2.5.2 explains why analysis by intent to treat should be used to control for bias that arises due to differences in lymph node staging practices between radiotherapy and surgical patients. Section 2.5.3 describes how restricting study populations in observational studies can be beneficial, and describes treatment practice guidelines in the prostate cancer population that can be used to restrict study populations. Sections 2.5.4 to 2.5.6 describe analytic methods that can be used to control for confounding by indication, with a description of each method including its benefits and limitations.

### 2.5.2 Intent to Treat

When comparing treatment regimes in prostate cancer, a bias may be present which favours prostatectomy over radiotherapy. Upon surgical exploration, positive lymph nodes may be discovered which leads to subsequent abandonment of the planned prostatectomy. A similar pattern is not seen in men treated by radiotherapy, as their lymph nodes are not commonly dissected prior to radiotherapy. This difference in approach between the two modalities leads to an underestimation of regional disease in radiotherapy patients compared to surgical patients. Because those with regional disease are not included in the survival analysis of prostatectomy, this may make the survival rate of clinically localized prostate cancer in the surgical group appear better than it actually is compared to those treated by radiotherapy. In order to reduce this bias favouring radical prostatectomy, an “intent to treat” approach can be taken in which treatment groups are compared based on the original treatment plan, which is not necessarily the treatment that was received.

A study conducted by Lu-Yao and Yao illustrated the potential effect this bias can have, by comparing disease-specific survival analyzing by intent to treat versus treatment received in prostate cancer patients. While survival differences between analysis by intent to treat and treatment received were not seen for those treated with radiotherapy or conservative management, they did vary for prostatectomy. When analyzing the data by treatment received, the 10-year
disease-specific survival was 89%, while when analyzing by intent to treat only 83% survival was seen (p<0.0001). This study demonstrates the importance of using analysis by intent to treat method when comparing survival patterns between radical prostatectomy and radiotherapy.

2.5.3 Restriction

In a randomized trial, entry criteria are used to limit the heterogeneity of the subjects. Patients may be excluded due to poor prognosis, comorbid illnesses, atypical disease and/or ineligibility for one of the treatments. By restricting subjects, one ensures that the patients in the study population are more comparable to each other than the patients in the unrestricted study population. A similar approach can be taken in observational studies, by restricting the study population to those persons who are eligible for both treatments. Restriction removes patients who have strong indications or counter indications for one of the treatments, which better ensures that the results of the study are actually due to the treatments and not the differences in the prognostic factor distribution of patients treated with one treatment versus another.

One way to determine eligibility for treatment is through reference to practice guidelines. The National Comprehensive Cancer Network (NCCN) has developed guidelines on the types of patients who should be treated with particular modalities based on their risk and life expectancy. Surgery is provided as an option for all risk groups, however it is only recommended for low volume tumours with no fixation. The BC Cancer Agency Guidelines state that only low and intermediate risk patients are indicated for surgery. Similarly, the Cancer Care Nova Scotia guidelines recommend surgery for low and intermediate risk patients, however they only recommend surgery to high risk patients with low volume tumours. High risk patients include those with Gleason score ≥8, T categories ≥T3a, and/or PSA >20 ng/mL. All guidelines state that patients with cancers that have lymph node involvement or metastases are not candidates for curative radiotherapy or radical prostatectomy. Other relevant authorities such as the
Canadian Urological Association and Cancer Care Ontario do not currently have practice guidelines for the treatment of localized prostate cancer.\textsuperscript{66, 67} Therefore, only the BC Cancer Agency and Cancer Care Nova Scotia guidelines are suitable for restricting a population to those eligible for both treatments.

\textbf{2.5.4 Multivariate Analysis}

Any observational study may have a number of confounding variables which affect both the exposure and outcome and that is certainly the case of treatment effectiveness in prostate cancer where the exposure is the treatment. The study population can be stratified based on those variables in order to assess their impact on the association of interest.\textsuperscript{68} However, with increasing number of variables the size of some subgroups can be small, even with a large overall sample size.\textsuperscript{68} This problem can be overcome by using multivariate analysis, in which all covariates are included in the model.\textsuperscript{68} By simultaneously including variables that affect exposure and outcome within the model, the impact of each of the covariates can be assessed and a less biased estimate of the effect of the variable of interest can be calculated.\textsuperscript{68} A limitation of multivariate analysis is that the statistical assumptions that must be made are sometimes hard to verify, and of course you can only control for those variables that you have information on.\textsuperscript{68}

\textbf{2.5.5 Instrumental Variables}

Instrumental variables may be used to address residual differences between two comparison groups after controlling for known confounders. An instrumental variable is a factor that is correlated with an exposure but is not directly correlated with an outcome in a study.\textsuperscript{69} The ideal instrumental variable is the “coin toss” of a randomized trial, in which the exposure is 100% dependent on the instrumental variable and the outcome is completely independent.\textsuperscript{70} In an observational study, using instrumental variables is analogous to a “natural experiment” in which subjects are placed into groups that differ in their probability of exposure. The effect estimate
generated using this method is the average “marginal effect” which reflects the average treatment
effect only in the group of patients whose treatment choice was influenced by the instrumental
variable.\textsuperscript{71} This estimate may not describe the treatment effect in the remainder of patients whose
treatment choice was not affected by the instrument.\textsuperscript{72}

For example, in a study conducted by McClellan et al, invasive procedures for acute
myocardial infarction were investigated in order to estimate their effect on survival.\textsuperscript{71} The
instrumental variable of differential distance between the nearest hospital that routinely
performed cardiac catheterizations or revascularizations and the nearest hospital of any type was
used in order to divide the subjects into groups depending on that distance.\textsuperscript{71} They assumed that
people living closer to a hospital which routinely performed these invasive procedures would
have a higher chance of having these procedures done than people with the same characteristics
who lived farther from a hospital that routinely performed invasive procedures.\textsuperscript{71} Therefore the
instrumental variable, distance to a catheterization or revascularization hospital, is associated with
the exposure, having an invasive procedure, but is only associated with the outcome through its
influence on the chance of catheterization or revascularization.

There are two assumptions that are made when using instrumental variables.\textsuperscript{69} The
instrument is assumed to be associated with the treatment and the instrument is assumed not to be
correlated with the error, or the unobservable confounders.\textsuperscript{69} Because there is no concrete way to
assess the latter assumption, it is often satisfied by assessing if the observable covariates are
similar, as then it is likely that the unobservable determinants are also similar.\textsuperscript{69} As well as
satisfying these assumptions, in order to get a precise estimate of the effect of the exposure on the
outcome, the sample size must be large (in order to reduce sampling error) or there must be large
variations in the exposure.\textsuperscript{69}

Stukel et al examined the effect of different methods for removing treatment selection
bias in a cohort of patients with acute myocardial infarction who were eligible for cardiac
catheterization.\textsuperscript{73} Multivariable model risk adjustment, propensity score risk adjustment,
propensity-based matching, and instrumental variable analysis were employed to estimate the effect of cardiac catheterization on mortality rates. They found that the results from the instrumental variable analysis fell within the range of survival benefit seen in randomized clinical trials, whereas the propensity score methods and multivariable model risk adjustment overestimated the survival benefit of cardiac catheterization. This study showed that in observational studies, using an appropriate instrumental variable may provide less biased results than other methods that account for measured confounding.

2.5.6 Propensity Scores

Propensity scores can be used to control for known confounders within a study by replacing all of the covariates with one value. Propensity scores are calculated using logistic regression where the treatment is the dependent variable and the known confounders are the independent variables. A score is generated for each subject in the study, which represents the estimated probability that a patient received a particular treatment given his covariates. If two people have the same propensity score their likelihood of receiving one treatment versus the other is the same.

When assessing treatment effectiveness, propensity scores can be used to control for covariates in three ways: matching, stratification and regression adjustment. In the matching technique, a person in one treatment group is matched to a person in the other treatment group based on their propensity scores. In stratification, the entire study population is divided into quantiles based on their propensity scores and analyses are conducted within each quantile. Rosenbaum and Rubin contend that just as stratifying any covariate into five strata removes 90% of the bias caused by that variable, dividing propensity scores into quintiles can control for over 90% of the overt biases within observational studies. Finally, in the regression adjustment, the actual propensity score is used as a covariate within a statistical regression technique.
approach reduces the degrees of freedom needed for covariate adjustment compared to traditional multivariate analysis.

The best way to determine the propensity score model is still under debate. Brookhart et al. performed a simulation experiment where they incorporated variables into the model that were confounders, associated with only the outcome or associated with only the exposure. They showed that using a variable that was a confounder and one that was associated with only the outcome produced the least variance and mean squared error. However, Rubin argues that the efficiency gains of leaving out a variable that is strongly related to an exposure but only weakly related to the outcome do not compensate for the possible biasing effects of removing it. It is often stressed that researchers should remember that the test of a good propensity score is not how parsimonious it is or the goodness of fit, but how well it balances covariates within propensity score ranges.

Most studies that compare the use of propensity scores to traditional multivariate analysis have found that the two methods produce similar results. A recent study comparing propensity score methods to traditional regression showed that only 10% of the 78 exposure-outcome comparisons in 43 studies differed in statistical significance between the two methods. Similarly, another study showed that the results of only 13% of 69 studies differed by more than 20% between conventional methods and propensity scores.

However, one advantage of using the propensity score method over the traditional method of multivariate analysis is that it does not rely on the assumptions of standard statistical models. For example, when using propensity scores it does not matter if the dependent variable has a normal distribution around the independent variables, or that the variation in the dependent variable is the same for each independent variable value. Also, it makes no assumption about the relationship between the dependent and independent variables (such as a linear relationship when using linear regression). Finally, when using a regression model with many confounding variables, small differences in each variable may compile to a large overall difference that is not
accounted for. Another advantage of using propensity scores is that the distribution of covariates between treatments within each score-based strata can be compared to determine the success of the control of covariates. There should be very few differences in the distribution of covariates if the propensity score adequately controls for the covariates. Another advantage of using a stratified propensity score analysis is that treatment effects can be examined as the indication for the treatment varies. For example, a study done by Kurth et al examined the effect of tissue plasminogen activator and death in stroke patients. They found that the odds ratio for death increased substantially as the propensity for treatment (compared to no treatment) decreased with a hazard ratio of 0.25 for those who were in the 99-100th percentile to 25.11 for those who were in the 10-25th percentile and this difference was statistically significant (p=0.008). The overall hazard ratio for death for those receiving treatment compared to those who received no treatment was 3.35. Clearly, there was effect modification by indication for the treatment.

There are several limitations to the use of propensity scores. Like other methods of statistical control, propensity scores can only reduce the impact of observable confounders and not unobservable ones. Propensity scores are more efficient with larger samples, as the distribution of covariates is more balanced with increased sample size. Therefore, small sample sizes may result in imbalanced distribution of covariates and the use of propensity scores may not be appropriate. Finally, covariates that are related to treatment assignment but are minimally related to outcome are treated in the same manner as those that are strongly related to outcome. The inclusion of covariates that are not strongly related to outcome reduces the efficiency of the control of those that are strongly related.

### 2.6 Summary

Prostate cancer is the leading non-skin cancer diagnosed in Canadian men, and most men are diagnosed with localized or regional disease. Treatment of localized prostate cancer is
controversial for many reasons, however the primary cause is the lack of randomized clinical trial evidence demonstrating the superiority or equivalency of surgery compared to radiation. Observational studies have been used in lieu of randomized clinical trials in order to provide information on treatment effectiveness. Although an observational study is not the ideal study design to compare treatment effectiveness, the only other alternative is not to study treatment effectiveness in this population at all. There are a number of ways to reduce the inherent biases of observational studies including using analysis by intent to treat, restricting the study population, and using statistical methods to control for confounding by indication.

2.7 References


56. PR06 Collaborators. Early closure of a randomized controlled trial of three treatment approaches to early localised prostate cancer: the MRC PR06 trial. BJU International 2004; 94(9):1400-1401.


Chapter 3: Methods

3.1 Introduction

The data for this study were previously collected to examine the role that 1) comorbid illnesses play in other cause mortality following curative radiotherapy or surgery and 2) to explore the difference in effectiveness between these treatments. This was a case-cohort study in which the subcohort and cases were randomly sampled stratified by region of residence. Related to its two objectives, the two case groups were defined 1) as those who died of other causes and 2) as those who died of prostate cancer within the study period.

The current study addressed objective 2. More specifically, the objective was to explore differences in treatment effectiveness between radiotherapy and surgery for localized prostate cancer, using different methods to control for confounding by indication. Through the conduct of this work, issues arose with the instrumental variable approach that led to an investigation of one of its primary assumptions. The following is an explanation of the study design, population and analyses used in this study.

3.2 Study Objectives

3.2.1 Primary Objectives

1. To examine the difference in treatment effectiveness between radiotherapy and surgery for localized prostate cancer
2. To examine the consistency of treatment effectiveness differences between radiotherapy and surgery across groups whose indication for radiotherapy differs, using propensity scores to calculate the probability of receiving radiotherapy
3. To explore if the assumption of equal distribution of covariates within instrumental variable groups holds when using a prostate cancer administrative database with sparse information on covariates.
3.2.2 Secondary Objectives

1. To determine if the same conclusions in treatment effectiveness differences between radiotherapy and surgery for prostate cancer were reached using analysis by intent to treat versus analysis by treatment received.

2. To determine if the same conclusions in treatment effectiveness differences between radiotherapy and surgery for prostate cancer were reached using an unrestricted versus a restricted population defined using published surgical candidacy guidelines.

3.3 Study Design

The case-cohort design was first presented by Prentice as an efficient way to conduct a cohort study in which the cohort is large but the number of events is small. In a typical case-cohort study, all of the cases are sampled and a random sample or stratified random sample of the entire cohort, called the subcohort, is chosen. Due to the sampling method there is often overlap between the case group and subcohort group, where patients are selected for both the case group and subcohort. In our study, cases were defined as patients who died of prostate cancer by December 31, 1999. A second case group, which was not used in our study, was defined as patients who died of causes other than prostate cancer by December 31, 1999. The subcohort consisted of a stratified random sample of prostate cancer patients by Cancer Care Ontario Region who met the inclusion criteria listed in Section 3.4.

3.4 The Study Population

The target population was identified using the Cancer Care and Epidemiology (CCE) database at Queen’s University. 45,035 residents of Ontario were diagnosed with adenocarcinoma of the prostate between January 1, 1990 and December 31, 1998. Inclusion criteria were prostatectomy or lymph node dissection within seven months of diagnosis or radiotherapy within nine months of diagnosis. Exclusion criteria were 1) patients who had a lymph node dissection
only and who also had a bladder cancer diagnosis; 2) those whose total curative radiation dose was less than 200cGy; 3) those whose curative radiation treatment was longer than 90 days; 4) those who were listed as only having had radiotherapy but had no record of a curative dose; 5) those whose treatment occurred more than 30 days prior to diagnosis. The population meeting these inclusion and exclusion criteria numbered 17,934.

For this study, cases were defined as those patients who died of prostate cancer by December 31, 1999. The parent study also included another case group of patients who died of other causes by December 31, 1999. Both case groups and the subcohort were chosen based on a stratified random sample from each of the Cancer Care Ontario Regions (CCORs). Up to 75 prostate cancer death cases and up to 75 other cause death cases were randomly selected from each of the eight CCORs. Up to 150 subcohort patients were randomly chosen from each CCOR, and CCORs that did not have 75 prostate cancer death cases and/or 75 other cause death cases sampled more than 150 subcohort patients to ensure adequate numbers from that region. Because of its large population size, the Central East region had a sample size that was doubled, with approximately 150 prostate cancer deaths, 150 other cause deaths and over 300 subcohort subjects. By design, the sampling of cases was done independently of the subcohort and this resulted in an overlap of some patients between the subcohort and cases. The sampling based on region is shown in Table 3.1 and a diagram illustrating the overall sampling of the study population is shown in Figure 3.1. For the current study, we excluded patients who were chosen as only other cause death cases, those with histology other than adenocarcinoma, those who were not candidates for prostatectomy or curative radiotherapy, patients whose surgical candidacy could not be determined, high risk patients, and patients with unknown T category and/or Gleason score or grade, as shown in Figure 3.2. In all, we were able to include 81.7% of eligible subjects.
Table 3.1: Sampling distribution based on Cancer Care Ontario Region (CCOR) for the subcohort, prostate cancer (PC) death cases and other cause (OC) death cases (reproduced with permission)

<table>
<thead>
<tr>
<th>CCOR</th>
<th>Subcohort (sampling %)</th>
<th>PC deaths (sampling %)</th>
<th>OC deaths (sampling %)</th>
<th>Overlapping PC death*</th>
<th>Overlapping OC death*</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Central East</td>
<td>336 (4.55)</td>
<td>150 (38.66)</td>
<td>150 (30.06)</td>
<td>3</td>
<td>6</td>
<td>627</td>
</tr>
<tr>
<td>Central West</td>
<td>173 (6.98)</td>
<td>75 (43.60)</td>
<td>150 (35.71)</td>
<td>8</td>
<td>5</td>
<td>310</td>
</tr>
<tr>
<td>East</td>
<td>169 (9.47)</td>
<td>75 (68.18)</td>
<td>75 (66.37)</td>
<td>7</td>
<td>7</td>
<td>305</td>
</tr>
<tr>
<td>North East</td>
<td>186 (19.52)</td>
<td>67 (100)</td>
<td>75 (68.81)</td>
<td>13</td>
<td>16</td>
<td>299</td>
</tr>
<tr>
<td>North West</td>
<td>280 (82.11)</td>
<td>30 (100)</td>
<td>35 (100)</td>
<td>19</td>
<td>30</td>
<td>296</td>
</tr>
<tr>
<td>South</td>
<td>212 (29.78)</td>
<td>46 (100)</td>
<td>71 (100)</td>
<td>17</td>
<td>21</td>
<td>291</td>
</tr>
<tr>
<td>South East</td>
<td>173 (14.13)</td>
<td>73 (100)</td>
<td>75 (75.76)</td>
<td>8</td>
<td>11</td>
<td>302</td>
</tr>
<tr>
<td>South West</td>
<td>174 (5.69)</td>
<td>75 (41.21)</td>
<td>75 (24.92)</td>
<td>6</td>
<td>8</td>
<td>310</td>
</tr>
<tr>
<td>Total</td>
<td>1703 (9.50)</td>
<td>591 (55.34)</td>
<td>631 (43.91)</td>
<td>81</td>
<td>104</td>
<td>2740</td>
</tr>
</tbody>
</table>

*Due to the sampling method of the case-cohort design, some PC deaths and OC deaths were randomly selected for both the subcohort and their respective case groups.
Figure 3.1: Sampling of the study population in the original study. The addition of the prostate cancer death cases, other cause death cases and the subcohort does not total 2740 due to overlap between the case groups and the subcohort.
3.5 Data Collection

The CCE database, which was used to determine the sampling frame for this study, uses data from the Ontario Cancer Registry (OCR), which is composed of a variety of sources. The OCR is a passive registry that receives cancer diagnosis information from all hospital admissions from the Canadian Institute for Health Information (CIHI) when cancer is recorded as a discharge.
diagnosis, electronic records from the radiotherapy cancer centres as well as Princess Margaret Hospital (PMH), pathology reports of all cancer diagnoses in the province, and records of deaths from the Registrar General of Ontario.\textsuperscript{5} It is estimated that the OCR captures 95\% of all incident cases within Ontario.\textsuperscript{6} Each data source that contributes to the OCR is given a unique identifier for each patient, and CCE then links the cancer diagnosis from the OCR to surgical data from CIHI and radiotherapy administration data from the cancer centres and PMH using this unique identifier.

