AN EXPLORATION OF LIFE EXPECTANCY CALCULATION METHODS TO AID IN PROSTATE CANCER SCREENING AND TREATMENT DECISION-MAKING

by

Dylan Jeffrey Wykes

A thesis submitted to the Department of Community Health & Epidemiology

In conformity with the requirements for

the degree of Master of Science

Queen’s University

Kingston, Ontario, Canada

(April, 2011)

Copyright ©Dylan Jeffrey Wykes, 2011
Abstract

**Background:** Life expectancy estimation is an important part of both screening and treatment decision-making for potentially curable prostate cancer. Clinicians’ estimation of patient life expectancy is typically made using population-based life tables and intuition and it is often inaccurate. This study explores methods to improve life expectancy prediction by formally considering patient age and co-morbid illness status, as opposed to age alone, in the development of a life expectancy prediction tool.

**Methods:** We conducted a population-based retrospective cohort study of patients from the Ontario Cancer Registry who were curative treatment candidates, identified between 1990-1998. We analyzed data on three sub-populations of this cohort, and we used life expectancy estimates from the Ontario Life Tables. Each model utilized Cox proportional hazards analysis, and/or the declining exponential approximation of life expectancy, or both to estimate the survival experience of potential curative treatment candidates, including the impact due to both age and co-morbid illness status. We developed five separate models, tested them using a random subset of the cohort study sample, and compared their predictive accuracy by measuring both discriminative ability and calibration to determine the ‘best’ model. Further analysis was conducted to demonstrate the clinical usefulness of the ‘best’ model and to develop life expectancy reference tables. We also conducted a supplementary analysis using logistic regression to develop a model to predict the probability of 10-year survival.

**Results:** The ‘best’ of our models, the Sub-Cohort LE prediction model, demonstrated a c-index of 0.65 and very good calibration performance. Two other models demonstrated slightly higher
discrimination ($C=0.66$), but comparatively poor calibration. Further analysis revealed that our ‘best’ model violated the Cox PH assumption for age and it’s predictions consistently over-estimated observed life expectancy. Supplementary analysis of the logistic regression prediction model demonstrated a $c$-index of 0.70.

**Conclusions:** Our exploration of methods to predict life expectancy resulted in modest predictive accuracy. However, based on the results of the logistic regression model, we conclude that the results of our life expectancy prediction models are reasonable, and obtaining a high level of predictive accuracy may not be possible given just age and co-morbid illness status as predictors. Further studies should continue to explore these and other methods for life expectancy prediction, being cautious of the Cox PH assumption and its implications when violated. External validation of the ‘best’ model from the current study is required before the model and its accompanying life expectancy reference tables can be recommended for use in a clinical setting for screening or treatment decision-making.
Co-Authorship

This thesis is the research work of Dylan Wykes in collaboration with his supervisors Dr. Paul Y. Peng and Dr. Patti A. Groome.
Acknowledgements

First I would like to thank my supervisors, Dr. Patti Groome and Dr. Paul Peng, for all of the time, knowledge, and patience they lent to this project. I learned an unbelievable amount from both of them, and I cannot imagine a more knowledgeable, diverse, and supportive team of supervisors.

I would also like to acknowledge Dr. Groome’s other students – John, Julie, John, Coleen, Felicia and Matt for all of their suggestions and help with my project. Thanks also to everyone at the Cancer Care & Epidemiology Division of Queen’s Cancer Research Institute for their support and knowledge throughout the project, especially Sue Rohland and Jina Zhang-Salomens.

I would also like to extend a thank you to Kat Cook and Lee Watkins in the Department of Community Health & Epidemiology for their help in completing this process.

Finally, a big thank you to my partner, Francine, and my family for their constant support and encouragement.
Table of Contents

Abstract ................................................................................................................................................................. ii  
Co-Authorship ...................................................................................................................................................... iv  
Acknowledgements ................................................................................................................................................. v  
Table of Contents .................................................................................................................................................. vi  
List of Figures ......................................................................................................................................................... ix  
List of Tables ......................................................................................................................................................... xi  
List of Equations .................................................................................................................................................... xiii  

Chapter 1: Introduction ........................................................................................................................................... 1  
  1.1 Background and Rationale ............................................................................................................................ 1  
  1.2 Study Objectives ........................................................................................................................................... 2  
  1.3 Outline ........................................................................................................................................................... 3  
  1.4 References ..................................................................................................................................................... 4  

Chapter 2: Literature Review ............................................................................................................................... 7  
  2.1 Prostate Cancer .............................................................................................................................................. 7  
  2.2 Co-morbidity ................................................................................................................................................ 10  
  2.3 Screening ....................................................................................................................................................... 12  
  2.4 Treatment ....................................................................................................................................................... 13  
  2.5 Life Expectancy ........................................................................................................................................... 14  
  2.6 Life Expectancy Calculation Tools ............................................................................................................. 16  
  2.7 Summary ....................................................................................................................................................... 19  
  2.8 References ..................................................................................................................................................... 20  

Chapter 3: Methods ............................................................................................................................................. 26  
  3.1 Overview ....................................................................................................................................................... 26  
  3.2 Study Design ................................................................................................................................................. 27  
  3.2.1 Target Population ................................................................................................................................. 27  
  3.2.2 Case-Cohort Population ....................................................................................................................... 31  
  3.2.3 Sub-Cohort Population ....................................................................................................................... 39  
  3.2.4 Ontario Life Tables ............................................................................................................................. 41  
  3.3 Analysis Strategy ......................................................................................................................................... 42  
  3.3.1 Overview ................................................................................................................................................. 42
3.3.2 Life Expectancy Prediction Model Development ......................................................... 42  
3.3.3 Comparing Accuracy of Life Expectancy Predictions ................................................ 45  
3.3.4 Clinical Usefulness of 'Most Accurate' Model ............................................................. 47  
3.3.5 Life Expectancy Reference Tables .............................................................................. 47  
3.3.6 Supplementary Logistic Regression Analysis ............................................................. 48  
3.4 References ........................................................................................................................... 48  

Chapter 4: Results ....................................................................................................................... 51  
4.1 Overview ............................................................................................................................. 51  
4.2 Description of Study Populations .................................................................................. 51  
4.2.1 Target Population ....................................................................................................... 51  
4.2.2 Sub-Cohort and Case-Cohort Populations ................................................................. 52  
4.3 Life Expectancy Prediction Models: Methods and Results ........................................... 56  
4.3.1 Model #1: Sub-Cohort Model ..................................................................................... 57  
4.3.2 Model #2: Case-Cohort Target Model ....................................................................... 61  
4.3.3 Model #3: Simplified Case-Cohort Target Model ....................................................... 66  
4.3.4 Comparing the Cox PH Type LE Prediction Models .................................................. 69  
4.3.5 Assessing Proportional Hazards Assumption ............................................................. 69  
4.3.6 Model #4: OLT DEALE Model .................................................................................... 74  
4.3.7 Model #5: Non-Constant OLT Model ...................................................................... 78  
4.3.8 Determining the 'Best' LE Prediction Model ............................................................... 83  
4.4 Clinical Usefulness ............................................................................................................. 85  
4.5 Tables of Model-Predicted Life Expectancy Values ...................................................... 92  
4.6 Supplementary Analysis: Logistic Regression Prediction Model ................................. 96  
4.7 References .......................................................................................................................... 101  

Chapter 5: Discussion ............................................................................................................... 103  
5.1 Overview ......................................................................................................................... 103  
5.2 Interpretation of Results ................................................................................................. 104  
5.3 Model Performance ......................................................................................................... 110  
5.4 Assumptions .................................................................................................................... 116  
5.5 Limitations ....................................................................................................................... 119  
5.6 Strengths .......................................................................................................................... 121  
5.7 Conclusions ...................................................................................................................... 124
5.8 References ............................................................................................................................... 124

Appendices .................................................................................................................................. 128

