

**VALIDATION OF QUALITY INDICATORS  
FOR RADICAL PROSTATECTOMY**

by

Ellen Oi Man Chan

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

August, 2007

Copyright © Ellen Oi Man Chan, 2007

## ABSTRACT

**Background:** Radical prostatectomy is the surgical procedure performed on men with clinically localized prostate cancer. In recent years, radical prostatectomy quality indicators have been recommended, but the feasibility and validity for many of these listed surgical quality indicators have yet to be examined. We tested the convergent construct validity of these quality indicators by assessing their associations with hospital volume, a variable repeatedly associated with the quality of surgical care, for prostate cancer patients treated with radical prostatectomy.

**Objectives:** (1) To assess variations in quality indicators by hospital volume; and (2) To investigate whether certain explanatory variables account for some of the variation observed in Objective 1.

**Methods:** This was a retrospective cohort study using medical chart review data that had already been collected as part of a parent study. The study population consisted of a stratified random sample of prostate cancer patients diagnosed between 1990 and 1998 in Ontario, who were treated by radical prostatectomy with curative intent within six months of diagnosis (n = 645). The feasibility of using this data to assess a number of quality indicators was explored, and where possible, variables were developed for analysis. Ultimately, detailed analyses were performed for the quality indicators: total blood transfusions of three units or greater, length of hospital stay, and use of non-nerve-sparing surgical technique.

**Results:** Even using high-quality chart data, it was not feasible to evaluate all of the quality indicators that were explored. For blood transfusions of three units or greater, length of hospital stay, and use of non-nerve-sparing surgical technique, worse outcomes were generally apparent with decreasing hospital volume, both before and after adjusting for the effect of explanatory variables.

**Conclusions:** We demonstrated convergent construct validity for three quality indicators (blood transfusions, length of hospital stay, and non-nerve-sparing surgery). If their validity is further demonstrated in future studies, these indicators could be used for quality assessment and could provide feedback to surgeons, surgical department heads, hospital administrators, and quality councils by suggesting areas for quality improvement in surgical care, such that future outcomes can be optimized.

## **STATEMENT OF CO-AUTHORSHIP**

This thesis is an embodiment of the research of Ellen O.M. Chan in collaboration with her thesis supervisors Dr. Patti A. Groome and Dr. D. Robert Siemens.

## ACKNOWLEDGEMENTS

I wish to acknowledge both my supervisors, Dr. Patti Groome and Dr. Rob Siemens, for their exceptional guidance throughout the development, analysis, and writing of my thesis. I am grateful to Dr. Groome for bringing me into this excellent graduate program and for her mentorship and support during my studies. Her expertise, advice, and editorial input were invaluable to the completion of this thesis. I would also like to thank Dr. Siemens for his ideas, clinical advice, and enthusiasm. I am grateful to have learned from them, as well as from all the other outstanding teachers in the department, the building blocks to pursue future endeavours in epidemiology and public health.

My appreciation is extended to parent study coordinator Sue Rohland for helping me understand the data and for ethics advice, Zhi Song for help with data access, Suzanna Keller, and Marguerite Laudanski. Thanks also to Dr. Paul Peng and Dr. Miu Lam for their biostatistics advice.

This research could not have been completed without funding provided by the Ontario Graduate Scholarship, Empire Life Fellowship, and the National Cancer Institute of Canada.

Finally I wish to acknowledge my family, friends, and classmates for their encouragement and support.

## TABLE OF CONTENTS

|   |            |
|---|------------|
| <b>ABSTRACT.....</b>  | <b>ii</b>  |
| <b>STATEMENT OF CO-AUTHORSHIP.....</b>                                      | <b>iv</b>  |
| <b>ACKNOWLEDGEMENTS .....</b>   | <b>v</b>   |
| <b>TABLE OF CONTENTS .....</b>  | <b>vi</b>  |
| <b>LIST OF TABLES .....</b>   | <b>x</b>   |
| <b>LIST OF FIGURES .....</b>  | <b>xi</b>  |
| <b>LIST OF ABBREVIATIONS .....</b>  | <b>xii</b> |
| <br>  |            |
| <b>CHAPTER 1: INTRODUCTION.....</b>   | <b>1</b>   |
| 1.1 Prostate cancer .....   | 1          |
| 1.2 Quality of care.....  | 2          |
| 1.3 Volume.....   | 3          |
| 1.4 Rationale, construct validity, and overview of study .....              | 4          |
| <br>  |            |
| <b>CHAPTER 2: LITERATURE REVIEW.....</b>                                    | <b>6</b>   |
| 2.1 Scope and search strategy .....   | 6          |
| 2.2 Prostate cancer .....   | 6          |
| 2.2.1 Epidemiology and natural history .....                                | 6          |
| 2.2.2 Diagnosis .....   | 8          |
| 2.2.3 Treatment options .....   | 8          |
| 2.2.4 Radical prostatectomy.....  | 9          |
| 2.3 The structure-process-outcome paradigm.....                             | 11         |
| 2.4 Volume.....   | 12         |
| 2.4.1 Introduction.....   | 12         |
| 2.4.2 Radical prostatectomy volume literature .....                         | 13         |
| 2.4.3 Explaining the volume effect .....                                    | 15         |
| 2.4.4 Evaluating process and outcome quality indicators against volume..... | 17         |
| 2.5 Quality indicators.....   | 18         |
| 2.5.1 Developing quality indicators for prostate cancer surgery .....       | 18         |
| 2.5.2 Follow-up studies on recommended quality indicators .....             | 20         |

|  |           |
|--|-----------|
| 2.6 Background on individual quality indicators .....                          | 23        |
| 2.6.1 Acute complications.....   | 23        |
| 2.6.2 30-day mortality.....  | 25        |
| 2.6.3 Blood loss and transfusion .....   | 25        |
| 2.6.4 Length of hospital stay.....   | 26        |
| 2.6.5 Nerve-sparing surgery.....   | 27        |
| 2.6.6 Urinary incontinence and erectile dysfunction .....                      | 28        |
| 2.6.7 Positive surgical margins and pathology reporting.....                   | 30        |
| 2.6.8 PSA levels after surgery.....  | 31        |
| 2.6.9 Biochemical disease-free survival .....                                  | 32        |
| 2.7 Testing the convergent construct validity of quality indicators .....      | 33        |
| <b>CHAPTER 3: METHODS .....</b>  | <b>34</b> |
| 3.1 Study purpose and objectives.....  | 34        |
| 3.2 Study design.....  | 34        |
| 3.3 Data collection overview .....   | 35        |
| 3.4 Study population .....   | 36        |
| 3.5 Data quality and validity.....   | 40        |
| 3.5.1 Medical chart abstraction.....   | 40        |
| 3.5.2 Electronic data .....  | 41        |
| 3.6 A priori study variables.....  | 42        |
| 3.7 Data processing and exploration.....                                       | 46        |
| 3.7.1 Diagnosis and surgery dates.....   | 46        |
| 3.8 Quality indicators.....  | 47        |
| 3.8.1 Acute complications.....   | 47        |
| 3.8.2 Length of hospital stay.....   | 49        |
| 3.8.3 Urinary incontinence and erectile dysfunction .....                      | 49        |
| 3.8.4 Urinary incontinence: Data processing and exploration .....              | 50        |
| 3.8.5 Erectile dysfunction: Data processing and exploration.....               | 51        |
| 3.8.6 Urinary incontinence and erectile dysfunction: Data quality issues ..... | 52        |
| 3.8.7 Nerve-sparing surgery.....   | 53        |
| 3.8.8 Positive surgical margins .....  | 53        |
| 3.8.9 Undetectable PSA levels at first follow-up.....                          | 54        |
| 3.8.10 Biochemical disease-free survival .....                                 | 55        |

|  |           |
|--|-----------|
| 3.9 Explanatory variables.....   | 56        |
| 3.9.1 Post-surgery hormone therapy .....                                 | 56        |
| 3.9.2 Additional non-prostatectomy initial treatment .....               | 57        |
| 3.9.3 Surgical approach .....  | 57        |
| 3.9.4 Age.....   | 58        |
| 3.9.5 Comorbidity .....  | 58        |
| 3.9.6 Year of diagnosis .....  | 59        |
| 3.9.7 Socioeconomic status.....  | 59        |
| 3.9.8 Disease severity: Pre-treatment PSA .....                          | 60        |
| 3.9.9 Disease severity: Clinical stage .....                             | 60        |
| 3.9.10 Disease severity: Biopsy Gleason score .....                      | 61        |
| 3.10 Hospital volume .....   | 62        |
| 3.10.1 Determining hospital volume.....                                  | 62        |
| 3.10.2 Hospital amalgamations.....                                       | 64        |
| 3.10.3 Categorizing hospital volume .....                                | 64        |
| 3.11 Summary: Final set of study variables.....                          | 65        |
| 3.12 A priori power calculations.....                                    | 67        |
| 3.13 Statistical analysis strategy .....                                 | 68        |
| 3.14 Ethical considerations .....  | 70        |
| <br>   |           |
| <b>CHAPTER 4: RESULTS .....</b>  | <b>71</b> |
| 4.1 Description of the study population .....                            | 71        |
| 4.2 Quality indicators by volume categories.....                         | 73        |
| 4.3 Explanatory variables by hospital volume .....                       | 77        |
| 4.4 Multivariate analysis .....  | 81        |
| 4.4.1 Total blood transfusions of three units or greater .....           | 81        |
| 4.4.2 Length of hospital stay.....                                       | 84        |
| 4.4.3 Use of non-nerve-sparing surgical technique.....                   | 86        |
| <br>   |           |
| <b>CHAPTER 5: DISCUSSION .....</b>                                       | <b>89</b> |
| 5.1 Summary of study design.....   | 89        |
| 5.2 Overview of findings .....   | 89        |
| 5.3 Feasibility of using chart data to evaluate quality indicators ..... | 90        |

|   |            |
|---|------------|
| 5.4 Volume-quality indicator associations.....                          | 93         |
| 5.4.1 Blood transfusions of 3 units or greater.....                     | 94         |
| 5.4.2 Length of hospital stay.....                                      | 95         |
| 5.4.3 Use of non-nerve-sparing surgical technique.....                  | 97         |
| 5.5 Effect of explanatory variables.....                                | 98         |
| 5.6 Strengths and limitations.....                                      | 100        |
| 5.6.1 Stratified sampling strategy.....                                 | 100        |
| 5.6.2 Other Strengths.....  | 101        |
| 5.7.2 Other Limitations.....  | 102        |
| 5.8 Summary of common themes and implications.....                      | 103        |
| 5.9 Recommendations and future research.....                            | 104        |
| <b>REFERENCES.....</b>  | <b>107</b> |
| <b>APPENDIX I: Medical chart abstraction form (selected pages).....</b> | <b>116</b> |
| <b>APPENDIX II: Data processing.....</b>                                | <b>121</b> |
| <b>APPENDIX III: Ethics approvals.....</b>                              | <b>125</b> |
| <b>APPENDIX IV: Supplementary results.....</b>                          | <b>128</b> |

## LIST OF TABLES

|   |     |
|---|-----|
| Table 3.1: List of quality indicators recommended by Cancer Care Ontario .....  | 44  |
| Table 3.2: <i>A priori</i> list of other quality indicators for possible investigation .....                              | 45  |
| Table 3.3: <i>A priori</i> list of explanatory variables to be investigated .....   | 45  |
| Table 3.4: Final set of study variables .....   | 65  |
| Table 3.5. <i>A priori</i> power calculations and lowest odds ratios required to achieve 80% statistical<br>power.....    | 67  |
| Table 4.1: Description of the study population.....   | 72  |
| Table 4.2: Quality indicators by hospital volume categories.....  | 75  |
| Table 4.3: Explanatory variables by hospital volume categories.....   | 79  |
| Table 4.4: Logistic regression analysis on total blood transfusions of three units or greater by<br>hospital volume ..... | 83  |
| Table 4.4: Mean length of hospital stay (in days) by hospital volume, stratified by era of<br>diagnosis.....              | 85  |
| Table 4.6: Logistic regression analysis on use of non-nerve-sparing surgical technique by<br>hospital volume .....        | 88  |
| Table II.1: Acute complications and blood transfusion rates.....  | 121 |
| Table II.2: Urinary incontinence rates .....  | 121 |
| Table II.3: Erectile dysfunction rates.....   | 122 |
| Table II.4: Classification of verbatim erectile status descriptions into erectile dysfunction status<br>.....             | 122 |
| Table II.5: Classification of pathology report comments into positive surgical margins categories<br>.....                | 122 |
| Table II.6: Event distribution for 5-year disease-free survival .....   | 123 |
| Table II.7: Hormone therapy rates .....   | 123 |
| Table II.8: Additional non-prostatectomy initial treatment rates.....   | 123 |
| Table II.9: Type of radical prostatectomy rates .....   | 124 |
| Table II.10: Distribution of study subjects by year of diagnosis .....  | 124 |
| Table II.11: Gleason score frequencies before and after assigning scores using level of<br>differentiation data .....     | 124 |

## LIST OF FIGURES

|   |     |
|---|-----|
| Figure 3.1: Conceptual framework .....  | 35  |
| Figure 3.2: Derivation of the study population .....  | 39  |
| Figure 4.1: Distribution of quality indicator outcomes by hospital volume categories (Number of radical prostatectomies performed per month in treating hospital):.....   | 76  |
| Figure IV.1: Change in Pearson chi-square diagnostic from deleting individual observations versus the estimated probability from the fitted parsimonious model for total blood transfusions of three units or greater .....     | 129 |
| Figure IV.2: Change in deviance diagnostic from deleting individual observations versus the estimated probability from the fitted parsimonious model for total blood transfusions of three units or greater.....                | 129 |
| Figure IV.3: Confidence interval displacement diagnostic from deleting individual observations versus the estimated probability from the fitted parsimonious model for total blood transfusions of three units or greater ..... | 130 |
| Figure IV.4: Change in Pearson chi-square diagnostic from deleting individual observations versus the estimated probability from the fitted parsimonious model for use of non-nerve-sparing surgical technique.....             | 130 |
| Figure IV.5: Change in deviance diagnostic from deleting individual observations versus the estimated probability from the fitted parsimonious model for use of non-nerve-sparing surgical technique.....                       | 131 |
| Figure IV.6: Confidence interval displacement diagnostic from deleting individual observations versus the estimated probability from the fitted parsimonious model for use of non-nerve-sparing surgical technique.....         | 131 |

## LIST OF ABBREVIATIONS

AJCC = American Joint Committee on Cancer  
CCE = Division of Cancer Care & Epidemiology  
CCO = Cancer Care Ontario  
CCP = Canadian Classification of Procedures  
CI = Confidence interval  
CIHI = Canadian Institute for Health Information  
CIRS = Cumulative Illness Rating Scale  
CQCO = Cancer Quality Council of Ontario  
CQSI = Cancer Quality System Index  
DAD = Discharge Abstract Database  
DRE = Digital rectal examination  
EORTC = European Organisation for Research and Treatment of Cancer  
ICD-9 = International Classification of Diseases, 9<sup>th</sup> edition  
LHIN = Local Health Integration Network  
MOH = Ontario Ministry of Health and Long-Term Care  
OCR = Ontario Cancer Registry  
OR = Odds ratio  
PSA = Prostate-specific antigen  
RCC = Regional Cancer Center  
RP = Radical prostatectomy  
RPP = Radical perineal prostatectomy  
RRP = Radical retropubic prostatectomy  
SAS = Statistical Analysis Software  
SEER = Surveillance, Epidemiology, and End Results  
SES = Socioeconomic status  
TMN = Tumor, node, and metastases (staging system)  
TURP = Transurethral resection of the prostate

## CHAPTER 1: INTRODUCTION

### 1.1 Prostate cancer

Prostate cancer is the most commonly diagnosed cancer and the third most common cause of cancer death among Canadian men, and is a serious public health problem that causes a significant amount of morbidity in the population.<sup>1,2</sup> Treatment for prostate cancer often cures patients with clinically localized disease, defined as tumor confined within the prostate with no evidence of regional or distant metastasis.<sup>3,4</sup> For these patients, the three primary treatment strategies are surgery, radiotherapy, and watchful waiting.<sup>3,5</sup> The typical surgical treatment is radical prostatectomy (RP), a complex and challenging procedure which involves the total removal of the prostate and surrounding tissues. As the incidence of localized disease has increased dramatically with the advent of PSA testing, the number of radical prostatectomies performed has increased.<sup>3,6,7</sup> While this surgery can completely eradicate the cancer, there are concerns about the considerable post-operative and long-term morbidities that can significantly affect a patient's quality of life.<sup>3,6</sup>

Controversy over many aspects of surgical treatment has led to variations in practice and outcomes.<sup>8</sup> There is concern that variations in outcome may be due to variations in the quality of care provided, although there is little known even about which aspects of quality should be examined.<sup>9</sup>

## 1.2 Quality of care

Quality of care research evaluates the “degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge.”<sup>10</sup> In health sciences, research in quality assessment is based on the Donabedian conceptual framework, which categorizes quality of care measures into three domains – structure, process, and outcome.<sup>11,12</sup> Concerns over shortcomings in the quality of medical care over the past few decades in light of rising health care expenditures has raised serious concerns among both patients and payers about the value of their health care dollars. These concerns have led to consumer and payer driven efforts to better understand, measure, and influence medical care quality. As a result, independent councils with overall responsibility to monitor the quality of health care systems and to guide change have emerged in the U.K., U.S., and Australia.<sup>11,13</sup>

In Ontario, the provincial government created the Cancer Quality Council of Ontario (CQCO) in 2002, a body unique in Canada for its focus on quality within a specified disease context as opposed to assessing quality across the entire health care system. This impartial group provides advice to the province through Cancer Care Ontario (CCO), and is committed to improving the quality of all cancer services by enhancing the measurement of system performance, and using this information to motivate and coordinate quality improvement.<sup>13</sup> However, the work of this council, as well as other groups internationally, has been impeded by a lack of reliable and valid data for measuring quality and a failure to control for patient characteristics.<sup>11,13</sup>

In prostate cancer, both the RAND group in California and Cancer Care Ontario (CCO) have recommended quality indicators for surgical care provided to patients with clinically localized disease, with some overlap. Using an expert panel, recommendations were made based on the literature and/or consensus from the clinicians.<sup>9,14</sup> However, the lack of high quality evidence on many aspects of the surgical treatment (radical prostatectomy) means that it is important to assess whether the recommendations are valid in distinguishing between good and poor quality surgical care before their application in the quality-of-care process.<sup>10,11</sup> Follow-up studies to the RAND recommendations have tested the feasibility of some but not all of the quality indicators listed by RAND and CCO. A formal evaluation of validity has not yet been performed, but is necessary.<sup>9,15-17</sup>

### **1.3 Volume**

Volume refers to the number of procedures performed at individual hospitals or by individual surgeons.<sup>18</sup> The relationship between higher volume and lower rates of adverse outcomes following treatment is well established,<sup>19,20</sup> with the common explanation being that high volume surgeons generally improve their surgical techniques with experience.<sup>21</sup> In radical prostatectomy, studies using electronic claims data from the U.S. have found better short term outcomes with increasing hospital and surgeon volume. However, these American studies were limited by the data source, which did not allow for examination of all RAND and CCO listed quality indicators, and by incomplete control of covariates.<sup>18,20,22-25</sup> Also, there is a gap in knowledge due to the absence of radical prostatectomy volume studies performed in Canada.

#### **1.4 Rationale, construct validity, and overview of study**

This study attempted to address the gaps in the literature as highlighted above. To address the concern that variations in radical prostatectomy outcomes may be due to variations in the quality of care provided, RAND and CCO expert panels developed recommended quality indicators for prostate cancer surgery.<sup>9,14</sup> We continued this line of inquiry by examining whether radical prostatectomy quality indicators can be assessed using patient charts and whether certain indicators are valid in distinguishing quality in an Ontario population, taking into account the effect of important patient and disease characteristics that might affect comparisons. Although the definition of quality is quite broad and can include many aspects of medical care including prevention and continuity of care, we limited the scope of our study to the examination of technical surgical quality indicators.<sup>12</sup> Since it is well established that higher volume hospitals generally have better post-surgical outcomes for cancer surgeries in general as well as for radical prostatectomy,<sup>18,20,23,24</sup> we used the hospital prostatectomy volume, as a proven surrogate for surgeon volume,<sup>23</sup> to test the construct validity of the candidate quality indicators. At the same time, this allowed us to better understand the volume effect.

Construct validity refers to the validity of measures of unobservable constructs, and is used when there is no criterion against which to evaluate the validity of these measurements.<sup>26,27</sup> This is true for quality indicators, instruments proposed for measuring the unobservable construct of surgical quality, but for which there is no reference standard. In this study, we used a specific type of construct validity called convergent validity, which tests the extent to which two or more instruments that purport to be measuring the same topic agree with each other.<sup>26</sup> Our hypothesis is that if higher

volume hospitals have better outcomes as demonstrated by the quality indicators of interest, then we will have demonstrated their convergent construct validity: that hospital volume and the quality indicator are both measuring the underlying concept of surgical quality. Indicators that we found feasible to collect and to be valid could be used by the Ontario council and others to monitor the quality of prostate cancer surgical care.<sup>13</sup>

This was a retrospective cohort study that used chart-based data collected previously for an observational study designed to examine the role of comorbidity on early death after prostate cancer treatment, to compare the outcomes of radiotherapy and radical prostatectomy treatments, and to explore within-modality practice variations and their impact. The current study partially addresses this last objective. Dr. Groome was the Principal Investigator on the parent study and Dr. Siemens was a clinical co-investigator. The study population was drawn randomly from a study base consisting of all prostate cancer patients treated for cure in Ontario from 1990 through 1998 within six months of diagnosis.

Specifically, our study purpose was to test the construct validity of radical prostatectomy quality indicators by assessing their associations with hospital volume among prostate cancer patients treated with radical prostatectomy in Ontario. The objectives were to assess variations in quality indicators by hospital volume, and to investigate whether selected explanatory variables account for some of this variation.

## **CHAPTER 2: LITERATURE REVIEW**

### **2.1 Scope and search strategy**

This literature review covers four major areas. Firstly, an overview of prostate cancer, with a focus on surgical treatment of the disease, is provided based on clinical practice guidelines, surgical texts, and review articles from Medline. The chapter then reviews the concept of volume in relation to prostate cancer surgery, all quality indicator research in prostate cancer surgery, background on individual quality indicators, and the extent that each has been evaluated or used as a quality indicator. Peer-reviewed literature for these topics was obtained through searches of the Medline database from 1996 to June 2007. Articles preceding this time frame were reviewed if they were cited as important references in literature found through the Medline search.

### **2.2 Prostate cancer**

#### **2.2.1 Epidemiology and natural history**

Prostate cancer is the most commonly diagnosed cancer and the third most common cause of cancer death among Canadian men, according to the Canadian Cancer Society. In 2006, it was estimated that 20 700 men were diagnosed and 4 200 men in Canada died of this disease.<sup>1</sup> In addition, men survive longer with prostate cancer than people affected by other common cancers such as lung carcinoma. Therefore, prostate cancer is a serious public health problem that causes a significant amount of morbidity in the population.<sup>1,2</sup>

The most definitive risk factors for prostate cancer are age, family history, and ethnic group. The risk of developing this disease increases faster with age than any other major cancer, with both incidence and mortality rates increasing at a near exponential rate after age 50. Other probable risk factors include high intake of fat, meat, and dairy products, while protective factors include the consumption of soybean products, tomatoes, and selenium.<sup>2,28</sup> The presence of diabetes mellitus may be another probable risk factor, as those suffering from diabetes may be more likely to be obese and have hyperinsulinemia, and both may alter endogenous steroid metabolism.<sup>28</sup>

The natural history of prostate cancer progresses slowly over many years, with doubling times for local tumors estimated at two to four years. Early prostate cancer is often asymptomatic, since tumors usually arise from the peripheral zone of the prostate, which is distant from the urethra. The presence of voiding symptoms suggests locally advanced disease that has spread to the seminal vesicles, ureters, and/or the bladder base. Prostate cancer most commonly metastasizes to the lymph nodes and bones, but has also been found to metastasize to the lungs, bladder, liver, adrenal glands, and testes. Symptoms associated with these metastases include perineal pain, weight loss, cachexia, bone pain, and neurological complications.<sup>2,29</sup>

### **2.2.2 Diagnosis**

Prostate biopsies confirming the presence of cancer are typically recommended when abnormalities are found on digital rectal examination (DRE) or with serum prostate-specific antigen (PSA) levels above age-specific ranges.<sup>2,5</sup> The Gleason system is the most common histologic grading system for prostate cancer, and is based on the glandular pattern of the biopsy specimen at low magnification. Assigned grades can range from 2 to 10, with prognosis worsening with the numerical score.<sup>2</sup> The tumor, node, and metastases (TMN) staging system is used to describe the extent of disease as determined through DRE, PSA testing, biopsy, and imaging modalities.<sup>2,3</sup> The combination of clinical stage, PSA, and Gleason score provide an understanding of the severity of the disease to evaluate prognosis and to direct therapy. Preoperative nomograms based on these three measures of disease severity have been demonstrated to be prognostic of outcomes such as prostate cancer-specific mortality and biochemical failure after curative surgery and radiotherapy.<sup>3,30-33</sup>

### **2.2.3 Treatment options**

Treatment for prostate cancer often cures patients with clinically localized disease, but it is rarely curative for patients with locally extensive tumor, and is generally not curable for metastatic disease.<sup>3</sup> Clinically localized cancer is defined as tumor confined within the prostate with no evidence of regional or distant metastasis. It includes tumors that are not palpable by DRE (TNM category T1 N0 M0) or are palpable but do not extend outside the prostate (TNM category T2 N0 M0).<sup>4</sup> For these patients,

the three primary treatment strategies are radical prostatectomy, definitive radiation therapy, and watchful waiting.<sup>3,5</sup> However, due to the lack of high quality randomized studies directly comparing these options that have accounted for the effects of treatment selection factors, there is great diversity of opinion regarding the optimal management technique. Therefore, physicians currently make treatment recommendations based on disease severity, age, comorbidity, and patient preference.<sup>2-4</sup> In this thesis, we are only concerned with surgery for prostate cancer and thus, there will be no further discussion of the other treatment options or on the optimal choice of therapy.

#### **2.2.4 Radical prostatectomy**

As the incidence of localized disease has increased dramatically with the advent of PSA testing, the rates of surgeries performed for prostate cancer have also increased. The typical surgical treatment is radical prostatectomy (RP), which involves the total removal of the prostate and surrounding tissues.<sup>3,6</sup> It is a complex and challenging surgical procedure that is also one of the most common operations performed by urologists in the Western world.<sup>7</sup>

Radical prostatectomy is generally appropriate for younger men with tumors that are confined to the prostate (TNM categories T1 or T2) and without lymph node metastases. They should also have no major comorbid medical illnesses that would preclude major pelvic surgery and adversely affect post-surgical outcomes. Patients with more advanced disease are unlikely to be cured, and the risks and side effects of the procedure cannot usually be justified for them.<sup>2,3,6,7</sup> While radical prostatectomy can

completely eradicate the cancer, there are concerns of considerable post-operative and long-term morbidities than can significantly affect quality of life.<sup>3,6</sup>

During the surgical procedure, the first step is to resect the regional pelvic lymph nodes to determine the presence of metastasis. Some surgeons may elect to forgo the dissection altogether for early stage, low grade cancers where the probability of lymph node metastasis is very low in order to reduce complications and operating time. Following the lymph node dissection, radical prostatectomy is completed. While the retropubic (RRP) and perineal (RPP) approaches were the most common during the 1990s, the retropubic approach gained more acceptance due to urologists' greater familiarity with pelvic anatomy and the ability to perform lymph node dissection and radical prostatectomy in one session through a single incision.<sup>3,5,6,34</sup>

In recent years, improved understanding of the anatomy of the nerves, blood vessels, and structures surrounding the prostate has led to the development of surgical techniques aimed at preserving continence and erectile function, the two most common complications of treatment, as well as reducing blood loss and generally improving post-surgical outcomes.<sup>2,6,7,35</sup> Nerve-sparing techniques to preserve the neurovascular bundles that innervate the corpora cavernosa of the penis have been shown to improve post-operative erectile function, although this has not yet been confirmed in a randomized study.<sup>7,29,35,36</sup> Also, techniques involving careful apical dissection in the preparation of the vesicourethral anastomosis have resulted in reduced rates and severity of urinary incontinence.<sup>6,35</sup> Finally, careful attention to control of the dorsal venous complex and meticulous dissection and control of the many small vessels surrounding the seminal vesicles have reduced intraoperative blood loss and created a relatively bloodless field

where it is also easier to identify landmarks, thereby reducing the sequelae of the treatment.<sup>2,6,35,37</sup>

Immediately following the complete removal of the prostate, initial PSA levels should be undetectable, indicating that all cancer and normal prostate cells have been removed. PSA levels that remain detectable, or pathological findings such as positive surgical margins, may reflect persistent disease that was not entirely removed in surgery. These findings can suggest a high risk of recurrence, and thus adjuvant radiotherapy or hormonal therapy post-surgery may be considered for these patients.<sup>3,5,6</sup>

Otherwise, the general follow-up pattern is to see the surgeon at three-month intervals in the first year after surgery. Afterwards, the follow-up interval can be increased to six months and may be carried out by the patient's general practitioner.<sup>5,6</sup> During each follow-up appointment, DRE and PSA testing is recommended to detect recurrent disease such that salvage therapy may be administered early to cure or to prolong life.<sup>5</sup>

### **2.3 The structure-process-outcome paradigm**

In health sciences, the foundation for much emerging research in quality assessment is based on the Donabedian structure-process-outcome paradigm, which categorizes quality of care measures into three domains.<sup>11,12</sup> Firstly, structure of care includes the equipment, resources, and provider experience necessary to provide care, such as surgical volume. Secondly, process of care refers to the technical and interpersonal elements of care that transpire between doctor and patient. For example, the use of nerve-sparing surgery may be considered a technical process of care. Thirdly,

outcomes of care refer to complications, survival rates, and patient-centered outcomes, such as the health-related quality of life outcomes of erectile dysfunction and urinary incontinence.<sup>9,12</sup> The quality indicators recommended by the CCO panel are all process and outcome indicators. The next topic of review, volume, is a well-validated structural indicator of surgeon experience that will be used in this thesis to assess the validity of the recommended process and outcome indicators.