Data abstractors received one week of training at the co-coordinating office in Kingston, Ontario, which included use of training charts previously abstracted by the study coordinator. Abstractors learned how to properly assign variables by abstracting these charts as a group, and then in small groups and individually where abstractions were compared and differences discussed. Abstractors received a comprehensive manual that contained detailed information on how to accurately assign values to each variable. An electronic abstraction form was used to minimize error, using Microsoft Access format and MedQuest software. Relevant excerpts from that form are provided in Appendix II.

The data abstractors abstracted the treating chart: hospital charts for surgical candidates and cancer centre charts for radiotherapy candidates, and also cancer centre charts when available for surgical cases. All charts were abstracted at their respective hospitals or cancer centres. Missing information was obtained from secondary hospital, urologist and/or general practitioners’ charts. Electronic abstractions were sent to the co-coordinating office on a weekly basis. The pathology reports of the biopsies, the transrectal ultrasound reports, radiotherapy treatment records and surgical reports were photocopied and sent to the co-coordinating office.

Realtime logic and missing data checks were done on a weekly basis following electronic submission of information by abstractors. Information on any inconsistencies or missing variables were sent to the abstractor, and abstractors then either confirmed the illogical inconsistency was true or corrected the information, and sent any missing information that was
available. This process continued until all inconsistencies were resolved for each patient.

Therefore, at the end of every week there were patients with cleaned data that were ready for analysis. The study coordinator also ensured the quality of the chart review data by conducting on-site visits, where she abstracted random patient charts to compare the information she found to the information recorded by the regional abstractor. Based on these checks, the study co-ordinator provided remedial training as needed.

3.6 Study Variables

3.6.1 Treatment Candidacy

The original treatment intent for each patient was previously generated using information abstracted from the charts regarding the treatment actually given to the patient, the original treatment plan, the outcome of any lymph node dissection and the treatments offered to the patient, using the information from the chart abstraction. Figure 3.3 describes the algorithm used. Original treatment intent was used instead of actual treatment received in order to account for patients who had lymph node dissections, whose subsequent surgeries were abandoned due to their positive lymph node status, and who then went on to receive radiotherapy or have no treatment at all. This approach is analogous to that used in randomized clinical trials and the rationale for its use is the similar; not including these patients in the analysis may bias the treatment effectiveness result against radiotherapy because radiotherapy patients are less likely than surgical patients to receive a lymph node dissection prior to treatment. Therefore, the radiotherapy treated group includes some patients with undetected positive lymph nodes. Patients who had surgery with post-operative radiotherapy were considered to be surgical candidates.
Figure 3.3: Radiotherapy and surgery candidacy designation (adapted from flowchart by Susan Rohland)
3.6.2 Mortality Information

The Ontario Cancer Registry undergoes a death clearance using death certificate data (vital status and underlying cause of death) from the Ontario Registrar General. Deaths occurring by December 31, 1999 were included in this analysis. More recent death data could not be used because the case sampling was done with data available at the commencement of the chart review. We calculated time to prostate cancer death using the date of diagnosis and the date of prostate cancer death. Patients who did not die of prostate cancer were censored at date of other cause death or December 31, 1999, whichever came first. In keeping with usual practice when using passive registry data, we otherwise assumed that patients were alive up to December 31, 1999 if their vital status in the OCR or chart review did not indicate that they had died of prostate cancer or of another cause.

3.6.3 Covariates

The key disease-related prognostic factors in prostate cancer are Gleason score, T category and prostate specific antigen (PSA).\(^7\) Gleason score provides an indication of the aggressiveness of the prostate cancer.\(^8\) The two most common patterns of growth based on area are each given a score from 1 to 5, with 1 representing tumours that are well-differentiated and 5 representing tumours that are poorly differentiated.\(^8\) The two scores are then added together to give an overall Gleason score between 2 to 10.\(^8\) A prostate cancer with a score of 2 has glands and tissues that closely resemble a normal prostate while a prostate cancer with a score of 10 shows minimal structural organization of a normal prostate.\(^8\) In cases where the total Gleason score was unavailable, the primary and secondary patterns were summed to give a total score. If these were not available, grade was used to approximate Gleason Score with grade 1 (well differentiated) assigned a Gleason score of 3, grade 2 (moderately differentiated) assigned a Gleason score of 6, and grade 3-4 (poorly differentiated or undifferentiated) assigned a Gleason score of 9. Reviewed Gleason score or grade was used in preference to the original biopsy
Gleason score or grade respectively. Gleason score was analyzed using clinically meaningful categories of 2-4, 5-6, 7, and 8-10 as described by the American Joint Committee on Cancer (AJCC).9

The T category of a tumour describes the size of the tumour and how far it has spread locally. It ranges from T1 (tumour is not clinically apparent or palpable) to T4 (tumour invades local structures besides the seminal vesicles).9, 10 T categories of T1, T2, and T3 are further divided into subcategories based on the spread of the cancer and are named alphabetically from a to b (and c for the T category of T1).9, 10 In cases where the T category subgroup was not available, the T category was assigned to the lowest subgroup within that category, however this rarely occurred. For example, a T1 tumour was assigned to the T1a subgroup. This approach is in keeping with general rule 4 of the TNM classification system.10

Prostate specific antigen (PSA) is a protein that is produced by the epithelial cells of the prostate.11 The levels of PSA in the blood increase naturally with age, however in men with prostate cancer the level of PSA exceeds normal age-related levels, increasing slowly and then exponentially.12 In this study, the PSA value recorded closest to the beginning of treatment was used. PSA values that were recorded within 30 days post-biopsy were not included if another value was available to use instead because PSA values have been shown to be unstable at that time.13 PSA values were also excluded if they were taken after orchiectomy, which is the removal of the testes, or after initiation of hormone therapy, which prevents the testes from releasing testosterone or prevents testosterone from acting on the prostate.14 Hormone therapy acts to decrease PSA levels by preventing PSA from being produced and by causing prostate cell death.15 PSA was divided into commonly used clinical categories of ≤4, >4 to ≤10, >10 to ≤20 and >20.16,17

Age has not been found to be associated with biochemical relapse or survival.16, 18-21 However, since men who receive radiotherapy as their primary mode of treatment tend to be older than those who receive prostatectomy16, 21-23 and because analyses of survival in prostate cancer...
sometimes account for age, age was considered as a potential covariate. Age at diagnosis was already included in the dataset, and was calculated using date of birth and date of diagnosis.

We compared continuous and categorical variables for age in order to decide how it should be used in the regression models. The categories used for age were <55, 55-59, 60-64, 65-69, 70-74, and >74, with <55 as the referent. The wide ranges at the extremes were chosen due to the small numbers in both the youngest and eldest groups. As shown in Figure I.I in Appendix I, there was little difference between the continuous and categorical variables for the hazard ratios for age, and the continuous line fell within the 95% confidence intervals of each of the categories. Because there were no important differences between the categorical and continuous variables for age, we decided to use continuous variables for simplicity and to conserve degrees of freedom in analyses.

Patients with increased co-morbid illnesses have been shown to be more likely to die of prostate cancer and other causes.24 Because radiotherapy patients are likely to have more co-morbid illnesses, when co-morbidity is not accounted for in survival analysis it may make surgery appear to have a greater advantage over radiotherapy than it actually does.24 In this study, comorbid illnesses were scored using a modified version of the Cumulative Illness Rating Scale (CIRS-G) which provides a comorbid illness rating for 14 independent organ areas.25, 26 A five-point scale is used for each category to rate co-morbid illnesses or diseases on a scale of “0” for no problem to “4” for extremely severe conditions.25 Each morbidity category has its own individual scoring chart based on the general scoring guide.25 The total CIRS-G score is calculated by summing the individual scores from each category.25 In a pilot study done for this study, five different comorbid illness indices were compared in order to determine which system predicted the most variation in other cause death.27 Three of the five indices, including CIRS-G, predicted at least 7% of the variation and the authors suggested that any of these three methods would be appropriate to use in this population.27 CIRS-G was chosen because it was easy to use, its instructions were very clear, and the abstractors had strong opinions that it captured the
comorbid illness information better than the other methods.\textsuperscript{27} This scale has been used in other populations to predict mortality. For example, Parmelee et al performed a validation study in 439 patients from a large, multilevel care facility and found that the CIRS-G predicted higher 2-year mortality rates among patients with increasing total CIRS-G (p<0.001).\textsuperscript{28}

We compared continuous and categorical variables total CIRS-G in order to decide how it should be used in the regression models. A score of 0 was used as the referent for total CIRS-G and scores of 9 or more were grouped into the same category due to small numbers. Similar to age, little difference was found for the categorical and continuous CIRS-G variables, and the continuous line fell within the 95\% confidence intervals of the categories as shown in Figure I.II in Appendix I. Because there were no important differences between the categorical and continuous variables for CIRS-G, we decided to use continuous variables for simplicity and to conserve degrees of freedom in analyses.

\textbf{3.6.4 Risk Groups}

We stratified some analysis by risk group and investigated the impact of including the high risk group. Patients with a PSA value $\leq 10$, Gleason score $\leq 6$, and T category $\leq T2a$ are considered to be low risk patients.\textsuperscript{29} Patients who are not low risk patients and have a PSA value $\leq 20$, Gleason score $\leq 7$, and T category $\leq T2b$ are considered to be intermediate risk patients.\textsuperscript{29} Patients with PSA $>20$ and/or Gleason score $>7$ and/or a T category of T3 to T4 are considered to be high risk patients.\textsuperscript{29} Generally, high risk patients are not considered to be surgical candidates because of the high probability that the disease is not contained within the prostate.\textsuperscript{30} For instance, in Canada, the BC Cancer Agency and Cancer Care Nova Scotia guidelines do not recommend surgery for high risk patients with the exception of patients with low volume tumours.\textsuperscript{3, 4} These guidelines also state that both radiotherapy and surgery are curative treatment options for low and intermediate risk patients.\textsuperscript{3, 4}
3.6.5 Diagnostic Tests

We were concerned that the diagnostic tests done to inform TNM staging may have differed between radiotherapy and surgical candidates, and this differential staging may explain prostate cancer-specific survival differences between the two treatments. We compared the diagnostic tests done between the radiotherapy and surgical candidacy groups in the subcohort to explore if this may have occurred. For each diagnostic evaluation, the categories in the chart abstraction included “test not done”, “abnormal, suggestive of cancer”, “abnormal, not suggestive of cancer”, “normal”, “procedure attempted and incomplete”, “test done, results unknown” and “test ordered, unknown if completed”. Patients who had missing information on diagnostic tests were assumed not to have had that test, however missing information rarely occurred. Patients who had a procedure attempted and yet it was not completed were not considered have had that procedure done, as we were interested in tests that could inform TNM staging and incomplete tests would not be able to do this. Patients who had a test ordered but it was unknown if it was completed were assumed to have had the test done. We also examined the proportion of patients who had pre-operative or intra-operative lymph node dissections.

3.7 Analysis

3.7.1 Overview

Section 3.7.2 describes how we examined the baseline characteristics of the study population. Section 3.7.3 describes the traditional multivariate Cox-proportional hazard regression used to examine the difference in treatment effectiveness between radiotherapy and surgery. In order to test some of the decisions that we made when designing the study, we conducted two subanalyses using the multivariate approach. The first subanalysis was analyzed by treatment actually received, where patients who only received a lymph node dissection were removed from analysis. We also excluded radiotherapy patients whose treatment intent was curative, but who received a total dose of radiation less than 40 Gy. The second subanalysis
included high risk patients in the study population to examine the effect of restricting the patient population to low and intermediate risk patients. Section 3.7.4 describes how propensity scores were generated. Propensity scores were used to assess the consistency of treatment effects across groups of patients whose indication for one treatment versus another varies. The original analysis plan included using instrumental variables, however this was not possible due to a violation of a primary assumption of this approach. Therefore, our objective changed to an assessment of the validity of that assumption. A more detailed explanation of this process is provided in Section 3.7.5.

3.7.2 Descriptive Analysis

Age, total CIRS-G, T category, Gleason score and PSA were compared between the radiotherapy and surgery intent to treat groups in both the subcohort and case groups. Categorical variables were compared using chi-square tests while continuous variables were compared using t-tests. The proportion of patients actually receiving their original treatment intent was described. The proportion of patients receiving diagnostic tests were compared between surgical and radiotherapy candidates using chi-square tests.

3.7.3 Traditional Multivariate Analysis

Cox-proportional hazards regression was used to compare risk of prostate cancer death between radiotherapy and surgical candidates.\textsuperscript{31} The study subjects were not independent because the cases were over-sampled and were chosen based on their outcome status, therefore the number of events in the study is much higher than would have occurred if we had used a cohort study. This design issue is taken into account when calculating Cox-proportional hazard ratios by using a pseudolikelihood instead of the partial likelihood to calculate the hazard ratio point estimates.\textsuperscript{1} Using this method, a case that occurs outside of the subcohort is not considered to be
at risk until just prior to failure and is not included in these risk sets until just prior to failure. Prentice, Self and Prentice, and Barlow have proposed weighting schemes for the subcohort and different methods to handle nonsubcohort cases at failure.

These methods provide accurate point estimates, however the variance estimates are artificially smaller than they actually are due to the over-sampling of cases. Each method provides a way to calculate a robust standard error to adjust for the case-cohort sampling. In a study with a similar sized sampling frame to our study, the three methods to account for the case-cohort design gave identical results even when the sampling proportion was 1%. In our study the subcohort sampling proportion was close to 10%, therefore the three methods to account for the case-cohort design should provide the same point estimates. Therneau and Li provide a straightforward way to calculate point estimates and variance for Cox-proportional hazard regression for case-cohort studies using the method of Self and Prentice. Therefore, this was the method that was used in this study.

We decided a priori that T category, Gleason score and PSA would be forced in the model regardless of their statistical significance, as these variables have been shown to affect survival and treatment choice in prostate cancer. Age and total CIRS-G were considered as confounders in the regression analysis if they were associated with both the type of treatment and survival with a p-value of 0.25. Age and total CIRS-G remained in the regression model as confounders if they changed the treatment effectiveness estimate by greater than 10%. This analysis was repeated, stratified by low and intermediate risk group. Patients in the subcohort were censored at 10 years post-diagnosis, other cause death, or December 31, 1999, whichever came first. The very few patients in the subcohort who died of prostate cancer but were not chosen as a case were also censored at date of death, due to the case-cohort study design.
3.7.4 Propensity Scores

The probability of getting radiotherapy (the propensity score) was calculated using T category, PSA and Gleason score using logistic regression. The subcohort represents the average practice of prostate cancer treatment across Ontario, so it was used to generate the propensity score equation. This equation was then used to create the propensity scores for each member of the subcohort and case groups.\(^3\) We chose to do a stratified analysis by propensity scores instead of a regression including the propensity score for two reasons. First, studies that use multivariate analysis and propensity scores in a regression have been shown to frequently produce the same results.\(^3\) Second, we were interested to see if the treatment effectiveness hazard ratio varied as a patient’s indication for radiotherapy increased. Propensity scores were divided approximately into quintiles using the subcohort, and cases were placed into their corresponding quintiles based on their propensity scores. All covariates were compared between the radiotherapy and surgery groups for each quintile in the subcohort with an expectation that there should be little or no difference in covariate distributions between the treatment groups within each quintile.\(^4\) In comparing these distributions, chi-square tests were used for categorical variables and t-tests were used for continuous variables. Cox-proportional hazards regression, as described in Section 3.7.3, was then conducted within each quintile to compare risk of prostate cancer death between radiotherapy and surgical candidates.

3.7.5 Instrumental Variables

Originally, our analysis plan included using instrumental variables as a method to control for confounding by indication. However, this was not possible to do using our study population. A description of this investigation and its findings are provided in the “Supplemental Results” chapter (Section 7.3). Most covariates differed either marginally or statistically significantly between the instrumental variable groups we formed, which violates one of the primary assumptions of the instrumental variable method: that there are no differences in the distribution
of known or unknown covariates between the groups formed. Once we realized that we were not able to identify geographic regions whose treatment pattern varied but distribution of covariates remained similar, we wondered whether this might be a problem generally with this approach. Since our dataset contained detailed information on case mix, we used it to check this key assumption of the instrumental variables approach.

To mimic what would be done in an instrumental variables study using administrative data without detailed information on case mix, we used the original dataset describing our target population of 17,934 patients treated for cure from which the sample of 2740 patients was drawn. We determined the radiotherapy rates within each Cancer Care Ontario Region (CCOR) using these data. For these analyses, we focused on those who received radiotherapy or surgery as candidacy would not normally be determined in an instrumental variable study and, in any case, we were not able to assign it without chart information. Groups were formed based on radiotherapy rates in each CCOR. Using the subcohort from the case-cohort study, the covariates of the patients were compared using the groups of CCORs formed by the larger dataset. We used the inverse of the sampling proportion to weight each person within a CCOR in order to account for differences in covariate distributions between regions that were grouped together. Continuous covariates were compared using the Taylor expansion method to calculate variance while categorical variables were compared using chi-square tests, using the Rao-Scott method to account for the sampling weight. We used logistic regression, using the Taylor expansion method to calculate variance, to test for trends from lowest to highest regions of radiotherapy use using an ordinal region variable as the outcome for the continuous variables. We assessed trends in categorical variables using Pearson correlation. We also calculated the average age in each instrumental variable-derived region using the registry-based cohort. We were unable to calculate the distribution of other covariates in the registry-based cohort as this database does not carry this detailed information.
3.8 References


Chapter 4: Manuscript 1

This manuscript is intended for The Journal of Urology and was created in accordance with their submission guidelines.