Appendix I: Impact of Lymph Node Dissection on Treatment Candidacy ........................ 128
Appendix II: Cumulative Illness Rating Scale (CIRS-G) ....................................................... 130
Appendix III: Analysis of Fit, Continuous vs Categorical Variables for Age ................. 132
Appendix IV: Macro to Calculate Life Expectancy from Survival Function .................. 135
Appendix V: Adjustment for unmeasured CIRS-G_{pros} in Target Population ............... 137
Appendix VI: Excess Mortality Due to Prostate Cancer ...................................................... 139
Appendix VII: Treatment Effect Adjustment ............................................................................ 140
References ................................................................................................................................... 141
List of Figures

Chapter 3

Figure 3.1: Development of final Target population ................................................................. 29
Figure 3.2: Development of original Case-Cohort Study population ........................................ 33
Figure 3.3: Selection of final Case-Cohort population ............................................................... 35
Figure 3.4: Selection of Sub-Cohort population ........................................................................ 40

Chapter 4

Figure 4.1: Histogram: Frequency of CIRS-Gpros scores for Case-Cohort population ............. 56
Figure 4.2a: Calibration Plot for Sub-Cohort LE Prediction Model ........................................ 59
Figure 4.2b: Calibration Plot for ‘Age Only’ Sub-Cohort LE Prediction Model ..................... 61
Figure 4.3: Calibration Plot for Case-Cohort Target LE Prediction Model .............................. 65
Figure 4.4a: Calibration Plot for Simplified Case-Cohort Target LE Prediction Model .......... 67
Figure 4.4b: Calibration Plot for ‘Age Only’ Target LE Prediction Model ............................... 68
Figure 4.5a: Plot of Log(-log(K-M-survival)) vs Survival Time, by deciles of Age ............. 71
Figure 4.5b: Score Process Plot for Cox PH modeling of Age in Sub-Cohort data ............... 71
Figure 4.6: Plot of Log (-log(K-M survival)) vs Survival Time, for CIRS-Gpros .................... 73
Figure 4.7a: Calibration Plot for OLT DEALE LE Prediction Model .................................. 77
Figure 4.7b: Calibration Plot for ‘Age Only’ OLT DEALE LE Prediction Model..................... 78
Figure 4.8a: Calibration Plot for Non-Constant OLT LE Prediction Model ......................... 81
Figure 4.8b: Calibration Plot for ‘Age Only’ Non-Constant OLT LE Prediction Model ........ 82
Figure 4.9: Scatterplot of Sub-Cohort model Life Expectancy Predictions vs Observed
List of Tables

Chapter 3

Table 3.1: Original Case-cohort population sampling methods, by CCOR ................................. 34
Table 3.2: Modified case-cohort population sampling methods, by CCOR ................................. 34

Chapter 4

Table 4.1: Baseline characteristics of Target population .................................................................. 52
Table 4.2: Baseline Characteristics of sub-cohort and cases separately .......................................... 55
Table 4.3: Results of the development of inputs derived for use in the Case-cohort Target, Simplified Case-cohort Target, and Sub-cohort Life expectancy prediction models.......58
Table 4.4: C-statistics by Quintile of Predicted Probability of 10-year survival ................................. 74
Table 4.5a: Contingency Table for Sub-Cohort Model Predicted vs Observed Life Expectancy, treat decision window of 7-13 years ................................................................. 86
Table 4.5b: Contingency Table for Sub-Cohort Model Predicted vs Observed Life Expectancy, treat decision window of 10-15 years ........................................................................ 87
Table 4.6a: Contingency Table for Ontario Life Tables Predicted vs Observed Life Expectancy, treat decision window of 10-15 years ........................................................................ 89
Table 4.6b: Contingency Table for Ontario Life Tables Predicted vs Observed Life Expectancy, treat decision window of 10-15 years ........................................................................ 89
Table 4.7a: Sub-cohort Model-Predicted Probability Cutoffs for 10-Year Survival After Diagnoses, among sub-cohort population members .......................................................... 91
Table 4.7b: Ontario Life Tables-Predicted Probability Cutoffs for 10-Year Survival After Diagnoses, among sub-cohort population members .......................................................... 92
Table 4.8a: Life Expectancy Reference Table of Sub-Cohort Model predicted, post-treatment LE estimates................................................................. 94
Table 4.8b: Life Expectancy Reference Table of Sub-Cohort Model predicted, pre-treatment (adjusted) LE estimates................................................................. 95
Table 4.9: Logistic Regression Model, Analysis of Maximum Likelihood Estimates .................. 96

Table 4.10: Life Expectancy Reference Table of Predicted Probabilities of 10-year Survival .................................................................................................................. 100

Chapter 5

Table 5.1: Contingency Table of Model-Predicted vs Observed Life Expectancy, Sub-Cohort population members with age 50-59 years.......................................................106

Table 5.2: Contingency Table of Model-Predicted vs Observed Life Expectancy, Sub-Cohort population members with age 50-59 years and CIRS-\text{G}_{\text{pros}} \geq 5.........................107

Table 5.3: Contingency Table of Model-Predicted vs Observed Life Expectancy, Sub-Cohort population members with age 60-69 years ..........................................................108

Table 5.4: Contingency Table of Model-Predicted vs Observed Life Expectancy, Sub-Cohort population members with age 70-79 years ..........................................................108

Table 5.5: Contingency Table of Model-Predicted vs Observed Life Expectancy, Sub-Cohort population members with age 70-79 years and CIRS-\text{G}_{\text{pros}} \leq 3.........................108
List of Equations

Chapter 3

Equation 1: typical structure of Cox PH survival function ................................................................. 43
Equation 2: declining exponential approximation of life expectancy .................................................. 44
Equation 3: defining the contribution of each usable pair to the concordance .................................. 46
Equation 4: the concordance index (c-index) ..................................................................................... 46

Chapter 4

Equation 5: survival function for Model #1: the Sub-Cohort model .................................................. 57
Equation 6: survival function for Model #2: the Case-Cohort Target model ..................................... 61
Equation 7: survival function for Model #3: the Simplified Case-Cohort Target model .................. 66
Equation 8: equation of life expectancy for Model #4: the OLT DEALE model ............................... 75
Equation 9: survival function for Model #5: the Non-Constant OLT model .................................... 79
Equation 10: logistic regression equation for prediction of 10-year survival probability ............... 96

Appendices

Appendix V

Equation 11: adjustment for unmeasured confounding ....................................................................... 137

Appendix VII

Equation 12: calculation of treatment effect adjustment ....................................................................... 140
Equation 13: pre-treatment (adjusted) Sub-Cohort model ................................................................. 140
1.1 Background & Rationale

Prostate cancer is the leading non-skin cancer diagnosed in Canadian men,(1) and most men are diagnosed with localized disease.(2, 3) Prostate cancer is most commonly diagnosed in older men and is often experienced as slow growing and asymptomatic.(4-6) For men with prostate cancer co-morbid illnesses are common and have been shown to be important prognostic indicators.(7-10) Prognosis is favorable for men with localized prostate cancer; with 5 and 10-year relative survival rates as high as 100 and 95% respectively.(11) Several forms of curative treatment are available including external beam radiotherapy and radical prostatectomy.(12)

Since prostate cancer occurs primarily in older men, curing the disease may not actually increase life expectancy(LE). In addition to age at diagnosis, the often slowly progressing and asymptomatic nature of prostate cancer, suggest that it could remain undetected or cause few symptoms for the patient’s remaining lifespan.(5, 13-15) This fact, coupled with the increased prevalence of co-morbidities in elderly men, also means that many newly diagnosed men have other illnesses that will cause death prior to their prostate cancer.(16) In turn, in the presence of increased screening in the PSA-era, the over-diagnosis of prostate cancer, and it’s accompanying physical and psychological harms, have become a concern.(17,18) Furthermore, although curative treatment is recommended for men with localized disease, the harms of treatment may outweigh the benefits in many cases.(19-21) Therefore there is a need to better identify those individuals whose life expectancy is long enough to potentially benefit from prostate cancer screening and or treatment.