## **2.4 Volume**

### **2.4.1 Introduction**

Volume refers to the number of procedures performed at individual hospitals (hospital or institutional volume) or by individual surgeons (surgeon or provider volume).<sup>18</sup> In 2004, CIHI commissioned the Institute for Clinical Evaluate Sciences in Toronto to conduct a systematic review of studies on the volume-outcome relationship.<sup>19</sup> The literature showed that the relationship between higher volume and lower rates of adverse outcomes following treatment is well established, although the magnitude of the volume effect varies by procedure.<sup>18-20</sup> Strong and consistent trends have been noted for pancreatectomy and esophagectomy, while smaller trends have been noted for other procedures such as primary surgery for colon, breast, and lung cancer.<sup>18</sup> Research has shown that the size of the volume effect seems to depend on the level of surgical skill required to perform a particular procedure.<sup>20,38</sup> Accordingly, the common explanation for the volume phenomenon is that high volume surgeons generally improve their surgical techniques with experience.<sup>21</sup>

Based on observational evidence that patients treated at facilities performing fewer radical prostatectomies reported more surgical complications than higher volume institutions, RAND recommended volume as an important quality indicator. Panelists rated it as both a valid and feasible structural indicator, although they did not specify whether hospital or surgeon volume should be evaluated.<sup>9,10</sup> Most studies examining adverse outcomes following cancer surgery have typically focused on hospital volume, as a proxy for surgeon volume, although high hospital volume could very well be spread among a number of low volume surgeons.<sup>18</sup> However, it has been shown that while surgeon volume rather than hospital volume is the most important predictor of short-term radical prostatectomy outcomes, hospital volume can still provide a good approximation of surgeon volume if surgeon-specific data are unavailable.<sup>23</sup> A review article on urological volume literature found that improvements in acute surgical complication rates following radical prostatectomy were more significant for high surgeon volume as opposed to high hospital volume. The authors concluded that while both are independent predictors of outcome, models that include both surgeon and hospital volume suggest that high volume hospitals have better outcomes in part because of the effect of surgeon volume, and vice versa.<sup>20</sup>

#### **2.4.2 Radical prostatectomy volume literature**

Most of the radical prostatectomy volume literature has examined short term outcomes by hospital and surgeon volume using electronic Surveillance, Epidemiology, and End Results (SEER) cancer registries and Medicare claims data from the U.S.<sup>18,23-25</sup> Adverse outcomes examined include in-hospital or post-operative complications, length

of stay, in-hospital, 30-day, and 60-day mortality, and late urinary complication and long-term incontinence symptoms or procedures. These outcomes have shown better rates with increasing hospital and surgeon volume, with the associations with surgeon volume more likely to be significant.<sup>18,20,22-25</sup>

In one study with the Genito-Urinary Group of the European Organisation for Research and Treatment of Cancer (EORTC) by Van Poppel *et al.* in 2001, other quality indicators were examined in relation to surgeon volume. These included blood transfusion rate, positive surgical margins, undetectable PSA, and incontinence (as detected through a 24-hour pad test) at three months after RRP. Twenty-three surgeons in the Genito-Urinary group agreed to participate by completing questionnaires about the post-surgical outcomes of ten consecutive patients with T1c or T2 stage prostate cancer. Of these four indicators examined, only blood transfusions demonstrated an association with volume. It was found that low volume surgeons tended to give transfusions of  $\geq 3$  units more frequently than higher volume surgeons.<sup>39,40</sup>

The evidence for blood transfusion is further strengthened by another study, which found that patients treated by surgeons performing  $\leq 15$  RRP annually have an 8.6 increased odds of requiring blood transfusion when compared to surgeons with more experience. A volume association was also found with blood loss.<sup>41</sup> For positive surgical margins, the association with volume has been demonstrated in other studies. In two large cohorts, one based in two large urban centers in the U.S., and a subsequent study based on a single European institution, positive surgical margin rates were found to be significantly associated with surgeon volume, even after controlling for case mix.<sup>42,43</sup> On

the other hand, no studies have examined the effect of volume on nerve-sparing surgery, erectile dysfunction, and biochemical disease-free survival.

In addition to this gap in knowledge, there are other issues with the quality indicator-volume associations reported in previous radical prostatectomy literature. The electronic data studies used hospital discharge abstracts, SEER, and Medicare claims, which cannot provide detailed clinical information.<sup>18,21,23-25</sup> Also, they did not control completely for confounders, especially disease severity.<sup>18,20,21,23-25</sup> With the exception of one study performed in Quebec where the examination of volume was not a main objective,<sup>44</sup> all of the studies were performed in the U.S.<sup>18,21,23,25</sup> The EORTC study has a number of problems. No background was given to justify the selection of quality indicators that were examined, and there was no adjustment for any confounders, although variations in covariates were noted.<sup>39,40</sup> There was also selection bias, since surgeons could choose which patients they reported, making the group of study subjects unrepresentative of the population experience. The positive surgical margins studies were tested in a small number of large institutions rather than on a larger geographic scale, and again were not performed in Canada.<sup>42,43</sup> Finally, it is difficult to compare results across all of these studies, because the cutoff points for volume categorizations varied.<sup>20</sup>

### **2.4.3 Explaining the volume effect**

Several underlying explanations have been proposed to explain the association between volume and adverse outcomes in literature for surgery in general as well as for radical prostatectomy specifically. These include the level of surgeon skill, confounding

by disease severity and comorbidity, selective referral, and better supportive care.

Extensive research has shown that the size of the volume effect depends on the level of surgical skill required to perform a particular procedure.<sup>20,38</sup> Since radical prostatectomy is a complex and challenging surgery, high volume surgeons likely have better results because ‘practice makes perfect’.<sup>2,7,21,45</sup> Confounding by variations in patient and disease characteristics, including comorbidity, have also been proposed to explain the volume effect. However, it is unlikely that their adjustment could fully explain volume-outcome relationships.<sup>18,20,42,46</sup> Selective referral is another proposed explanation, as primary health care physicians tend to refer more patients to centers of excellence.<sup>20,45</sup> Finally, a 2002 study suggests that the volume-outcome effect is not specific to the volume of the specific procedure performed, but is rather related to the overall volume of complex surgeries performed at a hospital. Those authors suggested that the outcomes are related to better supportive care at tertiary care institutions, such as better equipment, fully staffed intensive care units, and a broader range of specialists and other resources, rather than to surgical skill.<sup>21,47</sup> Variations in the characteristics of patients sent to different hospitals due to selective referral may be largely accounted for with adjustment for confounders and adjusted associations between volume and adverse outcomes for radical prostatectomy are due to both variations in surgical skill and quality of supportive care. While it is not possible to separate the influence of these two explanations in the current study, together they demonstrate that volume describes variations in overall quality of care.

#### **2.4.4 Evaluating process and outcome quality indicators against volume**

In general, the strength of the literature has led experts to conclude that surgeon experience, or volume, an important indicator of good quality care.<sup>9,11,45</sup> As reviewed above, volume has been found to be associated with some processes of care and outcomes of radical prostatectomy. In addition, it is easily determined from readily available electronic data sources, and all stakeholder groups can understand the concept of volume. However, volume is an insufficient measure to pinpoint quality issues more specifically, as it only works well as an indicator on average. This is useful in large scale studies, but provider volume does not work well in predicting the quality of individual hospitals and surgeons.<sup>48</sup> For example, substantial variation in short-term outcomes and positive surgical margin rates have been found among high volume surgeons even after adjustment for hospital volume and other covariates.<sup>18,42,43</sup> It also does not give us a path forward beyond centralizing surgery, which can cause travel difficulties for patients in Canada given our geographical size. Therefore, it is useful to better understand the volume effect, and it is necessary to develop more proximate measures of quality to understand the variations in processes and outcomes of care that underlie this volume phenomenon.<sup>11</sup> The work of our study in validating these more proximate quality indicators will provide the knowledge necessary to eventually allow for problems in surgical quality to be identified and targeted for improvement.

## **2.5 Quality indicators**

### **2.5.1 Developing quality indicators for prostate cancer surgery**

In prostate cancer, controversy over many aspects of management has led to variations in practice and outcomes.<sup>8</sup> There is concern that variations in outcome may be due to variations in the quality of care provided, although there is little known even about which aspects of quality should be examined. In 2000, the RAND group of Santa Monica, California made the first effort to identify the components of high-quality surgical and radiotherapy care for men with early-stage prostate cancer.<sup>9</sup> This effort was followed by a similar effort undertaken by a Cancer Care Ontario (CCO) panel in 2005 that built on the RAND initiative in identifying quality indicators for prostate cancer surgery as part of a provincial oncology performance assessment program in Ontario.<sup>14</sup>

To identify processes and outcomes that reflect quality of care, both groups identified potential quality indicators using a formal consensus procedure called the modified Delphi method. Firstly, a list of indicators and covariates were identified through a literature search and through expert interviews. Next, nominated expert panels consisting of surgeons and other health care providers developed the final list of quality indicators through several rounds of rating and feedback. Finally, an in-person discussion involving all members of the panel led to the final list of quality indicators, which were chosen based on published evidence or expert consensus. Whereas the Ontario group was asked to create an “ideal” set of indicators regardless of perceived availability of administrative data by which they could be measured, the RAND indicators were selected based on rankings of their validity and feasibility by panel members. In the RAND exercise, indicators were ranked on the feasibility of obtaining

the relevant data from medical records. Panel members ranked indicators on their validity based on scientific evidence, professional consensus, or professional experience that an indicator distinguishes quality. The overlap in the final sets of indicators identified by the two groups included acute surgical complications, biochemical disease-free and overall survival at 5, 10, and 15 years after treatment, length of stay, loss of erectile function, and urinary continence.<sup>9,10,14</sup>

The RAND group also noted that if the indicators are used to assess quality, covariates such as pre-treatment PSA level, Gleason score, clinical stage, age, comorbidity, and socioeconomic status measures should be adjusted for in order to control for differences in the providers or institutions being compared. For example, patients with higher PSA and Gleason scores are more likely to have disease that has extended beyond the prostate and are therefore more likely to have biochemical failure or shorter survival.<sup>9</sup>

The advantages of the RAND and CCO studies are that the modified Delphi method is a validated and accepted avenue for developing quality indicator recommendations.<sup>10</sup> These evidence and consensus based quality indicators can be applied in any jurisdiction, as long as the data required are available.<sup>14</sup> However, the lack of high quality evidence, such as randomized controlled trials and large-scale epidemiologic studies, for many aspects of treatment for localized prostate cancer does undermine the validity of evidence-based recommendations despite their endorsement by the expert consensus panel.<sup>10,11</sup> The two groups reached different conclusions on appropriate quality indicators, which may be due to temporal trends in the importance of clinical topics, as well as variations in knowledge, attitudes, and beliefs across

professional groups and geographic regions. The CCO group did note that they built upon the work of RAND initiative, a much bigger effort, rather than starting from the beginning.<sup>14</sup> In the context of the current project, we considered the more recent CCO recommendations as our best source for identifying quality indicators for study, since they were developed by experts based in Ontario and we are concerned with the development of quality indicators appropriate for a population of Ontario patients.

### **2.5.2 Follow-up studies on recommended quality indicators**

While no other studies have been found in the literature to assess the feasibility and validity of the CCO recommended quality indicators, two surgically-relevant indicators are reported on the CCO website as part of the CQCO's Cancer System Quality Index 2007.<sup>49</sup> For radical prostatectomy, deaths within 30 days of surgery and completeness of pathology reporting is reported for the province as a whole and by Local Health Integration Network (LHIN). The other quality indicators reported for prostate cancer include 5-year relative survival, cancer surgery waits, and cancer deaths, none of which are specific to prostate cancer surgery and therefore, are not necessarily reflective of the quality of surgical technique. On the other hand, RAND rated their proposed quality indicators by feasibility and validity based on the opinion of the panel members, but did not formally evaluate them in an observational study.<sup>10</sup> However, several follow-up studies have assessed their feasibility.<sup>15-17</sup> Since there is overlap between the two sets of quality indicators, all studies completed to date based on the RAND surgical quality indicator recommendations will be described.

In 2003, Miller *et al.*<sup>15</sup> reviewed care provided to 168 patients diagnosed with localized prostate cancer either in 1995 or in 2000 at the University of Michigan through a combination of electronic database review and explicit paper chart review. They found that it was feasible to measure 19 of the 22 indicators assessed, and also observed that compliance on several of the 19 feasible indicators improved in 2000 compared to in 1995.

Subsequently, in 2005 Krupski *et al.*<sup>16</sup> evaluated the feasibility of measuring the subset of 16 quality indicators recommended by RAND that were selected based on expert opinion. A convenience sample of 55 men who received radical prostatectomy and 29 who received radiotherapy from various institutions was drawn from a state-funded program providing free prostate cancer treatment to uninsured California residents near or below the poverty line with untreated, early stage prostate cancer. Quality of care measurement was found to be feasible in 13 of the indicators examined using chart abstraction, administrative documents, and patient questionnaires, and some differences in compliance were noted by treatment type.

In 2007, the findings of these two studies were used by Miller *et al.* to select a subset of the RAND candidate indicators for the first nationwide assessment of the quality of localized prostate cancer care in the U.S.<sup>17</sup> A stratified random sample of prostate cancer diagnoses made during 2000-2001 was taken from the National Cancer Data Base and chart abstractions were performed for 1390 patients who received radical prostatectomy and 1385 who received radiotherapy as initial treatment in hundreds of hospitals throughout the country. Sampling weights were applied to obtain national estimates of quality indicator compliance, which ranged from 40-100%, although most

indicators had compliance rates approaching 100%. The study focused on comparisons between surgery and radiotherapy, and found that compliance to the recommended quality indicators was higher for men treated with radiotherapy.

The studies described above highlight three important points. First, they compared quality indicator compliance between treatment types rather than focusing on surgery-specific quality indicators. As a result, some of the CCO recommended quality indicators that are surgery-focused were not addressed in any of the three studies. Second, while the studies have found that quality indicator compliance changed over time and varied by treatment type, there has been no formal evaluation of the validity of the indicators. Third, they all suggested that consideration of covariates is necessary in quality assessment and improvement initiatives, but the adjustment for covariates in their studies was not done or was incomplete. Krupski *et al.*<sup>16</sup> described patient and disease characteristics but did not adjust for the differences between the surgery and radiotherapy groups when comparing the quality indicator compliance between them. In the Miller *et al.* studies,<sup>15,17</sup> adjustment for several covariates was considered, but comorbidity was not assessed.

One of the goals of CQCO is to develop indicators to measure the quality of cancer surgery throughout the province for all disease sites, such as those that measure clinical health and quality of life, and to subsequently monitor these quality indicators on a regular basis to address variations in the quality of care.<sup>13</sup> While feasibility and validity on some of these indicators has been evaluated in the U.S., it is not specific to the Canadian experience and also does not focus on practice variations among patients treated with radical prostatectomy. Of the indicators for prostate cancer care currently

reported by CQCO, only the completeness of pathology reporting was recommended by the CCO prostate cancer expert panel, and this relates to the quality of reporting by pathologists rather than the quality of care provided by surgeons. Before the CCO recommended quality indicators for prostate cancer surgery can be put into use, their feasibility and validity must be evaluated for use in an Ontario population, and the impact of covariates should be understood.

## **2.6 Background on individual quality indicators**

A subset of the process and outcome quality indicators recommended by Cancer Care Ontario in the Gagliardi *et al.* paper<sup>14</sup> was selected for investigation in this thesis, based on the data available for analysis and if the quality indicator is influenced by surgeon skill. The scope of this section, which will provide a brief background on each recommendation and a review of their use as quality indicators for prostate cancer surgical care in the literature, will be limited to this subset of CCO indicators as well as several additional process and outcome indicators that we considered important to surgical quality.

### **2.6.1 Acute complications**

Cancer Care Ontario recommended the occurrence of acute surgical complication (blood loss of 2.0 L or greater, cardiovascular complications, deep vein thrombosis, pulmonary embolism, infection, and long term anticoagulant therapy) as an outcome quality indicator, based on RAND expert panel consensus and evidence from

observational studies. Additionally, death (no time frame specified) was another acute surgical complication suggested by RAND.<sup>10,14</sup> Post-operative complications are commonly reported in the medical literature and are considered to be directly attributable to the treatment.<sup>10</sup> Decreases in these complication rates have been reported from academic medical centers with increasing surgeon experience along with the use of enhanced surgical techniques first disseminated in the 1990s.<sup>50</sup> With the exception of blood loss which will be discussed separately, the rates of these acute complications after radical prostatectomy are low, usually between 0-2%,<sup>6,35,51</sup> and rates increase with age and comorbidity.<sup>52</sup> Miller *et al* found that acute complications (including 30-day mortality) could be feasibly assessed through detailed chart review.<sup>15</sup> In an Australian study by Ansari *et al.*, acute adverse events including urinary complications, haemorrhage or haematoma complicating a procedure, postoperative infection, and cardiovascular complications were examined as an indicator of poor quality of care.<sup>51</sup> Rates were actually found to increase over time (1989-1995), but the study population was restricted only to patients with no proven cancer and 99% of the patients received transurethral resection of the prostate (TURP) rather than open prostatectomy (i.e. retropubic and suprapubic approaches). With such low rates, the authors suggested that this was not likely to reflect a decline in the quality of care, but more likely to do with the accuracy of the routine electronic database used for analysis and changes in thresholds for coding diagnoses.

### **2.6.2 30-day mortality**

Operative mortality can also be defined as death within 30 days of surgery.<sup>35</sup> Although not recommended by Gagliardi *et al.*,<sup>14</sup> Cancer Care Ontario states that 30-day mortality is an important measurement of quality of care in complex operations such as radical prostatectomy,<sup>49</sup> since the risk of operative death is related to surgical skill as well as perioperative care and case mix.<sup>35,49,53</sup> The risk of death in hospital or within 30 days after surgery for all patients receiving prostatectomy in Ontario in 2002-2004 was reported by CCO as a quality indicator in the CQCO's Cancer Service Quality Index 2007. The overall rate was 0.23%, with rates by LHIN ranging from 0-0.95%.<sup>49</sup> This is consistent with operative mortality rates found in U.S. and other Canadian studies on national, provincial, or multi-state levels for surgeries performed in the 1990s, which ranged from 0.25-0.6%.<sup>18,24,25,44,52</sup> With such low rates, it is questionable whether this outcome indicator can really identify LHINs offering poor quality care, and the small variations observed may be due to confounding by age and comorbidity.<sup>49,52</sup>

### **2.6.3 Blood loss and transfusion**

Blood loss has been endorsed by RAND as a rough proxy for the surgeon's skills, although they note that it is important to control for covariates such as stage, which affect the technical difficulty of the procedure.<sup>9</sup> Intraoperative blood transfusion has been inversely associated with biochemical progression-free survival.<sup>34</sup> While there are problems in estimating blood loss due to inter-observer variability and inter-individual variance in the criteria for transfusion,<sup>9</sup> significant declines in blood loss and transfusions have been reported as surgeon experience and technique improved over the 1990s with

better understanding of prostatic blood vessel anatomy and as experience with the surgery has increased.<sup>2,6,34,35</sup> A review of the literature found that estimated blood loss during RRP ranged from 0.4 L to 1.0 L in the 1990s, and blood transfusion rates declined from 68-89% in the late 1980s to 1-3.5% in the late 1990s.<sup>34</sup> As a quality indicator, both Miller *et al.*<sup>15</sup> and Krupski *et al.*<sup>16</sup> found that operative blood loss data could be feasibly obtained from paper and electronic medical records as well as clinical databases. However, the use of blood transfusions as a process quality indicator has not been assessed in the RAND follow-up studies,<sup>15-17</sup> since it was not recommended by the RAND group.<sup>10</sup>

#### **2.6.4 Length of hospital stay**

The length of hospital stay may be reflective of surgical complications, for patients who have no problems should be discharged sooner than those who experience more complications. According to Klein,<sup>54</sup> the most important antecedents for successfully reducing length of hospital stay after RRP occur during surgery, where reduced blood loss and shortened operative time limit the need for anesthetic agents, opioids, and intravenous fluid, resulting in less sedation, less third-space fluid loss, and earlier return of bowel function. In a 5% random sample of Medicare beneficiaries who underwent radical prostatectomy from 1991-1998 in the U.S., the mean length of stay decreased from 8.1 to 5.1 days.<sup>50</sup> This might reflect improved quality of surgery over the years. It could also reflect the general push to decrease length of hospital stays over time,<sup>51,54</sup> which could make temporal comparisons problematic. In the Australian study by Ansari *et al.*, mean length of stay for TURP was also found to decrease from 1989 to

1995, and length of stay was significantly associated with acute complications, which is known to be related to surgical skill.<sup>51</sup> The feasibility of this outcome quality indicator has not been assessed by CCO<sup>49</sup> or by any of the RAND follow-up studies.<sup>15-17</sup>

### **2.6.5 Nerve-sparing surgery**

The nerve-sparing or anatomic radical prostatectomy technique introduced by Walsh and Donker involves the removal of the prostate with controlled hemostasis to allow visualization of the neurovascular bundles of the corpora cavernosa and the urethral sphincter mechanism.<sup>55</sup> The preservation of one or both neurovascular bundles can reduce erectile dysfunction rates after surgery without compromising the surgeon's ability to eradicate the cancer.<sup>6,37,55-57</sup> A recent review found that sexual potency was preserved in 31-86% of men with organ-confined disease after bilateral nerve-sparing RRP, whereas only 13-56% recovered erectile function after receiving a unilateral nerve-sparing procedure. In comparison, the potency rate was 11-40% in general urological practice irrespective of operation technique.<sup>56</sup> Erectile function rates are also higher when the nerve-sparing technique is used in academic vs. non-academic or community series, which suggests that the skill and experience of the surgeon is an important factor in obtaining good results even with better techniques.<sup>55,56</sup>

The use of nerve-sparing surgery also generally implies that other aspects of the surgery are performed with more care, including the preservation of the nerves that supply the external sphincter mechanism. A good dissection at the apex of the prostate can prevent injury to this neurological supply, which is important for urinary continence. While improved continence rates have been observed with the use of nerve-sparing

surgery, the evidence is not as strong as for the recovery of erectile function. This is likely because the apical dissection is the most difficult part of the operation to learn and requires a high level of surgeon skill in order for patients to benefit in terms of their urinary function.<sup>36,55,57</sup>

The use of nerve-sparing surgery was not recommended by the RAND or CCO panels as a process quality indicator.<sup>9,14</sup> While the CCO panel felt that it was an excellent indicator of quality, they thought that it should only be used with adjustment for patient and disease characteristics.<sup>58</sup> Indeed, the successful application of fine details of this surgical technique have been shown to increase the recovery of continence and erectile function after radical prostatectomy.<sup>2,17,36,37,55,56,59-61</sup> Thus, one could consider the use of nerve-sparing surgery, which is relatively easy to ascertain from operative reports, as a proxy for calculating these two adverse event rates directly in a context where the late morbidity data are unavailable.

### **2.6.6 Urinary incontinence and erectile dysfunction**

Urinary incontinence and erectile dysfunction are the most widely reported causes of morbidity following surgery and they can have severe impact on a patient's quality of life.<sup>3</sup> Wide variations in the rates of these two adverse outcomes have been reported over a number of population-based surveys, although rates are typically low for those who received care from experienced surgeons.<sup>2,35,36,62</sup> A review of the literature has found incontinence rates ranging from 0.3-65.6%, and erectile dysfunction rates from 11-87%.<sup>36</sup> Many reasons have been proposed to explain this wide variation in rates, including differing methods of data collection, lack of consensus on the duration of follow-up and

on the definitions of the conditions, data derived from the comparison of single vs. multi-institutional studies, patient selection methods, inclusion of patients that were incontinent or had no erectile function prior to surgery, and variation in patient age, comorbidities, disease severity across studies.<sup>3,6,35,36,60,61,63</sup> Most importantly, however, variations in the use of the surgical techniques described above may be the most important factor accounting for the wide variation in rates.<sup>37,60</sup>

In the quality indicator literature, the RAND expert panel opinion was that the feasibility of using their indicator “Patient assessment of their erectile dysfunction and urinary incontinence using a reliable, validated survey instrument” was variable depending when and where it is used.<sup>10</sup> Both Miller *et al.*<sup>15</sup> and Krupski *et al.*<sup>16</sup> found that these were feasible as quality indicators when assessed through patient questionnaires. However, they found that this information could not be obtained from administrative documents or charts.<sup>16</sup> Accordingly, Miller *et al.* suggested that the use of these potential outcome indicators of quality of care must include the use of a validated instrument.<sup>15</sup> Contrary to this advice, with an aim to evaluate quality Hu<sup>50</sup> identified urinary incontinence and erectile dysfunction using the procedure and diagnosis codes found in electronic medical claims data, which are subject to numerous limitations including coding errors, changes in coding standards, and underreporting of events. The only meaningful trend these investigators found was a decline in the diagnosis code identified 3-year incontinence rate from 20% in 1991 to 4% in 1995. While age, region, time, and race were adjusted for, other important covariates were not considered. On the whole, the use of electronic claims data does not seem to be the ideal method to effectively evaluate the incidence of urinary incontinence and erectile dysfunction. We

believe that these sensitive topics are not always brought up by the doctor and/or the patient at follow-up, and therefore are not diagnosed or treated. Consequently, these complications are not recorded in charts or subsequently coded into electronic data. In addition, the presence of incontinence recorded in claims data has been compared to patient questionnaire responses and the sensitivity was found to be low (39%), limiting the ability of claims data to detect true variations in rates.<sup>18</sup>

### **2.6.7 Positive surgical margins and pathology reporting**

Positive surgical margins are generally defined as the presence of neoplastic cells in contact with the ink on the margin of the tissue removed during surgery and sent to the pathologist for analysis. It suggests the presence of residual disease and increases the risk of biochemical recurrence.<sup>64,65</sup> In a multivariate analysis, two-year PSA failure rates were higher (45-55%) for patients with positive surgical margins compared to those with clear margins (11-25%).<sup>64</sup> The incidence of positive surgical margins is affected by disease severity, differences in specimen processing, and surgical skill.<sup>64,65</sup> The careful use of surgical modifications, such as those described by Alsikafi and Brendler in 1998,<sup>66</sup> have later been shown to reduce positive margin rates.<sup>67</sup> While considerable variation in rates has been found among different surgeons, experienced surgeons do tend to have lower rates.<sup>64,65</sup> Video analysis has shown that many positive surgical margins are due to technical errors such as a capsular tear during neurovascular bundle dissection.<sup>68</sup> For these reasons, positive surgical margins should be an excellent outcome indicator of the quality of surgical care. Although it was considered by both panels, it was not ultimately recommended by either group.<sup>10,14,58</sup> The CCO panel felt that positive surgical margins

was an important indicator of quality but interpretation can vary by pathologist, and results could be easily misinterpreted if not adjusted for confounders.<sup>58</sup> Since we are able to assess covariates, it is included here.

The *documentation* of surgical margin status has been evaluated and assessed as a process quality indicator. Miller *et al.*<sup>17</sup> evaluated the presence of surgical margin status documentation in the surgical pathology reports, not the actual status itself, as a quality indicator and found that 93.5% of reports were compliant. The completeness of pathology reporting is also currently reported in the Cancer System Quality Index 2007. CCO has adopted the standards recommended by the College of American Pathologists, which among other elements must include a statement of margin involvement or uninvolved, and a statement of the name(s) of the involved margin(s). In the May-June 2006 audit timeframe, 91% of radical prostatectomy reports contained all the elements required, a rise from 88% in 2005, with some variations by LHINs.<sup>49</sup>

### **2.6.8 PSA levels after surgery**

Post-surgical PSA levels are important outcomes for evaluating quality of care as they are indicators of disease control and the risk of recurrence.<sup>6,9,35</sup> After radical prostatectomy, PSA levels in the blood should decline to undetectable levels within four to six weeks if all PSA-producing prostate cells were removed during surgery.<sup>9,69</sup> Most contemporary series use an undetectable threshold level of > 0.2 ng/mL, although thresholds ranging from 0.1-4.0 ng/mL have also been used.<sup>70</sup> The measurement of PSA levels is the most sensitive way of detecting cancer recurrence many months or years before symptoms or radiographic evidence appear, and thus is an intermediate endpoint

commonly used to measure the success of prostate cancer treatments.<sup>9</sup> In a surgical series using the nerve-sparing technique, an undetectable nadir defined as  $\leq 0.01$  ng/mL was found to be an excellent prognostic factor of biochemical recurrence, with only 3% of patients with undetectable PSA relapsing compared to 75% among those with detectable PSA post-surgery.<sup>70</sup> As a quality indicator, undetectable PSA in the first test after surgery has not been tested for its feasibility. However, data for a related indicator, three consecutively increasing PSA values after treatment, could be feasibly found through detailed chart review.<sup>15</sup>

### **2.6.9 Biochemical disease-free survival**

Biochemical recurrence occurs when serum PSA rises to a detectable level after falling to undetectable levels post-surgery,<sup>69</sup> and is defined by the British Columbia Cancer Agency as two successive increases in serum PSA to a level  $> 0.3$  ng/mL.<sup>5</sup> Some published studies also consider the institution of secondary therapy as a sign of relapse.<sup>71,72</sup> In large radical prostatectomy series, 5-year biochemical recurrence rates ranged from 70-87%, and 10-year rates ranged from 52-68%.<sup>35</sup>

Biochemical disease-free survival is an important outcome for evaluating quality of care because it is the preferred surrogate marker for cancer control and survival after radical prostatectomy, and has been widely validated as such.<sup>9,35</sup> Since treatment approaches are evolving rapidly, long-term disease-free survival data are irrelevant by the time the data are collected and analyzed because they represent the outcomes of treatments that have been substantially revised or are no longer in use.<sup>9</sup> There is also some question about the prognostic value of biochemical disease-free survival, as it does

not always imply that a clinically significant cancer relapse will follow.<sup>9</sup> While it was recommended by both panels,<sup>10,14</sup> there has not been any subsequent reference to biochemical disease-free survival in other radical prostatectomy quality indicator literature.