Title: Prostate cancer-specific survival differences in patients treated by radical prostatectomy versus curative radiotherapy

Authors:

Julie M. DeGroot1,2
Michael Brundage1
Miu Lam2
Susan L. Rohland1
Jeremy Heaton3
William J. Mackillop1,2
D. Robert Siemens3
Patti A. Groome1,2

Affiliations:

1Division of Cancer Care and Epidemiology, Queen’s Cancer Research Institute, Queen’s University, Kingston, Ontario, Canada
2Department of Community Health and Epidemiology, Queen’s University, Kingston, Ontario, Canada
3Department of Urology, Queen’s University, Kingston, Ontario, Canada

Corresponding Author:

Dr. Patti Groome
Cancer Care and Epidemiology
Queen’s University
10 Stuart Street
Kingston, Ontario, Canada
K7L 3N6

Key Words: prostatic neoplasms, disease-specific mortality, radiotherapy, prostatectomy
Abstract

Purpose: The relative treatment effectiveness of surgery versus radiotherapy for prostate cancer is uncertain and randomized clinical trials are unlikely to be performed to provide conclusive evidence. This study describes the difference in cause-specific survival between patients treated with radiotherapy and those treated with surgery, using a number of design and analytic steps to mitigate confounding by indication within an observational study.

Materials and Methods: We conducted a population-based case-cohort study, sampling patients from the Ontario Cancer Registry who were treated or were candidates for cure by radiotherapy or surgery. Cases were defined as those who died of prostate cancer within 10 years. Cause-specific survival was analyzed using Cox-proportional hazard regression, with variance adjustment for the case-cohort sampling. Analysis using intent to treat was compared to that using treatment received. Propensity scores were also calculated and Cox-proportional hazard regression was conducted within each propensity score quintile to test for variation in hazard ratios across patients whose indication for radiotherapy differs.

Results: The adjusted hazard ratios for risk of prostate cancer death for radiotherapy compared to surgery were 1.44 (95% CI 0.86-2.40) and 1.84 (1.06-3.17) using intent to treat and treatment received respectively. Stratified hazard ratios comparing radiotherapy to surgery for death from prostate cancer from the lowest propensity quintile to the highest propensity quintile were 0.30 (0.04-2.28), 1.54 (0.35-6.77), 0.90 (0.29-2.82), 2.71 (1.01-7.31) and 1.08 (0.41-2.81). Differences among these hazard ratios were not statistically significant (p=0.13).

Conclusion: Analysis by intent to treat produced a hazard ratio closer to the null than the analysis by treatment received, indicating that uncontrolled confounding toward more serious cases
getting radiotherapy was present. Future studies should focus on obtaining sufficient power for subgroup analysis such as stratification by risk groups.

Introduction

Prostate cancer is the leading non-skin cancer diagnosed in Canadian men, accounting for over 25% of new cancer cases. In 2009, it is expected that 25 500 men in Canada will be diagnosed with prostate cancer and 4400 men will die from it, and approximately 90% of these cancers will be diagnosed with local or regional disease. Because prostate cancer is a slow-growing disease and is usually diagnosed in older men, many of whom have competing illnesses, survival is quite high for localized disease with a 5-year relative survival rate approaching 100% and a 10-year relative survival rate close to 95%.

Curative treatment for prostate cancer is controversial, with a lack of definitive evidence favoring survival for surgery over radiotherapy or confirming their equivalence. Using randomized clinical trials to compare these treatments has been unsuccessful as several trials have closed early due to poor recruitment and the one trial that ran to completion has been heavily criticized. In the absence of well designed and completed randomized clinical trials, observational studies have been relied on to provide information on treatment effectiveness. In the current study, we examined how observational studies can be used to inform treatment effectiveness in a situation where a randomized clinical trial is unlikely to be completed. We explored the difference in treatment effectiveness between curative radiotherapy and surgery for localized prostate cancer by using a restricted study population and an intent to treat approach, combined with traditional multivariate regression to control for confounding by indication. We also used propensity scores to assess the consistency of treatment effect across subgroups whose indication for one treatment versus the other varies.
**Materials and Methods**

We conducted a population-based case-cohort study designed to examine the difference in treatment effectiveness between curative radiotherapy and surgery. Queen’s University Research Ethics Board provided ethics approval with other participating institutions providing local research ethics board approval as needed.

The study population was drawn from a list in the Cancer Care and Epidemiology (CCE) database\(^7\) of 17,934 residents of Ontario diagnosed with adenocarcinoma of the prostate between 1990 and 1998 who were treated by surgery or had an exploratory lymph dissection within seven months of diagnosis (mainly surgical patients), or received radiotherapy within nine months of diagnosis. Radiotherapy patients whose treatment intent in the electronic record database was curative and who received at least one fraction (200 cGy) of their intended course were included. All patients received conventional external beam radiotherapy. In a case-cohort study, a random sample of patients, called the subcohort, is taken to represent the entire study population whereas cases, defined as patients who have the outcome of interest, are intentionally over-sampled.\(^8\) In our study, cases were defined as those patients who died of prostate cancer by December 31, 1999. A stratified random sample of approximately 10% of all patients was taken based on region of residence in Ontario to form the subcohort (n=1703) and approximately 55% of cases (n=591) were sampled separately in the same way for a total study population of 2213.

Patients were excluded if upon chart review: 1) their cancers had a histology other than adenocarcinoma (n=6); 2) they were not candidates for surgery or radiotherapy (n=50); 3) their candidacies could not be determined or treatment was not initiated for reasons other than pathologically positive lymph node status (n=58); 4) they had clinical nodal spread or metastases or were high risk patients (n=842); 5) they had unknown T category, Gleason score, and/or PSA (n=182). We restricted the population to those deemed eligible for either treatment, which excluded high risk patients, defined as patients with a PSA>20, Gleason score >7 and/or T
category >T2b (n=842).9-11 This left a final study population of 1075 patients including 106 cases and 985 subcohort patients, with 16 patients selected for both the subcohort and cases due to the case-cohort design.

The CCE database was used to identify the study population and corresponding charts were obtained from hospitals and regional cancer centers. Trained data abstractors used standardized forms and procedures in order to minimize error. Both cancer center charts and hospital charts were reviewed whenever possible due to the greater depth of detail for some data in the cancer center charts, and all charts were abstracted at their respective hospitals or cancer centers. Abstracted data were sent electronically to the coordinating office in Kingston on a weekly basis to conduct logic and missing checks. Electronic chart abstractions were checked for missing variables or dates, and findings regarding any missing data were sent back to the corresponding abstractor with instructions to find the missing information or confirm it was not present. If necessary, missing information was sought from secondary hospital, urologist and/or general practitioner charts. Information regarding illogical data were also sent to the abstractor corrected or confirmed the information.

During the time period of our study, lymph node dissections were often performed at the start of a prostatectomy and the surgery was abandoned if lymph nodes were positive. Because a similar approach is not used in radiotherapy patients, some radiotherapy patients would have had undetected nodal disease while those patients were excluded from the surgically treatment group. This may cause surgery to appear to have more of a survival advantage compared to radiotherapy than it actually does. Therefore, patients were analyzed by radiotherapy and surgical candidacy using their original treatment intent. The original treatment intent for each patient was based on information abstracted from chart review and was determined using a computerized algorithm that considered the types of treatments offered to the patient, the patient’s actual treatment, the original treatment plan, the outcome of any lymph node dissection, and the reason for any aborted surgery. In addition, patients who had surgery with post-operative radiotherapy were considered
to be surgical candidates (n<5). Our main analysis compared outcomes by treatment intent. The
treatment actually received was analyzed in a secondary analysis. In this case, patients who did
not receive radiotherapy or surgery were removed from analysis (n=10), as were radiotherapy
patients whose total administered dose was less than 40 Gy (n=29). This left a total of 1036
patients for the treatment received analysis.

The Ontario Cancer Registry undergoes a death clearance using death certificate data
(vital status and underlying cause of death) from the Ontario Registrar General. Deaths occurring
by December 31, 1999 were included in this analysis. More recent death data could not be used
because the case sampling was done with data available at the commencement of the chart
review. We calculated time to prostate cancer death using the date of diagnosis and the date of
prostate cancer death. Patients who did not die of prostate cancer were censored at date of other
cause death or December 31, 1999, whichever came first.

Gleason score, T category and PSA have been shown to be strong prognostic indicators
in prostate cancer-specific survival. For the 15% (n=159) of patients whose Gleason score was
not available, grade was used to approximate it. To assign Gleason score, reviewed pathologic
reports were used when available and the original was used otherwise. Gleason score was
analyzed using the clinically meaningful categories of 2-4, 5-6, and 7. We used the PSA level
recorded closest to the beginning of treatment but not within 30 days post-biopsy or after
initiation of hormone treatment as these actions influence its value. PSA was divided into
commonly used categories of \( \leq 4 \), >4 to \( \leq 10 \), and >10 to \( \leq 20 \). We also considered age at
diagnosis, androgen deprivation therapy (ADT), and comorbid illnesses, measured by the total
Cumulative Illness Rating Scale (CIRS-G), as potential covariates. The CIRS-G quantifies a
patient’s diseases or conditions based on 14 different organ systems. Each category is give a
score from 0, corresponding to mild or no diseases or conditions, to 4 for serious disease. The
total CIRS-G score is calculated by adding the score from each of the 14 categories.
To conduct a stratified analysis, we formed risk groups in accordance with the guidelines set by the Canadian Genitourinary Radiation Oncologist Group, which defines low risk patients as those with PSA ≤10, Gleason score ≤6, and T category ≤T2a. Intermediate patients were those who were not low risk and had PSA ≤20, and/or Gleason score ≤7, and/or T category ≤T2b. Higher risk patients are not always considered for surgical treatment and were, therefore, excluded from the study.

All statistical analysis was performed using SAS software (version 9.1, SAS Institute, Inc., Cary, NC). We compared age, total CIRS-G, T category, Gleason score and PSA between the radiotherapy and surgery groups in both the cases and subcohort. Categorical variables were compared using chi-square tests while continuous variables were compared using t-tests.

Cox-proportional hazard regression was performed to compare death from prostate cancer between radiotherapy and surgical candidates and for each variable considered as a covariate. We used the method proposed by Self and Prentice to calculate point estimates and variance that accounted for the case-cohort design. Variables that were associated with both treatment type and survival with a p-value of 0.25 were considered as covariates. These variables were considered in the adjusted model, however they were removed if their exclusion changed the treatment estimate by less than 10%. Because of their clinical importance, we forced T category, Gleason score and PSA into the model. We repeated our analysis using the treatment patients actually received and the intent to treat analysis was also repeated stratifying by risk group.

A propensity score is the probability that a patient receives a particular treatment given his prognostic indicators. This score can be used to control for overt bias within an observational study, as it is a composite score of a patient’s covariates. Propensity scores were generated based on a patient’s probability of receiving radiotherapy using logistic regression. The variables used to form risk groups (PSA, T category and Gleason score) were used to make the propensity score. Only the subcohort was used to determine the point estimates for each variable,
and the propensity score was assigned to each member of the subcohort and cases using the equation generated from this process. Propensity scores can be divided into quantiles in order to examine if the difference in effectiveness between treatments is stable as a patient’s indication for a treatment varies. Propensity scores were divided approximately into quintiles and all covariates were compared in the subcohort between the radiotherapy and surgery groups in each quintile in order to determine if there were any differences in covariates within each quintile group. This was done to ensure that the propensity score adequately balanced the covariates within propensity score quintiles. Cox-proportional hazards regression with adjustment for the case-cohort design was then conducted within each propensity score quintile, controlling for age, ADT and total CIRS-G as necessary. We did a likelihood ratio G test to assess the homogeneity of the hazard ratios across propensity score quintiles. This analysis used the intent to treat approach.

Results

The characteristics of the patients in the study are summarized in Table 1, stratified by subcohort and cases and then by radiotherapy and surgical candidacy. Compared to surgical patients, radiotherapy candidates were older and had more comorbid illnesses in both the subcohort and cases. The distribution of T category was statistically significantly different between the radiotherapy and surgical candidates in both the subcohort and cases with radiotherapy candidates having a greater probability of a high T category of T2b. The distribution of Gleason score did not statistically vary between radiotherapy and surgical candidates in both the subcohort and cases. PSA was statistically significantly different between radiotherapy and surgery candidates in the subcohort but was not statistically significantly different between radiotherapy and surgical candidates in the case groups.

In the subcohort, 12% of radiotherapy candidates and 88% of surgical candidates had lymph node dissections while in the cases, 28% of radiotherapy candidates and 72% of surgical
candidates had lymph node dissections. The subcohort and radiotherapy cases almost always received their original treatment plan, whereas only 81% of surgical cases actually received their original treatment plan. These surgical patients had positive lymph nodes upon lymph node dissection and either went on to receive radiotherapy or no treatment (n<5).

The median administered dose of radiotherapy for radiotherapy candidates was 64 Gy for both the analysis by intent to treat and the analysis by treatment received. By design, the range was lower in the analysis by intent to treat (2-70 Gy), since radiotherapy was abandoned in some patients, compared to the analysis by treatment received (40-70 Gy), where curative radiotherapy was defined as having received ≥40 Gy. There were 255 patients who received androgen deprivation therapy, with 6 patients receiving an orchiectomy, 47 patients receiving adjuvant hormone therapy, and 212 patients receiving neoadjuvant hormone therapy.

In the subcohort, 19 patients died of prostate cancer (including 16 patients who were both in the subcohort and case group due to the sampling method of the case-cohort design), 73 patients died of other causes, 881 were alive on December 31, 1999, and 10 were censored at the date of last contact recorded from the chart review due to unknown date of death. The median follow-up time for patients in the subcohort for the intent to treat analysis was 4.2 years with a maximum of 10 years, and the median time to death for cases was 4.2 years with a maximum of 9.3 years.

Table 2 displays the crude and adjusted hazard ratios for risk of prostate cancer death for radiotherapy compared to surgical candidates for the intent to treat analysis, as well as the adjusted hazard ratios for the actual treatment received analysis. Radiotherapy was marginally associated with worse survival in the adjusted model for the intent to treat analysis but was statistically significantly associated with worse survival in the actual treatment received analysis. Total CIRS-G was not associated with prostate cancer-specific survival, and ADT did not change the treatment effect point estimate, therefore neither variable was included in the final adjusted models. Age changed the effect estimate in the low risk group by greater than 10%, therefore it
remained in all regression models. Table 2 also displays the hazard ratios for death from prostate cancer for the intent to treat analysis stratified by risk group and adjusted for covariates. Although the point estimates for both risk groups indicate increased risk with radiotherapy, they are extremely imprecise, with confidence limit as low as 0.37 and as high as 6.67 for the low risk group.

The distribution of covariates between radiotherapy and surgical candidates in each propensity score quintile is shown in Table 3. In the second quintile, the proportion of patients with a Gleason score of 7 was marginally higher in the radiotherapy candidates than the surgical candidates (p=0.07). In the third and fourth quintiles the proportion of patients with a PSA >10 to 20 was marginally higher in the surgical group (p=0.14 for both quintiles). None of the differences in the distribution of covariates in the first and fifth quintiles achieved a p-value of <0.25. The number of cases increased with increased probability of receiving radiotherapy, as was expected. As shown in Table 3, the hazard ratios for radiotherapy compared to surgery for death from prostate cancer from the lowest quintile to the highest quintile displayed no pattern. Variables that differed between radiotherapy and surgical groups within the propensity score quintiles were adjusted for in these models, however these covariates did not change the point estimates by >10% and were not included in the final analyses. The test of homogeneity indicated the differences in the hazard ratios between the propensity score quintiles were marginally statistically significant (p=0.13).

**Discussion**

Our study showed a possible decrement for radiotherapy compared to surgery with a hazard ratio of 1.44 (95% CI 0.86-2.40) using traditional multivariate analysis and an intent to treat approach restricted to patients whose disease status made them candidates for either treatment. We believe that this hazard ratio would have been statistically significant in a more highly powered study. However, this difference is not substantial and with modern dose-
escalated radiotherapy a even lower hazard ratio than we observed may be produced. When stratifying by risk group, the difference in treatment effectiveness was not statistically significant with hazard ratios of 1.58 (0.37-6.67) and 1.34 (0.77-3.32) for low and intermediate risk groups respectively. Although the hazard ratios were not consistent across propensity score quintiles, the hazard ratio was statistically significant in only one quintile and there was no statistically significant difference in the hazard ratios between quintiles.

Other observational studies on this topic have shown mixed results: some indicate a survival benefit for surgery over radiotherapy while others have found no statistically significant difference in survival. One study analyzing intent to treat showed a 13-year cause-specific survival decrement for radiotherapy with a hazard ratio of 2.2 (1.6-3.1). Another prostate cancer-specific survival study similarly showed a 10-year survival decrement for radiotherapy analyzing actual treatment received with a hazard ratio of 2.3 (1.2-4.3). Both of these studies included high risk patients in their analyses. Several studies that use biochemical relapse-free survival (bRFS) have shown no statistically significant difference in survival. A study completed by Kupelian et al found that there was no difference in biochemical free survival between radiotherapy and surgery with a hazard ratio of 1.01 (95% CI 0.77-1.32). However, in a future study they stratified radiotherapy patients by total dose, and found that radiotherapy had a bRFS decrement compared to surgery with a hazard ratio of 2.24 (1.83-2.73) when examining those patients with a total dose <72 Gy but there was no difference in bRFS for those with a dose ≥72 Gy with a hazard ratio of 1.08 (0.78-1.50). This study indicates that doses less than 72 Gy may not provide the same survival advantage as radical prostatectomy. Another study found that the 7-year biochemical relapse free survival between radiotherapy compared to surgery was not statistically significant with an adjusted hazard ratio of 1.18 (0.86-1.62). Finally, another study found no difference in the 7-year odds ratio for bRFS with an odds ratio of 0.98 (0.55-1.74) for surgery compared to radiotherapy. Because odds ratios do not take time to event and censoring
into account, and because follow-up was not complete for all patients, a hazard ratio may have been a better alternative to compare treatment effectiveness in this study.

When we analyzed by actual treatment received, which is how most studies comparing radiotherapy to surgery in prostate cancer are analyzed, the hazard ratio increased over our intent to treat approach and the result was statistically significant. Also, only the point estimate for radiotherapy compared to surgery changed whereas the covariate point estimates remained relatively unaffected. We used intent to treat in preference to actual treatment received in order to minimize the bias from underreporting of pathological positive lymph node status in radiotherapy patients compared to surgical patients, as only 12% of radiotherapy candidates in the subcohort received a lymph node dissection compared to 88% of surgical candidates. Lu-Yao and Yao showed that analyzing by intent to treat versus actual treatment received had little effect on 10-year survival rates for patients treated by radiotherapy but differed significantly for surgical patients with survival rates of 83% (95% CI 81-84%) versus 89% (87-91%) respectively. A similar pattern can be seen in our study, as nearly all radiotherapy candidate cases actually received radiotherapy whereas only 81% of surgical candidate cases actually received surgery. If radiotherapy candidates had received lymph node dissections in the same proportion as surgical candidates, we would expect that at least the same proportion of positive lymph node dissections would have been found. By analyzing patients by original treatment intent, which was not necessarily the treatment that was received, we present hazard ratios for death from prostate cancer for radiotherapy compared to surgery that can be used to help guide the initial treatment decision prior to intra-operative (or pre-operative) lymph node dissection. This means that these hazard ratios reflect the difference in treatment effectiveness between radiotherapy and surgery at the time of initial treatment decision-making before any knowledge of positive lymph nodes from lymph node dissection, which may cause the original treatment plan to be abandoned.