Doctors typically recommend curative treatment to men with clinically localized disease who are believed to have at least 10 years of remaining life.(22, 23) Although pre-diagnosis life expectancy guidelines are not as uniform, it has been suggested the same 10 year rule of thumb
should be applied, considering most men who undergo the screening process and are diagnosed with localized disease opt for curative treatment. (17) The screening and subsequent curative treatment process for localized prostate cancer have both positive and negative effects on life expectancy and quality of life. (17-25) Clinician’s estimation of patient life expectancy is typically based on population-based life tables and intuition, but is often inaccurate. (26-28) Several methods have been proposed to improve the accuracy of clinician’s life expectancy estimates. (9, 29-33) Both population-based life tables and other commonly used life expectancy prediction methods lack appropriate consideration of important predictors and may be limited by methods used to approximate excess disease hazards. (34-36)

Because the estimation of a patient’s life expectancy is such an important factor in the screening and treatment decision making process and co-morbidities have a large impact in the population most affected by prostate cancer, the use of a life expectancy calculation tool which formally considers both a patient’s age and co-morbid illnesses, as opposed to age alone, could aid doctors and patients in the decision making process. Such a calculation tool could provide far more accurate approximation of the patient’s remaining lifespan than the current – informal and often inaccurate – methods used in the clinical setting. Furthermore, a tool that is quick and easy enough to use at a patient’s bedside or during annual physical examination would be of greatest benefit.

1.2 Study Objectives

1) To develop a life-expectancy prediction model that considers patient’s age and co-morbid illness information, as opposed to age alone with application to both screening for prostate cancer and treatment decision-making for potentially curable prostate cancer. In a randomly sampled group of Ontario men, diagnosed with clinically localized prostate cancer from 1990 to 1998, who were indicated for curative treatment.
2) To compare the life expectancy prediction models developed in this thesis and existing methods based on age alone, in terms of the following measures of predictive accuracy (PA):

i. discrimination

ii. calibration

iii. level of agreement across clinically meaningful ‘time-windows’.

iv. optimal negative predictive value (NPV) of model derived 10-year survival probability cutoffs.

With the aim of providing better PA than methods based on the consideration of age alone.

3) To present LE predictions in an easy-to-use format for screening and as a treatment decision aid in the clinical setting, applicable to potential candidates for curative treatment of prostate cancer between 50 to 80 years of age.

1.3 Outline

The remaining chapters in this thesis are organized as follows. The second chapter reviews the prostate cancer and co-morbid illness status literature relevant to the current study’s objectives. This literature review also summarizes the methods employed in recent studies to develop life expectancy calculation tools that aim to improve prediction accuracy generally, and specifically for men with clinically localized prostate cancer. The third chapter provides a description of the study design, populations, structure of potential life expectancy prediction models, and analyses strategy used in this study. The fourth chapter provides detailed results of our models’ development, predictive accuracy, and assessment of clinical usefulness. This chapter also contains the ‘end products’ from this work, suggested for use in the clinical setting as an aid in the treatment
decision-making process. The fifth chapter discusses the overall results, conclusions, strengths and limitations of this thesis. The appendices provide detailed information on the methods and results of certain elements of model development, and excerpts from the co-morbid illness rating scale (CIRS-G) scoring process.

1.4 References


Chapter 2: Literature Review

2.1 Prostate Cancer

Prostate cancer continues to affect more men in Canada than any other type of cancer, accounting for a projected 27.3% (24,600 cases) of all cancer cases in 2010.(1) An incidence rate of 123.0 cases per 100,000 was estimated for 2009.(1) Trends in prostate cancer incidence have not been stable over the past 30 years and are often characterized by two peaks, in 1993 and to a lesser extent in 2001. The first peak, during which incidence rates reached 140.8 cases per 100,000 men, followed the introduction of PSA screening. The second is thought to be due to the publicity around the prostate cancer diagnosis of the (then) Canadian Minister of Health following multiple PSA tests.(1) After a moderate decline for several years following the 2001 peak, incidence rates have remained fairly stable to present. Prostate cancer accounted for approximately 4,300 deaths among Canadian men in 2010.(1) The mortality rate of prostate cancer in Canada peaked in 1991 at 31.2 deaths per 100,000 men. In recent years the mortality rate appears to be declining; 22.0 deaths per 100,000 men are projected for 2010.(1)

In the current era of PSA screening it has become increasingly likely that men will be diagnosed with early stage prostate cancer.(2, 3) Cooperberg, et al demonstrated that the proportion of early stage prostate cancers more than doubled from 29.9% to 78.3% from the early 1990’s to early 2000’s among a large group of men from across the United States followed from 1990 to 2006.(3) Evidence of similar increases exist in Canada and many European countries where the use of PSA screening proliferated in the early 1990’s. Currently, approximately 90% of men with prostate cancer are diagnosed with a clinically localized form of the disease.(4, 5) As such, the number and proportion of men diagnosed with prostate cancers at a curable stage have increased dramatically, making the treatment decision making process increasingly important.(6)
Prostate cancer effects older men, to a greater extent than other cancers, with 81.3% of cases occurring in men 60 years and older, and an average age at diagnosis of 68. (1) The disease is rare before the age of 50, but age-specific incidence rates increase thereafter. In fact, the incidence of prostate cancer increases faster with age than that of any other major cancers. (7) As such, age is the most important risk factor for the development of prostate cancer.

Prostate cancer is often experienced as a slowly progressing disease that will not cause symptoms or death within a patient’s lifetime. (8-10) This is especially true when considering the first 10-15 years following diagnosis. In a population-based, cohort study of men diagnosed with early stage disease, with an average of 21 years follow-up, Johansson et al. found that 76% did not experience metastases within 10-15 years. There is also strong evidence to support the fact that among men with localized prostate cancer mortality rates do not differ significantly when outcomes within the first 10-years post-diagnosis are compared to outcomes at 15-years post-diagnosis. In a retrospective, population-based, cohort study using Connecticut Tumor Registry (CTR) data, Albertsen et al. found that among men with localized prostate cancer, followed for an average of 24 years, the disease specific mortality rate was very low, and did not differ significantly from 15, to more than 15 years of follow-up (33 deaths per 1000 person years, vs 18 death per 1000 person years). (10) Also, prior to the widespread adoption of PSA screening, the prevalence of undiagnosed prostate cancer at death determined from autopsy studies ranged from 10% to 40%, depending on age and ethnicity (11, 12) indicating that many men die with prostate cancer rather than of prostate cancer.

Prognosis for men with clinically localized prostate cancer is quite good due to the long natural history of the disease. The 10-year disease specific survival rates vary from 43% to 94% depending on the severity of the disease. When compared to the population of men of the same age, without prostate cancer, patient’s with localized disease experience 5 and 10-year relative survivals of nearly 100% and 95% respectively. (13) Brenner and Arndt have demonstrated that
relative to the general population, men with prostate cancer often have 5 and 10-year survival rates above 100%.(14) These findings are true for several subsets of patients, including men of all ages who are diagnosed with clinically localized prostate cancer, and men between the ages of 64-75 regardless of clinical stage. Hakulinen and Brenner have also demonstrated this phenomenon when comparing very-long term survival rates (i.e. > 10-years) of prostate cancer patients, to the general public.(15) These results suggest that the population most commonly diagnosed with prostate cancer, men with clinically localized disease, experience no excess mortality due to prostate cancer, may be healthier and can expect to live at least as long as the general population. It obviously seems counterintuitive that the occurrence of prostate cancer would increase an individual’s chance of survival over the general public. Some have suggested this is due to a selection bias in PSA screening, whereby lower socioeconomic status men, who in general have higher mortality rates, were less likely to be screened and subsequently diagnosed.(16)

Controversy exists regarding the relationship between patient age and disease aggressiveness. Some studies have shown that younger men diagnosed with prostate cancer experience more aggressive tumor types and subsequent mortality compared to older men.(17, 18) However, others have found that survival rates do not differ across age(19), and that tumor stage and grade advance with age.(20, 21) In a meta-analysis of the negative impact of young age on prostate cancer mortality, Parker et. al. identified 34 studies of 27,551 patients from the Medline Database between 1966 and 2000. After controlling for TNM stage, tumor grade, and PSA, young age failed to prognosticate increased risk of mortality. Furthermore, Carter et al, studied a series of 492 men undergoing prostatectomy for early stage disease, demonstrating a lower probability of ‘curable disease’ with advancing age, as well as a trend toward higher Gleason score with older age. These uncertain findings highlight the need for clinicians to consider the possible harms and benefits of treatment for both young and elderly men diagnosed with localized prostate cancer.
2.2 Co-morbidity

In addition to the age of the population most affected by the disease and the slow progressing nature of the disease, the high burden of co-morbid illnesses experienced by men diagnosed with prostate cancer provides significant context for this study. Co-morbidity is defined as "diseases that coexist in a study participant in addition to the index condition that is the subject of the study"(22) and can affect patient survival both in direct relation to the co-morbid illness or in tandem with the index disease.(23) The number of co-morbid illnesses a patient has increases with advancing age(17, 24), predicting competing causes for death that are important contributors to all-cause mortality rates, and therefore life expectancy estimates, among prostate cancer patients.