## **2.7 Testing the convergent construct validity of quality indicators**

The next step in the development of recommended process and outcome indicators is to evaluate their validity in an Ontario population using high quality medical chart data, adjusting for the effects of disease severity, comorbidity, and other covariates. Although a reference standard with which to evaluate the validity of quality indicators is not available, we do know that higher volume generally indicates better overall quality of care. Therefore, volume may be used as the best tool available to test the validity of candidate quality indicators through the test of convergent construct validity, which determines whether two instruments proposed for measuring the same unobservable construct agree with each other.<sup>26</sup> That is, if higher volume hospitals perform better on the quality indicators of interest, then it demonstrates that both instruments are measuring the underlying concept of quality.

## CHAPTER 3: METHODS

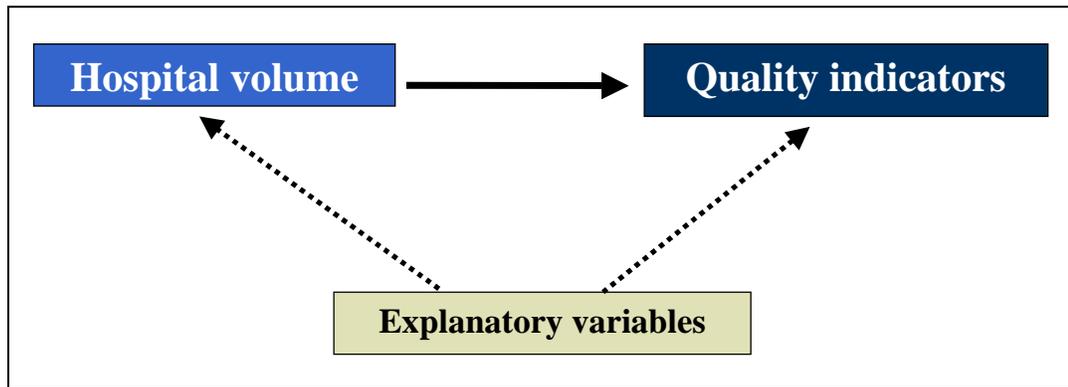
### 3.1 Study purpose and objectives

The study purpose was to test the construct validity of radical prostatectomy quality indicators by assessing their associations with hospital volume for prostate cancer patients treated with radical prostatectomy in Ontario. The objectives were:

- (1) To assess variations in quality indicators by hospital volume.
- (2) To investigate whether selected explanatory variables account for some of the variation observed in Objective 1.

### 3.2 Study design

This was a retrospective cohort study, which allowed for the examination of multiple surgical quality indicators. The study population was assigned to hospital volume categories representing the number of radical prostatectomies performed in their treating hospital during the year they received surgery, and followed after treatment for the outcomes that defined the quality indicators of interest. For Objective 1, each quality indicator for which we had sufficient and valid data was analyzed to determine the association between hospital volume and the quality indicator outcome. For Objective 2, we examined how these relationships changed with the consideration of explanatory variables (Figure 3.1).



**Figure 3.1: Conceptual framework**

### **3.3 Data collection overview**

Study subjects were previously identified and extensive data collected on them as part of a larger case-cohort study designed to examine the role of comorbidity on early death after prostate cancer treatment and to compare the outcomes of radiotherapy and radical prostatectomy treatments. Dr. Groome was the Principal Investigator of that parent study and Dr. Siemens was a clinical co-investigator. Only the sub-cohort group in the larger case-cohort study was examined here.

For the parent study, subjects were identified through a cancer treatment database housed at the Division of Cancer Care & Epidemiology (CCE) at the Queen's Cancer Research Institute. The CCE database links Ontario Cancer Registry (OCR) records with Canadian Institute for Health Information's (CIHI) Discharge Abstract Database (DAD), along with Cancer Care Ontario (CCO) radiotherapy treatment data and Canadian census area-level socioeconomic status (SES) indicators, thereby enhancing the data present in the OCR alone.<sup>73,74</sup>

Extensive data were collected on the study subjects during the larger study. Under the direction of the study coordinator, health care providers trained as data abstractors reviewed detailed hospital and cancer center medical charts, with subsequent missing data sought from other sources. Along with relevant CCE database elements, abstracted chart data was acquired from the already completed electronic study database and used for data processing, exploration, and analysis. Please refer to Appendix I for selected pages of the medical chart abstraction form.

### **3.4 Study population**

The study population consisted of a stratified random sample of patients diagnosed with prostate cancer from 1990 to 1998 in Ontario who were treated by radical prostatectomy with curative intent within six months of diagnosis. This was defined in the parent study, using the electronic CCE database.

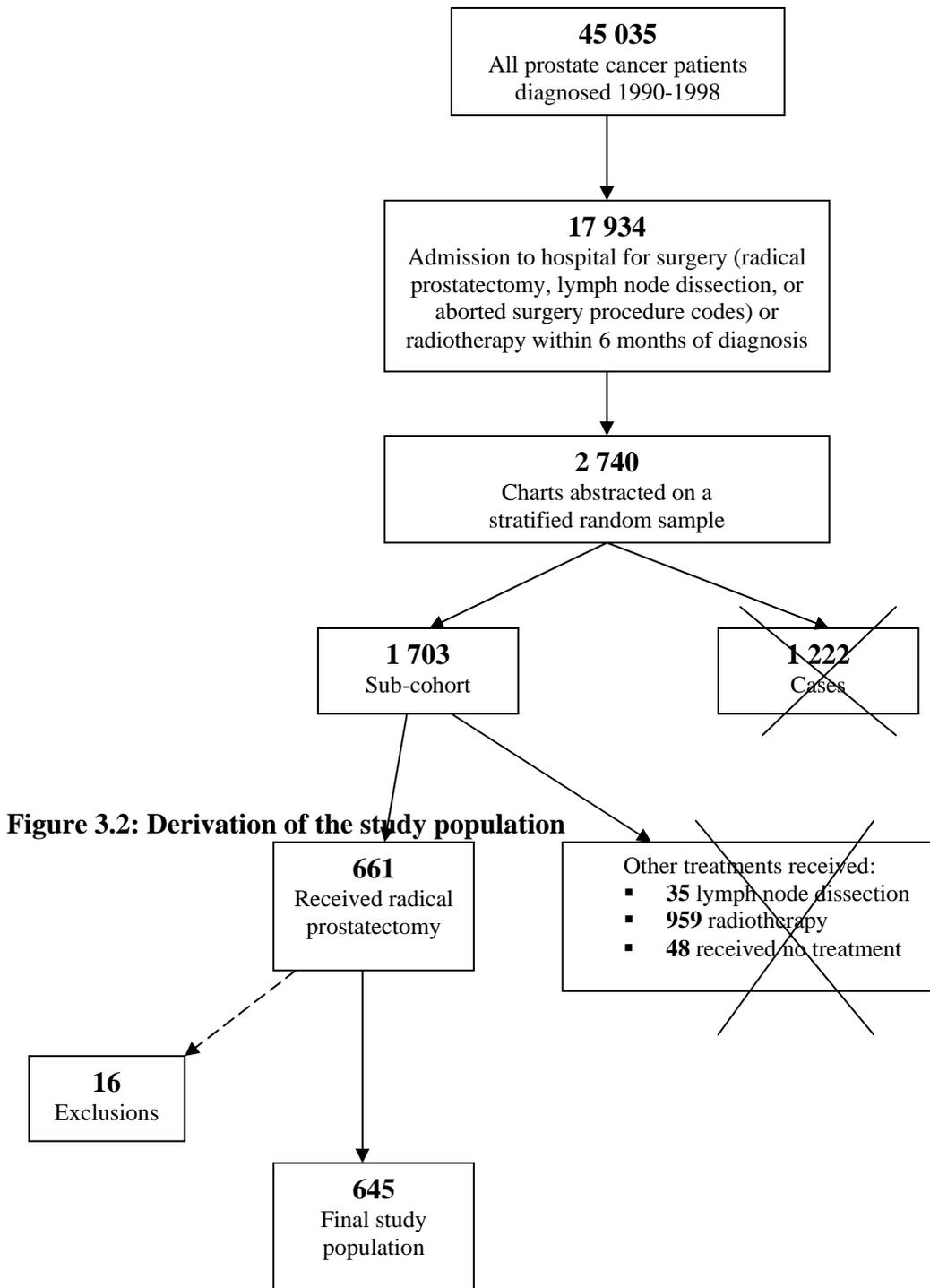
Data collection of 150 randomly selected patients from each of the 8 Ontario Regional Cancer Center (RCC) catchment areas was planned to create the sub-cohort group. This ensured that the study population spanned the province and allowed for representative region-specific analyses. However, this frequency was doubled when sampling Central East RCC to account for increased population density and heterogeneity of patients seen. Also, there was over-sampling of sub-cohort members in the smaller cancer centers due to small numbers of prostate cancer mortality cases, which was another group targeted by the larger study but not of interest here.

There were 45,035 men diagnosed with prostate cancer from January 1, 1990 and December 31, 1998 in Ontario, as indicated by the ICD-9 diagnosis code of 185 (prostate cancer)<sup>75</sup> in the Ontario Cancer Registry. Among these patients, surgical procedure codes recorded in the Discharge Abstract Database were used to identify patients admitted into hospital for radical prostatectomy within six months of diagnosis. This found 7,407 patients who received surgery, 7,921 patients who had lymph node dissection (potentially abandoned surgical cases), and 70 patients with aborted surgery. Because of overlap between these three groups of patients, the total ‘surgical treatment’ group came out to 10,293 patients. Together with 9,524 men who received high dose (curative) radiotherapy within six months of diagnosis, the resulting sampling frame of the parent study consisted of 17,934 men. Again, this number was less than the total of the surgical and radiotherapy groups because some patients received both treatments.

Through the stratified random sampling procedure detailed above, medical chart data was collected for 2,740 patients, 1,703 of whom formed the sub-cohort group. After the medical chart abstraction process, it was confirmed that 661 of these patients had actually received surgery as curative treatment. Therefore, these patients made up the initial dataset for this study. This process of subject selection is detailed in Figure 3.2.

We used the sub-cohort group for analyses, as this group was created through random sampling of the RCCs. The cases were not of interest here as they only consist of patients who died after surgery of other causes or from prostate cancer. It should be noted that due to the case-cohort design, some of the cases were included in the sub-cohort group. It is due to this overlap that the total number of charts abstracted is less than the total number of cases and sub-cohort patients.

During the data collection and chart abstraction process, some of these 661 patients were excluded if they were not residents of Ontario, their surgery was aborted or if there was insufficient information available on a patient (i.e. only pathology or histology reports were found). They were also excluded if the chart abstracted surgery date was greater than 190 days (six months + one extra week to allow for the time between hospital admission and surgery) after the date of diagnosis. Two patients were excluded due to insufficient information and 14 were excluded because the surgery date exceeded 190 days following diagnosis. Therefore, the final study population was 645.



**Figure 3.2: Derivation of the study population**

### **3.5 Data quality and validity**

#### **3.5.1 Medical chart abstraction**

In the medical chart data abstraction procedure, the treating hospital chart was reviewed first. Where available, cancer center charts were also reviewed in order to provide a greater depth of information about patients. If these two sources did not provide sufficient initial information or post-treatment follow-up, other charts were pursued, including those from secondary hospitals, urologists, and general practitioners. The use of multiple sources of patient information was also beneficial in allowing the verification of overlapping data elements.

The data abstractors were trained and monitored by the study coordinator. They were provided with a manual describing the variables to be measured and details regarding their abstraction, and were trained in-house by the study coordinator using previously abstracted patient charts and standardized abstraction forms. During the training period, abstractors completed multiple individual and group practice sessions with discussions and debriefings of their performance. After training, the data abstractors performed the data abstractions at the site of record storage (i.e. Regional Cancer Centers and hospitals). Photocopies of some pertinent sections of patient records were made by the abstractors and retained under locked storage at the Division of Cancer Care & Epidemiology for future reference. All of the information acquired was entered directly into an electronic database, in Microsoft Access format, using Medquest software.<sup>76</sup>

While no formal inter-rater or intra-rater reliability tests were performed, a number of measures were taken to standardize the abstraction procedure and minimize errors. The abstractors must have achieved 100% reliability on key study variables before they were sent into the field. Major errors in these key variables were reconciled with further training, remediation, and modification of the abstraction manual. Furthermore, the study coordinator made site visits and performed on-site reabstractions during the data collection process. Also, abstracted data was checked weekly for completeness and errors.

It should be noted that the medical chart abstraction form and the data collection process used were based on the results of a pilot study, which confirmed that the key variables essential to the parent study were available in the charts and that it was possible to access general hospital charts in a timely fashion.

### **3.5.2 Electronic data**

The Ontario Cancer Registry, a key source in the CCE database, has been found to capture information on at least 97% of all incident cancer cases diagnosed.<sup>77,78</sup> Therefore, given its completeness, the OCR was an excellent source for creating a sample of study subjects. In addition, the OCR date of diagnosis, date of death, and vital status were used to calculate the survival variables. Their use was preferred over the use of the corresponding variables obtained through chart abstractions. For date of diagnosis, the chart data were as good or better as the OCR data, but the OCR date of diagnosis was used because it all came from the same source of pathology reports. Date of death and vital status was not sought after in the charts because the OCR undergoes a death

clearance using death certificate data from the Registrar General of Ontario. The only concern with the registrar data were the patients who moved out of province after prostate cancer diagnosis, because even if they died, they would still be considered alive by this source. However, this number was likely to be very small.

The CCE database contains, through the CIHI Discharge Abstract Database, hospital discharge data for cancer patients whenever a diagnosis of cancer is recorded. By 1995, CIHI was receiving reports on all hospitalizations in Ontario, including records of all surgical procedures. Prior to that period, the number of radical prostatectomies missed should have been very small. Although data quality control is extensive and occurs at several different levels, DAD data quality documentation for 2001-2002 indicated that in comparison with medical charts, 7% discrepancies were found for the principal procedure recorded for the hospital stay and 19.7% for other procedures.<sup>79</sup> The patients included in this study did receive radical prostatectomy (i.e. true positives), since this was confirmed during the chart abstraction process. However, false negatives who were incorrectly coded as not having received prostatectomy in the DAD would have been missed and not included when deriving the study sample, thereby potentially causing some selection bias.

### **3.6 *A priori* study variables**

A subset of the quality indicators recommended by Cancer Care Ontario<sup>14</sup> was selected for investigation *a priori* (Table 3.1). Included quality indicators were those influenced by surgeon skill and for which data was abstracted in the parent study.

In addition, two other possible quality indicators for investigation were developed *a priori* and are listed in Table 3.2. Positive surgical margins and the use of nerve-sparing surgical techniques were thought by the CCO panel to be excellent indicators of quality, but could be easily misinterpreted without adjustment for patient and disease characteristics.<sup>58</sup> Since we are able to make these adjustments, we retained the option of using these two indicators as proxies for some of the CCO recommended quality indicators, in the event that those recommendations could not be analyzed. Both post-surgical PSA levels and positive surgical margins are indicators of disease control and risk of recurrence following surgery,<sup>6,35,64,65</sup> and the use of a nerve-sparing surgical technique is known to reduce the incidence of urinary incontinence and erectile dysfunction after surgery.<sup>7,29,35-37,60</sup>

To ensure that variations seen in surgical quality indicators by hospital volume were not due to other factors that might have influenced the rates of adverse outcomes following radical prostatectomy, we examined eight such explanatory variables (age, comorbidity, year of diagnosis, socioeconomic status, surgical approach, pre-treatment PSA level, clinical T category, and biopsy Gleason score). The analysis considered whether the inclusion of each of these variables reduced or otherwise changed the volume-quality associations seen. This could be used to provide advice to urologists by suggesting reasons why specific volume groups had worse outcomes. Evidence for the selection of these covariates is given in Table 3.3.

**Table 3.1: List of quality indicators recommended by Cancer Care Ontario**

---

**Quality indicators selected for investigation *a priori***

Proportion of prostate cancer patients with acute surgical complications:

- Blood loss of 2.0 L or greater
- Rectal injury
- Cardiovascular complications such as coronary heart disease, myocardial infarction, heart failure, or pulmonary edema
- Proximal deep vein thrombosis / pulmonary embolism
- Infection
- Placed on long term anticoagulant therapy

Proportion of prostate cancer patients with length of stay greater than four days following radical prostatectomy

Proportion of patients undergoing prostate cancer surgery who experience loss of potency, incontinence, or undergo procedures for bladder neck contracture or stenosis

Proportion of prostate cancer patients whose PSA level is undetectable at first follow-up

Biochemical disease-free and overall survival at 5, 10, and 15 years after primary treatment with radical prostatectomy

**Quality indicators *not* selected for investigation *a priori***

Proportion of patients with low risk prostate cancer who report they were informed about treatment options and adverse effects, and with documentation of involvement in treatment decision

Proportion of prostate cancer patients assessed after treatment for voiding function and potency

Proportion of prostate cancer patients dying without cancer less than five years after surgery

Proportion of cancer patients with clinically low risk disease age 75 or greater who undergo radical prostatectomy

Proportion of needle biopsy pathology reports for prostate cancer patients meeting Canadian national standards

Proportion of patients having undergone definitive therapy for prostate cancer who are followed at least twice in the first year and then at least annually thereafter

---

Source: Gagliardi *et al.*<sup>14</sup>

**Table 3.2: *A priori* list of other quality indicators for possible investigation**

---

|   |
|---|
| Proportion of prostate cancer patients with positive surgical margins following radical prostatectomy             |
| Proportion of prostate cancer patients who received radical prostatectomy with a nerve-sparing surgical technique |

---

**Table 3.3: *A priori* list of explanatory variables to be investigated**

---

| <i>Explanatory variable</i>         | <i>Evidence for relationship with surgical quality indicators</i>  |
|-------------------------------------|--|
| <b>Age</b>                          | Decline in general health with age may increase the risk of adverse outcomes: i.e. age has been shown to be related to acute complication, incontinence, and erectile dysfunction rates <sup>2,52,60,62</sup>  |
| <b>Comorbidity</b>                  | Due to competing risks, patients with many comorbid conditions tend to have worse surgical outcomes: <sup>35</sup> i.e. comorbidity is associated with acute complication rates <sup>52</sup>  |
| <b>Year of diagnosis</b>            | Improved techniques <sup>6,35</sup> might explain better results in later years: i.e. as seen for acute complications <sup>51,53</sup>   |
| <b>Socioeconomic status (SES)</b>   | SES may impact the patient’s ability to cope with treatment, which could be reflected in post-surgical outcomes: i.e. SES might be a risk factor for prostate cancer progression and death <sup>80</sup>   |
| <b>Surgical approach</b>            | There may be slight differences for some outcomes depending on the surgical approach used: i.e. retropubic approach has been found to lead to more blood loss, rectal injury, incontinence, and longer hospital stay, while the perineal approach has been found to lead to higher rates of erectile dysfunction <sup>6,7,81</sup> |
| <b>Disease severity:</b>            |  |
| ▪ <b>Pre-treatment PSA level</b>    | Patients with more severe disease tend to have worse outcomes: i.e. these descriptors of disease severity are commonly used to predict the risk of post-surgical outcomes such as PSA recurrence and prostate cancer death <sup>30,31,70,82</sup>  |
| ▪ <b>Clinical stage</b>             |  |
| ▪ <b>Biopsy Gleason score</b>       |  |
| <b>Post-surgery hormone therapy</b> | Evidence from randomized trials indicates that while neoadjuvant hormone therapy does not alter PSA recurrence rates, <sup>83</sup> disease recurrence rates are lower for patients who receive hormone therapy post-surgery <sup>84</sup>   |

---

### **3.7 Data processing and exploration**

All data processing and statistical analyses were performed using SAS (Statistical Analysis Software), version 8.2.<sup>85</sup> Sections 3.7.1 to 3.10.3 describe the data processing and exploration that was completed in order to develop variables for analysis. If it was decided that it was not possible to proceed with statistical analysis, the reasoning was described. Any relevant tables and univariate statistics for the variables being developed can be found in Appendix II.

#### **3.7.1 Diagnosis and surgery dates**

Although not direct quality indicators or explanatory variables themselves, the dates of diagnosis and surgery are important pieces of information that were used to help process and define many of the study variables. Therefore, work was done to create clean and accurate diagnosis and surgery date variables. Even though diagnosis dates were abstracted from the charts, the OCR date of diagnosis was used for the reason described previously (Section 3.5.2). On the other hand, surgery dates were not available from the Discharge Abstract Database, which only recorded date of hospital admission for the surgical procedure. Therefore, the date of surgery was defined using the charts by reviewing surgical reports and doctor's notes. Both the OCR date of diagnosis and the chart-abstracted date of surgery were 100% complete in the final study population.

## **3.8 Quality indicators**

### **3.8.1 Acute complications**

We were limited by the information captured on the medical chart abstraction form (Appendix I). From the list of acute complications suggested by CCO for use as quality indicators (Table 3.1),<sup>14</sup> only rectal injury, myocardial infarction, and thromboembolism (another term for proximal deep vein thrombosis / pulmonary embolism) were abstracted. Furthermore, statements of blood loss were not captured, as they were deemed too variable and subjective by the investigative team that developed the surgical abstraction form. However, units of blood transfused were abstracted, and were assessed as a surrogate for blood loss. Additionally, we decided to include deaths as another acute complication of interest. Although this was not recommended by the Ontario panel, it was listed by RAND as a potential quality indicator<sup>10</sup> and is currently reported in the Cancer System Quality Index 2007.<sup>49</sup> We felt that neglecting to include deaths would provide an incomplete picture of acute post-surgical outcomes.

Abstracted acute complications were captured if they occurred during or within 90 days after surgery, with the exception of myocardial infarction, which was only captured up to seven weeks post-surgery. To be consistent with the 90 day time frame used on the abstraction form, deaths were captured if they occurred within 90 days after surgery. Deaths were identified using the OCR vital status and date of death fields from the CCE database. Total units of blood transfused were initially calculated by adding the units of blood given during and after surgery.

The frequency distributions of these acute complications are presented in Table II.1. The rates of rectal injury, myocardial infarction, thromboembolism, and death within 90 days after surgery were low, ranging from 0.5-2.2%. In each case, the frequency of unknowns was greater than the frequency of events (1.2-28.1%). In contrast, the rate of any blood transfusion was 58.6%, with no missing values. Because of its much higher event rate, blood transfusions were developed into a separate study variable. To increase the number of events, the remaining acute complications were grouped into a summary study variable, where any indication of an event occurring was coded as 'Yes' and otherwise was coded as 'No', with those with an 'unknown' assignment assumed to have not experienced the event. This definition yielded 29 events, or a 4.5% acute complication rate.

Since we have blood transfusion rather than blood loss data, it was not possible to use CCO's recommended cut-point of 2.0 L blood loss or greater for dichotomizing blood transfusion. We felt that 1-2 units of total blood transfused is related to mild to moderate blood loss, and it is usually expected during radical prostatectomies that this amount would be transfused. However, total blood transfusions of three units or greater are considered an excessive amount of blood to use, and therefore, our variable for blood transfusions was defined as such. In the study population, 24.5% of the patients received three or more units of blood (Table II.1).

### **3.8.2 Length of hospital stay**

Although CCO recommended that a length of hospital stay greater than four days following radical prostatectomy is an indicator of poor surgical care in 2005,<sup>14</sup> we did not think that it was an appropriate cut-point to use for the era in which the patients in our study population were treated (1990-1998). This is supported when the frequencies for length of hospital stay were examined, with 81.4% of patients that stayed in hospital for longer than four days and only 15.3% that stayed for 4 days or less (3.3% missing data). Instead of arbitrarily defining our own cut-point, we decided to treat length of hospital stay as a continuous variable. We compared the mean and median for the length of stay variable and found that the data was sufficiently normally distributed to allow for the use of parametric analysis.

### **3.8.3 Urinary incontinence and erectile dysfunction**

For the CCO quality indicator “Proportion of patients undergoing prostate cancer surgery who experience loss of potency, incontinence, or undergo procedures for bladder neck contracture or stenosis,”<sup>14</sup> we found that only erectile dysfunction (potency) and urinary incontinence data were collected (Appendix I). We proceeded to explore these available data in detail.

### **3.8.4 Urinary incontinence: Data processing and exploration**

The chart abstraction form included questions that attempted to collect data on urinary incontinence before and after surgery. In the prostate project abstractor's manual, a patient was considered to be continent if they were "dry" or only had "occasional" incontinence. We examined incontinence rates prior to surgery, to ensure that we were not wrongly attributing an unfavourable outcome to the surgery when the condition was already present prior to the procedure. As shown in Table II.2, only 17 patients (2.6%) were noted to be incontinent prior to surgery. Among these patients, four were noted to be continent and six incontinent > 90 days following the cessation of initial treatment, and the continence status of the remaining seven subjects was unknown. Given this fairly well distributed and rather low number of events, we decided that it was not necessary to consider prior status in developing the incontinence study variable.

Post-treatment incontinence experience was simply defined based on a question on the abstraction form that asked if incontinence was present > 90 days following the cessation of initial treatment. Incontinence that was resolved prior to 90 days following treatment was not recorded as an event, since problems with urination frequently arise after radical prostatectomy. It was not possible to use the more detailed questions on the abstraction form (i.e. degree of severity, treatment, number of pads used per day, and wetness of pads), due to missing information among those known to be incontinent. We found that even using this simplest of definitions, post-treatment incontinence status was unknown for 32.3% of our study population (Table II.2).

### **3.8.5 Erectile dysfunction: Data processing and exploration**

Similarly to urinary incontinence, a series of questions on the abstraction form ask about the presence of erectile dysfunction before surgery and > 90 days following the cessation of initial treatment. In the abstractor's manual, erectile dysfunction was defined as the inability to initiate and maintain an erection for the purpose of sexual intercourse. When we examined erectile dysfunction rates prior to surgery, 13.5% of the sample was noted to have erectile dysfunction, and erectile status was unknown for 48.4% of the sample (Table II.3).

Despite the high rate of unknowns, we used the information on pre-surgical erectile dysfunction in developing the study variable for post-treatment erectile dysfunction, which was created by combining data from some of the more detailed questions on post-treatment function. Subjects were categorized as unknown if no data on post-treatment erectile function was present, and as having pre-treatment erectile dysfunction in accordance with the pre-surgical variable. Otherwise, subjects were considered to have erectile dysfunction if they: (i) used aids to achieve and/or maintain an erection for the purpose of sexual intercourse (Viagara, injections, penile implant, etc.); and/or if there was: (2) a verbatim description of erectile dysfunction post-treatment (see Table II.4 for a breakdown of how the statements were categorized). Despite this work, post-treatment erectile function was unknown for 55.7% of the sample. Since erectile function tends to be a sensitive topic for both patients and physicians, it is possible that the unknowns might represent patients for which erectile status was simply not mentioned or assessed during follow-up appointments.

### **3.8.6 Urinary incontinence and erectile dysfunction: Data quality issues**

We sought further explanation of the high percentages of unknown status for incontinence and erectile dysfunction from the study coordinator. There were three main reasons why this occurred. Firstly, there was limited follow-up for patients with no further hospitalizations recorded in the CIHI Discharge Abstract Database. Without subsequent hospitalizations, for which urinary incontinence and erectile dysfunction does not necessarily require, there were no secondary hospital charts available to provide further information. Secondly, at least five high volume urologists refused to provide access to their patient charts. In two of these cases, the urologists treated most or all prostate cancer patients in their areas. As a result, follow-up data for patients from entire areas is missing, and we were concerned that there may be systematic differences between the urologists who consented and those who declined chart access. Thirdly, Ontario experienced the SARS outbreak during the data collection period of the larger study, which virtually eliminated any further opportunities to access secondary hospital or physician records to obtain data missing from already abstracted primary charts before the end of the data collection period. To improve the follow up information needed for the parent study, the study coordinator and her staff sought information from treating physicians and/or family doctors on a minimal list of key variables. This list did not include incontinence and erectile dysfunction status. Given the large number of unknowns, many of which are due to inadequate access to chart information post-treatment, our rates of post-treatment urinary incontinence and erectile dysfunction are not meaningful. Therefore, these variables were excluded from statistical analysis.