A randomized clinical trial would be the ideal study design to end the treatment effectiveness debate, however multiple attempts have been unsuccessful due to poor recruitment.
Curative treatment for prostate cancer is controversial for other reasons as well. Many men diagnosed with prostate cancer are more likely to die of other causes because of older age at diagnosis and the long natural history of prostate cancer and this, combined with the potential for lower quality of life due to side effects following treatment, has led many to conclude that curative treatment may be inappropriate for some men.

Even screening for prostate cancer remains controversial. A recent European RCT study found that although those screened were 20% less likely to die of prostate cancer (p=0.04), 1410 men would need to be screened and 48 cancers would need to be treated to prevent 1 prostate cancer death after a median follow-up time of approximately 9 years.30 Another recent RCT which was completed in the United States has not found a survival advantage for those who were screened compared to those who were not screened at 7 years post-diagnosis, with a rate ratio of 1.13 (95% CI 0.75 to 1.70).31 Until more concrete evidence is provided, the choice to screen for prostate cancer and the subsequent treatment choice following diagnosis remains largely a personal preference for many patients and their physicians.

Our current study has several strengths. It was a population based study, with subjects accrued from all across Ontario, therefore the results are not influenced by the practice patterns or expertise of a few select hospitals. We were able to use detailed information on covariates due to the extensive chart review that was conducted on each patient, which provided more information than is available when administrative data are used to conduct population-wide analyses. We had the advantage of using death as the clinical outcome which is the definitive endpoint, compared to most studies which use biochemical free relapse, which is a surrogate endpoint. Our study used multiple methods to reduce the impact of confounding, including using a restricted study population and analyzing by intent to treat, which have not been used as stringently or in combination in past studies. Finally, the case-cohort design is an efficient design when extensive work is needed to abstract data from patient charts as it allowed for 106 prostate deaths to be included whereas a similar sized cohort study would have had only about 19 such deaths.
The main limitation of our study is that it was an observational study, and unmeasured confounders that distinguish those who get radiotherapy from surgery may be biasing our results. Also, residual confounding may be occurring as grade was used to approximate 15% of Gleason scores when they were not available. Because grade is less specific than Gleason score, some Gleason scores that were approximated by grade may be misclassified. Another important limitation is that conventional radiotherapy was used during this time period whereas dose-escalated radiotherapy is now commonly used. Dose-escalated radiotherapy produces better biochemical relapse rates compared to conventional radiotherapy.\textsuperscript{32, 33} As mentioned previously, an observational study which compare surgery to radiotherapy doses of $\leq 72$ Gy and $\geq 72$ Gy found that although there was a bRFS decrement for radiotherapy doses of $<72$ Gy compared to surgery, no differences were seen between surgery and radiotherapy doses $\geq 72$ Gy. Therefore, dose-escalated radiotherapy in the modern era may produce lower hazard ratios than we observed. Finally, the cases were chosen with deaths only up to 10 years post-diagnosis, and because of the slow-growing nature of prostate cancer, following patients for longer than this may be necessary to observe the full extent to which this treatment decision influences the natural history of the disease. The patients who died of prostate cancer in our study also may not be representative of most low or intermediate risk patients who die of prostate cancer, as the median time to death following diagnosis was 4.2 years. Therefore, these patients may have had more advanced prostate cancer than the average low or intermediate risk patient who will die of prostate cancer over a longer follow-up interval.

Conclusions

Our study suggests that analyzing survival by actual treatment received may overestimate the survival benefit of surgery over radiotherapy in localized prostate cancer and may be an inappropriate way to examine survival differences in this study population. Our analysis by
treatment received suggests a benefit for surgery over radiotherapy for localized prostate cancer, however the hazard ratio was substantially reduced when we used an analysis by intent to treat.

A randomized trial would provide more credible experimental evidence however, since these studies are unlikely to occur, well-designed observational studies that include dose-escalated radiotherapy are needed in order to provide more definitive answers for the relative treatment effectiveness of surgery compared to radiotherapy for localized prostate cancer. Future studies should focus on obtaining sufficient power for subgroup analysis, such as the stratification by risk group and by propensity score quintile, in order to have the power to detect differences between treatment modalities. Future studies would also benefit from using either the case-cohort or nested case-control design, perhaps nested within a larger cohort initiative, because of the rare outcome of prostate cancer deaths.

Acknowledgements

We thank Zhi Song for previously processing some of the variables in the database. This study was supported by the NCIC with funds from the Canadian Cancer Society. Dr. Groome is the Canada Research Chair in Cancer Care Evaluation.
References


Table 1: Baseline characteristics of the subcohort and cases based on radiotherapy and surgical candidacy. Note that there are 16 patients who overlap between the subcohort and cases.

<table>
<thead>
<tr>
<th>Subcohort</th>
<th>Radiotherapy (n=512)</th>
<th>Surgery (n=473)</th>
<th>Cases</th>
<th>Radiotherapy (n=69)</th>
<th>Surgery (n=37)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Age (SD)</td>
<td>69.2 (5.5)</td>
<td>62.7 (6.1)</td>
<td>67.4 (5.8)</td>
<td>62.3 (5.9)</td>
<td></td>
</tr>
<tr>
<td>p-value</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>Average CIRS-G (SD)</td>
<td>6.0 (3.7)</td>
<td>5.1 (3.3)</td>
<td>6.1 (4.2)</td>
<td>4.6 (2.7)</td>
<td></td>
</tr>
<tr>
<td>p-value</td>
<td>&lt;0.0001</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>T category % (n)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1a/b</td>
<td>9.4 (48)</td>
<td>6.1 (29)</td>
<td>8.7 (6)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>T1c</td>
<td>20.3 (104)</td>
<td>36.2 (171)</td>
<td>8.7 (6)</td>
<td>24.3 (9)</td>
<td></td>
</tr>
<tr>
<td>T2a</td>
<td>32.6 (167)</td>
<td>37.5 (177)</td>
<td>24.6 (17)</td>
<td>35.1 (13)</td>
<td></td>
</tr>
<tr>
<td>T2b</td>
<td>37.7 (193)</td>
<td>20.3 (96)</td>
<td>58.0 (40)</td>
<td>40.5 (15)</td>
<td></td>
</tr>
<tr>
<td>p-value</td>
<td>&lt;0.0001</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>Gleason Score % (n)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-4</td>
<td>28.3 (145)</td>
<td>24.3 (115)</td>
<td>26.1 (18)</td>
<td>13.5 (5)</td>
<td></td>
</tr>
<tr>
<td>5-6</td>
<td>48.6 (249)</td>
<td>52.6 (249)</td>
<td>37.7 (26)</td>
<td>56.7 (21)</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>23.1 (118)</td>
<td>23.0 (109)</td>
<td>36.2 (25)</td>
<td>29.7 (11)</td>
<td></td>
</tr>
<tr>
<td>p-value</td>
<td>0.37</td>
<td>0.17</td>
<td>0.17</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>PSA % (n)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤4</td>
<td>14.8 (76)</td>
<td>12.1 (57)</td>
<td>7.3 (5)</td>
<td>13.5 (5)</td>
<td></td>
</tr>
<tr>
<td>&gt;4 to ≤10</td>
<td>44.9 (230)</td>
<td>58.1 (275)</td>
<td>49.3 (34)</td>
<td>37.8 (14)</td>
<td></td>
</tr>
<tr>
<td>&gt;10 to ≤20</td>
<td>40.2 (206)</td>
<td>29.8 (141)</td>
<td>43.5 (30)</td>
<td>48.7 (18)</td>
<td></td>
</tr>
<tr>
<td>p-value</td>
<td>0.001</td>
<td>0.65</td>
<td>0.65</td>
<td>0.65</td>
<td></td>
</tr>
</tbody>
</table>
**Table 2:** Hazard ratios for death from prostate cancer using intent to treat and actual treatment received for radiotherapy compared to surgery and for study covariates (with 95% confidence intervals)

<table>
<thead>
<tr>
<th></th>
<th>Intent to Treat</th>
<th>Treatment Received</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overall unadjusted</td>
<td>Overall adjusted</td>
<td>Low risk adjusted</td>
<td>Intermediate risk adjusted</td>
<td>Overall adjusted</td>
<td></td>
</tr>
<tr>
<td>Subcohort (n)</td>
<td>985</td>
<td>985</td>
<td>373</td>
<td>612</td>
<td>954</td>
<td></td>
</tr>
<tr>
<td>Cases (n)</td>
<td>106</td>
<td>106</td>
<td>15</td>
<td>91</td>
<td>97</td>
<td></td>
</tr>
<tr>
<td>Radiotherapy</td>
<td>1.60 (1.02-2.51)</td>
<td>1.44 (0.86-2.40)</td>
<td>1.58 (0.37-6.67)</td>
<td>1.34 (0.77-2.32)</td>
<td>1.84 (1.06-3.17)</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>1.02 (0.99-1.06)</td>
<td>1.0 (0.96-1.04)</td>
<td>1.10 (0.99-1.22)</td>
<td>0.97 (0.93-1.01)</td>
<td>0.99 (0.95-1.03)</td>
<td></td>
</tr>
<tr>
<td>Total CIRS-G</td>
<td>1.03 (0.97-1.10)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>ADT</td>
<td>1.02 (0.59-1.76)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>T category</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1a/b</td>
<td>0.31 (0.12-0.85)</td>
<td>0.41 (0.15-1.13)</td>
<td>0.81 (0.15-4.39)</td>
<td>0.55 (0.13-2.32)</td>
<td>0.42 (0.15-1.19)</td>
<td></td>
</tr>
<tr>
<td>T1c</td>
<td>0.52 (0.27-0.98)</td>
<td>0.57 (0.29-1.12)</td>
<td>2.63 (0.62-11.15)</td>
<td>0.61 (0.27-1.37)</td>
<td>0.60 (0.30-1.21)</td>
<td></td>
</tr>
<tr>
<td>T2a</td>
<td>0.55 (0.33-0.91)</td>
<td>0.63 (0.37-1.10)</td>
<td>1.0</td>
<td>0.98 (0.53-1.80)</td>
<td>0.61 (0.35-1.09)</td>
<td></td>
</tr>
<tr>
<td>T2b</td>
<td>1.0</td>
<td>1.0</td>
<td>-</td>
<td>1.0</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Gleason Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-4</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>5-6</td>
<td>1.66 (0.93-2.94)</td>
<td>1.55 (0.87-2.78)</td>
<td>1.35 (0.39-4.71)</td>
<td>1.64 (0.83-3.23)</td>
<td>1.40 (0.76-2.57)</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>3.14 (1.71-5.76)</td>
<td>2.76 (1.49-5.12)</td>
<td>-</td>
<td>2.11 (1.04-4.28)</td>
<td>2.83 (1.50-5.33)</td>
<td></td>
</tr>
<tr>
<td>PSA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 to 4</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>&gt;4 to 10</td>
<td>1.76 (0.81-3.84)</td>
<td>1.51 (0.68-3.35)</td>
<td>0.48 (0.12-1.84)</td>
<td>2.84 (0.84-9.61)</td>
<td>1.62 (0.71-3.67)</td>
<td></td>
</tr>
<tr>
<td>&gt;10 to 20</td>
<td>2.47 (1.13-5.40)</td>
<td>1.76 (0.78-3.99)</td>
<td>-</td>
<td>2.43 (0.73-8.14)</td>
<td>1.74 (0.75-4.03)</td>
<td></td>
</tr>
</tbody>
</table>
Table 3: Distribution of covariates between radiotherapy and surgical candidates in the subcohort within each propensity score quintile (percentage)

<table>
<thead>
<tr>
<th></th>
<th>0.32-0.42</th>
<th>0.43-0.45</th>
<th>0.45-0.56</th>
<th>0.57-0.66</th>
<th>0.66-0.73</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RT</td>
<td>SX</td>
<td>RT</td>
<td>SX</td>
<td>RT</td>
</tr>
<tr>
<td>Subcohort (n)</td>
<td>52</td>
<td>121</td>
<td>95</td>
<td>102</td>
<td>100</td>
</tr>
<tr>
<td>Cases (n)</td>
<td>2</td>
<td>5</td>
<td>8</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>T category %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1a/b</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4.0</td>
</tr>
<tr>
<td>T1c</td>
<td>100</td>
<td>100</td>
<td>28.4</td>
<td>27.5</td>
<td>25.0</td>
</tr>
<tr>
<td>T2a</td>
<td>0</td>
<td>0</td>
<td>71.6</td>
<td>72.6</td>
<td>71.0</td>
</tr>
<tr>
<td>T2b</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Gleason Score %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-4</td>
<td>19.2</td>
<td>14.1</td>
<td>0</td>
<td>0</td>
<td>42.0</td>
</tr>
<tr>
<td>5-6</td>
<td>59.6</td>
<td>57.9</td>
<td>71.6</td>
<td>82.4</td>
<td>36.0</td>
</tr>
<tr>
<td>7</td>
<td>21.2</td>
<td>28.1</td>
<td>28.4</td>
<td>17.7</td>
<td>22.0</td>
</tr>
<tr>
<td>PSA %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤4</td>
<td>9.6</td>
<td>12.4</td>
<td>0</td>
<td>0</td>
<td>27.0</td>
</tr>
<tr>
<td>&gt;4 to ≤10</td>
<td>90.4</td>
<td>87.6</td>
<td>71.6</td>
<td>72.6</td>
<td>25.0</td>
</tr>
<tr>
<td>&gt;10 to ≤20</td>
<td>0</td>
<td>0</td>
<td>28.4</td>
<td>27.5</td>
<td>28.0</td>
</tr>
<tr>
<td>Hazard Ratio</td>
<td>0.30</td>
<td>1.54</td>
<td>0.90</td>
<td>2.71</td>
<td>1.08</td>
</tr>
<tr>
<td>95% CI</td>
<td>(0.04-2.28)</td>
<td>(0.35-6.77)</td>
<td>(0.29-2.82)</td>
<td>(1.01-7.31)</td>
<td>(0.41-2.81)</td>
</tr>
</tbody>
</table>
Chapter 5: Manuscript 2

This manuscript is intended for Journal of Clinical Epidemiology and was created in accordance with their submission guidelines.

**Title:** Assessing the instrumental variable method assumption of equal distribution of unmeasured covariates in a prostate cancer population

**Authors:**

Julie M. DeGroot\textsuperscript{1,2}
Michael Brundage\textsuperscript{1}
Miu Lam\textsuperscript{2}
Patti A. Groome\textsuperscript{1,2}

**Affiliations:**

\textsuperscript{1}Division of Cancer Care and Epidemiology, Queen’s Cancer Research Institute, Queen’s University, Kingston, Ontario, Canada

\textsuperscript{2}Department of Community Health and Epidemiology, Queen’s University, Kingston Ontario, Canada

**Corresponding Author:**

Dr. Patti Groome  
Cancer Care and Epidemiology  
Queen’s University  
10 Stuart Street  
Kingston, Ontario, Canada  
K7L 3N6

**Key Words:** prostatic neoplasms, instrumental variable, radiotherapy, prostatectomy
Abstract

**Background:** Instrumental variables are used in health research to control for unmeasured and unknown confounding when investigating intended treatment effects. One of the primary assumptions of this method is that there are no differences in unknown covariates, as shown by there being no difference in known covariates. We investigated this assumption in a population-based prostate cancer study.

**Methods:** Our study population consisted of two study cohorts. The first was a registry-based cohort of all patients diagnosed with prostate cancer between 1990 and 1998 who were treated for cure by radiotherapy or surgery. The second cohort was a chart-reviewed random sample of the registry-based cohort stratified by Cancer Care Ontario Region (CCOR). The instrumental variable chosen was the radiotherapy rate in each CCOR, and CCORs with similar radiotherapy rates were grouped using the radiotherapy rates generated from the registry-based cohort. Age, total CIRS-G, T category, Gleason score and PSA were compared across the groups using data on the chart-reviewed cohort. Logistic regression was used to assess trend for continuous variables while Pearson correlation was used for categorical variables.

**Results:** Radiotherapy rates in CCORs varied from 49.8% to 75.7%. The distribution of every prognostic indicator was statistically significantly different between instrumental variable groups. The proportion of patients with a T category of T2b (p=0.08) increased and average age (p=0.002) increased as radiotherapy use increased across instrumental variable groups. A high Gleason score of 7 decreased as use of radiotherapy increased (p=0.003). No trend was seen with PSA (p=0.55) or total CIRS-G (p=0.55).
**Conclusions:** The candidate instrumental variable, CCOR radiotherapy rate, could not be used for our population as one of the main assumptions, the equal distribution of unknown covariates across instrumental variable groups, was violated as all prognostic indicators differed across instrumental variable groups. Because the trends we saw were not convincing, we believe that radiotherapy use was not driven by prognostic factors. We believe that there are other factors that are contributing to the differences we saw across instrumental variable groups, such as differential staging and grading between CCORs.

1. **Background**

   Observational studies are generally considered inferior to randomized clinical trials due to the potential for unmeasured and unknown confounders. Instrumental variables have recently been used in the health literature to address this unknown confounding.\(^1\) An instrumental variable is a variable that is associated with an exposure of interest but is not associated with the outcome. The ideal instrument is the “coin flip” of a randomized clinical trial, where the exposure is 100% dependent on the coin toss and the outcome is completely independent from that assignment. There are two assumptions that must be met when using instrumental variables: 1) the instrument is associated with the exposure of interest and 2) the instrument is not correlated with the error, or the unknown confounders.\(^5\) Because the second assumption is impossible to verify, if the known confounders do not differ between the instrumental variable groups formed it is assumed that the unknown confounders also do not differ.\(^5\) This method is sometimes used with large administrative databases, which have limited information on patient and disease characteristics to test this second assumption.\(^4,6-8\)

   Confounders, those variables that distinguish who receive which treatment when treatment assignment is not random, play a role when comparing the treatment effectiveness of radiotherapy versus surgery for patients diagnosed with prostate cancer. Prostate cancer is the leading cancer diagnosed in Canadian men, accounting for over 25% of new cancer diagnoses.\(^9\)
Over 90% of these cancers are diagnosed locally or regionally, and radiotherapy and surgery are the two most common treatments. The relative treatment effectiveness between radiotherapy and surgery for localized prostate cancer is controversial, mainly due to the lack of randomized clinical trial evidence demonstrating the superiority or equivalency of surgery compared to radiotherapy. Observational studies have attempted to describe the difference in treatment effectiveness but they are susceptible to confounding by indication which can occur when treatment choice is influenced by the indications or a contraindications for the treatments which also affects the outcome of interest.