The co-morbidity burden of prostate cancer patients has been well documented in a variety of populations, including in Canada(25) and is an important predictor of survival for prostate cancer.(26-28) Eapen et al., demonstrated the combined effect of increased co-morbidities and increasing age on mortality among a group of 1,417 participants in the Canada Health Survey who were diagnosed with prostate cancer. Men aged 75-79, who suffered from 2 or more common co-morbidities were 5 to 7 times more likely to die during the observed follow-up, compared to a healthy 60-65 year old men with no co-morbidities. These results, demonstrate the prognostic importance of co-morbidities among prostate cancer patients in Canada.

In a case-cohort study of all curatively treated Ontario men diagnosed with prostate cancer between 1990-1998, Groome et al.(29) report strong trends in the relative risk of other-cause death due to the presence of co-morbid conditions, as documented by the cumulative illness rating scale (CIRS-G). In separate regression analyses performed on each of the organ system categories within the CIRS-G, they found that the risk increase was greatest for patients with cardiac and respiratory diseases. For patients with severe cardiac disease the risk of death is more than 2 times greater than patients without cardiac disease. For respiratory disease, the risk increase is more than 3-fold for
patients with severe respiratory disease compared to those with no respiratory disease. These results motivated the current thesis, specifically the objective of developing a life expectancy prediction method for prostate cancer using co-morbid illness status. The current thesis also utilized the same case-cohort population data to take advantage of the strong co-morbidity impact.

Furthermore, it has been demonstrated that co-morbidites have significant adverse effects on treatment outcomes among prostate cancer patients. Alibhai et al. (25) have documented the effect of co-morbidity on post-operative complications among radically treated patients in Ontario. Increasing co-morbidity was significantly associated with an increased risk of all categories of complications examined, with patients suffering from at least 2 co-morbid conditions being 5.2 times more likely to experience post-operative complications than patients with no co-morbid conditions, while men greater than 70 years old were only 1.3 times more likely to suffer from post-operative complications compared to men younger than 60 years old. These findings suggest that increasing co-morbidity is a better indicator of post-operative morbidities than age and therefore an important factor to consider in the treatment decision making process.

Many studies have demonstrated that the co-morbid illnesses suffered by men with prostate cancer put them at greater risk for death within 10 to 15 years than prostate cancer does. (30-32) In a retrospective cohort study of 767 men with localized prostate cancer from the Connecticut Tumor Registry, Albertsen et al, found that patients with two or more co-morbidities had significantly worse survival outcomes compared to men with one or fewer co-morbidities. Men with several co-morbidities were 1.9 times more likely to die of other causes than their healthier counterparts. Likewise, Soulie et al. found that among patients with localized prostate cancer, those with co-morbidities were more likely to die than men without co-morbidities. They found that men with ‘high’ Index of Coexisting Disease (ICED) (scores of 2 to 3) die of any cause within 10 years after diagnosis, while men with a score of 0 are much less likely to die within 10 years, and actually have better life expectancy, based on age at diagnosis, compared to the general
population. These findings demonstrate the high burden that co-morbid illness(es) place on men with prostate cancer, and highlight the fact that many of the early cases being detected today would not progress to a prostate cancer death and that curatively treating these men may be unnecessary.

2.3 Screening

Canadian Cancer Society guidelines (33) recommend that men over the age of 50 consult their doctor about being tested for prostate cancer. Digital Rectal Examination (DRE) and serum prostate-specific antigen (PSA) are the most common tests performed to screen for prostate cancer. If a patient presents with an irregular DRE and or elevated PSA level more tests, including a prostate biopsy, are performed in order to confirm a prostate cancer diagnosis.

Prostate cancer experts are currently split between those who advocate that every man should be screened and those who insist that no asymptomatic man should ever be screened. (34, 35) The purpose of screening should be to identify a group of asymptomatic men with early stage disease who could benefit from curative treatment. In order to “benefit” the group should experience prolonged survival and/or improved quality of life. There is strong evidence to support the fact that increased PSA screening has led to drastic increases in the incidence of prostate cancer diagnosed at an early stage. (36-38) However, the long awaited results of two recently published randomized controlled trial studies produced equivocal evidence that PSA screening saves lives. (39, 40) The Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial, conducted in the United States, found that the rate of death from prostate cancer was very low, and did not differ significantly between the screened group and the non-screened control group. In contrast, The European Randomized Study of Screening for Prostate Cancer (ERSPC) demonstrated a small survival advantage for screened men. Although the ERSPC results appear to support the use of wide spread screening, there is also evidence of significant over-diagnosis.
Over-diagnosis is the detection of latent disease that would not have been diagnosed in the patient’s lifetime in the absence of screening. Given the results of both of the aforementioned randomized controlled trials and other studies, over-diagnosis and subsequent overtreatment are recognized as the most important negative consequences of increased PSA screening. Rates of over-diagnosis in the United States and countries participating in the ERSPC have been estimated at 23 to 66%. The ERSPC study also concluded that it was necessary to screen 1,410 men and to have an additional 48 cancers diagnosed in order to prevent one prostate cancer death. Similarly, in their study of the effect of PSA screening on diagnosis and treatment in the USA, Welch and Albersten estimated that for 1 life to be saved due to the benefit of screening, at least 20 men would have to be diagnosed with the disease. They also concluded that over the past twenty years more than 1 million additional men have been diagnosed and treated for prostate cancer because of the introduction of PSA screening, further illuminating the issue of over-diagnosis. As a result of over-diagnosis, patients diagnosed with clinically insignificant tumors are subjected to unnecessary diagnostic tests and unneeded treatment, both of which can reduce a patient’s life expectancy and quality of life. Since evidence regarding the survival benefit of screening is inconclusive, issues of life expectancy and quality of life related to over-diagnosis and overtreatment are important in the decision-making process regarding whether to be screened.

2.4 Treatment

Curative treatment is available for those men whose disease is thought to be confined to the prostate at the time of diagnosis. Among men diagnosed with prostate cancer in the PSA era, approximately 85-90% are diagnosed with clinically localized or regional disease. This is a significant shift in the proportion of patients diagnosed with potentially curative prostate cancer, compared to the pre-PSA era. As a result, the sheer number of men diagnosed with prostate cancer undergoing curative treatment has dramatically increased. Two main forms of curative treatment exist, surgical removal of the prostate or high dose external beam radiotherapy.
Currently, there is no randomized trial evidence to prove that either treatment actually increases life expectancy, which is of particular concern when the cancer is diagnosed at a very early stage and/or when the patient’s life expectancy is short. Despite the lack of evidence for treatment efficacy approximately 90% of men with localized prostate cancer opt for some form of treatment.(47)

Although treatment related mortality rates are less than 1%(33, 46), both surgery and radiotherapy are associated with significant acute and long-term side effects, most commonly incontinence and impotence.(4, 48-51) In the population-based Prostate Cancer Outcomes Study, conducted in the United States, 8.4% of 1,291 men were incontinent and 59.9% were impotent at 18 or 24 months following radical prostatectomy. A systematic review of evidence of complications of radiation therapy shows that 20% to 40% developed erectile dysfunction and 2% to 16% of men developed urinary incontinence 12 to 24 months post-treatment. These treatment side effects are also experienced with greater frequency and intensity as patient’s age.(47, 52) Since neither radiotherapy or prostatectomy have been associated with an overall reduction in mortality and the disease may not become fatal in the patient’s remaining lifespan, the treatment decision making process is not straightforward; the complications associated with treatment must be weighed against the low risk of localized disease progressing to a prostate cancer death.