### **3.8.7 Nerve-sparing surgery**

Urinary incontinence and erectile dysfunction are very important post-surgical outcomes. They could not be analyzed directly, but since the use of a nerve-sparing surgical technique is known to reduce the incidence of urinary incontinence and erectile dysfunction after surgery,<sup>7,29,35-37,60</sup> we decided to examine nerve-sparing surgery, noted *a priori* as a possible quality indicator for investigation (Table 3.2), as a proxy for the related post-surgical outcomes. Patients were categorized as having received nerve-sparing radical prostatectomy if the use of this technique was considered and performed. Patients were categorized as having received non-nerve-sparing surgery if the technique was considered and not performed, or not considered at all. It was unknown whether the technique was used for 7.1% of the population.

### **3.8.8 Positive surgical margins**

We listed positive surgical margins in Table 3.2 as a quality indicator that may possibly be investigated, as it is also an indicator of disease control following surgery.<sup>6,64,65</sup> Since it was not possible to proceed with the CCO recommended indicator of disease control, undetectable PSA at first follow-up, we processed data abstracted from pathology reports after surgery (Appendix I) to create a positive surgical margins study variable. Patients were classified into one of four categories based on comments found on the pathology reports (Table II.5). In the case where multiple comments were made, the most extensive or specific involvement was used. With the resulting study variable, there are only three patients (0.5%) for which we have no pathology report data. While margin status was unknown for 74 (11.5%) patients, this was not because we did not have

their surgical pathology reports. It was because the pathologists writing the reports made no reference to margin status. Therefore, we decided to proceed with statistical analysis for positive surgical margins.

### **3.8.9 Undetectable PSA levels at first follow-up**

The levels and corresponding dates of serum PSA tests following surgery were recorded during the medical chart abstractions (Appendix I). However, there were several issues with the quality and availability of data for this indicator. Firstly, we had no record of any post-surgery PSA testing for 25.9% of patients. Much of this was likely due to missing follow-up as described previously. Secondly, because patients may have been initially seen in a doctor's office post-surgery, the completeness of the hospital charts that we based most of our data collection on was variable. That is, the first PSA test that was abstracted may not necessarily have been the first test performed after surgery. Thirdly, 37.4% of the PSA values could only be classified as undetectable since they were below the minimum sensitivity of the PSA tests being used, which ranged from 0.04 to 4.0 ng/mL. In the case of less sensitive tests, however, the actual PSA level might have been quite high, although it is impossible to tell based on the data we have. For these reasons, we decided to exclude this quality indicator from statistical analysis.

### **3.8.10 Biochemical disease-free survival**

CCO's recommended quality indicator for survival was "Biochemical disease-free and overall survival at 5, 10, and 15 years after primary treatment with radical prostatectomy."<sup>14</sup> However, the consistency of PSA testing varies, and thus we developed censoring and time-to-event variables in preparation for overall disease-free survival analysis instead. Although biochemical recurrence should always precede local failure detected by other means, as well as regional failure and distant failure, infrequent PSA testing might have resulted in other types of failure being identified first.

The variables used for 5-year disease-free survival analyses were defined as follows. The censoring date was 31 December 2001, the last date for which we have complete OCR vital statistics information, or five years after the date of surgery, whichever came first. Events occurring past the censoring dates were not counted. Patients noted on the abstraction form (Appendix I) as having residual disease at the completion of initial treatment were assigned the date of surgery as the failure date. For patients whose disease recurred, the earliest failure date was recorded and they were assigned to the censoring category corresponding to how they failed first (positive biopsy, positive digital rectal examination, PSA  $\geq 0.5$  ng/mL or any statement of PSA failure in the chart, regional failure, distant failure, or death). Deaths were assessed using the OCR vital status and date of death variables from the CCE database but, to ensure that we did not miss recurrent events, deaths were only counted if there was evidence of follow-up within the 18 months prior to the date of death otherwise, recurrence information was considered missing. Similarly, patients noted to have no residual or recurrent disease were only considered as censored if they were followed up within 18

months prior to 31 December 2001. Otherwise, their recurrence data was also considered missing.

Table II.6 presents the frequencies of the resulting recurrence variable, which had 188 patients with missing survival data (29.2%). Although not as extreme a problem as with incontinence and erectile dysfunction, we had decided *a priori* that less than 80% completeness would not be analyzable. Also, given that 5-year survival data were missing for 29% of the study population, we expected that 10 and 15-year survival would have even higher rates. Therefore, we did not proceed further with disease-free survival.

### **3.9 Explanatory variables**

#### **3.9.1 Post-surgery hormone therapy**

The variable for hormone therapy was developed from cleaning and processing several questions on the chart abstraction form. During the data collection process, the data abstractors were careful to record the dates of hormone therapy given before and/or after radical prostatectomy. We categorized hormone therapy as received pre-treatment, pre & post treatment, post-treatment, and none (see Table II.7 for rates). All dates were checked to ensure that they occurred in the correct time period (i.e. pre or post surgery). Post-surgery hormone therapy dates were checked to ensure that it was given as additional non-prostatectomy initial treatment rather than treatment for recurrence, and the study variable was corrected accordingly if that was not the case. This was done in collaboration with the study coordinator by reviewing entire electronic chart abstraction records for individual patients, and the decisions made were based on this information in conjunction with her knowledge and experience with the parent study. As shown on

Table II.8, a total of only 17 patients (2.6%) received post-surgery hormone therapy, which is known to lower disease recurrence rates.<sup>84</sup> Given the small frequency of events, we decided not to consider this explanatory variable in the statistical analysis.

### **3.9.2 Additional non-prostatectomy initial treatment**

In addition to hormone therapy, the frequencies of additional non-prostatectomy initial treatment received by the study population were examined to ensure that none of them should be included as additional explanatory variables. No patients were recorded as having received chemotherapy or other prostate cancer-related surgery, and only 1.4% received adjuvant radiotherapy (Table II.8). Again, given the lack of or very small proportion of patients receiving other treatment, we deemed that it was not necessary to consider this variable further.

### **3.9.3 Surgical approach**

We found that most of our study subjects (89.5%) received retropubic radical prostatectomy, with only 50 patients (7.8%) receiving perineal radical prostatectomy. Only one patient received an “other” type of radical prostatectomy, and type was unknown for 2.6% of patients (Table II.9). We noted that perineal radical prostatectomies were only performed by surgeons in two regions during the study era. Therefore, this variable would actually be describing the surgical quality of providers in these two areas, rather than how surgical quality differs by the type of radical

prostatectomy performed. For these reasons, and also because there was little variation, this explanatory variable was not used in the statistical analysis.

#### **3.9.4 Age**

Age was defined using the patient's date of birth and the OCR date of diagnosis. Since one of CCO's other recommended quality indicators was "proportion of cancer patients with clinically low risk disease age 75 or greater who undergo radical prostatectomy", we examined whether this was a suitable guideline to categorize the study population in terms of age. However, only 11 patients (1.71%) who received radical prostatectomy in our sample were aged 75 and above. For statistical analysis, age at diagnosis, for which we have 100% complete data, was kept in continuous form.

#### **3.9.5 Comorbidity**

Medical histories were reviewed in the 5-year period prior to surgery to assess comorbidity using the Cumulative Illness Rating Scale (CIRS), which assigns the severity of 15 disease categories on a scale from one to five, with five being the most severe illness. The categories of disease are: heart, vascular, hematopoietic, respiratory, EENT, upper GI, lower GI, liver, renal, genitourinary (excluding prostate cancer related conditions), musculoskeletal / integument, neurological, endocrine / metabolic and breast, psychiatric illness, and miscellaneous.<sup>86,87</sup> The chart-abstracted CIRS ratings had already been cleaned previously,<sup>74</sup> and only nine patients (1.4%) were found to have no comorbidity data. There are various ways of summarizing the 15 CIRS scores into one

variable representing comorbidity. The method we used was to add up the ratings to form an overall CIRS total score.<sup>87</sup> The continuous form of the CIRS total score was used for statistical analysis, as it had already been validated in a prior study. Hazard ratios for other-cause death at each CIRS increment using the single continuous effect were compared with those calculated using an indicator variable for each CIRS increment, and were found to match up well.<sup>88</sup>

### **3.9.6 Year of diagnosis**

When we examined the distribution of study subjects by year of diagnosis (Table II.10), we found that only 3.1% of subjects were diagnosed in 1990, 7.1% in 1991, and 7.3% in 1992 and higher proportions of subjects were diagnosed in the later years. To address the low numbers at the beginning of the study period, the explanatory variable for year of diagnosis was categorized as three eras (1990-1992, 1993-1995, and 1996-1998). As era was defined from the OCR date of diagnosis, it is 100% complete.

### **3.9.7 Socioeconomic status**

Socioeconomic status was already assigned for the study population in previous work, which was completed as follows.<sup>73,89</sup> The aggregate measure of SES used was median household income by Canadian census enumeration area. Subjects were assigned to enumeration areas based on the postal code of residence at the time of diagnosis, using a conversion file provided by Statistics Canada. Using median household income level information from the closest Canadian census, subjects were assigned into SES quintiles

based on the income distribution of the general population. Only 1.24% of the study population could not be assigned into an SES quintile.

### **3.9.8 Disease severity: Pre-treatment PSA**

The chart abstraction form (Appendix I) provided space to record up to four pre-treatment serum PSA (prostate specific antigen) levels and the corresponding dates of the tests. The following steps were taken to process this data into one pre-treatment PSA value for statistical analysis. The first choice was to use the most recent PSA test performed within one month prior to the biopsy date, to ensure that it reflected the patient's disease severity at the same time that clinical stage and Gleason score were assessed. In order to minimize the frequency of unknowns, the second choice was to take the PSA prior to biopsy that was closest in time. Otherwise, any post-biopsy, pre-surgery PSAs were read. We were unable to assign PSA values for 46 patients (7.1%). For statistical analysis, pre-treatment PSA was divided into clinically relevant quartiles (< 4, 4-10, 10-20, and  $\geq 20$ + ng/mL) that have been used in other literature.<sup>13,88</sup>

### **3.9.9 Disease severity: Clinical stage**

Clinical TNM staging was assigned using the sixth edition TNM classification.<sup>90,91</sup> When the stage was assigned in the chart using an alternate system that could not be automatically converted to this edition, the study coordinator or designate reviewed abstracted extent of disease information to assign stage. Examination of the N (node) and M (metastasis) status of the population found that the staging was

either zero or unknown, with the exception of one patient that was noted to be N1. We think that if patients had any positive lymph nodes or metastases, the appropriate stage would have been noted in the charts. Therefore, only T category, for which there were 2.8% subjects of unknown status, was needed to represent clinical stage. Patients with T1a and T1b disease were grouped into one category (6.2% total) due to small numbers, as were those with T3 disease and above (1.4% total). Another reason for grouping patients categorized as T3 and above is that this group all has advanced disease that has spread outside of the prostate and, therefore, were much less likely to be cured by surgery.<sup>6,7,35</sup>

### **3.9.10 Disease severity: Biopsy Gleason score**

The study variable for biopsy Gleason score, used to represent disease grade, was developed from processing a number of variables in the abstraction form. The first choice was to use the Gleason score that was abstracted from biopsy reports. However, as shown in Table II.11, this information was unavailable for 128 subjects (19.8%). To reduce the number of missing values, we assigned Gleason score using grade assignment from biopsy-based pathology reports. According to the 1997 AJCC Cancer Staging Manual,<sup>90</sup> subjects with well differentiated tumors correspond with Gleason scores of 2-4, those with moderately differentiated tumors correspond with scores 5-7, and poorly differentiated tumors correspond with scores from 8-10. We assigned the middle value of these three ranges, such that well differentiated tumors were assigned a Gleason score of 3, moderately differentiated a score of 6, and poorly differentiated to a score of 9. This strategy was used to complete the Gleason assignment for 88 of the 128 subjects missing

a biopsy Gleason score. The remaining 40 patients (6.2%) were missing both biopsy and assigned Gleason score (Table II.11).

For statistical analysis, Gleason score was divided into five categories (2-4, 5, 6, 7, 8-10) based on a relatively equal distribution of subjects and based on how Gleason has been categorized in previous work by D'Amico *et al.* in developing nomograms to predict biochemical failure and disease-free survival.<sup>31,82</sup>

### **3.10 Hospital volume**

Hospital volume was defined based on the total number of radical prostatectomies performed by a hospital in a given year. Since the total number of radical prostatectomies performed by each hospital varies by year, the same hospital can fall under different volume categories over the nine years of the study period (1990-1998). Each patient in our study population was assigned the volume category of their treating hospital in the year that they received surgery.

#### **3.10.1 Determining hospital volume**

Although we had planned to use a hospital volume variable that had already been assigned to the study population for a previous study using the CCE database,<sup>73</sup> this was not possible because we noticed that the programming code previously developed made a number of potentially incorrect assignments and also could not assign volume to some subjects. Since volume was the exposure variable in this study, it was very important to ensure that it was as complete and accurate as possible. Therefore, we went back to the

CCE database to understand why these problems occurred and to create a new hospital volume variable.

We suspected that the radical prostatectomy might not have been coded correctly into the Discharge Abstract Database by medical records staff who did not understand the nuances of the wording for this surgical procedure. Therefore, to calculate total hospital volume of all radical prostatectomies performed for each year of the study period, in addition to the Canadian Classification of Procedures (CCP) code 72.4 (radical prostatectomy), we also included prostate cancer patients who were assigned the codes 72.2 (suprapubic prostatectomy), 72.3 (retropubic prostatectomy), 72.52 (perineal prostatectomy), and 72.59 (other prostatectomy, including prostatectomy not otherwise specified, prostatic adenectomy not otherwise specified, and transcapsular prostatectomy).<sup>92</sup> Patients were only counted if they were given any of these other codes besides 72.4 if the procedure was received between seven days prior to six months after the OCR diagnosis date. This time frame was stipulated to increase the chance that these patients received curative treatment for prostate cancer, rather than for benign prostatic hypertrophy.

For the remaining 35 patients that could not be assigned volume through electronic data using these additional criteria, we sought the Ministry of Health code representing the hospital where radical prostatectomy was performed and the year of surgery from their paper charts to assign volume. We suspect that these remaining patients were coded as having received lymph node dissection only in the CIHI data.

### **3.10.2 Hospital amalgamations**

An additional problem with volume assignment during the study period was that many hospital amalgamations took place in Ontario. Using Ontario Ministry of Health and Long-Term Care records, we looked up all the hospital names corresponding to the MOH codes referenced as having performed radical prostatectomy, and corrected the coding and corresponding volume assignment to account for the amalgamations.<sup>93,94</sup> For example, if hospitals A and B amalgamated during 1996 and the new entity was assigned MOH code C, then from 1996 onwards patients who received surgery at hospital A or B were assigned code C instead. In the years prior to 1996, hospitals A and B were coded separately.

There is potential for misclassification of volume after amalgamations occur, as some hospital groups may have continued to operate as separate sites for the most part, in which case hospital volume would be overestimated, while others might have actually amalgamated their operations. The proximity between the amalgamated hospitals might have influenced whether one or the other was more likely to have happened.

### **3.10.3 Categorizing hospital volume**

Hospital volume was reformatted as the number of surgeries performed by a hospital per month. Radical prostatectomy volume in Ontario hospitals ranged from less than one to just under 15 surgeries per month during our study period. For analysis, these were collapsed into six categories ( $< 1/m$ ,  $< 2/m$ ,  $< 3/m$ ,  $< 4/m$ ,  $< 7/m$ , and  $\geq 7/m$ ). This categorization was based on a roughly equal distribution of subjects in the lowest four categories. The higher volumes were divided into two groups in order to examine the

differences between the truly high volume hospitals in the  $\geq 7$ /month category from moderately high volume hospitals found in the  $< 7$ /month category.

### 3.11 Summary: Final set of study variables

Table 3.4 summarizes the final set of study variables that were used for statistical analysis, along with the variable formatting and/or categorization.

**Table 3.4: Final set of study variables**

| <i>Study variable</i>   | <i>Format</i>     | <i>Categories (if applicable)</i>                 |
|---|-------------------|---|
| <b><u>Quality indicators</u></b>  |                   |   |
| Acute complications, any of: <ul style="list-style-type: none"> <li>▪ Rectal injury, thromboembolism, death (all during or within 90 days after surgery), or</li> <li>▪ Myocardial infarction (during or within seven weeks after surgery)</li> </ul> | Dichotomous       | Yes<br>No   |
| Total blood transfusions of three units or greater (during and/or within 90 days after surgery)   | Dichotomous       | Yes<br>No   |
| Length of hospital stay   | Continuous (days) | -   |
| Use of non-nerve-sparing surgical technique   | Dichotomous       | Yes<br>No<br>Unknown                              |
| Positive surgical margins   | Nominal           | Yes – Extensive<br>Yes – Minimal<br>No<br>Unknown |

| <i>Study variable</i>  | <i>Format</i>                    | <i>Categories<br/>(if applicable)</i>               |
|--|----------------------------------|---|
| <b><u>Explanatory variables</u></b>  |                                  |   |
| Age at diagnosis   | Continuous<br>(years)            | -   |
| Comorbidity  | Continuous<br>(CIRS total score) | -   |
| Era of diagnosis   | Ordinal                          | 1990-1992<br>1993-1995<br>1996-1998                 |
| Socioeconomic status<br>(Median household income by Canadian census<br>enumeration area)                             | Ordinal                          | 1 (Lowest)<br>2<br>3<br>4<br>5 (Highest)            |
| Pre-treatment PSA  | Ordinal<br>(ng/mL)               | < 4<br>4-10<br>10-20<br>> = 20                      |
| Clinical T category  | Ordinal                          | T1a and T1b<br>T1c<br>T2a<br>T2b<br>T3 and above    |
| Biopsy Gleason score   | Ordinal                          | 2-4<br>5<br>6<br>7<br>8-9                           |
| <b><u>Hospital volume</u></b><br>(Number of radical prostatectomies performed per<br>month by the treating hospital) | Ordinal                          | < 1/m<br>< 2/m<br>< 3/m<br>< 4/m<br>< 7/m<br>>= 7/m |

### 3.12 *A priori* power calculations

During the thesis proposal stage, we evaluated whether it was possible to use the study sample of approximately 660 prostate cancer surgery patients available to us from the parent study to find statistically significant results. Using the software Power and Precision,<sup>95</sup> the lowest odds ratios required to achieve at least 80% statistical power in the analysis of several of the candidate quality indicators by hospital volume were calculated, based on adverse outcome rates found in the literature (Table 3.5).

**Table 3.5. *A priori* power calculations and lowest odds ratios required to achieve 80% statistical power**

| <i>Quality indicators</i>            | <i>Rates found in literature</i> | <b>Hospital volume:</b><br>Compare highest and lowest volume quartiles |              |
|--------------------------------------|----------------------------------|--|--------------|
|                                      |                                  | <i>Odds ratio</i>  | <i>Power</i> |
| <b>Blood loss during surgery</b>     | 7% <sup>96</sup>                 | 2.92   | 83%          |
| <b>Acute surgical complications</b>  | 2% <sup>2</sup>                  | 6.06   | 84%          |
| <b>Loss of continence (low)</b>      | 14% <sup>36</sup>                | 2.27   | 82%          |
| <b>Loss of continence (medium)</b>   | 47% <sup>36</sup>                | 1.92   | 83%          |
| <b>Loss of continence (high)</b>     | 66% <sup>36</sup>                | 2.06   | 81%          |
| <b>Positive surgical margins</b>     | 30% <sup>65</sup>                | 1.91   | 80%          |
| <b>Erectile dysfunction (low)</b>    | 12% <sup>36</sup>                | 2.32   | 80%          |
| <b>Erectile dysfunction (medium)</b> | 44% <sup>36</sup>                | 1.91   | 83%          |
| <b>Erectile dysfunction (high)</b>   | 76% <sup>36</sup>                | 2.32   | 80%          |

Highest and lowest hospital volume quartiles were compared using the highest quartile as the reference, as those hospitals were expected to have fewer adverse outcomes. Since reported rates of incontinence and erectile dysfunction vary widely,

three estimates (based on low, medium, and high rates) were made for each. We concluded that the study population was of sufficient size to obtain statistically significant results for most of the candidate quality indicators.

### **3.13 Statistical analysis strategy**

The statistical analysis began with bivariate analyses for explanatory variables and quality indicators by each of the six volume categories. The significance of the association between volume and each of the explanatory variables was tested using the Pearson chi-square test for categorical explanatory variables, and one-way ANOVA for continuous explanatory variables. Explanatory variables must have been marginally associated with volume ( $p \leq 0.20$ ) for examination in the regression models. The presence of a significant linear trend in quality indicator outcomes with increasing hospital volume was assessed using the Cochran-Armitage linear test for trend where the quality indicator is dichotomous,<sup>97</sup> and one-way ANOVA with linear contrast coefficients where the quality indicator is continuous. However, if the quality indicator has nominal outcomes, it was not possible to test for a linear trend. Instead, the Pearson chi-square test for association was used. If more than 20% of the cells had counts less than five, the Monte Carlo simulation for the exact test was substituted.<sup>98</sup> Bivariate tests for association and linear trend with two-sided p-values below 0.05 were considered significant.

For quality indicators where a significant linear trend was found by hospital volume, simple regression was performed with volume being the only predictor variable included. For regression analyses, hospital volume was formatted as five indicator

variables representing each of the five lowest volume categories, using the highest volume category as the reference for each. The type of regression modeling used depended on the format of each quality indicator. Logistic regression was used for dichotomous quality indicator outcomes, linear regression for continuous outcomes, and polytomous logistic regression<sup>99</sup> for nominal outcomes. The five odds ratios (or six parameter estimates in linear regression) resulting from the inclusion of the five dummy variables representing hospital volume into the simple regression models provides the results to achieve Objective 1 of the study.

Subsequently, Objective 2 of the study required assessing the impact of each explanatory variable on the crude odds ratios or parameter estimates. For each quality indicator, this was completed using a series of regression models with each explanatory variable added to separate models. This provided sets of odds ratios or parameter estimates that were univariately adjusted for one explanatory variable at a time.

To address the concern that the association between hospital volume and quality indicators was different across the time period covered in this study, for each quality indicator that was associated with volume, we investigated the presence of an interaction effect between era and volume. Era-specific odds ratios or parameter estimates were reported for quality indicators where significant interaction was found ( $p \leq 0.05$ ).

Lastly, parsimonious regression models were developed in order to report final, multivariate adjusted odds ratios or parameter estimates for each quality indicator. To determine which explanatory variables were included in the parsimonious models, we used an adaptation of the 10% rule suggested by Rothman and Greenland,<sup>100</sup> where only explanatory variables that caused at least a 10% change in at least one of the odds ratios

or parameter estimates between the crude and univariate adjusted models were included. However, era was forced into all final models regardless of whether the 10% criteria was met in order to account for the unequal distribution of era by volume category and because changes in other variables were expected over time.<sup>50,51,53</sup> Regression diagnostics were performed on the final parsimonious models in order to evaluate the model assumptions and to examine the effect of influential observations.

### **3.14 Ethical considerations**

The Queen's University Health Sciences Research Ethics Board approved the study on November 10, 2006. In addition, CCO approved the use of OCR data housed at CCE on December 7, 2006 (Please refer to Appendix III for ethics approval documentation). The study involved no direct patient contact and there was no mechanism by which this research activity could have influenced patient care for this study group, directly or indirectly. Therefore, the only concern was confidentiality. The primary risk to patients or physicians was that their privacy could have been infringed upon through unauthorized access to confidential, non-anonymized medical information. To address this concern, the electronic dataset provided to the student was stripped of unique identifiers, and could only be accessed via the student's secure, password-protected account. The data was stored on a safe, isolated system behind the Kingston General Hospital firewall. The student has also signed pledges of confidentiality for CCE and OCR.

## CHAPTER 4: RESULTS

### 4.1 Description of the study population

The final study population consisted of 645 subjects. The characteristics of this group are presented in Table 4.1. Most of the patients in our study sample were diagnosed later on in the study period, with 46.5% diagnosed in the era 1996-1998 and only 17.5% diagnosed in the era 1990-1992. Socioeconomic status was fairly evenly distributed between the median household income quintiles, but with a few more patients in the middle three quintiles and less in the highest and lowest income categories.

The study population generally falls within the appropriate age range for radical prostatectomy (43.1-80.4 years) and have the disease characteristics for which the procedure is recommended. The mean CIRS total score representing comorbid status was 5.1, with a range of 0-20, out of a possible 75 points, indicating that the study population does not have major comorbid illnesses that would counter-indicate radical prostatectomy. Almost all of the patients (98.3%) are younger than 75 years, the threshold stated in the CCO quality indicator recommendations.<sup>14</sup> Finally, only 9.9% of patients had a pre-treatment PSA level  $\geq 20$  ng/mL, only 1.4% had a clinical stage of T3 or worse, and only 6.8% had biopsy Gleason scores of 8 or 9. According to the British Columbia Cancer Agency, these clinical disease characteristics represent high-risk disease for which radical prostatectomy is generally not recommended.<sup>5</sup>

**Table 4.1: Description of the study population**

|   | <i>Mean</i> | <i>SD</i> |
|---|-------------|-----------|
| <b>Age at diagnosis (Years)</b>                       | 63.5        | 6.1       |
| <b>Comorbidity, (CIRS total score)</b>                | 5.1         | 3.3       |
| Missing ( <i>N</i> )                                  | 9           |           |
|   | <i>N</i>    | <i>%</i>  |
| <b>Era of diagnosis</b>                               |             |           |
| 1990-1992   | 113         | 17.5      |
| 1993-1995   | 232         | 36.0      |
| 1996-1998   | 300         | 46.5      |
| <b>Socioeconomic status (Median household income)</b> |             |           |
| Lowest quintile                                       | 106         | 16.6      |
| 2   | 135         | 21.2      |
| 3   | 136         | 21.4      |
| 4   | 150         | 23.6      |
| Highest quintile                                      | 110         | 17.3      |
| Missing   | 8           |           |
| <b>Pre-treatment PSA (ng/mL)</b>                      |             |           |
| <4  | 66          | 11.0      |
| 4 to 10   | 298         | 49.8      |
| 10 to 20  | 176         | 29.4      |
| >=20  | 59          | 9.9       |
| Missing   | 46          |           |
| <b>Clinical T category</b>                            |             |           |
| T1a and T1b   | 39          | 6.2       |
| T1c   | 215         | 34.3      |
| T2a   | 225         | 35.9      |
| T2b   | 139         | 22.2      |
| T3 and above  | 9           | 1.4       |
| Missing   | 18          |           |
| <b>Biopsy Gleason score</b>                           |             |           |
| 2 to 4  | 146         | 24.1      |
| 5   | 101         | 16.7      |
| 6   | 186         | 30.7      |
| 7   | 131         | 21.7      |
| 8 to 9  | 41          | 6.8       |
| Missing   | 40          |           |

## 4.2 Quality indicators by volume categories

The distribution of each quality indicator outcome for the six hospital volume categories is given in Table 4.2 and illustrated in Figure 4.1. We found that only 29 subjects had any evidence of acute complications during and/or up to 90 days after surgery in their charts. Although acute complications most frequently occurred in the lowest volume category (6.2%) and least frequently in the highest (2.5%), the acute complication rates for the hospital volume categories were statistically similar to each other overall (Cochran-Armitage linear test for trend using Monte Carlo estimates for the exact test,  $p = 0.41$ ). Due to the small number of events and lack of statistical significance, we did not investigate this quality indicator any further.