There are many observable differences between patients who receive radiotherapy compared to those who have surgery. Radiotherapy patients tend to be older, have more comorbid illnesses and tend to have worse prognostic indicators: higher T category, higher PSA and higher Gleason score. There is also the possibility that there are other unknown confounders that may affect relative treatment effectiveness estimation when comparing radiotherapy to surgery for localized prostate cancer. This study examined the second assumption of the instrumental variable method to assess whether this method would be an appropriate course of action to take to control for unknown confounding in this setting. We decided that we would declare this assumption does not hold for our study population if the regions with higher radiotherapy were treating more severe cases.

2. Methods

This was a study designed to look at one of the assumptions of the instrumental variable methods using two cohorts: a registry-based cohort and a chart-reviewed cohort, a subset of the registry-based cohort, of patients diagnosed with prostate cancer. Queen’s University Ethics Board provided ethics approval.

The study population was drawn from a list in the Cancer Care and Epidemiology database of 45 035 residents of Ontario who were diagnosed with adenocarcinoma of the prostate
between 1990 and 1998. The inclusion criteria for the original study included those who were
treated by surgery and/or lymph dissection within seven months of diagnosis or curative
radiotherapy within nine months, for a total population of 17 934. Curative radiotherapy was
defined as a dose of at least 200 cGy with at least one record of curative intent in the electronic
database. For our study, patients were excluded if they were treated with lymph node dissections
but did not receive prostatectomy or radiotherapy (n=1443), those who could not be assigned a
Cancer Care Ontario Region (CCOR) of residence (n=22), and radiotherapy patients whose total
dose was less than 40Gy (n=66) for a total study population of 16 403. This study population will
be referred to as the registry-based cohort.

The second population was a random sample of approximately 10% of the
17 934 patients treated for cure stratified by CCOR of residence for a study population of 1703.
We excluded patients who only received lymph node dissections (n=82) and those radiotherapy
patients whose total dose was less than 40Gy (n=67). Patients were also excluded if they had
unknown PSA, Gleason score or T category (n=167) for a final study population of 1387 patients.
This study population will be referred to as the chart-reviewed cohort.

We identified patients using the Cancer Care and Epidemiology (CCE) database, which
includes diagnosis information from the Ontario Cancer Registry and has previously been
described in detail.14 Trained data abstractors reviewed corresponding hospital and/or cancer
centre charts on-site for the selected 1703 patients. Data abstractors used standardized forms and
procedures in order to minimize error. Both cancer centre charts and hospital charts were
reviewed whenever possible due to the greater amount of detail for some data in the cancer centre
chart, and secondary hospital, urologist and/or general practitioners’ charts provided information
on missing variables. Files were sent electronically to the coordinating office in Kingston on a
weekly basis where missing data and logic checks were subsequently completed. Information
regarding missing variables was sent back to the corresponding abstractor who then reviewed the
chart to find the missing variable or confirmed that it was not there. Similarly, information
regarding illogical or inconsistent variables or dates were sent back to the abstractor who corrected the information when possible.

The type of treatment that a patient received was recorded in the CCE database. Patients who received surgery and post-operative radiotherapy were included in the surgical group. Patient postal codes and/or Ontario Ministry of Health residence codes were also recorded in the OCR and these were used to determine CCOR of residence. CCOR is a geographical region of Ontario which contains adjacent counties that have at least one cancer centre between them and are meant to represent the service areas of the province’s regional cancer centres.

Gleason score, T category and PSA are strong prognostic indicators of prostate cancer-specific survival. We used grade to approximate Gleason score for the 13.5% (n=187) of patients whose Gleason score was not available. Reviewed pathologic reports were used when available and the original Gleason score was used otherwise. Gleason score was analyzed using the clinically meaningful categories of 2-4, 5-6, 7 and 8-10. We used the PSA level recorded closest to the beginning of treatment but not within 30 days post-biopsy or after initiation of hormone treatment as these actions influence its value. PSA was divided into commonly used categories of \( \leq 4, \) 4 to \( \leq 10, \) 10 to \( \leq 20, \) and >20. We also used age and comorbid illnesses as covariates. A patient’s comorbid illnesses were measured using the total Cumulative Illness Rating Scale-G (CIRS-G), which uses medical history information from patient charts. CIRS-G rates 14 organ systems by the severity of the disease or condition. These values range from 0, corresponding to a mild or no problem to 4, corresponding to severe disease. The total CIRS-G score is calculated by adding up the score from each of the 14 categories.

All statistical analyses were done using SAS software (version 9.1, SAS Institute, Inc., Cary, NC). We compared the distribution of covariates between radiotherapy and surgical patients using the chart-reviewed cohort. We used chi-square tests for categorical variables and t-tests for continuous variables. We used CCOR radiotherapy rate as the instrumental variable and
calculated the radiotherapy rate in each CCOR using both the registry-based and chart-reviewed cohorts to determine if the treatments in the sample reflected that in the whole population.

Regions were grouped into low, moderately low, moderately high, and high use of radiotherapy based on the radiotherapy rates in the registry-based cohort. The corresponding CCORs in the chart-reviewed cohort were also grouped together. We compared differences in covariates between instrumental variable-derived regions using the chart-reviewed data. A weight was given to each region based on the inverse of the sampling proportion from the original sampling frame when we compared covariates across instrumental variable-derived groups. We used a weight in order to account for differences in covariate distributions between regions that were grouped together. We calculated differences in continuous variables using the Taylor expansion method to account for the sampling weights. We used the Rao-Scott method to calculate chi-square tests for the categorical variables, again to account for the sampling weights. We used logistic regression, using the Taylor expansion method to calculate variance, to test for trends from lowest to highest regions of radiotherapy use using an ordinal region variable as the outcome for the continuous variables. We assessed trends in categorical variables using Pearson correlation. We also calculated the average age in each instrumental variable-derived region using the registry-based cohort. We were unable to calculate the distribution of other covariates in the registry-based cohort as this database does not carry this detailed information.

3. Results

Table 1 compares the distribution of covariates between radiotherapy and surgical patients. The distribution of every covariate was statistically significantly different. As expected, radiotherapy patients were older than surgical patients and had more comorbid illnesses. Radiotherapy patients were also more likely to have higher T categories, higher Gleason scores,
and high PSA values compared to surgical patients. Due to rare categories among those treated, the T categories of T1a and T1b were grouped together and T3 and T4 were grouped together.

Table 2 compares the rate of radiotherapy in each of the CCORs using the registry-based cohort and chart-reviewed cohort to determine the radiotherapy rate. The radiotherapy rate differed across CCORS (p<0.0001), with radiotherapy rates ranging from 44.7% to 75.7% for the registry-based cohort. All radiotherapy rates that were calculated using the registry-based cohort were similar to and fell within the 95% confidence intervals of the radiotherapy rates calculated using the chart-reviewed cohort.

Table 3 compares the prognostic indicators across the instrumental variable groups formed based on radiotherapy rates in the registry-based cohort using all patients in the chart-based cohort. The distribution of every variable differed between the instrumental variable groups. The test for trend was marginally statistically significant for T category and was statistically significant for age and Gleason score. As use of radiotherapy increased, patients had a higher proportion of a T category of T2b. Age was also found to increase slightly as use of radiotherapy increased. A high Gleason score of 7 decreased as radiotherapy use increased. No statistically significant trend was seen for total CIRS-G or PSA. Average age using the registry-based cohort mirrored what we saw using the chart data: from the lowest rate of radiotherapy group to the highest rate of radiotherapy group average ages were 65.5, 65.7, 66.9 and 66.9.

4. Discussion

Instrumental variables are a proposed method to control for known and unknown confounding within a study. One of the assumptions of this method is that there are no differences in the distribution of unknown covariates among groups formed. When we compared known prognostic indicators in prostate cancer across groups, we saw that the distribution of all prognostic indicators differed between instrumental variable groups. However, there was only a statistically significant trend for age and Gleason score, and a marginally statistically significant
trend for T category. More specifically, patients in the group with the highest rate of radiotherapy were more likely to have a T category of T2b and age increased slightly from the group with the lowest rate of radiotherapy to the highest rate of radiotherapy. There was no trend for an increase in total CIRS-G or PSA as radiotherapy use increased. Gleason score 7 decreased in regions with increased use of radiotherapy, which was not expected.

The two assumptions of the instrumental variable approach assess whether the instrument chosen is powerful and valid. In order to be powerful, the sample size must be large, in order to reduce sampling error, or there must be very large variations in the instrument. Although our sample size was not very large, we believe that our instrument was powerful because it predicted large variations in radiotherapy rates, with radiotherapy rates ranging from 44.7% to 75.7%. However, it appears that our instrument was not valid. The distribution of all prognostic factors were statistically significantly different across instrumental variable groups. Some variables behaved in a manner we would expect to see if an instrumental variable wasn’t used: the proportion of patients with a high T category of T2b increased and average age increased as the rate of radiotherapy increased.

The pattern in T category using the instrumental variable groups partially parallels that when simply comparing those who received radiotherapy and those who received surgery. Those who received radiotherapy were far more likely to have a high T category of T2b than surgical patients. However, no trend was apparent across the instrumental variable groups in patients with a T category of T3 or T4 as it was in the comparison of radiotherapy and surgical candidates. The statistically significant p-value for trend for Gleason score appears to be driven by the increase in Gleason score 7 as radiotherapy use decreases across instrumental variable groups as radiotherapy use increases. However, this was not the pattern we would expect to see as there was very little difference in the distribution of Gleason score 7 between patients who received radiotherapy and those who received surgery. Moreover, patients who were radiotherapy patients were more likely than surgical patients to have a Gleason score of 8-10 whereas this pattern was not seen as
radiotherapy use increased across instrumental variable groups. Patients who received radiotherapy were far more likely to have a PSA >20 than surgical patients. However, there was no pattern in PSA values >20 as radiotherapy use increased across instrumental variable groups, and in fact all values appear very similar. Age increased in the direction we would expect if the instrumental variable did not sufficiently balance prognostic indicators as those patients who received radiotherapy were older than those who received surgery. Although there was variation in the average total CIRS-G, there was no pattern seen as there was for radiotherapy patients compared to surgical patients where radiotherapy patients had a higher average total score. Because the trends in prognostic indicators in the instrumental variable groups do not parallel those seen when comparing prognostic indicators in radiotherapy and surgical patients, we believe that radiotherapy use is not necessarily associated with prognostic indicators across CCORs and other factors may be driving the differences seen between instrumental variable groups.

Stukel et al explored different methods to control for treatment selection bias in observational studies. They used a cohort of patients who had suffered a myocardial infarction to investigate the treatment benefit of cardiac catheterization.\(^1\) They also looked at the assumption of equal distribution of unknown prognostic factors before using this method, and although there were small differences in the distribution of prognostic factors they were confident that the distribution was satisfactory enough to infer that the distribution of unknown prognostic factors was likely similar.\(^1\) They found that while propensity scores and multivariate model risk adjustment overestimated the treatment benefit of cardiac catheterization, the instrumental variable method showed a benefit for cardiac catheterization that was within the range of benefits seen in randomized clinical trials.\(^1\) This gives strength to the belief that instrumental variables may be able to control for the unknown confounding within observational studies in some settings.
However, using a large administrative database may have disadvantages if confounding variables are not recorded within the database. A study by Setoguchi et al used an administrative database to study the effects of antipsychotic drugs on various causes of death. They used the instrumental variable method, however they acknowledged that certain important variables were missing from the database. If these variables were associated with both outcome and exposure and differed between instrumental variable groups, the assumption of equal distribution of unknown prognostic factors would be violated. In this case, using the instrumental variable method may provide little benefit over classic ways to control for observable covariates.

In our study, we believe that there are other factors that are associated with the differences in prognostic factor distribution other than radiotherapy use, as trends are not apparent in some variables and in Gleason score the trend is actually the opposite of what would be expected. The reasons for the trends seen require further explanation and exploration. However, it is not unreasonable to think that age might be related to radiotherapy use in CCORs. For example, those patients who live in one region may be younger than those who live in another region, and this difference may be associated with a small increase in radiotherapy use due to older ages in one region. The differences that we saw could also be partly explained by sampling error. For example, if we had grouped the regions based on radiotherapy rate using data from the chart-reviewed cohort, the South West region would have been in the moderately-low group instead of the moderately-high group, and the East region would have been in the moderately-high group instead of the moderately-low group. Although the radiotherapy rates were very similar between the registry-based cohort and the chart-reviewed cohort, sampling error may have caused some of the differences we saw in prognostic factors.

There were quite a few issues with our Gleason score categorization that may have contributed to differences between the instrumental variable groups. Grade was used to approximate Gleason score when Gleason score was not available and reviewed Gleason score was used in preference to original Gleason score. In some regions, grade may have been used
more often than others, and this may have caused differential Gleason score assignment between regions. Also, the average Gleason score has been changed over time, with an upgrading of Gleason score.\textsuperscript{15} Perhaps some regions, such as the Central East, adopted this new system before more rural CCORs, which would have caused a systematic upgrading of Gleason scores within some regions compared to others. Finally, some regions, for example the Northern regions, may have only a few pathologists and they may systematically underestimated Gleason score in comparison to areas that have a higher number of pathologists, where the Gleason score distribution is not dependent on only a few pathologists. It is interesting to note that of the three main prognostic factors (Gleason score, T category, and PSA) PSA is the least subjective and was not found to have a statistically significant trend over instrumental variable groups.

Our study has several strengths. First, we were able to use detailed information on patient prognostic factors due to the extensive chart review process. Next, this was a population-based study with patients accrued from all over Ontario and it was not influenced by the practice patterns, expertise or referral bias of a few select hospitals or cancer centres. Also, patients were over-sampled from less populous regions in order to provide enough numbers to be able to draw conclusions on treatment practices in these regions. Our chart-reviewed sample appears to represent the population from which it was drawn from as the age distribution of the sampling frame and the radiotherapy rates across CCORs are similar between the registry-based and chart-reviewed cohorts. This gives us more confidence that the prognostic factors distribution could be similar between cohorts, and that our chart-reviewed cohort is a valid representation of the registry-based cohort. A limitation of our study is that grade was used to approximate Gleason score when Gleason score was not available which may have caused misclassification between Gleason score categories. Also, we were using information from a small fraction of the registry-based cohort to infer prognostic indicator distribution for the entire cohort. Finally, another limitation is that we only explored only CCOR radiotherapy rate as the instrumental variable and
other variables, such as distance to a radiotherapy centre, may have been adequate instrumental variables in this study population.

In conclusion, we were not able to use CCOR radiotherapy rate as an instrumental variable in our population due to the differences in distribution for all prognostic indicators and the trends found in several of the prognostic indicators. We believe that caution should be used when trying to use the instrumental variable method to determine the differences in treatment effectiveness, as unknown confounders may not be as well controlled for as expected. Careful consideration should be used in studies that use administrative databases to ensure that the variable chosen is indeed a strong and valid instrument, especially for those studies using datasets with sparse information on covariates.

Acknowledgements

We would like to thank Zhi Song for his work in processing some of the variables used in our study. This study was supported by the NCIC with funds from the Canadian Cancer Society. Dr. Groome is the Canada Research Chair in Cancer Care Evaluation.
References


Table 1: Comparison of prognostic indicator distribution between radiotherapy and surgical patients in the chart-reviewed cohort

<table>
<thead>
<tr>
<th></th>
<th>Radiotherapy (n=829)</th>
<th>Surgery (n=558)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average age (SD)</td>
<td>69.1 (5.8)</td>
<td>62.8 (5.8)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Average total CIRS-G (SD)</td>
<td>5.9 (3.6)</td>
<td>5.1 (3.3)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>T category % (n)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1a/b</td>
<td>6.6 (55)</td>
<td>5.7 (32)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>T1c</td>
<td>16.9 (140)</td>
<td>35.0 (194)</td>
<td></td>
</tr>
<tr>
<td>T2a</td>
<td>26.4 (219)</td>
<td>35.1 (196)</td>
<td></td>
</tr>
<tr>
<td>T2b</td>
<td>37.5 (311)</td>
<td>22.8 (127)</td>
<td></td>
</tr>
<tr>
<td>T3/4</td>
<td>12.6 (104)</td>
<td>1.4 (8)</td>
<td></td>
</tr>
<tr>
<td>Gleason Score % (n)</td>
<td></td>
<td></td>
<td>0.004</td>
</tr>
<tr>
<td>2-4</td>
<td>21.4 (177)</td>
<td>22.6 (126)</td>
<td></td>
</tr>
<tr>
<td>5-6</td>
<td>42.2 (350)</td>
<td>47.7 (266)</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>23.0 (191)</td>
<td>22.4 (125)</td>
<td></td>
</tr>
<tr>
<td>8-10</td>
<td>13.4 (111)</td>
<td>7.4 (41)</td>
<td></td>
</tr>
<tr>
<td>PSA % (n)</td>
<td></td>
<td></td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>≤4</td>
<td>13.4 (111)</td>
<td>10.0 (56)</td>
<td></td>
</tr>
<tr>
<td>&gt;4 to ≤10</td>
<td>31.2 (259)</td>
<td>51.8 (289)</td>
<td></td>
</tr>
<tr>
<td>&gt;10 to ≤20</td>
<td>30.8 (255)</td>
<td>26.9 (150)</td>
<td></td>
</tr>
<tr>
<td>&gt;20</td>
<td>24.6 (204)</td>
<td>11.3 (63)</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Radiotherapy rates by Cancer Care Ontario Region (CCOR) using the registry-based cohort and the chart-reviewed cohort