2.5 Life Expectancy

Life expectancy is the average number of years an individual of a given age is expected to live if current mortality rates continue to apply.(22) Tables of vital statistics, such as the Life Tables: Canada, Provinces, and Territories(53), provide the age-sex-specific average life expectancies of persons in a specified population. Life expectancy values are considered an important factor in the clinical decision making process,(54-56) as the discussion of expected remaining lifespan allows clinicians and patients to more fully explore a patient’s long-term prognosis.
Currently, there is no absolute cut-off for age when curative treatment for prostate cancer might be of benefit. Generally doctors only recommend treatment if they think the patient’s life expectancy is greater than 10 or 15 years, a life expectancy ‘window’ in which the benefits of treatment are believed to outweigh the harms. This life expectancy window has been adopted by several professional associations and appears in their treatment guidelines. Since the average, newly diagnosed, prostate cancer patient in Canada is 68 years old and is expected to live for an additional 14.8 years, he will fall within the recommended life expectancy time window to be considered for treatment (based on his age alone). In the treatment decision-making process, clinicians must define the patient’s life expectancy.

Traditionally, doctors use population-based, age-specific life tables combined with informal consideration of the patient’s general health and other illnesses to assess life expectancy. This approach is quick and easy, but is also often inaccurate and can lead to selecting the wrong group of patients for curative treatment. Koch et al examined the accuracy of clinicians selection of treatment candidates among a group of 261 patients who received radical prostatectomy and found that 20% had life expectancies of less than 10-years. A study by Krahn et al. to determine the accuracy of clinicians’ estimates of life expectancy in patients with localized prostate cancer found that clinicians’ estimates were accurate within one year 31% of time. The average prediction error of the clinicians’ was 2.4 to 5.2 years. Walz et al, examined the accuracy of clinicians life expectancy estimates by presenting them with the characteristics of 50 patients with localized disease. They found clinicians demonstrated an average accuracy, indicated by the c-index, of 0.68, with some accuracy estimates as low as 0.52 (where a c-index score of 0.50 = chance and 1.0 = perfect prediction).

Of particular concern, the traditional methods employed by doctors do not formally account for the effect of co-morbid illnesses. When a patient presents with some illness or disease that limits survival, their life expectancy is subsequently limited to some extent. Welch et al.
were able to demonstrate that formal consideration of a new, hypothetical condition, with a 5-year mortality rate of 5%, lowered life expectancy estimates by almost 2 years for a 65 year old male. These results make it clear that formal consideration of co-morbid information could potentially have a huge impact on the accuracy of life expectancy estimates. This is especially true considering the population most commonly diagnosed with prostate cancer - elderly men between the ages of 60-69 - are also likely to suffer from co-morbid illness(es) because of their old age.

2.6 Life Expectancy Calculation Tools

In the context of calculating the remaining life span of a patient, life expectancy is often defined as the average life span of a cohort of individuals who are observed after an inception point.(64) Calculating the life expectancy of a cohort of individuals involves following all members of the cohort until they have died, summing the total amount of time, e.g. life-months or years, that the cohort accrued and then dividing by the number of individuals in the cohort.(64) The area beneath the survival curve, for this cohort, translates directly to the average life expectancy of the cohort. Alternatively, the survival time associated with the median survival probability of the cohort, also translates to the average life expectancy of the cohort.(64) In most clinical settings, it is not practical or feasible for the doctors to collect data on a cohort of patients that is representative of their entire patient population. Therefore, they depend on life-tables, nomograms, and life-expectancy calculation tools that are developed using large cohorts of data representative of their patient population.

Advances in patient-specific life expectancy prediction involve calculation tools that employ commonly available patient data relevant to the population of interest. Many of these tools also model the hazard of death, using varying statistical methods in order to estimate life expectancy. The most useful calculation tool in a clinical setting is one that will produce a valid life expectancy estimate simply (i.e. can be evaluated using a hand-held calculator or spreadsheet) and with relatively few data inputs.
For example, a simple calculation tool used to estimate life expectancy involves a closed-form parametric approach and employs the exponential survival function to model the hazard of death. These tools have been found to be relatively accurate in estimating the actual remaining lifespan in several populations, but do not normally involve the consideration of other health status information other than age and sex.

There are life expectancy calculation tools that allow for the inclusion of information related to patient characteristics other than age and sex. Beck et al. have proposed a life expectancy calculation tool that extends the use of the exponential model to estimate the hazard of death, called the declining exponential approximation of life expectancy (DEALE). This model separates the excess hazard of death into several components, allowing for the consideration of additional prognostic information such as co-morbid status. The DEALE is currently the most widely used life expectancy calculation tool in clinical settings. The calculation of life expectancy following the DEALE format is possible with a hand held calculator, making it a convenient tool for use in the clinical setting.

The use of a simple exponential function to model life expectancy, such as that of the DEALE, assumes a constant mortality rate across time. In general, however, mortality rates increase over time with a doubling time of approximately 8.3 years. This assumption also does not fit the observed mortality for many diseases, for which the risk of death decreases with time. Benbassat et al. analyzed data for a group of 66,113 subjects in the Surveillance, Epidemiology, and End Results (SEER) tumor registry with bronchial cancer diagnosed from 1973 to 1985, to demonstrate the inaccuracies in life expectancy estimation that results from assuming constant excess mortality. They compared the life expectancy estimates obtained by a model that assumed constant disease specific “excess” mortality rates to those obtained by a time-variant model (allowing flexible mortality rates). The results show the model using constant mortality overestimated life expectancy by as much as 70% relative to the time-variant model.
Therefore, a life expectancy model that assumes a constant hazard overestimates the remaining lifespan for older populations, and conversely underestimates remaining lifespan for younger populations. This is of particular concern in the current thesis because we are dealing with an older patient population. By over-estimating LE in older patients there is the potential to misclassify a greater number of patients as treatment candidates. We would prefer LE estimates confer the opposite, more conservative, decision in either the screening or treatment decision making process. The use of a flexible mortality rate has been suggested as a method to overcome this lack of precision.

This issue has been addressed by Van den Hout with the development of the gamma mixed exponential (GAME) life expectancy calculation tool.(70) As its name suggests, the GAME models the hazard of death using the gamma function and mixed exponential modeling. These functions allow for the estimation of non-constant mortality rates, and as such may be more applicable to estimating life expectancy for prostate cancer patients. Yet, the GAME still involves making strong parametric assumptions about the form of the hazard function that if incorrect, could lead to biases and inaccuracies in the estimation of life expectancy.

In recent years there have been several new life expectancy calculation tools, developed specifically for the prediction of life expectancy in men with clinically localized prostate cancer.(71-73) These predictive tools model the hazard of death using Cox Proportional Hazards (PH) regression modeling.(74) The Cox PH model is a semi-parametric model; meaning that the form of the underlying hazard of death for the population of interest does not need to be specified in order to model the hazard function. Therefore, life expectancy prediction tools that utilize the Cox PH model may have an advantage over models that assume the hazard of death follows a parametric form. These calculation tools also allow for the consideration of important prognostic information, including co-morbid illness status. Furthermore, the tools are presented in easy-to-use formats that clinicians are familiar with, such as look-up style life tables and nomograms. Thus,
these tools have the potential to replace more traditional life tables and calculation tools in the clinical setting.

2.7 Summary

Prostate cancer is the leading non-skin cancer diagnosed in Canadian men, and most men are diagnosed with localized disease. Prostate cancer affects primarily older men and is often experienced as slow growing disease that will not cause symptoms or death in a patient’s lifetime. Prognosis for men with localized prostate cancer is fairly good and several forms of curative treatment are available. PSA screening and subsequent treatment for localized prostate cancer have both favorable and negative outcomes related to patient life expectancy and quality of life that are important considerations in the screening and treatment decision-making process. Doctors typically recommend curative treatment to men with clinically localized disease who are believed to have at least 10 years of remaining life. Clinicians’ estimation of patient life expectancy is typically based on intuition and experience with the population most affected by prostate cancer, but is often inaccurate. Co-morbid illnesses are of high burden and have been shown to be important prognostic indicators for men with prostate cancer. Life expectancy calculations that formally consider the effect of co-morbid illness are potentially more accurate than traditional estimation methods.