Significant linear trends were found for total blood transfusions of three units or greater during and/or within 90 days after surgery (Cochran-Armitage linear test for trend,  $p < .0001$ ), length of hospital stay (One-way ANOVA with linear contrast coefficients,  $p < .0001$ ), and use of non-nerve-sparing surgical technique (Cochran-Armitage linear test for trend,  $p < .0001$ ). As seen in Table 4.2 and Figure 4.1, the general trend seen is that outcomes tend to improve from lowest to highest volume groups. The rate of blood transfusions decreased sharply from  $< 1$  to  $< 3$  surgeries per month before declining at a slower rate in the higher volume categories. The exception to the decreasing trend is the  $< 7/m$  group, which at 15.3% was just slightly higher than the next lowest volume category of  $< 4$  surgeries per month (15.1%). The  $< 1/m$  and  $< 2/m$  groups had higher transfusion rates (46.0% and 30.5%, respectively) than the overall study population (24.5%). For length of hospital stay, two small anomalies in the overall decreasing trend were observed, where the hospital volumes of  $< 3/m$  and  $\geq 7/m$  both

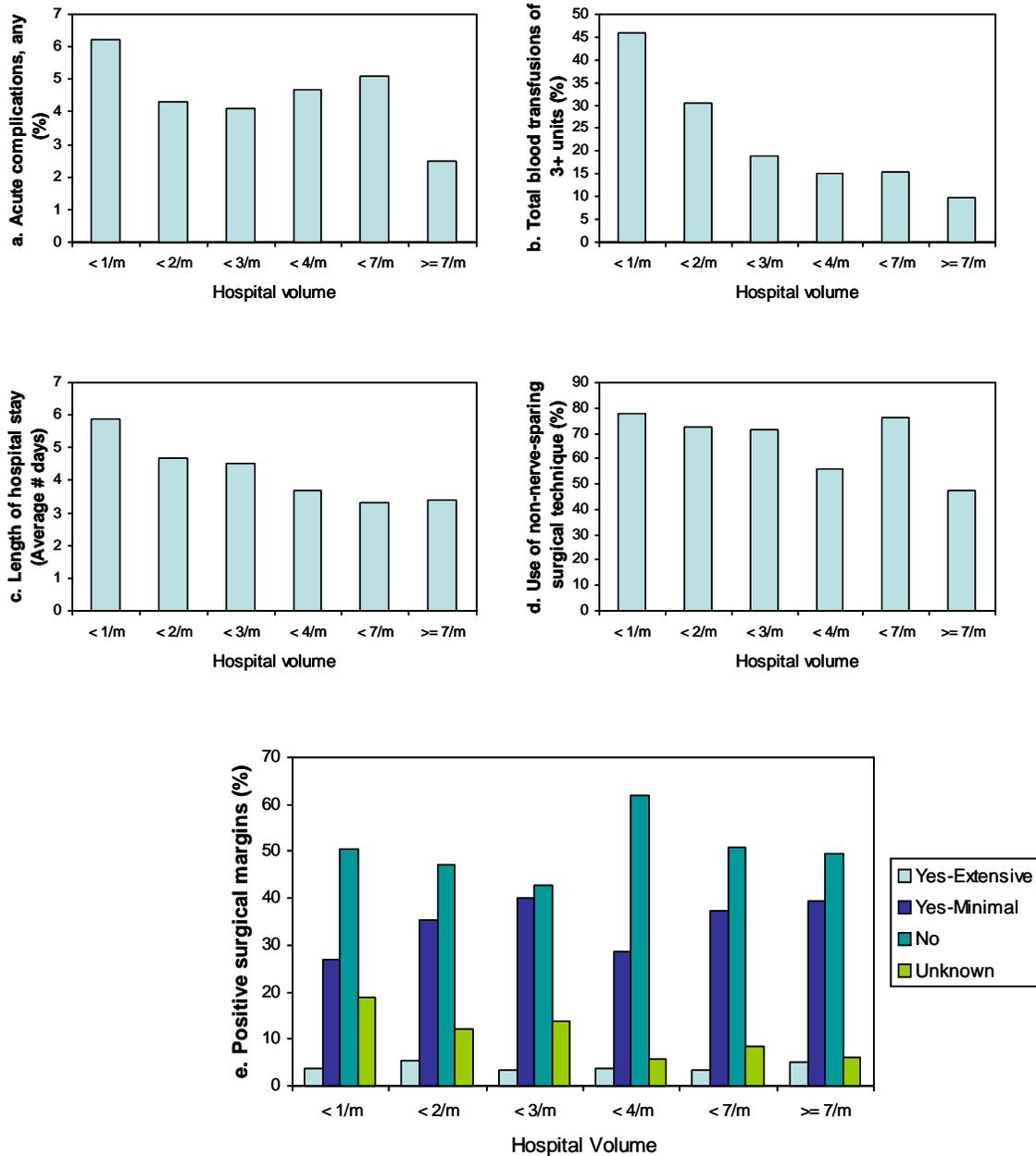
had slightly longer hospital stays than their next lowest volume categories. It can be seen in Figure 4.1 (c) that the mean length of stay was clustered into three groups ( $< 1/m$ ,  $< 2/m$  to  $< 4/m$ , and  $< 7/m$  to  $\geq 7/m$ ), as sharp drops were seen between  $< 1/m$  and  $< 2/m$ , and between  $< 3/m$  and  $< 4/m$ . The  $< 1/m$ ,  $< 2/m$ , and  $< 3/m$  groups had longer mean lengths of stay than the average for all subjects combined (8.6 days). For use of non-nerve-sparing surgical technique, the rate decreased slowly between  $< 1/m$  (77.6%) and  $< 3/m$  (71.2%) before dropping sharply to 55.8% in the  $< 4/m$  group. Unexpectedly, the rate then rose sharply to 76.0% in the next highest category before declining to the lowest rate of 47.4% in the highest volume group. All volume categories except for the  $< 4/m$  and  $\geq 7/m$  groups had worse rates than the total study population (67.5%).

Positive surgical margins were not statistically significantly associated with hospital volume (Pearson chi-square using Monte Carlo estimates for the exact test,  $p = 0.10$ ) and, as shown in Figure 4.1 (e), no linear trend by hospital volume was apparent among subjects for which surgical margin status was known. However, surgical margin status was more likely to be absent from the surgical pathology reports for men treated at the lower volume hospitals, indicating differences in the quality of reporting by hospital volume. Given the influence of unknown margin status on the results, it was not possible to proceed with further investigation this quality indicator.

**Table 4.2: Quality indicators by hospital volume categories**

|   | HOSPITAL VOLUME: Number of radical prostatectomies performed per month in treating hospital |           |             |           |             |           |             |           |             |           |             |           | p-value     |           |         |
|---|---|-----------|-------------|-----------|-------------|-----------|-------------|-----------|-------------|-----------|-------------|-----------|-------------|-----------|---------|
|   | All   |           | < 1/m       |           | < 2/m       |           | < 3/m       |           | < 4/m       |           | < 7/m       |           |             | ≥ 7/m     |         |
| <i>N</i>  | 645   |           | 113         |           | 164         |           | 122         |           | 106         |           | 59          |           | 81          |           |         |
|   | <i>N</i>  | %         | <i>N</i>    | %         | <i>N</i>    | %         | <i>N</i>    | %         | <i>N</i>    | %         | <i>N</i>    | %         | <i>N</i>    | %         |         |
| <b>Acute complications, any</b>                           | 29  | 4.5       | 7           | 6.2       | 7           | 4.3       | 5           | 4.1       | 5           | 4.7       | 3           | 5.1       | 2           | 2.5       | 0.41    |
| <b>Total blood transfusions of three units or greater</b> | 158   | 24.5      | 52          | 46.0      | 50          | 30.5      | 23          | 18.9      | 16          | 15.1      | 9           | 15.3      | 8           | 9.9       | < .0001 |
|   | <i>Mean</i>   | <i>SD</i> | <i>Mean</i> | <i>SD</i> | <i>Mean</i> | <i>SD</i> | <i>Mean</i> | <i>SD</i> | <i>Mean</i> | <i>SD</i> | <i>Mean</i> | <i>SD</i> | <i>Mean</i> | <i>SD</i> |         |
| <b>Length of hospital stay (Days)</b>                     | 8.6   | 4.8       | 12.0        | 5.9       | 8.7         | 4.7       | 9.0         | 4.5       | 6.8         | 3.7       | 6.3         | 3.3       | 7.5         | 12.0      | < .0001 |
| Missing ( <i>N</i> )                                      | 18  |           | 4           |           | 4           |           | 7           |           | 1           |           | 1           |           | 1           |           |         |
|   | <i>N</i>  | %         | <i>N</i>    | %         | <i>N</i>    | %         | <i>N</i>    | %         | <i>N</i>    | %         | <i>N</i>    | %         | <i>N</i>    | %         |         |
| <b>Use of non-nerve-sparing surgical technique</b>        | 404   | 67.5      | 83          | 77.6      | 109         | 72.2      | 84          | 71.2      | 53          | 55.8      | 38          | 76.0      | 37          | 47.4      | < .0001 |
| Missing   | 46  |           | 6           |           | 13          |           | 4           |           | 11          |           | 9           |           | 3           |           |         |
| <b>Positive surgical margins</b>                          |   |           |             |           |             |           |             |           |             |           |             |           |             |           |         |
| Yes-Extensive   | 27  | 4.2       | 27          | 4.2       | 4           | 3.6       | 9           | 5.5       | 4           | 3.3       | 4           | 3.8       | 2           | 3.4       | 0.10    |
| Yes-Minimal   | 221   | 34.3      | 221         | 34.3      | 30          | 27.0      | 58          | 35.4      | 49          | 40.2      | 30          | 28.6      | 22          | 37.3      | 0.53*   |
| No  | 320   | 49.6      | 320         | 49.6      | 56          | 50.5      | 77          | 47.0      | 52          | 42.6      | 65          | 61.9      | 30          | 50.9      |         |
| Unknown   | 74  | 11.5      | 74          | 11.5      | 21          | 18.9      | 20          | 12.2      | 17          | 13.9      | 6           | 5.7       | 5           | 8.5       |         |
| Missing   | 3   |           | 3           |           | 2           |           | 0           |           | 0           |           | 1           |           | 0           |           |         |

\* After excluding unknowns



**Figure 4.1: Distribution of quality indicator outcomes by hospital volume categories (Number of radical prostatectomies performed per month in treating hospital):**

- a) Proportion of patients with any evidence of acute complications during and/or up to 90 days following surgery
- b) Proportion of patients with total blood transfusions of three units or greater during and/or within 90 days after surgery
- c) Mean length of hospital stay, in days.
- d) Proportion of patients whose surgeons did not use a nerve-sparing surgical technique
- e) Proportion of patients with extensive surgical margin involvement, minimal surgical margin involvement, no positive surgical margins, and unknown margin status

### 4.3 Explanatory variables by hospital volume

The distribution of each explanatory variable for the six hospital volume categories is provided in Table 4.3. The mean age differences were marginally statistically significant (One-way ANOVA,  $p = 0.16$ ), ranging from 62.8 years in the highest volume category to 64.5 years in the  $< 1$  and  $< 7$  surgeries per month categories. Comorbidity, as measured by CIRS total score, was also marginally significant (One-way ANOVA,  $p = 0.15$ ), ranging from a mean total score of 4.6 in  $< 3/m$  hospitals to 5.5 in  $< 4/m$  hospitals. However, no clear volume trend was observed for either of these variables.

The era of diagnosis was highly associated with hospital volume (Pearson chi-square,  $p < .0001$ ), with patients more likely to be diagnosed in the last era of the study period (1996-1998) in the higher hospital volume categories. Only the  $< 1/m$  and  $< 3/m$  hospitals had their highest proportion of patients diagnosed in the other two eras (45.1% were diagnosed in 1990-1992 and 50.8% in 1993-1995 respectively). Meanwhile, socioeconomic status, as measured by area-level median household income, had a marginally significant relationship with hospital volume (Pearson chi-square,  $p = 0.13$ ).

The disease severity variables of pre-treatment PSA, clinical T category, and biopsy Gleason score were all associated with hospital volume although the significance level for PSA was marginal (Pearson chi-square,  $p = 0.09$ ). For pre-treatment PSA, the highest percentage of subjects with serum PSA levels  $\geq 20$  ng/mL were found in the  $< 1/m$  and  $< 3/m$  groups (14.4% and 15.2% respectively) and the lowest percentage was seen in the  $< 7/m$  group (3.6%). For clinical T category, a general decreasing trend was observed where a greater proportion of patients treated in the lower volume hospitals had

T1a and T1b stage disease (10.0% in the < 1/m category) as well as T2b stage disease (35.5% in the < 1/m group). Finally, the higher volume hospitals treated greater proportions of men with higher biopsy Gleason scores than the lower volume hospitals. For example, in the highest volume category 41.0% of the patients treated had a biopsy Gleason score of 7, whereas in the lowest category 36.5% of patients treated had scores of 2-4. On the other hand, the proportion of men with a biopsy Gleason score of 8-9 generally increased as hospital volume decreased.

From these observations, it appears that pre-treatment PSA and clinical T category demonstrate more severe cases being treated in the lower volume groups. However, the opposite trend was observed with biopsy Gleason score, although this might have been due to the way we assigned Gleason score when it was missing on the biopsy-based pathology reports. If we focus on the patients for which radical prostatectomy is generally not recommended (biopsy Gleason score above 7),<sup>5</sup> then inappropriately treated cases are more common in the lower volume categories, with the exception of the < 7/m group.

In conclusion, all of the explanatory variables had p-values of under 0.20 for association with hospital volume. Therefore, they were all subsequently considered in the regression analyses for the quality indicators blood transfusions of 3+ units, length of hospital stay, and use of non-nerve-sparing surgical technique.

**Table 4.3: Explanatory variables by hospital volume categories**

|   | HOSPITAL VOLUME: Number of radical prostatectomies performed per month in treating hospital |           |             |           |             |           |             |           |             |           |             |           | p-value |
|---|---|-----------|-------------|-----------|-------------|-----------|-------------|-----------|-------------|-----------|-------------|-----------|---------|
|   | < 1/m   |           | < 2/m       |           | < 3/m       |           | < 4/m       |           | < 7/m       |           | ≥ 7/m       |           |         |
| <i>N</i>  | 113   |           | 164         |           | 122         |           | 106         |           | 59          |           | 81          |           |         |
|   | <i>Mean</i>   | <i>SD</i> | <i>Mean</i> | <i>SD</i> | <i>Mean</i> | <i>SD</i> | <i>Mean</i> | <i>SD</i> | <i>Mean</i> | <i>SD</i> | <i>Mean</i> | <i>SD</i> |         |
| <b>Age at diagnosis (Years)</b>                       | 64.5  | 6.7       | 63.7        | 5.7       | 62.9        | 6.5       | 62.9        | 6.6       | 64.4        | 5.2       | 62.8        | 5.4       | 0.16    |
| <b>Comorbidity (CIRS total score)</b>                 | 5.1   | 3.3       | 5.5         | 3.6       | 4.6         | 2.7       | 5.5         | 3.3       | 4.7         | 2.8       | 4.7         | 3.5       | 0.15    |
| Missing ( <i>N</i> )                                  | 3   |           | 1           |           | 1           |           | 3           |           | 0           |           | 1           |           |         |
|   | <i>N</i>  | %         | <i>N</i>    | %         | <i>N</i>    | %         | <i>N</i>    | %         | <i>N</i>    | %         | <i>N</i>    | %         |         |
| <b>Era of diagnosis</b>                               |   |           |             |           |             |           |             |           |             |           |             |           |         |
| 1990-1992   | 51  | 45.1      | 20          | 12.2      | 22          | 18.0      | 3           | 2.8       | 6           | 10.2      | 11          | 13.6      | < .0001 |
| 1993-1995   | 43  | 38.1      | 53          | 32.3      | 62          | 50.8      | 30          | 28.3      | 25          | 42.4      | 19          | 23.5      |         |
| 1996-1998   | 19  | 16.8      | 91          | 55.5      | 38          | 31.2      | 73          | 68.9      | 28          | 47.5      | 51          | 63.0      |         |
| <b>Socioeconomic status (Median household income)</b> |   |           |             |           |             |           |             |           |             |           |             |           |         |
| Lowest quintile                                       | 22  | 19.6      | 22          | 13.5      | 26          | 21.9      | 13          | 12.4      | 10          | 17.0      | 13          | 16.5      | 0.13    |
| 2   | 23  | 20.5      | 40          | 24.5      | 23          | 19.3      | 17          | 16.2      | 12          | 20.3      | 20          | 25.3      |         |
| 3   | 34  | 30.4      | 35          | 21.5      | 24          | 20.2      | 20          | 19.1      | 9           | 15.3      | 14          | 17.7      |         |
| 4   | 23  | 20.5      | 38          | 23.3      | 26          | 21.9      | 30          | 28.6      | 19          | 32.2      | 14          | 17.7      |         |
| Highest quintile                                      | 10  | 8.9       | 28          | 17.2      | 20          | 16.8      | 25          | 23.8      | 9           | 15.3      | 18          | 22.8      |         |
| Missing   | 1   |           | 1           |           | 3           |           | 1           |           | 0           |           | 2           |           |         |

(cont'd...)

(..cont'd)

|                                  | <b>HOSPITAL VOLUME: Number of radical prostatectomies performed per month in treating hospital</b> |      |                 |      |                 |      |                 |      |                 |      |                  |      | <b>p-value</b> |
|----------------------------------|--|------|-----------------|------|-----------------|------|-----------------|------|-----------------|------|------------------|------|----------------|
|                                  | <b>&lt; 1/m</b>  |      | <b>&lt; 2/m</b> |      | <b>&lt; 3/m</b> |      | <b>&lt; 4/m</b> |      | <b>&lt; 7/m</b> |      | <b>&gt;= 7/m</b> |      |                |
|                                  | <i>N</i>   | %    | <i>N</i>        | %    | <i>N</i>        | %    | <i>N</i>        | %    | <i>N</i>        | %    | <i>N</i>         | %    |                |
| <b>Pre-treatment PSA (ng/mL)</b> |  |      |                 |      |                 |      |                 |      |                 |      |                  |      |                |
| <4                               | 9  | 9.3  | 21              | 13.4 | 11              | 9.8  | 12              | 11.9 | 5               | 8.9  | 8                | 10.5 | 0.09           |
| 4 to 10                          | 39   | 40.2 | 77              | 49.0 | 50              | 44.6 | 60              | 59.4 | 27              | 48.2 | 45               | 59.2 |                |
| 10 to 20                         | 35   | 36.1 | 46              | 29.3 | 34              | 30.4 | 22              | 21.8 | 22              | 39.3 | 17               | 22.4 |                |
| >=20                             | 14   | 14.4 | 13              | 8.3  | 17              | 15.2 | 7               | 6.9  | 2               | 3.6  | 6                | 7.9  |                |
| Missing                          | 16   |      | 7               |      | 10              |      | 5               |      | 3               |      | 5                |      |                |
| <b>Clinical T category</b>       |  |      |                 |      |                 |      |                 |      |                 |      |                  |      |                |
| T1a and T1b                      | 11   | 10.0 | 12              | 7.5  | 9               | 7.6  | 4               | 4.0  | 1               | 1.7  | 2                | 2.5  | 0.0004         |
| T1c                              | 19   | 17.3 | 65              | 40.4 | 40              | 33.6 | 40              | 40.0 | 19              | 32.8 | 32               | 40.5 |                |
| T2a                              | 39   | 35.5 | 47              | 29.2 | 36              | 30.3 | 41              | 41.0 | 31              | 53.5 | 31               | 39.2 |                |
| T2b                              | 39   | 35.5 | 35              | 21.7 | 32              | 26.9 | 14              | 14.0 | 6               | 10.3 | 13               | 16.5 |                |
| T3 and above                     | 2  | 1.8  | 2               | 1.2  | 2               | 1.7  | 1               | 1.0  | 1               | 1.7  | 1                | 1.3  |                |
| Missing                          | 3  |      | 3               |      | 3               |      | 6               |      | 1               |      | 2                |      |                |
| <b>Biopsy Gleason score</b>      |  |      |                 |      |                 |      |                 |      |                 |      |                  |      |                |
| 2 to 4                           | 39   | 36.5 | 37              | 24.3 | 35              | 31.0 | 10              | 10.2 | 16              | 28.1 | 9                | 11.5 | < .0001        |
| 5                                | 11   | 10.3 | 26              | 17.1 | 17              | 15.0 | 22              | 22.5 | 17              | 29.8 | 8                | 10.3 |                |
| 6                                | 35   | 32.7 | 41              | 27.0 | 37              | 32.7 | 33              | 33.7 | 15              | 26.3 | 25               | 32.1 |                |
| 7                                | 13   | 12.2 | 37              | 24.3 | 17              | 15.0 | 29              | 29.6 | 3               | 5.3  | 32               | 41.0 |                |
| 8 to 9                           | 9  | 8.4  | 11              | 7.2  | 7               | 6.2  | 4               | 4.1  | 6               | 10.5 | 4                | 5.1  |                |
| Missing                          | 6  |      | 12              |      | 9               |      | 8               |      | 2               |      | 3                |      |                |

## 4.4 Multivariate analysis

### 4.4.1 Total blood transfusions of three units or greater

Logistic regression was used to perform multivariate analyses on the quality indicator ‘total blood transfusions of three units or greater.’ Table 4.4 gives the results from the logistic regression modeling, where the highest volume category ( $\geq 7$  surgeries performed per month in the treating hospital) was used as the reference. For all of the models shown, the two lowest volume categories ( $< 1/m$  and  $< 2/m$ ) had significantly higher odds of receiving blood transfusions when compared to highest volume group.

The crude odds ratios showed a generally decreasing trend, where the odds of receiving blood transfusions generally declined as volume increased. The exception was that the  $< 7/m$  group had a similar but slightly higher odds ratio than the next lowest volume group of  $< 4/m$  (1.64 and 1.62, respectively).

The odds ratios shown in the second half of Table 4.4 were used to determine which explanatory variables should be included in the final model. Only explanatory variables for which there was at least a 10% change in at least one of the five odds ratios when compared to the crude estimates were included. For blood transfusions, only age at diagnosis and socioeconomic status did not fit this criteria. For comorbidity, the odds ratios increased in some volume groups and decreased in others when compared to the crude estimates. For era of diagnosis, the odds ratios all decreased except for the  $< 4/m$  category, which became higher than the adjacent groups. The greatest declines were in the  $< 1/m$  (29%) and  $< 3/m$  (17%) categories. For pre-treatment PSA, all of the estimates were lower than the crude rates and no there were no exceptions to the decreasing trend by volume. The largest decrease was found in the  $< 1/m$  (12%) group.

Since the many small cell sizes caused model convergence problems, the five clinical T category categories were collapsed into two (T1 and T2a vs. T2b and above). We used this dichotomous format for all of the quality indicator regression analyses. Here, all of the ORs increased from the crude estimates. The largest increases were found in the categories < 7/m (17%), followed by < 3/m (13%) and < 4/m (11%). Finally, for biopsy Gleason score, all of the ORs decreased from the crude estimates, with the largest decreases found in categories < 3/m (13%) and < 2/m and < 4/m (both 12%).

The final adjusted parsimonious model for blood transfusions had the same pattern as the crude ORs, where the < 7/m group had a higher OR, at 1.80 (95% CI: 0.59, 5.45), than the next lowest category, although this was not a significant difference. The ORs for all of the other volume categories decreased, with the biggest change found in the categories < 1/m (21%) and < 3/m (16%). Overall, the 95% confidence intervals for the < 3/m, < 4/m, and < 7/m groups indicate that these three insignificant adjusted odds ratios are essentially equal.

The likelihood ratio testing for interaction between hospital volume and era of diagnosis was not significant ( $p = 0.30$ ). The Hosmer-Lemeshow goodness-of-fit test ( $p = 0.87$ ) indicated adequate model fit. Furthermore, logistic regression diagnostics recommended by Hosmer and Lemeshow<sup>99</sup> did not identify any particularly ill-fitted or influential observations (Appendix IV, Tables IV.1-IV.3).

**Table 4.4: Logistic regression analysis on total blood transfusions of three units or greater by hospital volume**

|   | <b>HOSPITAL VOLUME: Number of radical prostatectomies performed per month in treating hospital</b> |               |                 |               |                 |               |                 |               |                 |               |                        |               |
|---|--|---------------|-----------------|---------------|-----------------|---------------|-----------------|---------------|-----------------|---------------|------------------------|---------------|
|   | <b>&lt; 1/m</b>  |               | <b>&lt; 2/m</b> |               | <b>&lt; 3/m</b> |               | <b>&lt; 4/m</b> |               | <b>&lt; 7/m</b> |               | <b>&gt;= 7/m (REF)</b> |               |
|   | <i>OR</i>  | <i>95% CI</i> | <i>OR</i>       | <i>95% CI</i> | <i>OR</i>       | <i>95% CI</i> | <i>OR</i>       | <i>95% CI</i> | <i>OR</i>       | <i>95% CI</i> | <i>OR</i>              | <i>95% CI</i> |
| <b>Crude</b><br>(Volume only)                                 | <b>7.78</b>  | 3.43, 17.62   | <b>4.00</b>     | 1.79, 8.92    | <b>2.12</b>     | 0.90, 5.01    | <b>1.62</b>     | 0.66, 4.00    | <b>1.64</b>     | 0.59, 4.54    | <b>1.00</b>            | -             |
| <b>Adjusted</b><br>(Volume + selected explanatory variables*) | <b>6.14</b>  | 2.41, 15.63   | <b>3.63</b>     | 1.49, 8.82    | <b>1.79</b>     | 0.68, 4.69    | <b>1.45</b>     | 0.52, 4.02    | <b>1.80</b>     | 0.59, 5.45    | <b>1.00</b>            | -             |
| <b>Volume + Age</b>   | <b>7.65</b>  | 3.37, 17.37   | <b>3.97</b>     | 1.78, 8.85    | <b>2.12</b>     | 0.90, 5.00    | <b>1.62</b>     | 0.66, 4.00    | <b>1.62</b>     | 0.58, 4.48    | <b>1.00</b>            | -             |
| <b>Volume + Comorbidity</b>                                   | <b>8.01</b>  | 3.52, 18.24   | <b>3.84</b>     | 1.72, 8.59    | <b>2.15</b>     | 0.91, 5.10    | <b>1.47</b>     | 0.59, 3.66    | <b>1.64</b>     | 0.59, 4.54    | <b>1.00</b>            | -             |
| <b>Volume + Era</b>   | <b>5.54</b>  | 2.39, 12.84   | <b>3.97</b>     | 1.77, 8.91    | <b>1.76</b>     | 0.73, 4.21    | <b>1.79</b>     | 0.72, 4.46    | <b>1.54</b>     | 0.55, 4.29    | <b>1.00</b>            | -             |
| <b>Volume + SES</b>   | <b>7.51</b>  | 3.29, 17.16   | <b>4.00</b>     | 1.79, 8.95    | <b>1.93</b>     | 0.81, 4.62    | <b>1.68</b>     | 0.68, 4.18    | <b>1.66</b>     | 0.60, 4.61    | <b>1.00</b>            | -             |
| <b>Volume + PSA</b>   | <b>6.85</b>  | 2.96, 15.88   | <b>3.59</b>     | 1.60, 8.09    | <b>2.08</b>     | 0.87, 4.95    | <b>1.61</b>     | 0.65, 4.00    | <b>1.47</b>     | 0.53, 4.12    | <b>1.00</b>            | -             |
| <b>Volume + T category</b>                                    | <b>8.45</b>  | 3.56, 20.07   | <b>4.44</b>     | 1.91, 10.36   | <b>2.39</b>     | 0.97, 5.88    | <b>1.69</b>     | 0.65, 4.42    | <b>1.92</b>     | 0.67, 5.52    | <b>1.00</b>            | -             |
| <b>Volume + Gleason score</b>                                 | <b>7.66</b>  | 3.30, 17.77   | <b>3.52</b>     | 1.55, 8.00    | <b>1.84</b>     | 0.76, 4.48    | <b>1.42</b>     | 0.56, 3.61    | <b>1.63</b>     | 0.58, 4.62    | <b>1.00</b>            | -             |

\* Adjusted for: Comorbidity, era of diagnosis, pre-treatment PSA, clinical T category, and biopsy Gleason score

#### 4.4.2 Length of hospital stay

Linear regression was initially used to analyze the quality indicator length of hospital stay. Crude, univariately adjusted, and parsimonious linear regression models were developed, and residual analysis was carried out on the final parsimonious model. However, a significant interaction ( $p = 0.03$ ) between hospital volume and era of diagnosis was found using the extra sums of squares principle. Thus, the reported results for length of hospital stay are stratified by era. Due to small sample size in the era-specific strata, especially in the earliest period (1990-1992), the effects of the explanatory variables were not examined. We report, therefore, mean length of hospital stay (in days) rather than era-specific crude linear regression coefficients (Table 4.4). In each era, the one-way ANOVA test with linear contrast coefficients was used to detect the significance of a linear trend by hospital volume.

The length of stay generally decreased over the study period in each volume category. A trend of decreasing length of stay with increasing hospital volume was evident in each era, but became more apparent with time. In the earliest era (1990-1992), the mean length of stay decreased to a low of 7.7 days in the  $< 4$  surgeries per month group before increasing again back up to 11.5 days in the highest volume group. There was no statistically significant linear trend in this era ( $p = 0.13$ ). From 1993 to 1995, the mean length of stay declined to a low of 5.6 days in the second highest volume category before jumping back up to 10.2 days in the highest volume group, and a significant linear trend was found ( $p = 0.003$ ). In the latest era (1996-1998), the test for trend was also significant ( $p = 0.007$ ). Here, mean length of hospital stay decreased from 7.6 days in the lowest volume category to 5.7 days in the  $\geq 7/m$  group.

**Table 4.4: Mean length of hospital stay (in days) by hospital volume, stratified by era of diagnosis**

|                   | <b>HOSPITAL VOLUME: Number of radical prostatectomies performed per month in treating hospital</b> |           |                 |           |                 |           |                 |           |                 |           |                  |           | <b>p-value*</b> |
|-------------------|--|-----------|-----------------|-----------|-----------------|-----------|-----------------|-----------|-----------------|-----------|------------------|-----------|-----------------|
|                   | <b>&lt; 1/m</b>  |           | <b>&lt; 2/m</b> |           | <b>&lt; 3/m</b> |           | <b>&lt; 4/m</b> |           | <b>&lt; 7/m</b> |           | <b>&gt;= 7/m</b> |           |                 |
|                   | <i>Mean</i>  | <i>SD</i> | <i>Mean</i>     | <i>SD</i> | <i>Mean</i>     | <i>SD</i> | <i>Mean</i>     | <i>SD</i> | <i>Mean</i>     | <i>SD</i> | <i>Mean</i>      | <i>SD</i> |                 |
| <b>All eras**</b> | <b>12.0</b>  | 5.9       | <b>8.7</b>      | 4.7       | <b>9.0</b>      | 4.5       | <b>6.8</b>      | 3.7       | <b>6.3</b>      | 3.3       | <b>7.5</b>       | 3.4       | < .0001         |
| <b>1990-1992</b>  | <b>14.2</b>  | 6.6       | <b>11.6</b>     | 4.4       | <b>10.4</b>     | 3.3       | <b>7.7</b>      | 0.6       | <b>11.0</b>     | 0.9       | <b>11.5</b>      | 2.5       | 0.13            |
| <b>1993-1995</b>  | <b>11.5</b>  | 5.1       | <b>10.2</b>     | 5.2       | <b>9.4</b>      | 3.2       | <b>9.3</b>      | 4.5       | <b>5.6</b>      | 3.0       | <b>10.2</b>      | 2.5       | 0.003           |
| <b>1996-1998</b>  | <b>7.6</b>   | 2.6       | <b>7.1</b>      | 3.7       | <b>7.6</b>      | 6.2       | <b>5.8</b>      | 2.7       | <b>6.0</b>      | 3.1       | <b>5.7</b>       | 2.3       | 0.007           |

\* One-way ANOVA test with linear contrast coefficients

\*\* As reported in Table 4.2 and previously described

#### 4.4.3 Use of non-nerve-sparing surgical technique

Logistic regression was used to perform multivariate analyses on the quality indicator 'use of non-nerve-sparing surgical technique,' and Table 4.6 gives the results. For all of the models shown, only the < 4 surgeries per month group did not have significantly higher odds of receiving non-nerve-sparing surgery when compared to highest volume reference group.