<table>
<thead>
<tr>
<th>CCOR</th>
<th>Registry-based cohort (n=16403)</th>
<th>Chart-reviewed cohort (n=1387)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Radiotherapy Rate</td>
<td>Radiotherapy Rate</td>
</tr>
<tr>
<td>Central East</td>
<td>6715 44.7 (43.5-45.9)</td>
<td>257 48.6 (42.4-54.9)</td>
</tr>
<tr>
<td>Central West</td>
<td>2208 63.6 (62.6-66.6)</td>
<td>143 68.5 (60.2-76.0)</td>
</tr>
<tr>
<td>East</td>
<td>1697 58.9 (56.5-61.2)</td>
<td>137 60.6 (51.9-68.8)</td>
</tr>
<tr>
<td>North East</td>
<td>890 75.7 (72.8-78.5)</td>
<td>156 79.5 (72.3-85.5)</td>
</tr>
<tr>
<td>North West</td>
<td>320 66.3 (60.8-71.4)</td>
<td>214 63.1 (56.2-69.6)</td>
</tr>
<tr>
<td>South</td>
<td>661 49.8 (45.9-53.7)</td>
<td>173 49.7 (42.0-57.4)</td>
</tr>
<tr>
<td>South East</td>
<td>1141 57.4 (54.5-60.3)</td>
<td>154 58.4 (50.2-66.3)</td>
</tr>
<tr>
<td>South West</td>
<td>2837 62.3 (60.4-64.1)</td>
<td>153 57.5 (49.3-65.5)</td>
</tr>
<tr>
<td>Table 3: Comparison of prognostic indicators between instrumental variable groups</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------------</td>
<td>-------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>Moderately low</td>
</tr>
<tr>
<td>Radiotherapy rate</td>
<td>48.8</td>
<td>59.7</td>
</tr>
<tr>
<td>Average Age</td>
<td>65.8</td>
<td>65.4</td>
</tr>
<tr>
<td>Average CIRS-G</td>
<td>5.8</td>
<td>5.2</td>
</tr>
<tr>
<td>T category %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1a/b</td>
<td>1.9</td>
<td>6.5</td>
</tr>
<tr>
<td>T1c</td>
<td>28.9</td>
<td>28.3</td>
</tr>
<tr>
<td>T2a</td>
<td>35.3</td>
<td>31.8</td>
</tr>
<tr>
<td>T2b</td>
<td>25.0</td>
<td>25.9</td>
</tr>
<tr>
<td>T3/4</td>
<td>8.9</td>
<td>7.5</td>
</tr>
<tr>
<td>Gleason Score %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-4</td>
<td>12.2</td>
<td>16.1</td>
</tr>
<tr>
<td>5-6</td>
<td>45.2</td>
<td>42.5</td>
</tr>
<tr>
<td>7</td>
<td>33.3</td>
<td>27.3</td>
</tr>
<tr>
<td>8-10</td>
<td>9.3</td>
<td>14.1</td>
</tr>
<tr>
<td>PSA %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤4</td>
<td>8.5</td>
<td>18.7</td>
</tr>
<tr>
<td>&gt;4 to ≤10</td>
<td>42.5</td>
<td>40.4</td>
</tr>
<tr>
<td>&gt;10 to ≤20</td>
<td>29.5</td>
<td>23.8</td>
</tr>
<tr>
<td>&gt;20</td>
<td>19.5</td>
<td>17.1</td>
</tr>
</tbody>
</table>
Chapter 6: Supplemental Results

6.1 Variable Formation

We used grade to approximate Gleason score when Gleason score was not available. Reviewed Gleason score or grade was used in preference to the original biopsy Gleason score or grade respectively. Reviewed Gleason score was used for 15% of patients (n=164), original Gleason score was used for 70% of patients (n=752), reviewed grade was used for 1% of patients (n=12) and original grade was used for 14% of patients (n=147). Using patients who had both reviewed and original Gleason score, reviewed Gleason score agreed with original Gleason score in 66% of patients. If we assume that reviewed Gleason score is superior to original Gleason score, then original Gleason score overestimated 12% of Gleason scores and underestimated 22% of Gleason scores.

There were many patients with unknown PSA, Gleason score and/or T category and 8.2% of the study population was excluded for this reason. Table 6.1 displays the number and proportion of patients with unknown covariates in the subcohort and case groups by radiotherapy and surgical candidacy. It appears that this exclusion was non-differential in the subcohort between radiotherapy and surgical candidates. However, in the case groups it appears that more surgical cases were excluded than radiotherapy cases due to unknown confounders.

Table 6.1: Proportion of patients with unknown covariates in the subcohort and case groups by candidacy % (n)

<table>
<thead>
<tr>
<th></th>
<th>Subcohort</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Radiotherapy</td>
<td>Surgery</td>
<td>Radiotherapy</td>
</tr>
<tr>
<td>T category</td>
<td>0.1 (1)</td>
<td>1.6 (18)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Gleason score</td>
<td>1.1 (12)</td>
<td>3.4 (38)</td>
<td>8.6 (8)</td>
</tr>
<tr>
<td>PSA</td>
<td>3.4 (38)</td>
<td>5.2 (59)</td>
<td>18.3 (17)</td>
</tr>
</tbody>
</table>

Table 6.2 examines all patients excluded due to unknown PSA, Gleason score, and/or T category to those included in the study population. The average age and average total CIRS-G were similar between those included and excluded. However, surgical candidates were more
likely to be excluded than radiotherapy patients with 17.2% of surgical candidates compared to 8.9% of radiotherapy candidates excluded due to unknown covariates.

**Table 6.2:** Comparison of included patients and patients excluded in the subcohort due to unknown PSA, Gleason score, and/or T category

<table>
<thead>
<tr>
<th>Candidacy</th>
<th>Excluded (n=148)</th>
<th>Included (n=985)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgery % (n)</td>
<td>17.2 (98)</td>
<td>82.8 (473)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Radiotherapy % (n)</td>
<td>8.9 (50)</td>
<td>91.1 (512)</td>
<td></td>
</tr>
<tr>
<td>Average Age (SD)</td>
<td>66.0 (6.3)</td>
<td>66.1 (6.7)</td>
<td>0.92</td>
</tr>
<tr>
<td>Average CIRS-G (SD)</td>
<td>5.8 (3.6)</td>
<td>5.5 (3.6)</td>
<td>0.48</td>
</tr>
</tbody>
</table>

**6.2 Diagnostic Tests**

We compared the diagnostic tests done between the radiotherapy and surgical candidacy groups in the subcohort to explore if differential staging may explain prostate cancer-specific survival differences between the two treatments. Table 6.3 displays the distribution of patients who had a test done, did not have a test done, those in who had a test ordered but it was unknown if the test was completed, and those who had missing information. As shown in Table 6.4, the proportion of patients who received a specific diagnostic test nearly always differed between the radiotherapy and surgical candidates in the subcohort. Those who were radiotherapy candidates were more likely to receive CT scans of the abdomen, CT scans of the pelvis, and bone scans, and while those who were surgical candidates were more likely to receive lymph node dissections, MRIs, ultrasounds of the abdomen/pelvis and/or liver, and chest x-rays.
Table 6.3: Proportion (%) of patients in the subcohort who had specific diagnostic tests done, not done, unknown if done, or who had information missing

<table>
<thead>
<tr>
<th>Diagnostic Test</th>
<th>Test Done</th>
<th>Test Not Done</th>
<th>Test ordered, results unknown</th>
<th>Information Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digital rectal exam</td>
<td>97.0</td>
<td>0.3</td>
<td>2.5</td>
<td>0.2</td>
</tr>
<tr>
<td>Transrectal ultrasonography</td>
<td>17.6</td>
<td>77.3</td>
<td>4.9</td>
<td>0.2</td>
</tr>
<tr>
<td>MRI</td>
<td>1.5</td>
<td>96.3</td>
<td>2.2</td>
<td>0.1</td>
</tr>
<tr>
<td>CT abdomen</td>
<td>25.5</td>
<td>71.0</td>
<td>3.4</td>
<td>0.2</td>
</tr>
<tr>
<td>CT pelvis</td>
<td>42.9</td>
<td>52.9</td>
<td>4.0</td>
<td>0.2</td>
</tr>
<tr>
<td>US abdomen/pelvis</td>
<td>17.3</td>
<td>79.1</td>
<td>3.5</td>
<td>0.2</td>
</tr>
<tr>
<td>US liver</td>
<td>1.6</td>
<td>95.2</td>
<td>3.1</td>
<td>0.2</td>
</tr>
<tr>
<td>Bone scan</td>
<td>80.2</td>
<td>12.6</td>
<td>7.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Chest x-ray</td>
<td>60.5</td>
<td>35.8</td>
<td>3.7</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Table 6.4: Proportion (%) of patients in the subcohort who had diagnostic tests or had test recommended grouped by radiotherapy and surgical candidacy

<table>
<thead>
<tr>
<th>Diagnostic Test</th>
<th>Radiotherapy %</th>
<th>Surgery %</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digital rectal exam</td>
<td>99.4</td>
<td>99.2</td>
<td>0.65</td>
</tr>
<tr>
<td>Transrectal ultrasound</td>
<td>16.8</td>
<td>20.5</td>
<td>0.15</td>
</tr>
<tr>
<td>Lymph node dissection</td>
<td>88.0</td>
<td>12.0</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>MRI</td>
<td>1.3</td>
<td>6.1</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>CT abdomen</td>
<td>33.1</td>
<td>24.2</td>
<td>0.002</td>
</tr>
<tr>
<td>CT pelvis</td>
<td>62.8</td>
<td>29.9</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>US abdomen/pelvis</td>
<td>16.9</td>
<td>24.9</td>
<td>0.002</td>
</tr>
<tr>
<td>US liver</td>
<td>2.3</td>
<td>7.1</td>
<td>0.0003</td>
</tr>
<tr>
<td>Bone scan</td>
<td>93.5</td>
<td>80.7</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Chest x-ray</td>
<td>58.4</td>
<td>70.3</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

6.3 High Risk Patients

We examined the high risk patients that were excluded from our main study to explore what our conclusions would have been if we had included these patients. Table 6.5 displays the distribution of T category, Gleason score and PSA in the subcohort using the same study population described in Manuscript 1 plus the high risk patients. Only 9% of patients with a T category of T3 or T4 were surgical candidates. For Gleason groups 2-4, 5-6 and 7, approximately 40% to 52% of the patients were surgical candidates but for Gleason 8-10 only 25% of patients were surgical candidates. Similarly, for PSA approximately 37% to 52% of patients with a PSA
≤20 were surgical candidates but this proportion declined to 24% for those patients with PSA >20. Therefore, the majority of high risk patients were not considered surgical candidates.

**Table 6.5:** Characteristics of radiotherapy and surgical patients in the subcohort when high risk patients are included, and the proportion of surgical patients for each variable

<table>
<thead>
<tr>
<th></th>
<th>Radiotherapy (n=512)</th>
<th>Surgery (n=473)</th>
<th>Proportion surgical patients (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T category % (n)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1a/b</td>
<td>6.8 (58)</td>
<td>5.5 (32)</td>
<td>35.6</td>
</tr>
<tr>
<td>T1c</td>
<td>17.2 (148)</td>
<td>34.1 (198)</td>
<td>57.2</td>
</tr>
<tr>
<td>T2a</td>
<td>25.9 (223)</td>
<td>35.0 (203)</td>
<td>47.7</td>
</tr>
<tr>
<td>T2b</td>
<td>37.9 (326)</td>
<td>23.5 (136)</td>
<td>29.4</td>
</tr>
<tr>
<td>T3/4</td>
<td>20.7 (106)</td>
<td>2.3 (11)</td>
<td>9.4</td>
</tr>
<tr>
<td>Gleason Score % (n)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-4</td>
<td>22.2 (191)</td>
<td>22.9 (133)</td>
<td>41.0</td>
</tr>
<tr>
<td>5-6</td>
<td>41.0 (253)</td>
<td>47.1 (273)</td>
<td>52.9</td>
</tr>
<tr>
<td>7</td>
<td>22.5 (194)</td>
<td>22.6 (131)</td>
<td>40.3</td>
</tr>
<tr>
<td>8-10</td>
<td>14.3 (123)</td>
<td>7.4 (43)</td>
<td>25.9</td>
</tr>
<tr>
<td>PSA % (n)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤4</td>
<td>12.5 (108)</td>
<td>10.7 (62)</td>
<td>36.5</td>
</tr>
<tr>
<td>&gt;4 to ≤10</td>
<td>31.1 (268)</td>
<td>50.5 (293)</td>
<td>52.2</td>
</tr>
<tr>
<td>&gt;10 to ≤20</td>
<td>30.3 (261)</td>
<td>26.0 (151)</td>
<td>36.7</td>
</tr>
<tr>
<td>&gt;20</td>
<td>26.0 (224)</td>
<td>12.8 (74)</td>
<td>24.8</td>
</tr>
</tbody>
</table>

Table 6.6 displays the adjusted hazard ratios for death from prostate cancer using only the high risk population, as well as the entire population including the high risk patients. The hazard ratio for death from prostate cancer for radiotherapy candidates compared to surgery candidates for the entire study population including high risk patients was 1.74 (95% CI 1.22-2.47).
Table 6.6: Hazard ratios for death from prostate cancer for radiotherapy compared to surgery and for covariates

<table>
<thead>
<tr>
<th></th>
<th>High Risk Only</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radiotherapy</td>
<td>1.75 (1.07-2.84)</td>
<td>1.74 (1.22-2.47)</td>
</tr>
<tr>
<td>Age</td>
<td>0.98 (0.95-1.01)</td>
<td>0.98 (0.96-1.00)</td>
</tr>
<tr>
<td>T category</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1a/b</td>
<td>0.55 (0.22-1.33)</td>
<td>0.48 (0.25-0.91)</td>
</tr>
<tr>
<td>T1c</td>
<td>0.73 (0.42-1.27)</td>
<td>0.68 (0.45-1.03)</td>
</tr>
<tr>
<td>T2a</td>
<td>0.62 (0.36-1.07)</td>
<td>0.60 (0.41-0.87)</td>
</tr>
<tr>
<td>T2b</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>T3a</td>
<td>1.08 (0.60-1.94)</td>
<td>1.37 (0.80-2.36)</td>
</tr>
<tr>
<td>T3b</td>
<td>1.12 (0.60-2.04)</td>
<td>1.36 (0.74-2.48)</td>
</tr>
<tr>
<td>T4</td>
<td>1.53 (0.44-5.30)</td>
<td>1.33 (0.36-4.93)</td>
</tr>
<tr>
<td>Gleason Score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-4</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>5-6</td>
<td>1.20 (0.70-2.07)</td>
<td>1.38 (0.87-2.17)</td>
</tr>
<tr>
<td>7</td>
<td>1.61 (0.89-2.91)</td>
<td>2.18 (1.42-3.32)</td>
</tr>
<tr>
<td>8-10</td>
<td>3.35 (1.86-6.04)</td>
<td>5.59 (3.63-8.60)</td>
</tr>
<tr>
<td>PSA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤4</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>&gt;4 to ≤10</td>
<td>1.23 (0.55-2.74)</td>
<td>1.15 (0.69-1.94)</td>
</tr>
<tr>
<td>&gt;10 to ≤20</td>
<td>0.72 (0.32-1.61)</td>
<td>1.05 (0.62-1.80)</td>
</tr>
<tr>
<td>&gt;20</td>
<td>1.43 (0.70-2.90)</td>
<td>2.49 (1.50-4.12)</td>
</tr>
</tbody>
</table>

6.4 The Instrumental Variable Method

The study population that was used for this analysis is the same one described in Manuscript 1. We first considered using Cancer Care Ontario Region (CCOR) radiotherapy rate as the instrumental variable. We thought that if there were no significant variations in the proportion of patients who were radiotherapy candidates between counties that make up the CCORs, we could justify using CCOR radiotherapy rate as the instrumental variable. However, as shown in Table 6.7, three out of the eight CCORs have significant county-level variation in their radiotherapy rates.

Next we explored using county radiotherapy rate as the instrumental variable and we plotted the radiotherapy rates in each county and CCOR with their 95% confidence limits, as shown in Figure 6.1a. We wanted to determine if groups of similar counties could be formed...
based on the confidence intervals. However, many of the confidence limits were quite large and this was not useful in determining cut-off values for using county as the instrumental variable.

Our third approach was to assign each patient in the subcohort the radiotherapy rate of his respective county. This is analogous to approaches used in the literature.1, 2 We divided the patients approximately into tertiles based on the radiotherapy rate in order to identify groups of counties that had low, medium and high rates of radiotherapy. When we compared these groups of counties, they differed significantly in their radiotherapy rates, as desired, however as shown in Table 6.8 they also varied in every potential confounder, which is not desired when using instrumental variables. However, the tests for trend only showed a statistically significant increase for age and T category as use of radiotherapy increased.

Lastly, despite the intra-regional variations (Table 6.7) we chose to use CCOR radiotherapy rate as the instrumental variable, since the county rates were too unstable and may have been driven by patient characteristics rather than practice patterns. We examined Figure 6.1b, and grouped the Central East, South West and South regions were together to make the “low” use of radiotherapy group as the point estimates for each of these regions fell within the 95% CI of the other regions. Similarly, we grouped the North West, South East, Central West, and East regions together to form the “moderate” radiotherapy use group. We kept the North East group by itself as it did not fall within the 95% CI of any of the other regions and was deemed to have “high” use of radiotherapy. The potential confounders were compared between these three groups formed and, as illustrated in Table 6.9, all variables except for age differed at least moderately between the two groups. However, there were no statistically significant trends for the severity of cases increasing as radiotherapy use increased.
Table 6.7: Percentage of patients in the subcohort who were radiotherapy candidates based on their Cancer Care Ontario Region (CCOR) radiotherapy rate of residence and county.