2.8 References


Chapter 3: Methods

3.1 Overview

The objective of this study was to develop a life expectancy prediction model that accurately captured the survival experience of men considered potential candidates for curative treatment of prostate cancer. Our aim was to develop a model that considered the impact of age and co-morbidity (as opposed to age alone), demonstrated good predictive accuracy, and would be a practical and easy-to-use aid for clinicians and patients. By considering the impact of age and co-morbid illness status (and not prostate cancer disease severity measures, commonly included in prostate cancer outcome prediction tools) the model may be applicable to either the screening or treatment decision-making process.

We analysed data from two main sources in our exploration of methods to predict life expectancy for men with clinically localized prostate cancer: 1) Statistics Canada Life Tables for Ontario males (1) and 2) a population-based retrospective cohort study. Using this data set we explored several approaches to the development of an LE prediction model with the above objectives in mind and tested each model on members of the sub-cohort population (a sub-set of the cohort data). The analysis of predictive accuracy focused on quantifying discriminative ability and calibration of the models we produced to determine the ‘best’ model. Supplementary to the development of the life expectancy prediction model, we also explored the development of a model that accurately predicted the probability of 10-year survival.

The remainder of this chapter is organized as follows. First, information on the study design, including details on each of the data sources, study populations and variables used in this project is given. Next, we provide broad details on the methods used to develop each of the life expectancy prediction models, as well as methods to test and compare the predictive accuracy of
the models, and to develop the ‘end product’ or ‘tool’ for clinical use. Lastly, we describe the methods employed for our supplementary analysis.

3.2 Study Design

The following sections provide a detailed description of populations used in the current study including the definitions, data sources, and data collection processes used to derive each of the populations. We also outline the variables derived and used from each population in the context of their application to LE prediction model development and testing.

3.2.1 Target Population

The target population was derived as part of a population-based retrospective cohort study, from data available via the Queen’s University Cancer Care & Epidemiology (CCE) Cancer database (described below). All Ontario prostate cancer patients diagnosed from 1990 to 1998 who were treated for cure with either radiotherapy or prostatectomy (N=45,035) were identified. Patients were included in the target population if they had either a prostatectomy or lymph node dissection within seven months of diagnosis, or radiotherapy within nine months of diagnosis, but did not have any of the following: a lymph node dissection without prostatectomy or radiotherapy in the presence of a bladder cancer diagnosis, a total curative radiation dose of less than 200cGy, curative radiation treatment lasting longer than 90 days, missing information on curative radiation dose, radiotherapy or prostatectomy or lymph node dissection more than 30 days prior to diagnosis. Patients in this population were initially followed for a maximum of 10 years (up until December 31, 1999). As shown in Figure 3.1, a total of 17,934 men met the initial inclusion criteria.

We subsequently updated the vital status information for this population using the most recent update of the CCE database (containing a death clearance date of December 31, 2008) to derive the final target population for the current study. There were some (n=64) patients who were present in the original CCE database that did not appear in the updated version of the database. It is
most likely that these patients were deleted from the Ontario Cancer Registry, which is a key component of the CCE database, because either the original prostate cancer diagnosis was incorrect, they were duplicate cases, or they may have lived outside the province. Patients were also excluded from the updated target population if at the time of diagnosis they were less than 50 years of age or greater than 79 years of age. We excluded these patients because we believed that a life expectancy tool to aid in the screening or treatment decision would not be applicable to men less than 50 or greater than 79 years of age. In the majority of cases of prostate cancer diagnosed in men less than 50, the decision to treat is clear; likewise for men older than 79 the decision not to treat is clear and does not require refined life expectancy estimates to guide this decision.(2, 3)
Figure 3.1 Development of final Target population

Prostate Cancer diagnosed between 1990-1998
N=45,035

Patients Treated with Curative Intent (Survival Information to 31 December, 1999)
N=17,934

Patients not in Updated OCR
N=64

Patients with Updated Survival Information (to 31 December, 2008)
N=17,870

Patients Excluded:
Age <50 or >79
N=439

Target Population
N=17,431
Our definition of the target population assumes that all prostate cancer patients who had a lymph node dissection alone in the absence of bladder cancer were prostate cancer treatment candidates. We performed an investigation to determine how this definition affected the development of our target population and subsequent case-cohort study population, details of which can be found in Appendix I. Using the more extensive data collected for the case-cohort study population we determined that approximately 18% of men who received a lymph node dissection were not curative treatment candidates. Since this misclassified group was small in terms of the entire cohort, we thought that their inclusion was unlikely to create any bias in our results.

Data Sources

The sampling frame used to form the target population was derived using the Queen’s University Cancer Care & Epidemiology (CCE) Cancer Database. The CCE Cancer Database gathers information from Ontario Cancer Registry (OCR) Composite Records. The OCR is a passive registry that identifies cancer cases on the basis of information from: 1) hospital discharges records received from the Canadian Institute for Health Information(CIHI) when cancer is recorded as a discharge diagnosis, 2) pathology reports of all cancer diagnoses from all acute care hospitals in Ontario, 3) electronic records from radiotherapy cancer centres and Princess Margaret Hospital (PMH), and 4) death certificates from the Registrar General of Ontario. The OCR records contain a unique identifier for each patient, which allows staff at CCE to link the case records in the OCR with electronic radiotherapy files from each radiation oncology department in Ontario and surgical procedure and discharge files from CIHI. Records for each patient therefore include information on date of diagnosis, age at diagnosis, vital status information, cause of death, and treatment. Further details of this database have been described elsewhere.(4-6)
Variables

The following variables describing the target population experience were considered in the life expectancy prediction models:

Survival time (i.e. time to death, t) was calculated by considering patient specific date of diagnosis and vital status information available from the CCE Cancer database. Study inception date was derived using information on date of prostate cancer diagnosis available from the CCE database. We identified patients who died within the study follow-up period, which ended on December 31, 2008, using vital status information. Each patient’s study termination date was set at the date of death. In keeping with the usual practice when using passive registry data, we otherwise assumed that patients were alive up to December 31, 2008 (the date of death clearance). Patient age at diagnosis was used to estimate the age effect in several of the life expectancy prediction models developed in this study and to describe the population’s baseline characteristics. Lastly, information on treatment type for each subject from the CCE database was used to describe this population, but was not included in model development.

3.2.2 Case-Cohort Study Population

The case-cohort study population consists of a sub-cohort and a set of cases which were randomly selected respectively from the target population. All cases died by December 31, 1999. This population was previously collected in the parent study 1) to examine the role that co-morbid illnesses play in other causes of death following curative radiotherapy or surgery and 2) to explore the difference in effectiveness between these treatments on cause-specific death.

The case-cohort design was first introduced by Prentice(7) as an alternative to the cohort design, to address the situation when confronted with a large cohort but a small number of events. In a typical case-cohort study all of the events are sampled and a random sample of the entire cohort is chosen. The design allows for more efficient sampling and is less resource-intensive than
classical cohort designs because it requires data collection for only a small fraction of the target population. (7) This was important in the parent study as the chart abstraction was an extremely labour intensive process. Because of the case-cohort sampling method there is often some overlap between the case group and the sub-cohort, where patients are selected into both groups.