With the exception of the < 7/m group, the crude odds ratios showed a decreasing trend in the use of non-nerve sparing surgery as volume increased. For the < 7/m category, the odds ratio of 3.51 (1.60, 7.71) was much higher than the next lowest category of < 4/m.

Five of the seven explanatory variables did not change any of the odds ratios (second half of Table 4.6) more than 10% and thus were not included in the final parsimonious model. The excluded explanatory variables included age at diagnosis (ranging from 3% increase to 6% decrease in odds ratios when compared to the crude estimates), comorbidity (from 4% increase to 2% decrease), socioeconomic status (2% increase to 4% decrease), clinical T category (3% increase to 4% decrease), and era of diagnosis (6% increase to 3% decrease). Although era of diagnosis should have been excluded from the final model, it was forced into the multivariate regression due to its potential importance (see Section 3.12). Along with era of diagnosis, pre-treatment PSA and biopsy Gleason score were included in the final model. For pre-treatment PSA, all of the ORs declined when compared to the crude ORs, with the exception of the < 7/m group, where the odds ratio increased by 11%. The greatest decreases in the ORs were found in the < 3/m (24%) and < 1/m (16%) categories. For biopsy Gleason score, the

odds ratios increased in some volume groups and decreased in others. The largest increase was 15% (< 1/m) and the largest decrease was 12% (< 4/m).

The parsimonious model for the use of non-nerve-sparing surgery had the same pattern as the crude ORs, where the < 7/m group had a much higher OR, at 4.72 (95% CI: 1.92, 11.59), than the next lowest category. This represented a 34% increase from the crude estimate. The ORs for the other volume categories both increased and decreased, with the next biggest changes found in the categories < 4/m (17% decrease) and < 3/m (11% decrease).

The likelihood ratio test was used to test for interaction between hospital volume and era of diagnosis, and a marginally significant interaction was found ( $p = 0.07$ ). The Hosmer-Lemeshow goodness-of-fit test ( $p = 0.70$ ) indicated adequate model fit. Furthermore, logistic regression diagnostics recommended by Hosmer and Lemeshow<sup>99</sup> did not identify any particularly ill-fitted or influential observations (Tables IV.4-IV.6).

**Table 4.6: Logistic regression analysis on use of non-nerve-sparing surgical technique by hospital volume**

|   | <b>HOSPITAL VOLUME: Number of radical prostatectomies performed per month in treating hospital</b> |               |                 |               |                 |               |                 |               |                 |               |                        |               |
|---|--|---------------|-----------------|---------------|-----------------|---------------|-----------------|---------------|-----------------|---------------|------------------------|---------------|
|   | <b>&lt; 1/m</b>  |               | <b>&lt; 2/m</b> |               | <b>&lt; 3/m</b> |               | <b>&lt; 4/m</b> |               | <b>&lt; 7/m</b> |               | <b>&gt;= 7/m (REF)</b> |               |
|   | <i>OR</i>  | <i>95% CI</i> | <i>OR</i>       | <i>95% CI</i> | <i>OR</i>       | <i>95% CI</i> | <i>OR</i>       | <i>95% CI</i> | <i>OR</i>       | <i>95% CI</i> | <i>OR</i>              | <i>95% CI</i> |
| <b>Crude</b><br>(Volume only)                                 | <b>3.83</b>  | 2.03, 7.24    | <b>2.88</b>     | 1.63, 5.08    | <b>2.74</b>     | 1.51, 4.97    | <b>1.40</b>     | 0.77, 2.55    | <b>3.51</b>     | 1.60, 7.71    | <b>1.00</b>            | -             |
| <b>Adjusted</b><br>(Volume + selected explanatory variables*) | <b>4.20</b>  | 1.99, 8.89    | <b>3.01</b>     | 1.61, 5.62    | <b>2.43</b>     | 1.24, 4.76    | <b>1.17</b>     | 0.61, 2.24    | <b>4.72</b>     | 1.92, 11.59   | <b>1.00</b>            | -             |
| <b>Volume + Age</b>   | <b>3.59</b>  | 1.89, 6.83    | <b>2.79</b>     | 1.57, 4.96    | <b>2.82</b>     | 1.54, 5.16    | <b>1.40</b>     | 0.76, 2.57    | <b>3.36</b>     | 1.52, 7.41    | <b>1.00</b>            | -             |
| <b>Volume + Comorbidity</b>                                   | <b>3.75</b>  | 1.99, 7.10    | <b>2.87</b>     | 1.62, 5.09    | <b>2.82</b>     | 1.55, 5.14    | <b>1.45</b>     | 0.79, 2.67    | <b>3.51</b>     | 1.60, 7.72    | <b>1.00</b>            | -             |
| <b>Volume + Era</b>   | <b>3.70</b>  | 1.90, 7.18    | <b>2.93</b>     | 1.66, 5.20    | <b>2.83</b>     | 1.54, 5.22    | <b>1.44</b>     | 0.79, 2.64    | <b>3.71</b>     | 1.68, 8.22    | <b>1.00</b>            | -             |
| <b>Volume + SES</b>   | <b>3.75</b>  | 1.96, 7.19    | <b>2.84</b>     | 1.59, 5.05    | <b>2.68</b>     | 1.46, 4.93    | <b>1.42</b>     | 0.77, 2.63    | <b>3.38</b>     | 1.52, 7.49    | <b>1.00</b>            | -             |
| <b>Volume + PSA</b>   | <b>3.22</b>  | 1.63, 6.37    | <b>2.74</b>     | 1.51, 4.99    | <b>2.07</b>     | 1.11, 3.87    | <b>1.29</b>     | 0.69, 2.42    | <b>3.88</b>     | 1.66, 9.05    | <b>1.00</b>            | -             |
| <b>Volume + T category</b>                                    | <b>3.94</b>  | 2.06, 7.55    | <b>2.90</b>     | 1.63, 5.16    | <b>2.64</b>     | 1.44, 4.83    | <b>1.36</b>     | 0.74, 2.52    | <b>3.53</b>     | 1.60, 7.77    | <b>1.00</b>            | -             |
| <b>Volume + Gleason score</b>                                 | <b>4.42</b>  | 2.26, 8.67    | <b>2.98</b>     | 1.64, 5.40    | <b>2.86</b>     | 1.52, 5.36    | <b>1.24</b>     | 0.66, 2.31    | <b>3.71</b>     | 1.64, 8.40    | <b>1.00</b>            | -             |

\* Adjusted for: Era of diagnosis, pre-treatment PSA, and biopsy Gleason score

## CHAPTER 5: DISCUSSION

### 5.1 Summary of study design

We performed a retrospective cohort study on prostate cancer patients diagnosed between 1990 and 1998 in Ontario, who were treated by radical prostatectomy with curative intent within six months of diagnosis. We sought to test the convergent construct validity of quality indicators recommended by Cancer Care Ontario and several other indicators that were also related to surgical skill by testing them against hospital prostatectomy volume, a surrogate measure of surgeon experience and an accepted structural indicator of quality surgical care. We expected that if higher volume hospitals have better outcomes as demonstrated by the quality indicators of interest, then we will have demonstrated the convergent validity of the candidate indicator in measuring the underlying concept of surgical quality. Using a database of high-quality medical chart data abstracted for a larger study, the objectives were to assess variations in quality indicators by hospital volume, and to investigate whether selected explanatory variables account for some of this variation.

### 5.2 Overview of findings

Even using high-quality chart data, we could not evaluate all of the quality indicators we had wished to study *a priori*. For several reasons, we could not assess the post-treatment outcomes acute surgical complications, urinary incontinence, erectile dysfunction, undetectable PSA level at first follow-up, positive surgical margins, and biochemical disease-free survival by hospital volume. We demonstrated convergent

construct validity for blood transfusions of 3 units or greater, length of hospital stay, and use of non-nerve-sparing surgical technique. In general, worse outcomes were apparent with decreasing hospital volume, both before and after adjusting for the effect of explanatory variables, providing evidence that these three may be valid surgical quality indicators that could be used for quality assessment in the future.

### **5.3 Feasibility of using chart data to evaluate quality indicators**

During the data processing and exploration process, we were able to determine the feasibility of obtaining information on each *a priori* listed quality indicator from the medical chart abstraction. While some of these indicators were evaluated for feasibility in follow-up studies for the RAND indicators<sup>15-17</sup>, this study provides new information on the feasibility of using the remaining quality indicators and for most, it is also the first time that their feasibility has been tested in the Ontario population.

It was not feasible to evaluate undetectable PSA level at first follow-up after surgery and biochemical disease-free survival using our data source, due to missing follow-up information and variable undetectable thresholds in PSA assays. Obtaining such follow-up information in a surgical series would require reviewing multiple data sources beyond the treating institution's chart. Also, the undetectable PSA threshold varied widely depending on which assay was used, which would have caused considerable outcome misclassification.

Quality councils might want to consider consistently capturing PSA test results shortly after surgery as a measurement of biochemical recurrence. The 5, 10, and 15-year time frames recommended by both the RAND and CCO expert panels to assess

biochemical disease-free survival<sup>9,14</sup> make it less useful for quality improvement purposes, as practice may have changed by the time these results were available, making the findings less relevant to current practice. However, if PSA test results can be recorded consistently throughout the province shortly after surgery, then undetectable PSA may be a good quality indicator for disease control and risk of recurrence that can also be evaluated in a very proximate time frame to the treatment date that is necessary for change.

We also found that the late morbidities of urinary incontinence and erectile dysfunction post-treatment could not be feasibly assessed in our chart data due to missing follow-up and/or lack of detail. This finding is consistent with the RAND follow-up studies that patient satisfaction with their urinary and erectile status after surgery could be assessed using validated patient questionnaires but not through administrative documents or charts.<sup>15,16</sup> According to Miller *et al.*, the systematic use of validated survey instruments for preoperative and postoperative urinary, sexual, and bowel function are necessary to truly measure and compare the occurrence of these important quality-of-life outcomes following radical prostatectomy.<sup>15</sup> Such an approach would address the data quality issues accounting for the wide variability in late morbidity rates in the literature.<sup>3,6,35,36,60,61,63</sup> Since incontinence and erectile dysfunction are considered very important outcomes of prostate surgery and reflect quality of surgical care,<sup>2,36,37,55,56,59-61</sup> it would be extremely useful for Cancer Care Ontario and other bodies responsible for cancer care quality to consider moving towards the province wide application of a validated survey instrument to evaluate these indicators. With adequate control for

confounders, this could provide relatively timely information for incorporation into hospital and LHIN quality improvement processes.

It was feasible to assess acute surgical complications (including 30-day mortality) in our chart review, supporting the conclusion reached by Miller *et al.*<sup>15</sup>. In bivariate analysis, the rates ranged from 6.2-2.5% between the lowest and highest volume categories. Although there was no statistically significant linear trend, this may have been due to a lack of study power rather than the absence of a real trend by volume. We had 40% power to detect this difference.<sup>95</sup> Since the rates were too low for us to conduct multivariate analysis, it is possible that differences in comorbidity, disease severity, and other covariates might explain some or all of this volume trend and thus, it was not possible to determine the construct validity of acute surgical complications. Cancer Care Ontario is currently reporting 30-day mortality rates by LHIN (0-0.95% in 2002-2004).<sup>49</sup> With such low rates and without control for covariates, it is questionable whether 30-day mortality and acute surgical complication rates are very useful indicators of surgical care, especially if applied for men currently receiving radical prostatectomy, for whom these rates are likely even lower than those we studied, as surgical quality should have improved over the past decade.

We were able to access pathology reports for almost all of the patients in our study population, and therefore it was feasible to evaluate positive surgical margin status if it was stated in the report. Among the patients for whom margin status was known, no linear trend by volume was observed, although this finding was likely affected by the 11.5% of patients whose margin status was not stated on their pathology reports. Because of this high proportion of unknowns, it was not possible to determine the

construct validity of positive surgical margins. It is interesting to note that lower volume hospitals were more likely have missing documentation of margin status on pathology reports, another quality indicator that was actually found to be feasible and is currently reported by CCO.<sup>17,49</sup> While the reporting of margin status is not related to quality of surgical care, our findings further validate its use by CCO as an indicator of overall quality of care. These findings also demonstrate the importance of consistent pathology reporting to allow for quality assessment of the important post-surgical outcome positive surgical margins.

We demonstrated feasibility for use of blood transfusions, length of stay, and non-nerve-sparing surgery as quality indicators. This evidence supports others' findings for blood transfusions,<sup>15,16</sup> but the feasibility of the other two quality indicators have not been tested in previous studies. These three indicators are further discussed in the following section.

#### **5.4 Volume-quality indicator associations**

We demonstrated the convergent construct validity of blood transfusions, length of stay, and use of non-nerve-sparing surgery, as we found that lower volume hospitals generally performed worse when compared to the highest volume institutions of the province. These findings were consistent with U.S. studies that examined blood transfusions and length of stay, and provide new knowledge on the use of non-nerve-sparing surgery in relation to surgical volume.

An exception to these general trends was observed in the < 7 surgeries per month volume category. Patients in this group sometimes fared more poorly than expected, especially in the use of non-nerve-sparing surgery. This exception likely reflects localized practice and might also demonstrate the effect of volume misclassification, as almost half of the patients in this category were treated at a group of hospitals that amalgamated during the study period. It is possible that these amalgamated hospitals (which would be in a higher volume category than the constituent hospitals previously were) continued to operate as separate institutions. Thus, the average volume of this category was likely overestimated, thereby explaining their unexpectedly poor results.

#### **5.4.1 Blood transfusions of 3 units or greater**

Patients in the two lowest volume categories had higher risks of receiving blood transfusions of 3 units or greater during and/or after surgery than those in the highest volume institutions of the province. Adjustment for explanatory variables clearly accounted for some of the variation seen in the crude odds ratios.

These findings are largely consistent with research conducted in the U.S. and in Europe. In a multi-institution study conducted over a similar time period of 1994-2000, 1123 consecutive RRP cases treated at the University of Michigan,<sup>41</sup> surgeons who performed  $\leq 15$  surgeries per year had an 8.6 times increased odds of their patients requiring any blood transfusion compared to those who performed  $> 15$  surgeries. Whereas we only counted transfusions of 3+ units or greater, these investigators counted any blood transfusion regardless of the amount, adjusting for banked autologous blood. Despite this difference in measurement, their finding was of similar magnitude to both

our crude and adjusted odds ratios (7.78 and 6.14 respectively) for our corresponding volume category of < 1/m. However, our findings were adjusted for more covariates than this previous study, which controlled for Gleason score and age but did not adjust for comorbidity, era of diagnosis, and pre-treatment PSA level, explanatory variables responsible for reducing the crude variation for this indicator in our study.

The investigators of the EORTC study<sup>39,40</sup> also found that blood transfusions of 3+ units were given more frequently by lower volume surgeons. However, their reported rates were all lower than ours. For example, the EORTC study reported that 17.4% of patients treated by surgeons performing less than 12 RRP annually required transfusion compared to 46.0% in our corresponding < 1/m group. This difference might be explained by the voluntary participation of surgeons and the self-report of their performance, which created opportunities for selection and response bias. In comparison, the chart review-based nature of our study is a strength of our work. Finally, in the EORTC study, there was no adjustment for covariates although differences were noted, and the nature of the participating institutions was unclear (i.e. academic centers vs. community hospitals).

#### **5.4.2 Length of hospital stay**

The average length of hospital stay decreased over the three eras of the study, in the meantime becoming more reflective of surgical quality. The phenomenon of declining length of stay is consistent with expectations, as economic pressures for early discharge in the cost-cutting climate of the 1990s led to evolutions in perioperative management that were successful at reducing length of stay without increasing acute

complication and mortality rates. For example, practice changes included assessment in pre-admission clinics, admitting patients directly into the operating room on the day of surgery, and earlier post-operative ambulation.<sup>51,54</sup> Furthermore, the introduction of more advanced and less morbid procedures may have also contributed to the decline.<sup>54</sup> This means that only those patients not doing well after surgery would be allowed to stay longer, as compared to the past where even patients doing well may have been able to remain in hospital for a few more days than necessary.<sup>101</sup> With the clearest volume trend seen in the last era of our study (1996-1998), it seems as though length of stay became more reflective of the patient's post-surgical condition and thus a more valid and useful indicator of surgical quality.

While we did see a volume trend in the last era, the many hospital amalgamations that were taking place in Ontario may have worked against our ability to see the true strength of association between volume and length of stay. Many hospital groups may not have fully combined the processes from their different sites at that time, resulting in misclassification of hospital volume. In the absence of this misclassification, it is possible that the volume-length of stay trend may have been even stronger than observed.

Our findings are consistent with U.S. studies that examined the volume-length of stay relationship using the Medicare and SEER databases.<sup>23,25</sup> These authors found that length of stay decreased over time and displayed a significant increasing trend with lower hospital as well as surgeon volume. They found lower lengths of stay than our findings in similar eras, perhaps because the pressure to reduce costs might have begun much earlier in the American system. While these studies did not adjust for covariates in their assessment of hospital stay, we were unable to improve upon this limitation, because we

did not have enough study power to examine the effect of explanatory variables after stratifying by era to account for the interaction effect.

### **5.4.3 Use of non-nerve-sparing surgical technique**

In our data, the use of non-nerve-sparing surgery was able to identify poor quality care even before controlling for covariates, although we do not advocate omitting this step. The crude and adjusted odds ratios were of similar magnitude and displayed the same trend by volume. Contrary to expectations, adjustment for explanatory variables actually moved the odds ratios further from unity, indicating that the distributions of the explanatory variables were masking some of the variation truly present. However, given that treatment decisions are also based on severity of disease,<sup>5</sup> the use of non-nerve-sparing surgery as a quality indicator should ideally be used only where patient and disease characteristics can be adjusted for, in order to assure surgeons and other stakeholders that results truly identify quality problems rather than differences in covariate distributions.

Our evidence on the validity of non-nerve-sparing surgery as a quality indicator represents new knowledge. Both expert panels considered but ultimately rejected its use during their quality indicator development processes,<sup>9,58</sup> and it has not been examined at all in the quality indicator and volume literature. Our findings suggest that following further investigation, non-nerve-sparing surgery could become a useful surrogate for evaluating the presence of late morbidities following further investigation, since the use of the nerve-sparing technique has been shown to increase the recovery of continence and

erectile function after radical prostatectomy in the literature,<sup>2,17,36,37,55,56,59-61</sup> and because it is much easier to capture in the absence of system-wide use of validated questionnaires.

It should be noted that the unusually high ORs for the < 7/m volume category reflect the observation that patients in this group received many non-nerve-sparing surgeries from institutions where volume is believed to have been overestimated.

### **5.5 Effect of explanatory variables**

We were able to assess the impact of each explanatory variable on the volume-quality indicator relationships for blood transfusions and use of non-nerve-sparing surgery, but not for length of hospital stay, due to our finding of an interaction effect by era of diagnosis.

For blood transfusions, the odds ratios adjusted for explanatory variables were closer to unity than the crude ORs. Era of diagnosis was responsible for the greatest impact in the crude estimates, followed by biopsy Gleason score and pre-treatment PSA. Era may have confounded the volume-blood transfusion relationship because of better techniques developed in later years that may have required fewer blood transfusions on average. In addition, confounding by Gleason score and PSA may have occurred because surgeries performed on patients with more severe disease may have been more technically difficult, resulting in more complications and a greater chance that blood transfusions were required. This confounding demonstrates that it is important to separate the effect of disease severity in order to observe the impact of surgeon experience on blood transfusion requirement.

For the use of non-nerve-sparing surgery, the odds ratios adjusted for explanatory variables were closer or further from unity than the crude ORs, depending on the volume category. These changes were due to only two of the seven explanatory variables that were assessed, biopsy Gleason score and pre-treatment PSA, reflecting the impact of disease severity on the surgeon's decision as to whether it is possible and/or useful to apply the nerve-sparing technique. In general, the lack of association with most explanatory variables suggests that the use of non-nerve-sparing surgery is much less dependent on patient and disease characteristics than blood transfusions, and might be more a matter of surgeon choice and skill.

Four of the explanatory variables examined had little effect on the crude odds ratios of either quality indicator. Socioeconomic status did not appear to have affected the quality of care or the recovery course. Age and comorbidity may not have had an effect because most of the patients sampled met the recommended criteria to receive surgery on these variables. Finally, clinical T category may not have confounded the relationships because patients with more severe disease among those receiving surgery are better indicated by higher PSA levels and Gleason scores than by the T classification. However, it should be noted that in bivariate analyses, a greater proportion of patients treated in the lower volume hospitals had higher T category disease than in higher volume hospitals. This may be because radiotherapy is centralized and is unavailable in lower volume hospitals. Patients with more severe disease for which radiotherapy is more appropriate may have received surgery because it was not possible to travel to a regional cancer center offering radiotherapy, and/or were not given a choice to do so. This system quality issue has been demonstrated previously.<sup>102</sup>

We think that it is necessary to control for confounders in quality assessments. Our examination of explanatory variables found similar trends in quality before and after adjustment, and these changes were only attributed to a few covariates. However, we do not advocate the use of these quality indicators without proper consideration of patient and disease characteristics, as the similar results that we observed may have been due to chance. Adjustment should always be performed to account for the different patient groups for whom different surgeons provide treatment before assessing quality.

## **5.6 Strengths and limitations**

### **5.6.1 Stratified sampling strategy**

The Regional Cancer Center-based stratified random sampling strategy used to create the study population in the parent study (see Section 3.4) has implications on the study results and their interpretation.

Since sampling fractions varied by Ontario RCC, overall results are not representative of the province. Oversampling in the smallest RCCs was an advantage in providing more subjects treated in low volume hospitals for study. However, as univariate descriptive statistics on our entire study population were not weighted by the sampling fractions used, those results cannot be interpreted as representing practice and covariate distributions in Ontario.

A further consequence of the variable sampling fractions was that patients within each of our volume categories were not a random sample of all patients seen at Ontario hospitals falling within that surgical volume range. Still, there was random sampling within each RCC, where patients were not more or less likely to be included depending

on the quality of care received. Therefore, we do not think that our results were biased but again, results within each volume category cannot be interpreted as the provincial experience. Since our purpose was to evaluate validity and not to describe provincial rates, this is not a concern, but readers should take note that a limitation of our study is that results are not representative of the Ontario experience.

### **5.6.2 Other Strengths**

This was the first Canadian study examining the feasibility and validity of quality indicators in localized prostate cancer surgery. This was also the first study in the field looking at the relationship between quality indicators and volume that considered the effect of explanatory variables in detail, and was not based on U.S. administrative data, which have little information on covariates. The OCR was a high quality and complete source to identify cases and together with the DAD, to calculate overall hospital volume. Non-participation was not an issue because patient recruitment was not required, and our study population captured 97.6% of the sample of patients confirmed to have received radical prostatectomy as curative treatment. Finally, we used high-quality, medical chart data, many from more than one chart, with a high capture rate in a carefully controlled abstraction process to assess quality indicators. Chart abstractors were likely unbiased, as they were unaware of the study objectives.

### 5.7.2 Other Limitations

Even using high-quality, work-intensive medical chart abstraction data, it was not feasible for us to assess all of the candidate quality indicators we had wished to examine *a priori*. Because of missing follow-up and infrequent events, we could not perform regression analyses on adverse outcomes including acute complications, undetectable PSA at first follow-up after surgery, positive surgical margins, erectile dysfunction, urinary incontinence, and biochemical disease-free-survival. For the three volume-process quality indicator associations we were able to analyze in detail, there were several limitations. The first was the volume misclassification caused by hospital amalgamations during the study period, as discussed in Section 5.4. Also, we analyzed hospital volume as the more proximate measure of surgeon volume was unavailable. As demonstrated in a previous study,<sup>23</sup> this likely reduced the strength of our findings, particularly for blood transfusions and use of non-nerve-sparing surgery, which are likely to be more related to the surgery rather than hospital practice. Length of stay is likely more influenced by hospital policies than the other two indicators and thus, the impact of using hospital instead of surgeon volume is uncertain for this indicator. Next, the inaccuracy caused by the assignment of some Gleason scores based on level of cellular differentiation, and the dichotomization of clinical T category, may have caused incomplete control of disease severity. Lastly, our study sample may have excluded some patients who were incorrectly coded as only having lymph node dissection in the DAD rather than the radical prostatectomy that they actually received. However, this number is likely to have been very small, and the strength of our results and the volume

trend observed make it unlikely that this minor selection bias could account for all of the variations seen.

## **5.8 Summary of common themes and implications**

We examined the feasibility of recommended quality indicators for radical prostatectomy. It was not feasible to assess urinary incontinence, erectile dysfunction, undetectable PSA level at first follow-up, and biochemical disease-free survival using our data. While it was feasible to assess acute surgical complications and positive surgical margins, it was ultimately not possible to demonstrate their construct validity due to low study power for acute complications and a high proportion of pathology reports that did not state margin status. However, we found that the actual pathology reporting of positive margin status was more common in higher volume hospital, further validating its current use by CCO as an indicator of overall quality of care<sup>17,49</sup> and demonstrating the importance of consistent pathology reporting to allow for the assessment of positive surgical margins.

We demonstrated that blood transfusions, length of hospital stay, and use of non-nerve-sparing surgery were valid indicators of surgical quality by testing their convergent construct validity against hospital volume. Although the odds ratios for blood transfusions and use of non-nerve-sparing surgery showed similar volume trends and magnitudes before and after explanatory variable adjustment, quality assessment should always control for confounders to ensure appropriate comparisons. That these quality indicators identified worse outcomes in < 7/m group caused by hospital amalgamation

further validated their ability to detect problems with quality, despite the fact that none were in the exact format recommended by the expert panels.

The advantages of the three validated quality indicators are that they do not represent long-term outcomes, and organizations such as Cancer Care Ontario do not need to review medical charts in order to evaluate them. They all relate to events that occur during or in the days shortly after surgery, unlike a much longer-term outcome such as biochemical disease-free survival, where data may no longer be relevant by the time it is collected and analyzed more than 5 years after surgery, as past treatment approaches may already be out of date.<sup>9</sup> Also, it should be possible to track these outcomes electronically for timely, continuous quality assessment and improvement.

Finally, we were also able to examine volume as a quality indicator. The worse outcomes in the < 7/m higher volume group identified by the three validated quality indicators illustrate the concern that volume, while a good indicator of average quality, is an insufficient measure to pinpoint quality issues more specifically. This finding provides further justification towards our efforts in validating better, more proximate quality indicators that will identify problems in surgical quality and suggest areas for improvement.

## **5.9 Recommendations and future research**

Although we found that blood transfusions of 3+ units or greater, length of hospital stay, and use of non-nerve-sparing surgery converge with hospital volume, these quality indicators should be further validated through other tests before their implementation. For example, the predictive validity of a quality indicator such as the

use of non-nerve-sparing surgery could be tested to see whether it predicts another related quality indicator that occurs later after surgery, such as incontinence.<sup>103</sup> Demonstration that the non-nerve-sparing technique predicts urinary function may further validate its use as a suitable proxy for late morbidities.

If their validity is further demonstrated, these three indicators can be applied in the quality of care assessment and improvement process, which was shown to be effective by the increased compliance in pathology reporting after its implementation by CCO as an area for quality improvement.<sup>49</sup> Recent electronic data can be monitored to assess quality using validated indicators in a timely manner, providing descriptions of prostate cancer surgical care and identifying areas in need of quality improvement, and to detect changes after interventions. Length of hospital stay is already recorded in the Canadian Institute for Health Information's Discharge Abstract Database. The use of nerve-sparing surgery and total blood transfusions required may already be recorded electronically. If not, systems could be developed to track these events in the future, as they are concrete events and could be easy to report within a short time frame after surgery. Long lengths of stay and high rates of blood transfusions can be markers of hospitals where quality requires improvement, and hospitals with high non-nerve-sparing surgery rates may be targets for surgical mentoring programs to familiarize those surgeons with the most effective techniques.

More work remains to be done for the candidate quality indicators that we could not feasibly assess in our study, including urinary incontinence and erectile dysfunction. In the absence of follow-up data on these two late morbidities, the use of non-nerve-sparing surgery was used as a surrogate measure. However, as important patient

outcomes following surgery, urinary incontinence and erectile dysfunction are more preferable and proximate indicators of the impact of surgeon skill. Therefore, widespread implementation of validated questionnaires to assess these late morbidities is ideal in order to consistently measure and to compare those outcomes across the province.

We also advise further study on undetectable PSA after surgery and positive surgical margins as potential quality indicators. These measures are important because they could describe the effect of surgical quality on cancer control and risk of recurrence. A PSA test performed consistently throughout the province shortly after surgery would provide a quality indicator within a short time frame after treatment. However, the presence of positive surgical margins might be more feasible to capture than undetectable PSA, as margin status is already consistently reported in the post-surgical pathology reports found in medical charts.