<table>
<thead>
<tr>
<th>CCOR</th>
<th>RT Rate (95% CI)</th>
<th>County</th>
<th>RT rate (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Name (n)</td>
<td></td>
<td>Name (n)</td>
<td></td>
</tr>
<tr>
<td>North West</td>
<td>59.4 (52.2-67.3)</td>
<td>Kenora (10)</td>
<td>90 (55.5-99.8)</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rainy River District (8)</td>
<td>87.5 (47.4-99.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thunder Bay District (154)</td>
<td>56.5 (48.3-64.5)</td>
<td></td>
</tr>
<tr>
<td>North East</td>
<td>73.4 (64.1-81.4)</td>
<td>Algoma District (31)</td>
<td>93.6 (78.6-99.2)</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cochrane District (11)</td>
<td>27.3 (6.0-61.0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Manitoulin District (0)</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nipissing District (28)</td>
<td>75.0 (55.1-89.3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Parry Sound District (11)</td>
<td>72.7 (39.0-94.0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Greater Sudbury (17)</td>
<td>64.7 (38.3-85.8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sudbury District (5)</td>
<td>60.0 (14.7-94.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Timiskaming District (6)</td>
<td>83.3 (35.9-99.6)</td>
<td></td>
</tr>
<tr>
<td>East</td>
<td>52.1 (41.6-62.4)</td>
<td>Lanark*</td>
<td>-</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ottawa (83)</td>
<td>51.8 (40.6-62.9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Prescott &amp; Russell (6)</td>
<td>50 (11.8-88.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Renfrew*</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stormont, Dundas &amp; Glengarry*</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>South East</td>
<td>53.6 (43.9-63.2)</td>
<td>Frontenac (29)</td>
<td>34.5 (17.9-54.3)</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hastings (17)</td>
<td>58.8 (32.9-81.6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Leeds &amp; Grenville (15)</td>
<td>73.3 (44.9-92.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lennox &amp; Addington (9)</td>
<td>77.8 (40.0-97.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Northumberland (5)</td>
<td>60 (14.7-94.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Peterborough (30)</td>
<td>50 (33.3-68.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Prince Edward (5)</td>
<td>60 (14.7-94.7)</td>
<td></td>
</tr>
<tr>
<td>Central East</td>
<td>37.9 (31.1-45.0)</td>
<td>Haliburton (0)</td>
<td>-</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Halton R.M. (16)</td>
<td>43.8 (19.8-70.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Toronto (90)</td>
<td>41.1 (30.8-52.0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Muskoka D.M. (5)</td>
<td>60 (14.7-94.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Durham R.M. (28)</td>
<td>42.9 (24.5-62.8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Peel R.M. (26)</td>
<td>19.2 (6.6-39.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Simcoe (8)</td>
<td>37.5 (8.5-75.6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kawartha Lakes (8)</td>
<td>62.5 (24.5-91.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>York R.M. (15)</td>
<td>20.0 (4.3-48.1)</td>
<td></td>
</tr>
<tr>
<td>Central West</td>
<td>59.0 (48.4-68.9)</td>
<td>Brant (14)</td>
<td>71.4 (41.9-91.6)</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dufferin*</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Haldimand &amp; Norfolk (13)</td>
<td>61.5 (31.6-86.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Niagara R.M (21)</td>
<td>52.4 (29.8-74.3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wellington (16)</td>
<td>43.8 (19.8-70.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hamilton (27)</td>
<td>70.4 (49.8-86.3)</td>
<td></td>
</tr>
<tr>
<td>South West</td>
<td>44.9 (35.2-54.8)</td>
<td>Bruce*</td>
<td>57.1 (18.4-90.1)</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>-----------------</td>
<td>---------</td>
<td>-----------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Elgin (7)</td>
<td>28.6 (3.7-71.0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Grey (7)</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Huron*</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chatham-Kent*</td>
<td>73.7 (48.8-90.9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lambton (19)</td>
<td>31.3 (16.1-50.0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Middlesex (32)</td>
<td>80 (28.4-99.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oxford (5)</td>
<td>20 (0.5-71.6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Perth (5)</td>
<td>39.1 (19.7-61.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Waterloo R.M. (23)</td>
<td>40.5 (31.8-49.6)</td>
<td></td>
</tr>
</tbody>
</table>

*numbers suppressed due to sample size <5

Table 6.8: Comparison of characteristics across instrumental variable groups, using county radiotherapy rate as the instrument.

<table>
<thead>
<tr>
<th></th>
<th>High % Radiotherapy (n=268)</th>
<th>Moderate% Radiotherapy (n=344)</th>
<th>Low % Radiotherapy (n=372)</th>
<th>p-value for trend</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Radiotherapy (95% CI)</td>
<td>74.0 (0.69-0.79)</td>
<td>52.0 (0.47-0.57)</td>
<td>36.0 (0.31-0.41)</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td>Average Age (SD)</td>
<td>67.0 (6.3)</td>
<td>66.1 (6.8)</td>
<td>65.5 (6.7)</td>
<td>0.02</td>
<td>0.006</td>
</tr>
<tr>
<td>Average CIRS-G (SD)</td>
<td>5.3 (3.1)</td>
<td>5.9 (3.8)</td>
<td>5.4 (3.6)</td>
<td>0.07</td>
<td>0.84</td>
</tr>
<tr>
<td>T category</td>
<td></td>
<td></td>
<td></td>
<td>0.003</td>
<td>0.05</td>
</tr>
<tr>
<td>T1a/b</td>
<td>7.8 (21)</td>
<td>11.1 (38)</td>
<td>4.8 (18)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1c</td>
<td>29.1 (78)</td>
<td>25.0 (86)</td>
<td>29.8 (111)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2a</td>
<td>28.0 (75)</td>
<td>37.2 (128)</td>
<td>37.6 (140)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2b</td>
<td>35.1 (94)</td>
<td>26.7 (92)</td>
<td>27.7 (103)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gleason Score</td>
<td></td>
<td></td>
<td></td>
<td>0.08</td>
<td>0.46</td>
</tr>
<tr>
<td>2-4</td>
<td>20.9 (56)</td>
<td>30.5 (105)</td>
<td>26.6 (99)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-6</td>
<td>54.9 (147)</td>
<td>49.1 (169)</td>
<td>48.7 (181)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>24.3 (65)</td>
<td>20.4 (70)</td>
<td>24.7 (92)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSA</td>
<td></td>
<td></td>
<td></td>
<td>0.007</td>
<td>0.81</td>
</tr>
<tr>
<td>≤4</td>
<td>11.6 (31)</td>
<td>17.7 (61)</td>
<td>11.0 (41)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;4 to ≤10</td>
<td>47.4 (127)</td>
<td>52.9 (182)</td>
<td>52.4 (195)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;10 to ≤20</td>
<td>41.0 (110)</td>
<td>29.4 (101)</td>
<td>36.6 (136)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 6.9: Comparison of characteristics across instrumental variable groups, using CCOR radiotherapy rate as the instrument.

<table>
<thead>
<tr>
<th></th>
<th>High % Radiotherapy (n=111)</th>
<th>Moderate% Radiotherapy (n=458)</th>
<th>Low % Radiotherapy (n=416)</th>
<th>p-value</th>
<th>p-value for trend</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Radiotherapy</td>
<td></td>
<td></td>
<td></td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td>Average Age (SD)</td>
<td>66.8 (6.2)</td>
<td>66.0 (6.6)</td>
<td>66.0 (6.8)</td>
<td>0.49</td>
<td>0.41</td>
</tr>
<tr>
<td>Average CIRS-G score (SD)</td>
<td>4.7 (2.9)</td>
<td>5.7 (3.7)</td>
<td>5.6 (3.6)</td>
<td>0.03</td>
<td>0.18</td>
</tr>
<tr>
<td>T category</td>
<td></td>
<td></td>
<td></td>
<td>0.0002</td>
<td>0.24</td>
</tr>
<tr>
<td>T1a/b</td>
<td>7.2 (8)</td>
<td>10.2 (50)</td>
<td>4.6 (19)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1c</td>
<td>19.8 (22)</td>
<td>27.5 (126)</td>
<td>30.5 (127)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2a</td>
<td>29.7 (33)</td>
<td>32.8 (150)</td>
<td>38.7 (161)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2b</td>
<td>43.2 (48)</td>
<td>28.8 (132)</td>
<td>26.2 (109)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gleason Score</td>
<td></td>
<td></td>
<td></td>
<td>0.13</td>
<td>0.77</td>
</tr>
<tr>
<td>2-4</td>
<td>20.7 (23)</td>
<td>27.7 (127)</td>
<td>26.4 (110)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-6</td>
<td>62.1 (69)</td>
<td>48.7 (223)</td>
<td>49.5 (206)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>17.1 (19)</td>
<td>23.6 (108)</td>
<td>24.0 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSA</td>
<td></td>
<td></td>
<td></td>
<td>0.11</td>
<td>0.48</td>
</tr>
<tr>
<td>≤4</td>
<td>17.1 (19)</td>
<td>15.1 (69)</td>
<td>10.8 (45)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;4 to ≤10</td>
<td>42.3 (47)</td>
<td>51.1 (234)</td>
<td>53.9 (224)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;10 to ≤20</td>
<td>40.5 (45)</td>
<td>33.8 (155)</td>
<td>35.3 (147)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 6.1: Proportion of patients who were radiotherapy (RT) candidates based on (a) county and (b) CCOR with 95% confidence intervals.
6.5 References


Chapter 7: General Discussion

7.1 Discussion of Key Findings

7.1.1 Treatment Effectiveness

The hazard ratio for death from prostate cancer for radiotherapy candidates compared to surgical candidates using intent to treat was 1.44 (95% CI 0.86-2.40). When stratified by risk group, the hazard ratios were 1.58 (0.37-6.67) and 1.34 (0.77-2.32) for the low and intermediate risk groups respectively. When we analyzed by actual treatment received, the hazard ratio for death from prostate cancer for radiotherapy compared to surgery was 1.84 (1.06-3.17). From these results we concluded that analyzing by actual treatment received might overestimate the survival benefit of surgery compared to radiotherapy because nodal spread is underestimated in radiotherapy candidates compared to surgical candidates when analyzing actual treatment received. Actual treatment received excludes surgical candidates who have a positive lymph node dissection prior to treatment, and there is no comparable group of radiotherapy candidates, as radiotherapy candidates typically do not receive lymph node dissections prior to treatment.

We also used propensity scores to assess whether the treatment effect estimate differed among groups whose indication for radiotherapy varied. When we analyzed using propensity scores, the hazard ratios for death from prostate cancer for radiotherapy compared to surgery from the lowest quintile to highest quintile (for the probability of receiving radiotherapy) were 0.30 (95% CI 0.04-2.28), 1.54 (0.35-6.77), 0.90 (0.29-2.82), 2.71 (1.01-7.31), and 1.08 (0.41-2.81). While these differences were not statistically significant, if they were to hold in a future, more highly powered study, differences of this magnitude would indicate that hazard ratios are not consistent across groups whose indication for radiotherapy varies. In other words, there would be some groups of patients who would benefit more from surgery than other groups within the range of case presentations who are candidates for that treatment.

A study completed by Albertsen et al found a prostate cancer-specific survival decrement for radiotherapy compared to surgery, with a 13-year hazard ratio for prostate cancer-specific
death of 2.2 (1.6-3.1). This study used analysis by intent to treat, however it included patients with a T category of T3 and T4. Similarly, Merglen et al found a 10-year prostate-cancer specific decrement for radiotherapy compared to surgery with a hazard ratio of 2.3 (1.2-4.3). However, this study did not use intent to treat, and also included patients with a T category of T3. A study by Tewari et al also showed a 15-year survival benefit for surgery over radiotherapy for localized prostate cancer, with a hazard ratio of 0.60 (0.39-0.92). However, this study did not take T category or PSA into account, which are both important prognostic indicators in prostate-cancer specific survival. Another study by Tewari et al, which examined high grade tumours, found a 15-year prostate cancer-specific survival hazard ratio of 0.51 (0.26-1.01) for prostatectomy compared to radiotherapy. However, once again they did not control for T category or PSA. These two studies also included patients diagnosed as early as 1980, which is earlier than any other studies comparing the treatment modalities, and may not reflect treatment patterns seen in later eras.

Several studies have used biochemical relapse free survival (bRFS) to compare the relative treatment effectiveness of radiotherapy to surgery. A study completed by Kupelian et al found that there was no difference in biochemical relapse-free survival (bRFS) between radiotherapy and surgery with a hazard ratio of 1.01 (95% CI 0.77-1.32). However, when they stratified radiotherapy patients by those receiving <72 Gy and those ≥72 Gy, in a stratified analysis by favourable (T category of T1 or T2a, Gleason score ≤6, and PSA ≤10) and unfavourable cancers, they found that there was not difference in survival between radical prostatectomy and radiotherapy with a dose ≥72 Gy (p=0.16 and p=0.96 respectively). However, they did find that radical prostatectomy provided a survival advantage compared to doses <72 Gy in both the favourable (p<0.001) and unfavourable (p<0.001) groups. In an extension of this study, radiotherapy was divided into doses <72 Gy and ≥72 Gy for the multivariate analysis. They found that radiotherapy had a bRFS decrement compared to surgery with a hazard ratio of 2.24 (1.83-2.73) when examining those patients with a total dose <72 Gy
but there was no difference in bRFS for those with a dose $\geq 72$ Gy with a hazard ratio of 1.08 (0.78-1.50). This indicates that doses less than 72 Gy may not provide the same survival advantage as radical prostatectomy.

D’Amico et al found that the 8-year estimates for bRFS were significantly higher for patients treated with radical prostatectomy compared to radiotherapy when analyzing low-risk patients and intermediate risk patients with low-volume tumours, however radiotherapy was not associated with decreased 8-year biochemical relapse free survival in intermediate risk patients with high-volume tumours or high risk patients. A limitation of this study is that although they stratified patients by risk group (and tumour volume for intermediate risk patients), differences in prognostic indicators within each risk group could be contributing to the differences in survival. Potters et al found that the difference in 7-year biochemical relapse free survival between radiotherapy compared to surgery was not statistically significant with an adjusted hazard ratio of 1.18 (0.86-1.62). Martinez et al found no difference in the 7-year odds ratio for bRFS with an odds ratio of 0.98 (0.55-1.74) for surgery compared to radiotherapy. Because odds ratios do not take time to event and censoring into account, and because follow-up was not complete for all patients, a hazard ratio may have been a better alternative to compare treatment effectiveness in this study.

It is interesting to note that most studies that have found a statistically significant survival advantage for surgery over radiotherapy have analyzed prostate cancer-specific survival rather than bRFS. Using bRFS as a surrogate endpoint for prostate cancer-specific death may be an issue when comparing treatment effectiveness in radiotherapy to surgery for two reasons. First, using bRFS has not been extensively validated as a surrogate endpoint in any prognostic group or for any treatment. Second, when examining survival curves of radiotherapy compared to surgery using survival curves, the differences in curves are constant when using prostate cancer-specific survival. However, it appear that when using bRFS the survival curves diverge early between radiotherapy and surgery but differences in survival cease after approximately 6-9 years.
7.1.2 The Instrumental Variable Method

We attempted to use the instrumental variable method to assess the difference in treatment effectiveness between radiotherapy and surgery for localized prostate cancer. However, as shown in Section 6.3, the distribution of nearly every variable was either marginally or significantly statistically different across groups of regions formed using both Cancer Care Ontario Region (CCOR) and county radiotherapy rate as the instrumental variable. This violated one of the key assumptions of the instrumental variable method: that there are no differences in unknown covariates, as shown by there being no difference in known covariates. Although only age and T category had a statistically significant trend when using county as the instrumental variable, we were not comfortable using either county or CCOR radiotherapy rate as an instrumental variable due to the large variations in the distributions of covariates.

We then calculated radiotherapy rates in CCORs using radiotherapy data we had available on the entire target population, grouping regions based on those radiotherapy rates. We compared covariates across the instrumental variable groups thus formed using data from the subcohort and found again that every covariate was statistically significantly different among the groups. However, only three variables had marginally or statistically significant trends. As radiotherapy use increased, patients were more likely to likely to have a T category of T2b and age slightly increased. Gleason score decreased as use of radiotherapy increased. We saw no pattern in total CIRS-G or PSA. We concluded that there might have been some systematic differences between CCORs in T category and Gleason score were assignment because trends were not convincing when comparing them to the differences in distribution of prognostic indicators between radiotherapy and surgery. We concluded that our choice of CCOR radiotherapy rate was not an adequate instrumental variable in this population. Other instrumental variables, such as distance to a radiotherapy centre, would have to be explored to determine if they meet the assumptions of the instrumental variable method.
7.2 Discussion of Supplemental Results

7.2.1 High Risk Patients

We restricted our analyses to those patients who appeared to be candidates for both treatments based on these clinical guidelines.\textsuperscript{11, 12} We included these patients in a subanalysis to examine what our conclusions would have been if we had not excluded them. Only 9% of patients with a T category of T3 or T4 were surgical candidates. Therefore, even without imposing clinical guidelines to define the study population, it appears that surgeons usually do not consider men with a T category of T3 or T4 as surgical candidates. A similar pattern occurs to a less dramatic extent for Gleason score and PSA. For Gleason groups 2-4, 5-6 and 7, approximately 40% to 52% of the patients were surgical candidates but for Gleason 8-10 only 25% of patients were surgical candidates. Similarly, for PSA approximately 37% to 52% of patients with a PSA $\leq 20$ were surgical candidates but this proportion declined to 24% for those patients with PSA $>20$. Therefore, the majority of high risk patients were not considered surgical candidates.

The hazard ratio for death from prostate cancer for radiotherapy candidates compared to surgery candidates for the entire study population including high risk patients was 1.74 (95% CI 1.22-2.47). The hazard ratio using only the low and intermediate risk groups, as stated in Manuscript 1, was 1.44 (0.86-2.40). There are two possible reasons for this increase in hazard ratio. First, if we had included high risk patients in our study population, it is possible that we may have overestimated the survival benefit of surgery compared to radiotherapy due to unmeasured confounding. Horwitz et al assessed the mortality reduction of $\beta$-Blockers following a heart attack using the results from a previous randomized clinical trial (RCT) and using a restricted and unrestricted observational study, where the restricted study used the same exclusion criteria as the RCT.\textsuperscript{13} The RCT indicated a mortality reduction of 21% using $\beta$-Blockers.\textsuperscript{13} Similarly, the restricted study population found a mortality reduction of 22% whereas the unrestricted population overestimated the mortality reduction at 29%.\textsuperscript{13} In a similar way, it is
possible that unknown confounders may be biasing the results away from the null when we included the high risk patients. Another explanation is that there is no difference in the hazard ratios as the point estimate using the high risk patients did fall within the 95% confidence interval of the hazard ratio found using only low and intermediate risk patients. If this is true, a statistically significant result may have been shown due to an increase in study population and prostate cancer death numbers.

The hazard ratio for death from prostate cancer for radiotherapy candidates compared to surgical candidates using only the high risk patients was 1.75 (1.07-2.84). It appears that the point estimate for the overall population, including high risk patients, was driven by the high risk patients as the point estimate using the overall population and the point estimate using only the high risk patients were very similar.

In conclusion, the majority of high risk patients were not surgical candidates. If we had used high risk patients in our study population we may have overestimated the survival benefit of surgery compared to radiotherapy as controlling for case-mix did not keep the hazard ratio the same as when only low and intermediate patients were used. Therefore, there may be confounders that are not controlled for that are distorting the hazard ratio estimate when high risk patients are included in the analysis.

7.2.2 Diagnostic Tests

Radiotherapy candidates were statistically significantly more likely to receive CT scans of the abdomen, CT scans of the pelvis, and bone scans, while those who were surgical candidates were statistically significantly more likely to receive MRIs, ultrasounds of the abdomen/pelvis, ultrasound of the liver, and chest x-rays. One explanation for the difference in diagnostic tests done in the surgical and radiotherapy candidates is that the charts for the surgical group are not as complete as those for the radiotherapy group, and the abstractors may have recorded a specific test as “test not done” because there was no reference to this test, when in
actual fact it was done and the information was missing. Assuming that the data are accurate, although we have restricted the study to low and intermediate risk patients, the radiotherapy group still has slightly worse disease, and may have had more tests, for example bone scans, to rule out the spread of the disease. However, surgical patients were more likely to have chest x-rays to check for spread. An alternate explanation is that the diagnostic benefit of some of the tests may overlap: for example CT scans of the abdomen/pelvis and lymph node dissections. While surgical patients were more likely to have their lymph node status checked upon surgery, radiotherapy patients usually do not have these done prior to treatment and may have had CT scans instead in order to check for regional spread. If the data are accurate and the difference is not due to different distribution in prognostic indicators, the amount of undetected spread to the nodes and metastases may have been differential between the surgical and radiotherapy group, although it is difficult to determine which direction that differential may be.