Sampling for the case-cohort population was performed separately for prostate cancer cases and other-cause cases. Up to 75 prostate cancer death cases and up to 75 other-cause death cases were randomly selected from each of the eight Cancer Care Ontario Regions (CCORs). Up to 150 sub-cohort patients were randomly chosen from each CCOR, and CCORs that did not have 75 prostate cancer death cases and/or 75 other cause death cases sampled more than 150 sub-cohort patients to ensure adequate numbers from that region. Because of its large population size, the Central East region had a sample size that was doubled, with approximately 150 prostate cancer deaths, 150 other cause deaths and over 300 sub-cohort subjects. By design, the sampling of cases was done independently of the sub-cohort and this resulted in an overlap of some patients between the sub-cohort and cases. This sampling resulted in 55% of all prostate cancer deaths, 43% of all other cause deaths, and 10% of all patients being selected for the prostate cancer cases, other cause cases, and sub-cohort respectively. Details of the selection of the original study population are shown in Figure 3.2.
Figure 3.2 Development of Original Case-Cohort Study Population

- Prostate Cancer diagnosed between 1990-1998
  - N=45,035

  - Curative Treatment Candidates
    - n=17,934

  - Prostate Cancer deaths within 10 years
    - N=1,068
    - Prostate Cancer Cases
      - N=591

  - Other Cause deaths within 10 years
    - N=1,437
    - Other Cause Cases
      - N=631

    - Sub-Cohort
      - N=1,703
      - (Other Cause Cases n=162, Prostate Cancer Cases n=89)

- Total Case-Cohort Study Population
  - N= 2,740
For the current study all cases were combined to form one case group defined as patients who died of any cause, within 10 years of diagnosis (by 31 December, 1999). The case groups were combined to incorporate the competing risk of death due to prostate cancer when considering the effect of co-morbid illness status on all cause mortality. Therefore, it was necessary to modify the case group sampling because the original sampling fraction was higher in the prostate cancer cases group. Prostate cancer deaths were randomly removed from the original prostate cancer death group, by CCOR, to match the percentage of other-cause deaths originally sampled from each CCOR. Details of the original and modified sampling methods by CCOR are shown in Tables 3.1 and 3.2.

### Table 3.1 Original Case-cohort population sampling methods, by CCOR

<table>
<thead>
<tr>
<th>Region</th>
<th>Total Number</th>
<th>Sampling Number</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PD&lt;sup&gt;1&lt;/sup&gt;</td>
<td>OD&lt;sup&gt;2&lt;/sup&gt;</td>
<td>PD</td>
</tr>
<tr>
<td>CE</td>
<td>388</td>
<td>499</td>
<td>150</td>
</tr>
<tr>
<td>CW</td>
<td>172</td>
<td>210</td>
<td>75</td>
</tr>
<tr>
<td>E</td>
<td>110</td>
<td>113</td>
<td>75</td>
</tr>
<tr>
<td>NE</td>
<td>67</td>
<td>109</td>
<td>67</td>
</tr>
<tr>
<td>NW</td>
<td>30</td>
<td>35</td>
<td>30</td>
</tr>
<tr>
<td>S</td>
<td>46</td>
<td>71</td>
<td>46</td>
</tr>
<tr>
<td>SE</td>
<td>73</td>
<td>99</td>
<td>73</td>
</tr>
<tr>
<td>SW</td>
<td>182</td>
<td>301</td>
<td>75</td>
</tr>
</tbody>
</table>

1. PD = prostate cancer death, 2. OD = other cause death.

### Table 3.2 Modified case-cohort population sampling methods, by CCOR

<table>
<thead>
<tr>
<th>Region</th>
<th>Total Number</th>
<th>Original OD Percentage (%)</th>
<th>Modified PD Sample (#)</th>
<th># of PD randomly removed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PD</td>
<td>OD</td>
<td>PD/OD</td>
<td>PD</td>
</tr>
<tr>
<td>CE</td>
<td>388</td>
<td>499</td>
<td>30.06</td>
<td>117</td>
</tr>
<tr>
<td>CW</td>
<td>172</td>
<td>210</td>
<td>35.71</td>
<td>61</td>
</tr>
<tr>
<td>E</td>
<td>110</td>
<td>113</td>
<td>66.37</td>
<td>73</td>
</tr>
<tr>
<td>NE</td>
<td>67</td>
<td>109</td>
<td>68.81</td>
<td>46</td>
</tr>
<tr>
<td>NW</td>
<td>30</td>
<td>35</td>
<td>100</td>
<td>30</td>
</tr>
<tr>
<td>S</td>
<td>46</td>
<td>71</td>
<td>100</td>
<td>46</td>
</tr>
<tr>
<td>SE</td>
<td>73</td>
<td>99</td>
<td>75.76</td>
<td>55</td>
</tr>
<tr>
<td>SW</td>
<td>182</td>
<td>301</td>
<td>24.92</td>
<td>45</td>
</tr>
</tbody>
</table>
For the final case-cohort study population we excluded patients who (after review of their chart) were not candidates for prostatectomy or curative radiotherapy (n=50), and patients with unknown co-morbidity information (n=30). Patients were also excluded if at the time of their prostate cancer diagnosis they were less than 50 years of age (n=22) or older than 79 years of age (n=60) for the same reason we restricted age in the target population. These exclusion criteria result in a final case-cohort study population of 2,460 patients, as shown in Figure 3.3.

**Figure 3.3 Selection of final Case-Cohort population**
Data Sources

The case-cohort data were obtained from the CCE database and through a province-wide retrospective chart review. For the case-cohort population we were able to collect the same information from the CCE database as was obtained for the target population.

For the retrospective chart review seven data abstractors underwent a training process in Kingston, Ontario that was concluded when only a few minor errors remained in abstractions from standardized charts. An electronic abstraction form was used to minimize error, using Microsoft Access format and MedQuest software.(8)

The data abstractors abstracted the treating chart: hospital charts for surgical candidates and cancer centre charts for radiotherapy candidates, and also cancer centre charts when available for surgical cases. All charts were abstracted at their respective hospitals or cancer centres. If the data were incomplete on key elements, charts were accessed from secondary hospitals, Urologists and/or general practitioners’ charts from doctors’ offices. The pathology reports of the biopsies, the transrectal ultrasound reports, radiotherapy treatment records and surgical reports were photocopied and sent to the co-coordinating office. Information on any inconsistencies or missing variables were sent to the abstractor, and abstractors then either confirmed the illogical inconsistency was true or corrected the information, and sent any missing information that was available. This process continued until all inconsistencies were resolved for each patient.

Variables

The following variables were used in the development of the co-morbid illness status effect parameter estimate in the life expectancy models. Survival time (i.e. time to death, t) and censoring were derived using the same methods applied to the target population. However, due to restrictions of the case-cohort study design, date of death clearance for case-cohort analyses was December 31, 1999.
Patient age was derived as a descriptive variable as well as a confounding variable in analysis performed to derive the co-morbid illness effect input for the life expectancy prediction models.

Co-morbid illness burden was measured using a revised version of the Cumulative Illness Rating Scale for Geriatrics (CIRS-G); details of this instrument are provided in Appendix II. (9, 10) This scale was chosen based on the results of a pilot study that showed it to be the best overall performer of the five chart-based co-morbid illness scales tested. (11) Three of the five indices, including CIRS-G, predicted at least 7% of the variation and the authors suggested that any of these three methods would be appropriate to use in this population. The CIRS-G was the most comprehensive: capturing 59.2% of the pilot study cases’ listed cause of death as an identifiable co-morbidity at diagnosis. (11) CIRS-G was also easy to use, its instructions were very clear, and the abstractors had strong opinions that it captured the co-morbid illness information better than the other methods piloted. (11)

The CIRS-G is a medical history-based index of co-morbidity that has been used before for retrospective chart reviews, such as the one performed for this study. A patient’s full medical history over a 5-year pre-diagnosis period was used to gather information for this scale. Past and present illness and disease information is rated and summed across 14 categories (independent organ systems) to derive this scale. A five-point scale is used for each category to rate co-morbid illnesses or diseases on a scale of “0” for no problem to “4” for extremely severe conditions. Each morbidity category has its own individual scoring chart based on the general scoring guide. The total CIRS score is calculated by summing the individual scores from each category. The chart reviews collected detailed medical history information specific to the calculation of the CIRS index.

In an analysis performed by Dr. Groome, results of which were described in the literature review chapter, a revised version of the CIRS-G was developed with specific application to the
The revised score considers only organ systems that presented a significant finding with regard to the prediction of deaths from causes other than prostate cancer. It also combines some severity categories based on the similarity of effect determined by regression coefficients. The model fit of the continuous form of this revised score was also investigated. Adding 2 to all revised CIRS-G\textsubscript{pros} scores of 1 and above achieved a better fit to the categorical data for values up to about 10 (CIRS-G 8+2).(12)

The following variables were used to describe the case-cohort study populations’ characteristics, but were not included in LE prediction model development. All information on these variables was obtained from the chart abstraction.