Finally, future studies may wish to assess the feasibility and validity of the other recommended quality indicators that we did not consider in our study. It was not within the scope of this study to examine all of the indicators recommended by RAND and CCO, as our focus was on surgical quality.

Ultimately, the goal of the entire quality process is to improve patient outcomes following radical prostatectomy through feedback to surgeons, heads of surgical departments, hospital administrators, and quality councils. Identifying groups with poorer outcomes will allow surgeons to modify their surgical care such that post-surgical outcomes can be optimized not only in tertiary care institutions, but in all hospitals throughout the province.

## REFERENCES

1. Canadian Cancer Society/National Cancer Institute of Canada. Canadian Cancer Statistics 2006. Toronto: 2006.
2. Campbell's Urology. 7th ed. Philadelphia: W.B. Saunders Company; 1998.2489-565 p.
3. National Cancer Institute. Prostate Cancer Treatment: Health Professional Edition 2007 Mar 14 [cited 2007 Jun 6] Available from <http://www.cancer.gov/cancertopics/pdq/treatment/prostate/healthprofessional>.
4. Nam RK, Jewett MA, Krahn MD. Prostate cancer: 2. Natural history. Canadian Medical Association Journal 1998 Sep 22;159(6):685-91.
5. BC Cancer Agency. Cancer Management Guidelines: Prostate 2006 Jun 30 [cited 2007 Jun 6] Available from <http://www.bccancer.bc.ca/HPI/CancerManagementGuidelines/Genitourinary/Prostate/default.htm>.
6. Goldenberg SL, Ramsey EW, Jewett MA. Prostate cancer: 6. Surgical treatment of localized disease. Canadian Medical Association Journal 1998 Nov 17;159(10):1265-71.
7. Linton KD, Hamdy FC. Early diagnosis and surgical management of prostate cancer. Cancer Treatment Reviews 2003 Jun;29(3):151-60.
8. Wennberg J, Gittelsohn A. Variations in medical care among small areas. Scientific American 1982 Apr;246(4):120-34.
9. Litwin MS, Steinberg M, Malin J, Naitoh J, McGuigan KA, Steinfeld R, Adams J, Brook RH. Prostate Cancer Patient Outcomes and Choice of Providers: Development of an Infrastructure for Quality Assessment. Santa Monica, CA: RAND; 2000.
10. Spencer BA, Steinberg M, Malin J, Adams J, Litwin MS. Quality-of-care indicators for early-stage prostate cancer. Journal of Clinical Oncology 2003 May 15;21(10):1928-36.
11. Miller DC, Montie JE, Wei JT. Measuring the quality of care for localized prostate cancer. Journal of Urology 2005 Aug;174(2):425-31.
12. Donabedian A. Evaluating the quality of medical care. Milbank Q. 1966;83(4):691-729.

13. Cancer Quality Council of Ontario. Strengthening the Quality of Cancer Services in Ontario. Toronto: Canadian Healthcare Association Press; 2003 Oct.
14. Gagliardi AR, Fleshner N, Langer B, Stern H, Brown AD. Development of prostate cancer quality indicators: a modified Delphi approach. *Canadian Journal of Urology* 2005 Oct;12(5):2808-15.
15. Miller DC, Litwin MS, Sanda MG, Montie JE, Dunn RL, Resh J, Sandler H, Wei JT. Use of quality indicators to evaluate the care of patients with localized prostate carcinoma. *Cancer* 2003 Mar 15;97(6):1428-35.
16. Krupski TL, Bergman J, Kwan L, Litwin MS. Quality of prostate carcinoma care in a statewide public assistance program. *Cancer* 2005 Sep 1;104(5):985-92.
17. Miller DC, Spencer BA, Ritchey J, Stewart AK, Dunn RL, Sandler HM, Wei JT, Litwin MS. Treatment choice and quality of care for men with localized prostate cancer. *Med Care* 2007 May;45(5):401-9.
18. Begg CB, Riedel ER, Bach PB, Kattan MW, Schrag D, Warren JL, Scardino PT. Variations in morbidity after radical prostatectomy. *New England Journal of Medicine* 2002 Apr 11;346(15):1138-44.
19. Urbach DR, Stukel TA, Croxford R, MacCallum NL. Analysis of Current Research Related to the Impact of Low-Volume Procedures/Surgery and Care on Outcomes of Care. Toronto: Institute for Clinical Evaluative Sciences; 2004.
20. Nuttall M, van der MJ, Phillips N, Sharpin C, Gillatt D, McIntosh G, Emberton M. A systematic review and critique of the literature relating hospital or surgeon volume to health outcomes for 3 urological cancer procedures. *Journal of Urology* 2004 Dec;172(6:Pt 1):t-52.
21. Bianco FJ, Jr., Riedel ER, Begg CB, Kattan MW, Scardino PT. Variations among high volume surgeons in the rate of complications after radical prostatectomy: further evidence that technique matters. *Journal of Urology* 2005 Jun;173(6):2099-103.
22. Joudi FN, Konety BR. The impact of provider volume on outcomes from urological cancer therapy. *Journal of Urology* 2005 Aug;174(2):432-8.
23. Hu JC, Gold KF, Pashos CL, Mehta SS, Litwin MS. Role of surgeon volume in radical prostatectomy outcomes. *Journal of Clinical Oncology* 2003 Feb 1;21(3):401-5.
24. Ellison LM, Heaney JA, Birkmeyer JD. The effect of hospital volume on mortality and resource use after radical prostatectomy. *Journal of Urology* 2000 Mar;163(3):867-9.

25. Yao SL, Lu-Yao GL. Population-based study of relationships between hospital volume of prostatectomies, patient outcomes, and length of hospital stay. *J Natl Cancer Inst* 1999;91:1950-6.
26. McDowell I, Newell C. *Measuring Health: A Guide To Rating Scales and Questionnaires*. 2nd ed. New York: Oxford University Press; 1996.33-7 p.
27. *Understanding Health Care Outcomes Research*. Gaithersburg, Maryland: 1997.
28. Crawford ED. Epidemiology of prostate cancer. *Urology* 2003 Dec 22;62(6:Suppl 1):Suppl-12.
29. Mazhar D, Waxman J. Prostate cancer. *Postgraduate Medical Journal* 2002 Oct;78(924):590-5.
30. D'Amico AV, Moul J, Carroll PR, Sun L, Lubeck D, Chen MH. Cancer-specific mortality after surgery or radiation for patients with clinically localized prostate cancer managed during the prostate-specific antigen era. *Journal of Clinical Oncology* 2003 Jun 1;21(11):2163-72.
31. D'Amico AV, Whittington R, Malkowicz SB, Wu YH, Chen M, Art M, Tomaszewski JE, Wein A. Combination of the preoperative PSA level, biopsy gleason score, percentage of positive biopsies, and MRI T-stage to predict early PSA failure in men with clinically localized prostate cancer. *Urology* 2000 Apr;55(4):572-7.
32. D'Amico AV. Combined-modality staging for localized adenocarcinoma of the prostate. *Oncology (Williston.Park)* 2001 Aug;15(8):1049-59.
33. Kattan MW, Eastham JA, Stapleton AM, Wheeler TM, Scardino PT. A preoperative nomogram for disease recurrence following radical prostatectomy for prostate cancer. *J Natl Cancer Inst* 1998 May 20;90(10):766-71.
34. Stav K, Rahimi-Levene N, Lindner A, Siegel YI, Zisman A. Retropubic radical prostatectomy: associated blood loss and transfusion requirements--a two-decade perspective review. *Israel Medical Association Journal* 2005 Feb;7(2):103-6.
35. *Prostate Cancer: Principles & Practice*. 1st ed. Philadelphia: Lippincott Williams & Wilkins; 2002.175-316 p.
36. Alivizatos G, Skolarikos A. Incontinence and erectile dysfunction following radical prostatectomy: a review. *ScientificWorldJournal*. 2005 Sep 13;5:747-58.
37. Meuleman EJ, Mulders PF. Erectile function after radical prostatectomy: a review. *European Urology* 2003 Feb;43(2):95-101.

38. Hillner BE, Smith TJ, Desch CE. Hospital and physician volume or specialization and outcomes in cancer treatment: importance in quality of cancer care. *Journal of Clinical Oncology* 2000 Jun;18(11):2327-40.
39. Van Poppel H, Collette L, Kirkali Z, Brausi M, Hoekstra W, Newling DW, Decoster M. Quality control of radical prostatectomy: a feasibility study. *European Journal of Cancer* 2001 May;37(7):884-91.
40. Van Poppel H, Boulanger SF, Joniau S. Quality assurance issues in radical prostatectomy. *European Journal of Surgical Oncology* 2005 Aug;31(6):650-5.
41. Dash A, Dunn RL, Resh J, Wei JT, Montie JE, Sanda MG. Patient, surgeon, and treatment characteristics associated with homologous blood transfusion requirement during radical retropubic prostatectomy: multivariate nomogram to assist patient counseling. *Urology* 2004 Jul;64(1):117-22.
42. Eastham JA, Kattan MW, Riedel E, Begg CB, Wheeler TM, Gerigk C, Gonen M, Reuter V, Scardino PT. Variations among individual surgeons in the rate of positive surgical margins in radical prostatectomy specimens. *Journal of Urology* 2003 Dec;170(6 Pt 1):2292-5.
43. Chun FK, Briganti A, Antebi E, Graefen M, Currlin E, Steuber T, Schlomm T, Walz J, Haese A, Friedrich MG, et al. Surgical volume is related to the rate of positive surgical margins at radical prostatectomy in European patients. *BJU International* 2006 Dec;98(6):1204-9.
44. Karakiewicz PI, Bazinet M, Aprikian AG, Tanguay S, Elhilali MM. Thirty-day mortality rates and cumulative survival after radical retropubic prostatectomy. *Urology* 1998 Dec;52(6):1041-6.
45. Canadian Institute for Health Information. *Health Care in Canada 2005*. Ottawa: 2005 Jun 8. 45-68 p.
46. Ellison LM, Trock BJ, Poe NR, Partin AW. The effect of hospital volume on cancer control after radical prostatectomy. *Journal of Urology* 2005 Jun;173(6):2094-8.
47. Urbach DR, Baxter NN. Does it matter what a hospital is "high volume" for? Specificity of hospital volume-outcome associations for surgical procedures: analysis of administrative data. *BMJ* 2004 Mar 27;328(7442):737-40.
48. Dimick JB, Birkmeyer JD, Upchurch GR, Jr. Measuring surgical quality: what's the role of provider volume? *World Journal of Surgery* 2005 Oct;29(10):1217-21.

49. Cancer Care Ontario/Cancer Quality Council of Ontario. Cancer Service Quality Index 2007 2007 [cited 2007 Jun 10] Available from <http://www.cancercare.on.ca/qualityindex2007/>.
50. Hu JC, Gold KF, Pashos CL, Mehta SS, Litwin MS. Temporal trends in radical prostatectomy complications from 1991 to 1998. *Journal of Urology* 2003 Apr;169(4):1443-8.
51. Ansari MZ, Costello AJ, Jolley DJ, Ackland MJ, Carson N, McDonald IG. Adverse events after prostatectomy in Victorian public hospitals. *Australian & New Zealand Journal of Surgery* 1998 Dec;68(12):830-6.
52. Alibhai SM, Leach M, Tomlinson G, Krahn MD, Fleshner N, Holowaty E, Naglie G. 30-day mortality and major complications after radical prostatectomy: influence of age and comorbidity. *J Natl Cancer Inst* 2005 Oct 19;97(20):1525-32.
53. Ansari MZ, Ackland MJ, Jolley DJ, Carson N, McDonald IG. Inter-hospital comparison of mortality rates. *International Journal for Quality in Health Care* 1999 Feb;11(1):29-35.
54. Klein EA, Grass JA, Calabrese DA, Kay RA, Sargeant W, O'Hara JF. Maintaining quality of care and patient satisfaction with radical prostatectomy in the era of cost containment. *Urology* 1996 Aug;48(2):269-76.
55. Catalona WJ, Carvalhal GF, Mager DE, Smith DS. Potency, continence and complication rates in 1,870 consecutive radical retropubic prostatectomies. *Journal of Urology* 1999 Aug;162(2):433-8.
56. Dubbelman YD, Dohle GR, Schroder FH. Sexual function before and after radical retropubic prostatectomy: A systematic review of prognostic indicators for a successful outcome. *European Urology* 2006 Oct;50(4):711-8.
57. Tobisu K. Function-preserving surgery for urologic cancer. *International Journal of Clinical Oncology* 2006 Oct;11(5):351-6.
58. Siemens DR. Personal communication. 2004 Mar 2.
59. Montorsi F, Briganti A, Salonia A, Rigatti P, Burnett AL. Current and future strategies for preventing and managing erectile dysfunction following radical prostatectomy. *European Urology* 2004 Feb;45(2):123-33.
60. Eastham JA, Kattan MW, Rogers E, Goad JR, Ohori M, Boone TB, Scardino PT. Risk factors for urinary incontinence after radical prostatectomy. *Journal of Urology* 1996 Nov;156(5):1707-13.

61. Robinson JW, Dufour MS, Fung TS. Erectile functioning of men treated for prostate carcinoma. *Cancer* 1997 Feb 1;79(3):538-44.
62. National Cancer Institute. Prostate Cancer Treatment: Health Professional Edition 2006 May 25 [cited 2006 Jun 5] Available from <http://www.cancer.gov/cancertopics/pdq/treatment/prostate/healthprofessional/allpages>.
63. Shekarriz B, Upadhyay J, Wood DP. Intraoperative, perioperative, and long-term complications of radical prostatectomy. *Urologic Clinics of North America* 2001 Aug;28(3):639-53.
64. Thurairaja R, Osborn J, McFarlane J, Bahl A, Persad R. Radical prostatectomy with positive surgical margins: how are patients managed? *BJU International* 2006 Mar;97(3):445-50.
65. Chang SS, Cookson MS. Impact of positive surgical margins after radical prostatectomy. *Urology* 2006 Aug;68(2):249-52.
66. Alsikafi NF, Brendler CB. Surgical modifications of radical retropubic prostatectomy to decrease incidence of positive surgical margins. *Journal of Urology* 1998 Apr;159(4):1281-5.
67. Orvieto MA, Alsikafi NF, Shalhav AL, Laven BA, Steinberg GD, Zagaja GP, Brendler CB. Impact of surgical margin status on long-term cancer control after radical prostatectomy. *BJU International* 2006 Dec;98(6):1199-203.
68. Touijer K, Kuroiwa K, Saranchuk JW, Hassen WA, Trabulsi EJ, Reuter VE, Guillonneau B. Quality improvement in laparoscopic radical prostatectomy for pT2 prostate cancer: impact of video documentation review on positive surgical margin. *Journal of Urology* 2005 Mar;173(3):765-8.
69. Pound CR, Brawer MK, Partin AW. Evaluation and treatment of men with biochemical prostate-specific antigen recurrence following definitive therapy for clinically localized prostate cancer. *Reviews in Urology* 2001;3(2):72-84.
70. Doherty AP, Bower M, Smith GL, Miano R, Mannion EM, Mitchell H, Christmas TJ. Undetectable ultrasensitive PSA after radical prostatectomy for prostate cancer predicts relapse-free survival. *British Journal of Cancer* 2000 Dec;83(11):1432-6.
71. Lee KL, Marotte JB, Ferrari MK, McNeal JE, Brooks JD, Presti JC, Jr. Positive family history of prostate cancer not associated with worse outcomes after radical prostatectomy. *Urology* 2005 Feb;65(2):311-5.

72. Bhatta-Dhar N, Reuther AM, Zippe C, Klein EA. No difference in six-year biochemical failure rates with or without pelvic lymph node dissection during radical prostatectomy in low-risk patients with localized prostate cancer. *Urology* 2004 Mar;63(3):528-31.
73. Siemens DR, Schulze KM, Mackillop WJ, Brundage MD, Groome PA. A population-based study of the waiting times for prostatectomy in Ontario. *Canadian Journal of Urology* 2005 Apr;12(2):2568-74.
74. Boulos DL, Groome PA, Brundage MD, Siemens DR, Mackillop WJ, Heaton JP, Schulze KM, Rohland SL. Predictive validity of five comorbidity indices in prostate carcinoma patients treated with curative intent. *Cancer* 2006 Apr 15;106(8):1804-14.
75. Commission on Professional and Hospital Activities. The International classification of diseases; 9th revision, clinical modification ; ICD-9-CM. 1978. Ann Arbor, Mich., Commission on Professional and Hospital Activities. Ref Type: Serial (Book,Monograph)
76. MedQuest Software Version 32. Baltimore, MD: U.S. Centers for Medicare & Medicaid Services; 1998.
77. Holowaty EJ, Dale D. The hospital only project. *Health Reports* 1993;5:91-5.
78. Holowaty EJ, Moravan V, Lee G, Chong N, Dale D. A reabstraction study to estimate the completeness and accuracy of data elements in the Ontario Cancer Registry. 1996. 1-45 p.
79. Canadian Institute for Health Information. Data Quality Documentation: Discharge Abstract Database 2001–2002. Ottawa: 2003 Feb.
80. Fleshner N. Defining high-risk prostate cancer: current status. *Canadian Journal of Urology* 2005 Feb;12 Suppl 1:14-7.
81. Gillitzer R, Thuroff JW. Relative advantages and disadvantages of radical perineal prostatectomy versus radical retropubic prostatectomy. *Critical Reviews in Oncology/Hematology* 2002 Aug;43(2):167-90.
82. D'Amico AV, Whittington R, Malkowicz SB, Fondurulia J, Chen MH, Kaplan I, Beard CJ, Tomaszewski JE, Renshaw AA, Wein A, et al. Pretreatment nomogram for prostate-specific antigen recurrence after radical prostatectomy or external-beam radiation therapy for clinically localized prostate cancer. *Journal of Clinical Oncology* 1999 Jan;17(1):168-72.
83. Gleave ME, Goldenberg SL, Chin JL, Warner J, Saad F, Klotz LH, Jewett M, Kassabian V, Chetner M, Dupont C, et al. Randomized comparative study of 3

versus 8-month neoadjuvant hormonal therapy before radical prostatectomy: biochemical and pathological effects. *Journal of Urology* 2001 Aug;166(2):500-6.

84. Messing EM, Manola J, Sarosdy M, Wilding G, Crawford ED, Trump D. Immediate hormonal therapy compared with observation after radical prostatectomy and pelvic lymphadenectomy in men with node-positive prostate cancer. *New England Journal of Medicine* 1999 Dec 9;341(24):1781-8.
85. SAS (Statistical Analysis Software) Version 8.2. Cary, NC: SAS Institute, Inc.; 2000.
86. Linn BS, Linn MW, Gurel L. Cumulative illness rating scale. *Journal of the American Geriatric Society* 1968;16:622-6.
87. Miller MD, Paradis CF, Houck PR, Mazumdar S, Stack JA, Rifai AH, Mulsant B, Reynolds CF. Rating chronic medical illness burden in geropsychiatric practice and research: Application of cumulative illness rating scale. *Psychiatry Research* 1992;41:237-48.
88. Groome PA, Siemens DR, Mackillop WJ, Heaton JP, Brundage MD, Kawakami J, Rohland SL, Song Z. A co-morbidity adjusted life expectancy tool for use in the urology clinic. *Canadian Journal of Urology* 13[3], 3077. 2006.  
Ref Type: Abstract
89. Groome PA, Schulze KM, Keller S, Mackillop WJ, O'Sullivan B, Irish JC, Bissett RJ, Dixon PF, Eapen LJ, Gulavita SP, et al. Explaining socioeconomic status effects in laryngeal cancer. *Clinical oncology (Royal College of Radiologists (Great Britain))* 2006 May;18(4):283-92.
90. American Joint Committee on Cancer. *AJCC Cancer Staging Manual*. 5th ed. Philadelphia: Lippincott-Raven; 1997.
91. *TNM classification of malignant tumours*. 6th ed. New York: Wiley-Liss; 2002.
92. Statistics Canada - Health Division. *Canadian classification of diagnostic, therapeutic, and surgical procedures*. Ottawa: Minister of Supply and Services Canada; 1986.
93. Ontario Ministry of Health & Long-Term Care. *Master Numbering System (CIHI Institution Numbers)*. Toronto: 2005 Apr 1.
94. Ontario Ministry of Health & Long-Term Care. *Ontario Master Numbering System* 2006 [cited 2007 Mar 15] Available from <http://www.fimdata.com/masternumber/>.

95. Power and Precision Borenstein M, Cohen J, Rothstein H, Schoenfeld D, Berlin J, Lakatos E. Version 2.1. Englewood, NJ: Biostat, Inc.; 2001.
96. Chang SS, Duong DT, Wells N, Cole EE, Smith JA, Jr., Cookson MS. Predicting blood loss and transfusion requirements during radical prostatectomy: the significant negative impact of increasing body mass index. *Journal of Urology* 2004 May;171(5):1861-5.
97. SAS Institute Inc. The FREQ Procedure: Cochran-Armitage Test for Trend 2003 [cited 2007 Mar 18] Available from [http://support.sas.com/91doc/getDoc/procstat.hlp/freq\\_sect24.htm](http://support.sas.com/91doc/getDoc/procstat.hlp/freq_sect24.htm).
98. SAS Institute Inc. The NPAR1WAY Procedure: Exact Tests 2003 [cited 2007 Mar 18] Available from [http://support.sas.com/91doc/getDoc/statug.hlp/npar1way\\_sect19.htm#stat\\_npar1way\\_nplexmcl](http://support.sas.com/91doc/getDoc/statug.hlp/npar1way_sect19.htm#stat_npar1way_nplexmcl).
99. Hosmer DW, Lemeshow S. *Applied Logistic Regression*. 2nd ed. New York: John Wiley & Sons, Inc.; 2000.
100. Rothman KJ, Greenland S. *Modern epidemiology*. 2nd ed. Philadelphia: Lippincott-Raven; 1998.
101. Gardner TA, Bissonette EA, Petroni GR, McClain R, Sokoloff MH, Theodorescu D. Surgical and postoperative factors affecting length of hospital stay after radical prostatectomy. *Cancer* 2000 Jul 15;89(2):424-30.
102. Huang M. The influence of waiting time on treatment choices among prostate cancer patients [MSc. Kingston, ON: Department of Community Health & Epidemiology, Queen's University; 2006 Sep.
103. *A Dictionary of Epidemiology*. 4th ed. New York: Oxford University Press; 2001.

**APPENDIX I: Medical chart abstraction form (selected pages)**

***Initial Surgery and Morbidity***

**1. *Intraoperative Course***

Units of blood given during surgery: \_\_\_\_\_  
Duration of surgery (in minutes, from first cut to closure): \_\_\_\_\_  
Intraoperative hemorrhage? Y/N/U \_\_\_\_\_  
Intraoperative MI? Y/N/U \_\_\_\_\_

Operative report of prostatectomy present in chart? (Y/N) \_\_\_\_\_  
Operative report of prostatectomy photocopied? (Y/N) \_\_\_\_\_

Surgical pathology report of prostatectomy present in chart? (Y/N) \_\_\_\_\_  
Surgical pathology report of prostatectomy photocopied? (Y/N) \_\_\_\_\_

**2. *Post-operative Course***

Post-operative nadir hemoglobin: \_\_\_\_\_

Exit hemoglobin at end of hospital stay: \_\_\_\_\_

Was iron therapy initiated during the patient's hospital stay? Y/N/U \_\_\_\_\_

Units of blood given in the post-operative period: \_\_\_\_\_

Was the patient admitted to ICU during the hospital stay? Y/N/U \_\_\_\_\_

For how many days after surgery did the catheter remain in place? \_\_\_\_\_

Length of hospital stay: \_\_\_\_\_ days

**3. As a direct consequence of the prostatectomy, is there any record in the chart that the patient sustained:**

**Intraoperative rectal injury? Y/N/U \_\_\_\_\_**

**Permanent rectal injury? Y/N/U \_\_\_\_\_**

**Thromboembolism? Y/N/U \_\_\_\_\_**

**Urethral stricture? Y/N/U \_\_\_\_\_**

**MI < 7 weeks after surgery? Y/N/U \_\_\_\_\_**

**Ongoing urinary drainage from the Jackson-Pratt tube or urinary leakage (anastomotic leak) lasting more than 6 weeks? Y/N/U \_\_\_\_\_**

**Chart information incomplete \_\_\_\_\_**

***Presenting Symptoms***

- #1. Incontinence? Y/N/U \_\_\_\_  
If yes, degree of severity (see "Incontinence" in instruction manual for definitions):  
Mild \_\_\_\_  
Moderate \_\_\_\_  
Severe \_\_\_\_  
Cannot be determined \_\_\_\_
  
- #2. Erectile dysfunction (pre-tx)? Present/Absent/Unknown \_\_\_\_

***Late Morbidity (> 90 Days of Cessation of Radiotherapy Tx or Prostatectomy)***

- #1. Incontinence? Y/N/U \_\_\_\_  
If yes:
  - a) Degree of severity (see "Incontinence" in instruction manual for definitions):  
Mild \_\_\_\_  
Moderate \_\_\_\_  
Severe \_\_\_\_  
Cannot be determined \_\_\_\_
  - b) Was there chart documentation of any secondary procedure to treat the incontinence (i.e. collagen injection or artificial urinary sphincter)? Y/N \_\_\_\_
  - c) Number of pads per day (record the total per day if stated or U if unknown)? \_\_\_\_
  - d) How wet were the used pads?  
Damp, Wet, Saturated, Unknown, Other \_\_\_\_  
If other, specify \_\_\_\_

#2. **Erectile Dysfunction**

Is there any documentation in the medical record regarding sexual function > 90 day post-treatment? Y/N \_\_\_\_

- If yes:
- a) Did the patient have a sexual partner? Y/N/U \_\_\_\_
  - b) Were they sexually active? Y/N/U \_\_\_\_  
Briefly describe: \_\_\_\_\_  
\_\_\_\_\_
  - c) Were they sexually functional? Y/N/U \_\_\_\_  
Briefly describe: \_\_\_\_\_  
\_\_\_\_\_
  - d) What was the erectile status? Known/Unknown \_\_\_\_  
If the erectile status is known, please record verbatim description:  
\_\_\_\_\_  
\_\_\_\_\_

e) Were aids (Viagra, injections, penile implant, etc.) required to achieve and/or maintain an erection for the purpose of sexual intercourse?

Y/N/U \_\_\_\_\_

f) If aids (Viagra, injections, penile implant, etc.) were prescribed to achieve and/or maintain an erection for the purpose of sexual intercourse, please list date of evaluation, type of treatment and treatment success.

| Date Prescribed(dd/mmm/yyyy) | Type of Treatment | Success? Y/N/U |
|------------------------------|-------------------|----------------|
| _____                        | _____             | _____          |
| _____                        | _____             | _____          |
| _____                        | _____             | _____          |

***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

***Post-treatment PSAs***

(#) RECORD DATES AND LEVELS OF ALL POST-TREATMENT PSAs

**DO NOT RECORD ANY PSA VALUES THAT OCCUR AFTER FOUR CONSECUTIVE RISES. FULLY DOCUMENT THOSE FOUR RISES AND ALL POST-TREATMENT PSAs THAT OCCURRED PRIOR TO THE FOUR CONSECUTIVE RISES. IF THE PATIENT DOES NOT EXPERIENCE FOUR CONSECUTIVE RISES, RECORD ALL POST-TREATMENT PSA VALUES FROM COMPLETION OF TREATMENT THROUGH THE END OF FOLLOW-UP IN THE CHART.**

PSA level 1 \_\_\_\_\_ Date 1 (dd) \_\_\_ (MMM) \_\_\_(yyyy)\_\_\_\_\_

PSA level 2 \_\_\_\_\_ Date 2 (dd) \_\_\_ (MMM) \_\_\_(yyyy)\_\_\_\_\_

PSA level 3 \_\_\_\_\_ Date 3 (dd) \_\_\_ (MMM) \_\_\_(yyyy)\_\_\_\_\_

PSA level 4 \_\_\_\_\_ Date 4 (dd) \_\_\_ (MMM) \_\_\_(yyyy)\_\_\_\_\_

PSA level 5 \_\_\_\_\_ Date 5 (dd) \_\_\_ (MMM) \_\_\_(yyyy)\_\_\_\_\_

PSA level 6 \_\_\_\_\_ Date 6 (dd) \_\_\_ (MMM) \_\_\_(yyyy)\_\_\_\_\_

PSA level 7 \_\_\_\_\_ Date 7 (dd) \_\_\_ (MMM) \_\_\_(yyyy)\_\_\_\_\_

PSA level 8 \_\_\_\_\_ Date 8 (dd) \_\_\_ (MMM) \_\_\_(yyyy)\_\_\_\_\_

PSA level 9 \_\_\_\_\_ Date 9 (dd) \_\_\_ (MMM) \_\_\_(yyyy)\_\_\_\_\_

PSA level 10 \_\_\_\_\_ Date 10 (dd) \_\_\_ (MMM) \_\_\_(yyyy)\_\_\_\_\_

***Pathological Findings***

**4. Surgical Margins from pathology report (more than one can be checked):**

All margins grossly and microscopically negative \_\_\_\_\_

Equivocal \_\_\_\_\_

Macroscopic involvement \_\_\_\_\_

Microscopic involvement \_\_\_\_\_

Focally positive \_\_\_\_\_

Extensively positive \_\_\_\_\_

Margins involved, NOS or Positive unspecified \_\_\_\_\_

Apical margin positive \_\_\_\_\_

Margins other than apical margin involved \_\_\_\_\_

Single margin involved \_\_\_\_\_

More than one margin involved \_\_\_\_\_

Close \_\_\_\_\_

Other (specify) \_\_\_\_\_

Unknown whether margins were involved (positive) or negative \_\_\_\_\_

**Which margin(s) were positive (more than one can be checked)?**

Lateral \_\_\_\_\_

Apical \_\_\_\_\_

Posterolateral \_\_\_\_\_

Bladder neck \_\_\_\_\_

Other (specify) \_\_\_\_\_

Unknown \_\_\_\_\_

## ***Recurrence/Spread***

### ***Residual Disease***

1. (#) Was there a finding of residual disease after the completion of all initial treatment?  
Y/N/U \_\_\_\_\_

### ***Recurrence***

2. (#) Once initial treatment was complete, was there a record in the chart of failure either by clinical exam (positive DREs), histologically (positive biopsies), on diagnostic imaging (bone scan), or biochemically (PSA  $\geq$  1.5 ng/ml after nadir of  $<$  1.5 ng/ml for radiotherapy cases OR  $\geq$  0.5 ng/ml after prostatectomy for surgical cases) or any statement of PSA failure in the chart?  
Y/N/U \_\_\_\_\_

3. Was there a local failure as defined by:

Positive biopsy? Y/N/U \_\_\_\_\_ Date: (dd) \_\_\_(MMM) \_\_\_(yyyy)\_\_\_\_\_

Positive DRE? Y/N/U \_\_\_\_\_ Date: (dd) \_\_\_(MMM) \_\_\_(yyyy)\_\_\_\_\_

PSA  $\geq$  1.5 ng/ml for R/T (after nadir of  $<$ 1.5 ng/ml) OR  $\geq$  0.5 ng/ml for surgical cases (any time after prostatectomy) or any statement of PSA failure in the chart?