### 7.3 Statistical Power

We originally estimated that with 80% power and a significance level of 0.05, we would have the power to detect a hazard ratio as small as 1.25, using a power calculation designed for case-cohort studies. However, I greatly overestimated the amount of overlap in patients with a PSA>20, Gleason score >7 and T category >T2a. Therefore, many more patients were excluded than we had expected. We calculated the detectable difference with the expectation that there would be 1350 patients in the subcohort after necessary exclusions when in fact only 985 patients remained. The smallest detectable hazard ratio would be much higher with a much smaller subcohort. However, because it has been suggested that post-hoc power calculations are not appropriate, we have not recalculated our minimal detectable difference.

The hazard ratio for death from prostate cancer for radiotherapy patients compared to surgical patients was 1.44 (95% CI 0.86-2.40). With the lower confidence limit of 0.86 we would expect that if we had a greater sample size, this result may have become statistically significant.
When we stratified the population by low and intermediate risk patients, our power to detect a
difference became even lower. The hazard ratios for the low risk group had a very large
confidence interval, with 95% confidence limits of 0.37 to 6.67, indicating a lack of power. Our
power to detect a difference in the intermediate risk also decreased, however to a less dramatic
extent than the low risk patients as most patients who died were intermediate risk patients.

7.4 Limitations

The most important limitation of this study is that it was an observational study, and
unknown confounding may explain the results. We planned to use instrumental variables to limit
this unknown confounding, but we were unable to use this method as one of the primary
assumptions was violated. A randomized clinical trial (RCT) would be the ideal study design to
address this unknown confounding but an RCT is unlikely to be performed in this population and
if observational studies are not done, the alternative would be to continue to have limited, poor
quality information about treatment effectiveness.

There are several issues surrounding our Gleason score variable that may have caused
residual confounding. First, there was no central pathology review so Gleason score estimations
could vary by physician. Next, grade was used to approximate Gleason score when Gleason score
was not available, and 15% of Gleason scores were estimated this way. Because Gleason score
ranks differentiation based on the two most common patterns of growth and not the pattern of the
overall tumour, this may have caused misclassification for some of the patients in whom grade
was used to approximate Gleason score. Also, there was no grade to approximate Gleason score
7. Therefore, patients who used grade to approximate Gleason score who actually had a Gleason
score of 7 were likely placed into the Gleason score 5-6 group or the 8-10 group, again causing
misclassification. Next, reviewed Gleason score agreed with original Gleason score in only 66%
of patients. If we assume that reviewed Gleason score is superior to original Gleason score, then
original Gleason score overestimated Gleason score in 12% of patients and underestimated
Gleason score in 22% of patients. However, these underestimations and overestimations are likely inflated, as some hospitals in which the original biopsies were done were tertiary hospitals, and these hospitals would likely provide an accurate original Gleason scores without the need for a reviewed score. Finally, Gleason score grading has changed over time. A study by Ghani et al found that between 1991 and 2001, the proportion of reported Gleason score 2-4 significantly decreased while the proportion of Gleason score 5-7 significantly increased. They also randomly selected 50 patients to have their Gleason scores reviewed and found that all of the reviewed Gleason scores of 2-4 were upgraded to 5-7. This has two impacts for our study. First, the reporting of Gleason score could have changed over time, as our study had participants who were diagnosed between 1990 and 1998. This means that patients who had the same differentiation patterns could have different Gleason scores based on their year of diagnosis. Secondly, the distribution of Gleason scores and their associated hazard ratios should be interpreted with caution as they may not represent what is occurring in the current era.

There were many patients with unknown PSA, Gleason score and/or T category and 8.2% of the study population was excluded for this reason. As shown in Table 6.1 in the “Supplemental Results” chapter, it appears that surgical candidates were slightly more likely to have missing covariate values than radiotherapy candidates in the subcohort and were more likely to have missing covariate values than radiotherapy candidates in the case groups. This probably occurred because radiotherapy charts were more complete than surgery charts. This may have led to an overestimation of the treatment benefit of surgery compared to radiotherapy due to the higher proportion of cases excluded from the surgical candidates compared to the radiotherapy candidates.

Another limitation is that there may be differential staging between radiotherapy and surgical candidates. Most diagnostic tests done were statistically significantly different between radiotherapy and surgical candidates. Some tests were more likely to be done for surgical patients, while some were more likely to be done for radiotherapy patients. If this caused
differential nodal and metastases staging between the two groups, it is possible that we may have overestimated or underestimated the hazard ratio of death for radiotherapy patients compared to surgical patients.

Conventional radiotherapy was used during this time period while currently 3-dimensional conformal radiotherapy is the standard of care. 3-dimensional conformal radiotherapy allows for a higher radiotherapy dose to be administered to the prostate, and has shown improved outcomes compared to conventional radiotherapy.\textsuperscript{17, 18} Therefore, dose-escalated radiotherapy in the modern era may produce lower hazard ratios for radiotherapy compared to surgery than we observed.

Finally, cases were chosen with deaths only up to 10 years post-diagnosis. Because of the slow-growing nature of prostate cancer, following patients for more than 10 years may be necessary in order to detect a difference in effectiveness between treatment modalities. Also, cases included in our study may have had more advanced cancers and may not be representative of all patients who will die of prostate cancer. For example, a patient diagnosed in 1990 who died of prostate cancer 10 years after diagnosis could have been chosen for our study, whereas a case who was diagnosed in 1997 would have died of prostate cancer at only a maximum of only two years (as patients were chosen as of December 31, 1999). Therefore, a higher proportion of patients with more advanced disease who died early of prostate cancer would have been chosen.

A limitation of our instrumental variable component is that we used the covariates of the chart-reviewed patients to approximate the covariates of the study frame from which they were selected. However, the rate of radiotherapy use in each Cancer Care Ontario Region was similar between the chart-reviewed cohort and the administrative cohort. Also, the average age was similar in the instrumental variable groups formed using the chart-reviewed cohort and the administrative cohort. This gives us more confidence that the subcohort is an accurate representation of the study population sampling frame.
7.5 Strengths

Our study has several strengths. It was a population based study, with subjects accrued from all across Ontario, therefore the results were not influenced by the practice patterns, expertise or referral bias of a few select hospitals. We were able to use detailed information on covariates due to the extensive chart review that was conducted on each patient, which provided more information than is available in an administrative database. This detailed information also allowed us to use intent to treat, rather than actual treatment received, which reduced the bias favouring surgical treatment.

Most studies that compare treatment effectiveness between radiotherapy and surgery use biochemical relapse-free survival. We used death from the cancer as the outcome, which is the more definitive endpoint. Shortly after the discovery of PSA as a biomarker for prostate cancer, Killian et al described the benefit of PSA testing as a method of detecting treatment failure for as it precedes detection of clinical disease recurrence by about 12 months.\textsuperscript{19} Due to the long natural history of prostate cancer, bRFS has the potential to offer information on treatment effectiveness long before information on disease-free survival is available. However, PSA is not a validated surrogate endpoint for prostate cancer death for any prognostic group or treatments.\textsuperscript{10} Because the value of using PSA as a surrogate endpoint has not been proven, using the clinical endpoint of death may be a better endpoint to use. Therefore, our study adds to the relatively few studies that compare prostate cancer-specific survival of radiotherapy compared to surgery while controlling for known covariates.

We were able to determine each patient’s original treatment plan from the detailed information from the chart review. Nearly all studies have used actual treatment received, which may introduce bias favouring surgery over radiotherapy when studying this population. Our study confirmed the conclusions of Lu Yao and Yao that using actual treatment received may overestimate the survival benefit of surgery.\textsuperscript{20} Also, we restricted patients to only those who were low and intermediate risk patients, while other studies have not been as stringent.\textsuperscript{1,2}
Finally, the case-cohort design is an efficient study design when extensive work is needed to abstract patient information. Our study had 106 prostate deaths due to the case-cohort study design, whereas a similar sized cohort study would only have around 19 such deaths.

7.6 Future Directions

Because a randomized clinical trial is unlikely to be conducted to compare the effectiveness of radiotherapy to surgery for this disease, future studies should focus on designing optimal observational studies. Ideally, these would be prospective studies conducted so that data collection can be comprehensive, consistent and of optimal quality. Such studies should use multiple methodological and analytical approaches to deal with unmeasured confounding. For example, these studies could use intent to treat to account for surgical patients who have positive lymph node dissections and should be restricted to those patients who are candidates for both treatments. These studies should also concentrate on accruing enough patients for subgroup analysis, such as the stratification by low and intermediate risk groups and stratification by propensity score quintiles. Future studies should also place a focus on obtaining patients who received 3-dimensional conformal radiotherapy in order to better assess the treatment differences that may be occurring in the current treatment era. This technique has only become common practice within the last 2 years. Therefore, a population-based study similar to our study would not be available for approximately 10 years. Because of the rare outcome of prostate cancer deaths, future studies should also consider the case-cohort or nested case-control design, perhaps nested within a larger cohort initiative. Finally, our study methods and analytical approaches used for this observational study are of value in other settings where an RCT is unlikely to be performed because treatments cross medical specialties and where there are major differences in treatment modalities.
7.7 References


Appendices

Appendix I: Continuous versus Categorical Variables for Age and CIRS-G

Figure I.I: Unadjusted hazard ratios (HR) for death from prostate cancer by age as a continuous variable and as a categorical variable. Error bars show the 95% confidence intervals for the categorical variables.

Figure I.II: Unadjusted hazard ratios (HR) for death from prostate cancer by total CIRS-G as a continuous variable and as a categorical variable. Error bars show the 95% confidence intervals for the categorical variables.
Appendix II: Relevant Excerpts from Chart Abstraction Form

HISTOLOGY, GLEASON SCORE AND GRADE

Histology

Date of first positive biopsy (dd) (MMM) (yyyy)  
2. Was this an adenocarcinoma? Y/N/U  

Gleason Score

(#) 3. Is a Gleason Score available prior to treatment? Y/N  
4. Primary Grade:  
5. Secondary Grade  
6. Gleason Score  

Grade

(#) 7. Is a general statement about grade available prior to treatment? Y/N  
If yes:  
___ GX Grade cannot be assessed  
___ G1 Well differentiated (slight anaplasia)  
___ G2 Moderately differentiated (moderate anaplasia)  
___ G3-4 Poorly differentiated or undifferentiated (marked anaplasia)  

Histology Review

8. Date of histology review: (dd) (MMM) (yyyy)  
9. Was this an adenocarcinoma? Y/N  

Reviewed Gleason Score

(#) 10. Is a reviewed Gleason score available prior to treatment? Y/N  
11. Primary Grade from reviewed score:  
12. Secondary Grade from reviewed score:  
13. Gleason Score from reviewed score  

Reviewed Grade

(#) 14. Is a general statement about reviewed grade available prior to treatment? Y/N  
If yes:  
___ GX Grade cannot be assessed  
___ G1 Well differentiated (slight anaplasia)  
___ G2 Moderately differentiated (moderate anaplasia)  
___ G3-4 Poorly differentiated or undifferentiated (marked anaplasia)  

PRE-TREATMENT MEDS

Pre-treatment hormone therapy
Date exogenous hormone therapy began (pre-radiotherapy or pre-surgery):  
(dd) (MMM) (yyyy)
**COMORBIDITY (Cumulative Illness Rating Scale – CIRS)**

Example of CIRS-G scoring guide:

<table>
<thead>
<tr>
<th>CIRS - 0</th>
<th>No problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIRS - 1</td>
<td>Remote MI (&gt; 5 yrs ago)/ occasional angina treated with prn meds/ mild atherosclerotic heart disease/ detectable murmurs that indicate valvular pathology without activity restriction (<em>more severely compromising valvular disease would require a higher rating, see instruction manual</em>)</td>
</tr>
<tr>
<td>CIRS - 2</td>
<td>CHF compensated with daily meds/ daily anti-angina meds/ left ventricular hypertrophy/ ECG findings of atrial fibrillation/ right or left bundle branch block/ daily antiarrhythmic drugs/ moderate atherosclerotic heart disease/ placement of a pacemaker for an incidental finding of periods of bradycardia during a holter monitor</td>
</tr>
<tr>
<td>CIRS - 3</td>
<td>Previous MI within 5 yrs/ abnormal stress test/ status post percutaneous coronary angioplasty or coronary artery bypass graft surgery/ CHF (intermediate severity)/ severe atherosclerotic heart disease/ bifascicular block/ placement of a pacemaker for cariogenic syncope/ pericardial effusion/ pericarditis</td>
</tr>
<tr>
<td>CIRS - 4</td>
<td>Marked activity restriction secondary to cardiac status (ie. unstable angina or intractable congestive heart failure)/ extremely severe atherosclerotic heart disease</td>
</tr>
<tr>
<td>CIRS - 8</td>
<td>Heart condition not listed above or lacking sufficient information to score. Describe:</td>
</tr>
</tbody>
</table>

**Diagnostic Evaluation**

(#) CODE EACH ITEM IN THE FOLLOWING LIST. USE INFORMATION PRESENT IN THE CHART FROM ALL STAGING AND DIAGNOSTIC TESTS THAT WERE ADMINISTERED.

Code:
0 = test not done; 1=abnormal, suggestive of cancer; 2=abnormal, not suggestive of cancer; 3=normal; 4=procedure attempted and incomplete 8=test done, results unknown; 9=unknown if test done

DRE ___
Transrectal ultrasonography (TRUS) ___
TRUS with biopsy ___
MRI ___
CT scan (abdomen) ___
CT scan (pelvis) ___
Ultrasound (abdomen/pelvis) ___
Ultrasound (liver) ___
Ultrasound (other than abdomen/pelvis or liver) ___ Specify: ____________
Bone scan ___ Full body? Y/N ___
Chest x-ray ___
X-ray (other than chest) ___ Specify: ____________
Needle biopsy of the prostate____ 
TURP and/or Cystoscopy____ 
Lymphangiogram ____
Other____ Specify: _____________________

Labs:
Alkaline Phophatase (ALP) __
If ALP = 1, specify the actual value: ____
Aspartate Aminotransferase (AST) ____
If AST = 1, specify the actual value____
Alanine Aminotransferase (ALT) ___
If AST = 1, specify the actual value____
Serum creatinine ______
If Serum creatinine = 1, specify the actual value____
Prostatic Acid Phosphatase (PAP)
If PAP = 1, specify the actual value ______

PSA Pre-Treatment

(#) Were any PSA levels recorded in the chart from tests done up to one year prior to treatment? (Yes, No or not Mentioned) ________
First PSA level: (dd) ___(MMM) ____ (yyyy)____ 1st Score ____ ng/ml
Second PSA level: (dd) ___(MMM) ____ (yyyy)____ 2nd Score ____ ng/ml
Third PSA level: (dd) ___(MMM) ____ (yyyy)____ 3rd Score ____ ng/ml
Fourth PSA level: (dd) ___(MMM) ____ (yyyy)____ 4th Score ____ ng/ml

Pre-treatment Lymph Node Dissection (not as part of completed or aborted prostatectomy)

(#) Regional lymph node surgery? Y/N ____
Type of regional lymph node surgery:
Laparoscopic ____
Open ____
Number of lymph nodes resected: ______
Number of positive lymph nodes: ______
Details unavailable (verbatim description): ______

Clinical Disease Stage

LOOK FOR ANY STATEMENTS OF CLINICAL TNM OR ABCD STAGING MADE BY PHYSICIANS. THE ONLY STAGE ASSIGNMENT THAT CAN BE REPORTED HERE IS THAT MADE ON THE BASIS OF EVIDENCE ACQUIRED BEFORE DEFINITIVE TREATMENT. IF STAGE IS NOT PRESENT IN THE CHART OR THE STAGE CATEGORY DOES NOT APPEAR IN THE FOLLOWING LISTS, OR THE STAGE CATEGORY IS PRESENT IN THE LIST BUT MARKED WITH AN (*), YOU MUST COMPLETE THE EXTENSION OF THE TUMOUR SECTION OF THE ABSTRACTION FORM. RECORD THE N AND M STAGE AS THEY APPEAR IN THE CHART. YOU DO NOT HAVE TO COMPLETE THE EXTENT OF DISEASE SECTION IF ONLY N (NODES) OR M (METASTATIC) STAGE IS MISSING.

TNM or ABCD classification assigned by a clinician:
<table>
<thead>
<tr>
<th>TNM</th>
<th>ABCD</th>
</tr>
</thead>
<tbody>
<tr>
<td>T0*</td>
<td>A*</td>
</tr>
<tr>
<td>TX*</td>
<td>A1</td>
</tr>
<tr>
<td>T1*</td>
<td>A2</td>
</tr>
<tr>
<td>T1a*</td>
<td>B*</td>
</tr>
<tr>
<td>T1b*</td>
<td>B0</td>
</tr>
<tr>
<td>T1c</td>
<td>B1</td>
</tr>
<tr>
<td>T2*</td>
<td>B2</td>
</tr>
<tr>
<td>T2a*</td>
<td>C*</td>
</tr>
<tr>
<td>T2b*</td>
<td>C1</td>
</tr>
<tr>
<td>T2c*</td>
<td>C2*</td>
</tr>
<tr>
<td>T3*</td>
<td>D*</td>
</tr>
<tr>
<td>T3a</td>
<td>D1</td>
</tr>
<tr>
<td>T3b*</td>
<td>D2</td>
</tr>
<tr>
<td>T3c</td>
<td></td>
</tr>
<tr>
<td>T4</td>
<td></td>
</tr>
<tr>
<td>T4a</td>
<td></td>
</tr>
<tr>
<td>T4b</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Other stage* (record verbatim): ______

N Stage: _______ M Stage: _______

**INITIAL TREATMENT**

*Initial Surgery and Morbidity*

(#{})Regional lymph node surgery? Y/N ______

Type of regional lymph node surgery:
Laparoscopic ______
Open ______

**ADDITIONAL INITIAL TREATMENT**

*Hormone Treatment (started during or after the completion of radiotherapy or surgical treatment)*

Date of orchiectomy: (dd) ___ (MMM) ____ (yyyy)_____

**LATEST STATUS**

*Final Disease Status*

IDENTIFY THE MOST RECENT INFORMATION ON THE PATIENT

(#{}) (A) Vital status: Alive _____ Dead ______

(#{}) (B) Date of last contact/death: (dd) ___ (MMM) ____ (yyyy)_____

117