Y/N/U \_\_\_\_\_ Date: (dd) \_\_\_(MMM) \_\_\_(yyyy)\_\_\_\_\_

4. Was there a regional failure as defined by:

Positive lymph nodes detected using CT scan? Y/N/U \_\_\_\_\_

Date: (dd) \_\_\_(MMM) \_\_\_(yyyy)\_\_\_\_\_

Positive lymph nodes detected by L N sampling? Y/N/U \_\_\_\_\_

Date: (dd) \_\_\_(MMM) \_\_\_(yyyy)\_\_\_\_\_

Positive lymph nodes detected by other diagnostic tests ? Y/N/U \_\_\_\_\_

Date: (dd) \_\_\_(MMM) \_\_\_(yyyy)\_\_\_\_\_

5. Was there distant failure as defined by:

Distant metastasis on bone scan, CT, US or MRI? Y/N/U \_\_\_\_\_

Date: (dd) \_\_\_(MMM) \_\_\_(yyyy)\_\_\_\_\_

6. Treatment(s) administered for residual disease/recurrence:

Surgery? Y/N/U \_\_\_\_\_ Date: (dd) \_\_\_(MMM) \_\_\_(yyyy)\_\_\_\_\_

Chemotherapy? Y/N/U \_\_\_\_\_ Date: (dd) \_\_\_(MMM) \_\_\_(yyyy)\_\_\_\_\_

Radiotherapy? Y/N/U \_\_\_\_\_ Date: (dd) \_\_\_(MMM) \_\_\_(yyyy)\_\_\_\_\_

Hormones? Y/N/U \_\_\_\_\_ Date: (dd) \_\_\_(MMM) \_\_\_(yyyy)\_\_\_\_\_

Orchiectomy? Y/N/U \_\_\_\_\_ Date: (dd) \_\_\_(MMM) \_\_\_(yyyy)\_\_\_\_\_

Other (describe) \_\_\_\_\_? Y/N/U \_\_\_\_\_ Date: (dd) \_\_\_(MMM) \_\_\_(yyyy)\_\_\_\_\_

None \_\_\_\_\_

## APPENDIX II: Data processing

**Table II.1: Acute complications and blood transfusion rates**

|   | <i>Yes*</i> |          | <i>Unknown</i> |          |
|---|-------------|----------|----------------|----------|
|   | <i>N</i>    | <i>%</i> | <i>N</i>       | <i>%</i> |
| <b>Any blood transfusion</b>  | 378         | 58.6     | -              | -        |
| <b>Blood transfusion of 3+ units</b>  | 158         | 24.5     | -              | -        |
| <b>Any of thromboembolism, rectal injury, myocardial infarction, or death</b> | 29          | 4.5      | -              | -        |
| Thromboembolism   | 14          | 2.2      | 24             | 3.7      |
| Rectal injury   | 7           | 1.1      | 23             | 3.6      |
| Myocardial infarction<br>(within seven weeks after surgery only)              | 7           | 1.1      | 181            | 28.1     |
| Death   | 3           | 0.5      | 8              | 1.2      |

\* All within 90 days after surgery, unless otherwise specified.

**Table II.2: Urinary incontinence rates**

|  | <i>Yes</i> |          | <i>Unknown</i> |          |
|--|------------|----------|----------------|----------|
|  | <i>N</i>   | <i>%</i> | <i>N</i>       | <i>%</i> |
| <b>Before surgery</b>                              | 17         | 2.6      | 67             | 10.4     |
| <b>&gt; 90 days cessation of initial treatment</b> | 162        | 25.1     | 208            | 32.3     |

**Table II.3: Erectile dysfunction rates**

|  | <i>Yes</i> |          | <i>Unknown</i> |          |
|--|------------|----------|----------------|----------|
|  | <i>N</i>   | <i>%</i> | <i>N</i>       | <i>%</i> |
| <b>Before surgery</b>                              | 87         | 13.5     | 312            | 48.4     |
| <b>&gt; 90 days cessation of initial treatment</b> | 268        | 41.6     | 359            | 55.7     |
| Erectile dysfunction pre-surgery                   | 87         | 13.5     |                |          |
| Erectile dysfunction post-surgery only             | 181        | 28.1     |                |          |

**Table II.4: Classification of verbatim erectile status descriptions into erectile dysfunction status**

| Erectile dysfunction | Verbatim erectile status descriptions   |
|----------------------|---|
| <b>Yes</b>           | Clear statement of erectile dysfunction since surgery, not sexually active, use of vacuum device (i.e. Erecaid), ICI (intracavernosal injections), occasional aid with MUSE or Viagra, decreased potency, sought treatment for ED, partial/occasional erections, erections insufficient for intercourse |
| <b>No</b>            | Normal erections, improvement, adequate for intercourse, sexually active  |

**Table II.5: Classification of pathology report comments into positive surgical margins categories**

| Category               | Comments on pathology reports  |
|------------------------|--|
| <b>Yes – Extensive</b> | Macroscopic involvement, extensively positive  |
| <b>Yes – Minimal</b>   | Microscopic involvement, focally positive, apical margin positive, margins other than apical margin involved, single margin involved, more than one margin involved, margins involved, NOS or positive unspecified |
| <b>No</b>              | All margins grossly and microscopically negative, equivocal, close   |
| <b>Unknown</b>         | Unknown whether margins were involved (positive) or negative (No reference made to margin status on pathology reports)   |

**Table II.6: Event distribution for 5-year disease-free survival**

| First evidence of failure   | <i>N</i> | %    |
|---|----------|------|
| <b>1 – Residual disease</b>   | 181      | 28.1 |
| <b>2 – Positive biopsy post-surgery</b>   | 3        | 0.5  |
| <b>3 – Positive digital rectal exam post-surgery</b>  | 1        | 0.2  |
| <b>4 – Biochemical failure: PSA <math>\geq</math>0.5 ng/mL or any statement of PSA failure in the chart</b> | 58       | 9.0  |
| <b>5 – Regional failure</b>   | -        | -    |
| <b>6 – Distant failure</b>  | 4        | 0.6  |
| <b>7 – Death if none of 2-6 occurred first</b>  | 8        | 1.2  |
| <b>0 – Censored</b>   | 202      | 31.3 |
| <b>Missing</b>  | 188      | 29.2 |

**Table II.7: Hormone therapy rates**

|                        | <i>Post surgery</i> |     | <i>Pre &amp; Post surgery</i> |     | <i>Pre Surgery</i> |      | <i>None</i> |      |
|------------------------|---------------------|-----|-------------------------------|-----|--------------------|------|-------------|------|
|                        | <i>N</i>            | %   | <i>N</i>                      | %   | <i>N</i>           | %    | <i>N</i>    | %    |
| <b>Hormone therapy</b> | 11                  | 1.7 | 6                             | 0.9 | 163                | 25.3 | 465         | 72.1 |

**Table II.8: Additional non-prostatectomy initial treatment rates**

|  | <i>Yes</i> |     | <i>Unknown</i> |     |
|--|------------|-----|----------------|-----|
|  | <i>N</i>   | %   | <i>N</i>       | %   |
| <b>Post-surgery hormone therapy</b>          | 17         | 2.6 | -              | -   |
| <b>Other prostate cancer related surgery</b> | -          | -   | 11             | 1.7 |
| <b>Chemotherapy</b>                          | -          | -   | 12             | 1.9 |
| <b>Adjuvant radiotherapy</b>                 | 9          | 1.4 | -              | -   |

**Table II.9: Type of radical prostatectomy rates**

|                                      | <i>Retropubic</i> |          | <i>Perineal</i> |          | <i>Other</i> |          | <i>Unknown</i> |          |
|--------------------------------------|-------------------|----------|-----------------|----------|--------------|----------|----------------|----------|
|                                      | <i>N</i>          | <i>%</i> | <i>N</i>        | <i>%</i> | <i>N</i>     | <i>%</i> | <i>N</i>       | <i>%</i> |
| <b>Type of radical prostatectomy</b> | 577               | 89.5     | 50              | 7.8      | 1            | 0.2      | 17             | 2.6      |

**Table II.10: Distribution of study subjects by year of diagnosis**

| Year of diagnosis | <i>N</i> | <i>%</i> |
|-------------------|----------|----------|
| <b>1990</b>       | 20       | 3.1      |
| <b>1991</b>       | 46       | 7.1      |
| <b>1992</b>       | 47       | 7.3      |
| <b>1993</b>       | 76       | 11.8     |
| <b>1994</b>       | 90       | 14.0     |
| <b>1995</b>       | 66       | 10.2     |
| <b>1996</b>       | 103      | 16.0     |
| <b>1997</b>       | 93       | 14.4     |
| <b>1998</b>       | 104      | 16.1     |

**Table II.11: Gleason score frequencies before and after assigning scores using level of differentiation data**

| Assigned Gleason score | <i>BEFORE:<br/>Gleason score only</i> |          | <i>AFTER:<br/>Gleason score + assigned level of differentiation</i> |          |
|------------------------|---------------------------------------|----------|---|----------|
|                        | <i>N</i>                              | <i>%</i> | <i>N</i>  | <i>%</i> |
| <b>2</b>               | 15                                    | 2.3      | 15  | 2.3      |
| <b>3</b>               | 23                                    | 3.6      | 78  | 12.1     |
| <b>4</b>               | 53                                    | 8.2      | 53  | 8.2      |
| <b>5</b>               | 101                                   | 15.7     | 101   | 15.7     |
| <b>6</b>               | 159                                   | 24.7     | 186   | 28.8     |
| <b>7</b>               | 131                                   | 20.3     | 131   | 20.3     |
| <b>8</b>               | 26                                    | 4.0      | 26  | 4.0      |
| <b>9</b>               | 9                                     | 1.4      | 15  | 2.3      |
| <b>Unknown</b>         | 128                                   | 19.8     | 40  | 6.2      |

## APPENDIX III: Ethics approvals



OFFICE OF RESEARCH SERVICES

Fleming Hall, Jemmett Wing  
Queen's University  
Kingston, Ontario, Canada K7L 3N6  
Tel 613 533-6081  
Fax 613 533-6806  
ors@post.queensu.ca  
www.queensu.ca/vpr/

November 10, 2006

Ms. Ellen Chan  
Department of Community Health & Epidemiology  
Queen's University

**Re: "Radical prostatectomy quality indicators: Can we use routinely collected data?"**

Dear Ms. Chan,

I am writing to acknowledge receipt of your recent ethics submission for the above-named study. I have reviewed the materials and do not feel that it is necessary for the study to undergo a full REB review. I have therefore given the study an expedited review and an approval sheet is appended for your records. This study will be reported to the Research Ethics Board.

Yours sincerely,

Albert Clark, Ph.D.  
Chair  
Research Ethics Board

AFC/kr

c.c.: Dr. Patti Groome, Department of Community Health and Epidemiology  
Dr. Robert Siemens, Department of Urology

*think* Research  
*think* Queen's

PREPARING LEADERS AND CITIZENS FOR A GLOBAL SOCIETY

QUEEN'S UNIVERSITY HEALTH SCIENCES & AFFILIATED TEACHING  
HOSPITALS RESEARCH ETHICS BOARD



Queen's University, in accordance with the "Tri-Council Policy Statement, 1998" prepared by the Medical Research Council, Natural Sciences and Engineering Research Council of Canada and Social Sciences and Humanities Research Council of Canada requires that research projects involving human subjects be reviewed annually to determine their acceptability on ethical grounds.

**A Research Ethics Board composed of:**

|                          |  |
|--------------------------|--|
| <b>Dr. A.F. Clark</b>    | <b>Emeritus Professor, Department of Biochemistry, Faculty of Health Sciences, Queen's University (Chair)</b>  |
| <b>Dr. S. Burke</b>      | <b>Emeritus Professor, School of Nursing, Queen's University</b>   |
| <b>Rev. T. Deline</b>    | <b>Community Member</b>  |
| <b>Dr. M. Evans</b>      | <b>Community Member</b>  |
| <b>Mr. C. Kenny</b>      | <b>Community Member</b>  |
| <b>Ms. C. Knott</b>      | <b>Research &amp; Evaluation, Southeastern Regional Geriatric Program, Providence Continuing Care Centre – St. Mary's of the Lake Hospital Site</b>  |
| <b>Dr. J. Low</b>        | <b>Emeritus Professor, Department of Obstetrics and Gynaecology, Queen's University and Kingston General Hospital</b>  |
| <b>Dr. W. Racz</b>       | <b>Emeritus Professor, Department of Pharmacology &amp; Toxicology, Queen's</b>  |
| <b>Dr. H. Richardson</b> | <b>Assistant Professor, Department of Community Health &amp; Epidemiology Project Coordinator, NCIC CTG, Queen's University</b>  |
| <b>Dr. B. Simchison</b>  | <b>Assistant Professor, Department of Anesthesiology, Queen's University</b>   |
| <b>Dr. A.N. Singh</b>    | <b>WHO Professor in Psychosomatic Medicine and Psychopharmacology Professor of Psychiatry and Pharmacology Chair and Head, Division of Psychopharmacology, Queen's University Director &amp; Chief of Psychiatry, Academic Unit, Quinte Health Care, Belleville General Hospital</b> |
| <b>Dr. E. Tsai</b>       | <b>Assistant Professor, Department of Paediatrics and Office of Bioethics, Queen's University</b>  |
| <b>Ms. K. Weisbaum</b>   | <b>LL.B. and Adjunct Instructor, Department of Family Medicine (Bioethics)</b>   |

has examined the protocol for the project entitled "Radical prostatectomy quality indicators: Can we use routinely collected data?" as proposed by Ms. Ellen Chan and Dr. Patti Groome of the Department of Community Health and Epidemiology and Dr. Robert Siemens of the Department of Urology at Queen's University and considers it to be ethically acceptable. This approval is valid for one year. If there are any amendments or changes to the protocol affecting the subjects in this study, it is the responsibility of the principal investigator to notify the Research Ethics Board. Any unexpected serious adverse event occurring locally must be reported within 2 working days or earlier if required by the study sponsor. All other serious adverse events must be reported within 15 days after becoming aware of the information."

*Robert Clark*

Chair, Research Ethics Board

*Nov 10, 2006*

Date

ORIGINAL TO INVESTIGATOR - COPY TO DEPARTMENT HEAD - COPY TO HOSPITAL(S) - P&T - FILE COPY

EPID-235-06

EX



December 7, 2006

Ellen Chan  
Queen's Cancer Research Institute  
Division of Cancer Care & Epidemiology (DCCE)  
10 Stuart Street, Level 2  
Kingston, ON K7L 3N6

Dear Ms. Chan,

This letter is to confirm that our Data Access Committee has approved the use of Ontario Cancer Registry data (housed at DCCE) for your research study, entitled, "Radical prostatectomy quality indicators: Can they be used to assess and compare quality?". This permission is subject to our receipt of a signed Cancer Care Ontario Confidentiality Agreement.

We wish you the best of luck in pursuing this research project.

Sincerely,

A handwritten signature in black ink, appearing to read "K. Milnes".

Kamini Milnes  
Director, Informatics

## APPENDIX IV: Supplementary results

### Interpretation of figures in this section:

- For all plots, the x-axis represents the estimated probability of experiencing or not experiencing the dichotomous outcome (blood transfusions, non-nerve-sparing surgery)
- For Figures IV.1, IV.2, IV.4 and IV.5:
  - The y-axis measures the change in the Pearson chi-square (Figures IV.1, IV.4) or deviance (Figures IV.2, IV.5) when each observation is deleted
  - Ill-fitting points are those found in the top right and top left corner of the plot, at some distance from the balance of the data plotted, or with changes in the Pearson chi-square or deviance diagnostic above 4.0
- For Figures IV.3 and IV.6:
  - The y-axis measures the change in the confidence interval when each observation is deleted
  - Influential points have a confidence interval displacement diagnostic greater than 1.0
- In these figures, even though a few ill-fitted points are identified in Figures IV.1, IV.2, IV.8, and IV.9, there were no points with a confidence interval displacement diagnostic above 1.0 that were extremely influential to either the blood transfusions or non-nerve-sparing surgery parsimonious models

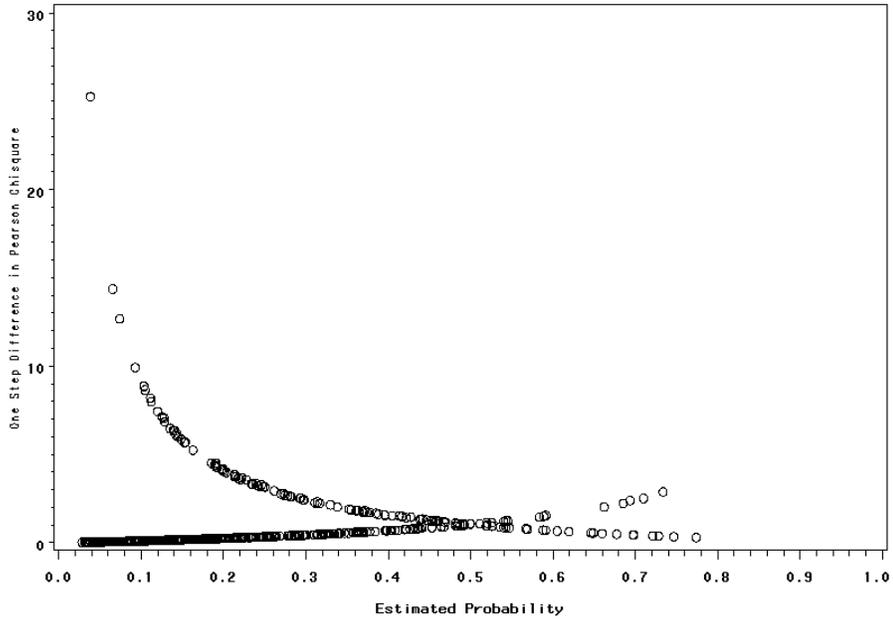


Figure IV.1: Change in Pearson chi-square diagnostic from deleting individual observations versus the estimated probability from the fitted parsimonious model for total blood transfusions of three units or greater

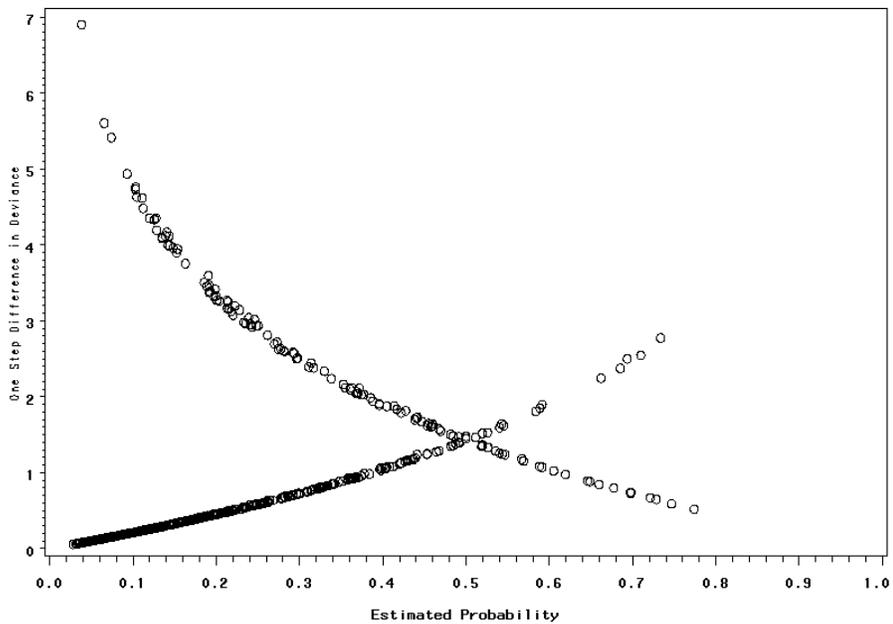


Figure IV.2: Change in deviance diagnostic from deleting individual observations versus the estimated probability from the fitted parsimonious model for total blood transfusions of three units or greater

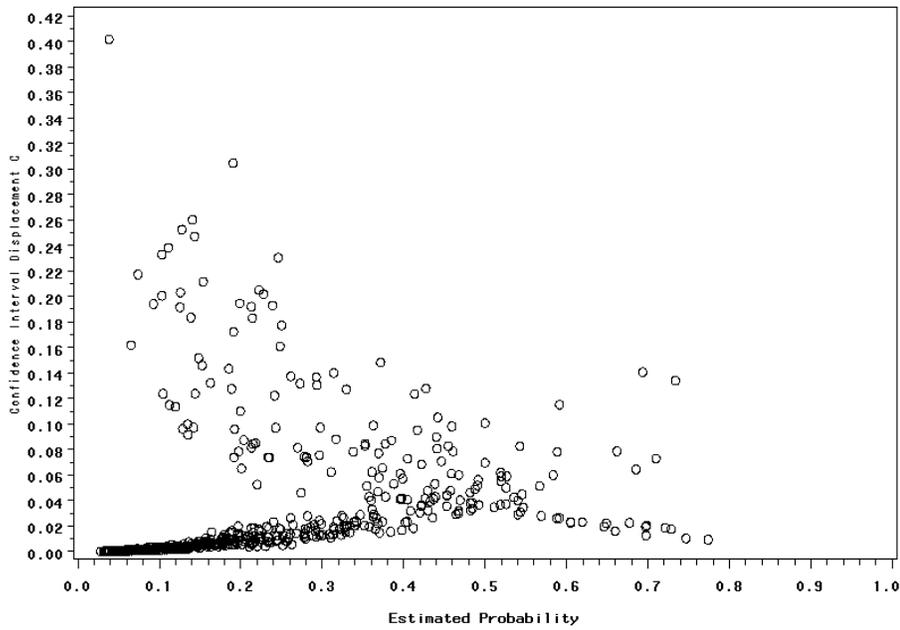


Figure IV.3: Confidence interval displacement diagnostic from deleting individual observations versus the estimated probability from the fitted parsimonious model for total blood transfusions of three units or greater

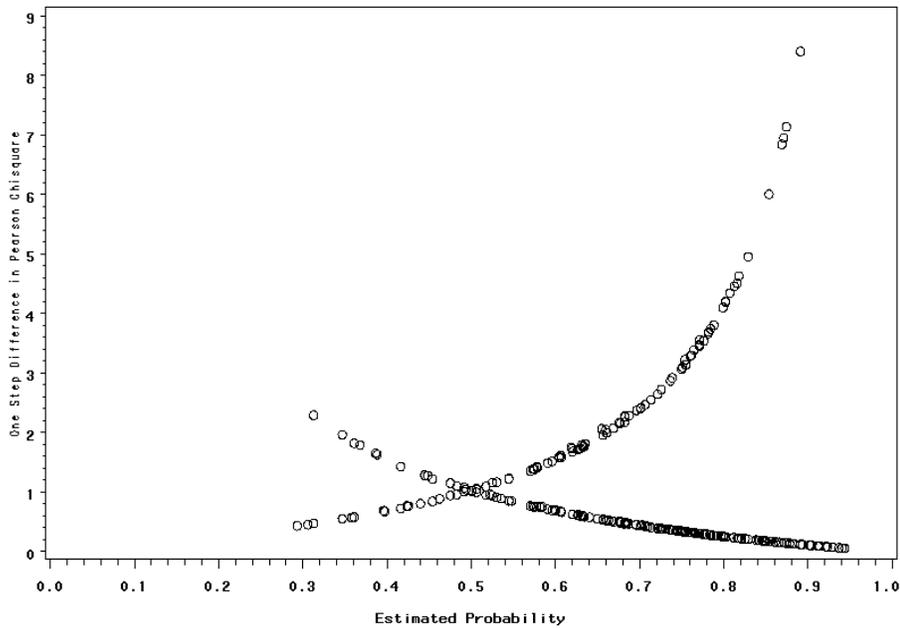


Figure IV.4: Change in Pearson chi-square diagnostic from deleting individual observations versus the estimated probability from the fitted parsimonious model for use of non-nerve-sparing surgical technique

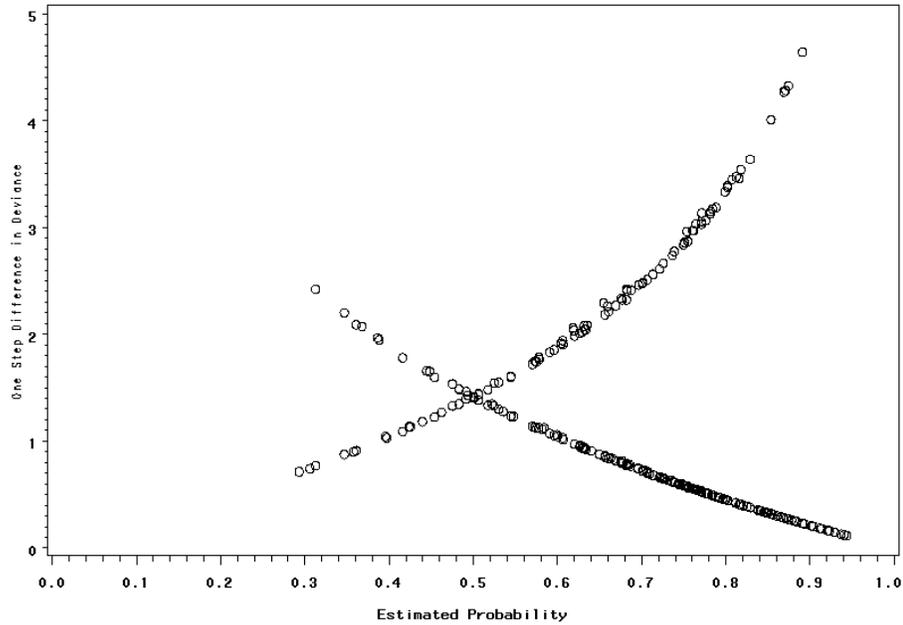


Figure IV.5: Change in deviance diagnostic from deleting individual observations versus the estimated probability from the fitted parsimonious model for use of non-nerve-sparing surgical technique

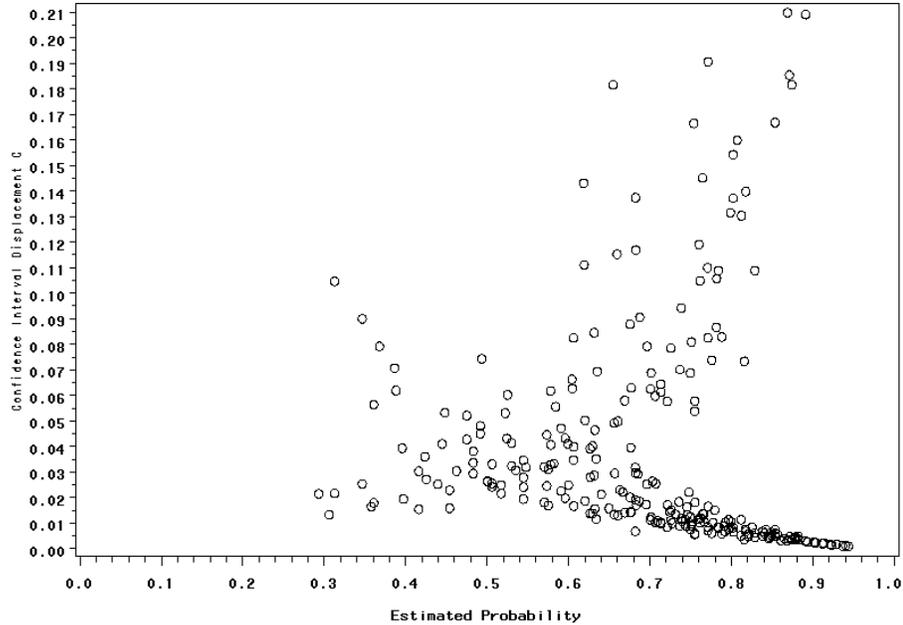


Figure IV.6: Confidence interval displacement diagnostic from deleting individual observations versus the estimated probability from the fitted parsimonious model for use of non-nerve-sparing surgical technique