Gleason score provides an indication of the aggressiveness of the prostate cancer.(13) The two most common patterns of growth based on area are each given a score from 1 to 5, with 1 representing tumors that are well-differentiated and 5 representing tumors that are poorly differentiated. The two scores are then added together to give an overall Gleason score between 2 to 10. Gleason score was analyzed using clinically meaningful categories of 2-4, 5-6, 7, and 8-10 as described by the American Joint Committee on Cancer (AJCC).(14)

The T category of a tumor describes the size of the tumor and how far it has spread locally. It ranges from T1 (tumor is not clinically apparent or palpable) to T4 (tumor invades local structures besides the seminal vesicles).(15) T categories of T1, T2, and T3 are further divided into subcategories based on the spread of the cancer and are named alphabetically from a to b (and c for the T category of T1).(15)

Prostate specific antigen (PSA) is a protein that is produced by the epithelial cells of the prostate. The levels of PSA in the blood increase naturally with age, however in men with prostate cancer the level of PSA exceeds normal age-related levels, increasing slowly and then exponentially.(16) For the current study, the PSA value recorded closest to the beginning of
treatment was used. PSA will be divided into commonly used clinical categories of $\leq 4$, $>4$ to $\leq 10$, $>10$ to $\leq 20$ and $>20$.

Information on treatment type for each subject was available from the CCE database and was verified using the chart review information. The main treatment modality was recorded, as was the use of planned adjuvant radiotherapy. Treatment types include: prostatectomy, external beam radiotherapy, a combination of prostatectomy and radiotherapy, and lymph node dissection.

3.2.3 Sub-Cohort Population

Patients who were originally selected into the case-cohort study population as sub-cohort members were selected to make-up the sub-cohort population. Details of this population are included in Figure 3.4, below. The population includes 1,596 patients.
Data Sources

The same vital status information collected for the target population, from the most recent version of the CCE cancer database (with death clearance as of December 31, 2008) was also applied to this population. Because all members of this group were also selected into the case-cohort study population, detailed chart abstracted co-morbidity information is also available for this group.

Variables

The following variables were used in the analysis to develop the baseline hazard, age effect, and co-morbid illness effect estimate inputs used in some of the life expectancy models developed in this study. We used the same survival time (t) information from the CCE database.
that was used for the Target population (i.e. death clearance date of December 31, 2008). We used the same age and CIRS-\(G_{\text{pros}}\) variables that we derived for the case-cohort population. To describe the sub-cohort population we used the same Gleason score, T-stage, PSA, and treatment variables derived for the case-cohort study population.

### 3.2.4 Ontario Life Tables

Statistics Canada periodically publishes age-sex specific tables of vital statistics, called Life Tables, for all provinces and territories. According to Statistics Canada “a life table represents a universally accepted demographic or actuarial model that synthesizes the mortality experience of a population”.(17) For this study we used the data available for Ontario men, during the period 1995-1997.(1) We used these Ontario life tables (OLT) to capture the mortality (and survival) experience of men who were potential candidates for curative treatment of prostate cancer. Many existing prognostic tools for prostate cancer utilize these (and other life tables developed using similar methodology) to estimate life expectancy.

Statistics Canada compiles these statistics by collecting data on the number of births and deaths at each age and for each sex during the time period indicated for the life tables. Both birth and death data were collected by the Health Statistics Division of Statistics Canada, by extracting birth and death data available from the Vital Statistics registrars for each province. Further details of the methodology used to produce these Life Tables have been detailed elsewhere.(17)

The analysis using this data included the following variables:

- \(\text{age}_x\) – age, in whole years
- \(p_x\) – the probability that person aged exactly \(x\) will survive to exact age \(x+1\). The entire column, \(s_p_x\), (for ages 0 to 104) represents the life table survival rate.
- \(e_x\) – life expectancy at age \(x\) (remaining lifetime), represents the average number of years remaining to be lived by persons surviving to exact age \(x\).
We chose to use data from the 1995-1997 publication of Life Tables (as opposed to earlier or more recent versions) because we felt the data available in those tables would best represent the men for whom we were attempting to predict life expectancy, i.e. men diagnosed with prostate cancer between January 1, 1990 and December 31, 1998.

3.3 Analysis Strategy

3.3.1 Overview

The following sections describe the methods of analysis performed in the current study. Details of the two general approaches to life expectancy model development are outlined, including the structure of the models and the estimation of model parameters. Then we describe the methods used to test and compare the predictive accuracy (PA) of the LE prediction models, as well as to determine their clinical usefulness. Next we outline the methods to develop life expectancy reference tables. And, lastly, we detail the supplementary analysis performed using logistic regression.

3.3.2 Life Expectancy Prediction Models

The life expectancy prediction models in this study took on one of two general forms (described below) based on either the Cox Proportional Hazards (PH) regression equation or the declining exponential function. Prediction models that follow the form of the Cox PH regression equation estimate a survival function, while the declining exponential function estimates life expectancy directly.

Accurately estimating the life expectancy of the men who are potential candidates for curative treatment of prostate cancer requires a statistical model that accurately explains the survival experience of the patients and the effect of age and co-morbid illness status (measured by CIRS-Gpros) on the survival experience. In this work we will consider traditional methods of LE
estimation using the population-based life table data available from Statistics Canada (described in section 3.2.4, above). As an alternative we will also consider LE estimation methods based on the Cox proportional hazards model. The PH model is considered because it does not require a strong parametric assumption and the model can be fit in some non-cohort based studies such as the case-cohort study considered in this work.

**Cox’s Proportional Hazards Regression Method:**

One approach to develop the life expectancy prediction models was based on the Cox proportional hazards model (Cox PH).(18) The Cox PH model consists of two main parts 1) the underlying or baseline survival function, describing how the probability of surviving beyond a specified time changes over time at baseline (i.e. mean) levels of included covariates, and 2) the hazard ratio describing how the hazard (risk) of death varies in response to included explanatory covariates.

Let $S_0(t)$ be the baseline survival function, and $\beta_1$ and $\beta_2$ be the hazard ratios for age and CIRS-G$_{pros}$ respectively, then Cox's PH model describes the effects of age and CIRS-G$_{pros}$ effects on the survival experience of patients by

$$S(t) = S_0(t)^{\exp((\beta_1 \cdot \text{AGE}) + (\beta_2 \cdot \text{CIRS-G}_{pros}))} \tag{1}$$

The advantage of the Cox PH model is that it requires no assumptions about the form of the baseline survival function, a problem associated with other proposed life expectancy calculation models.(19, 20) This method was also used to take advantage of the detailed vital status information available from the CCE database for all patients in the target population, and the co-morbidity information for patients in the case-cohort, and sub-cohort populations.

Given a fitted Cox’s PH regression model, the LE can be defined as the median of the survival distribution at given covariate values. That is, the time (t) when $S(t)=0.50$, by definition the median survival, can be considered the remaining life expectancy.(21) A macro was developed
to convert the survival function values into life expectancy estimates (details provided in Appendix IV).

Declining Exponential Method:

A second approach we considered in the development of our life expectancy prediction model was based on estimating survival using a simple exponential distribution, as originally proposed by Beck, et al.(22, 23) The model assumes a constant mortality rate as an approximation, and calculation of patient-specific life expectancy in this model requires the input of baseline and disease specific excess mortality rates for the population and disease(s) of interest. This method, referred to as the DEALE method, also allows for the input of mortality rates due to other causes or covariates of interest, such as age and co-morbid illnesses. These mortality rates can be combined by simple addition, ensuring easy application to clinical settings.(22) The life expectancy can be obtained by simply taking the reciprocal of the sum of all mortality rates input into the model, as shown in equation (2).

\[
LE = \frac{1}{\mu_{\text{baseline}} + \mu_{\text{age}} + \mu_{\text{CIRS-Gpos}}} \quad (2)
\]

Model Parameter Estimation

We used the PHREG procedure in SAS (version 9.2, SAS Institute, Inc., Cary, NC) to fit the Cox's PH model and to estimate the parameters in several LE prediction models we developed. The BASELINE statement of this procedure was used to calculate the probability of survival at mean values of included covariates.(24) Only covariates of interest, identified a priori were included in the MODEL statement of this procedure. This method of covariate selection is recommend by Harrell for optimal prediction, such that variables in the prediction rule are chosen on a conceptual basis rather than by variable selection algorithms.(25) The Cox PH model assumes a parametric form for the effects of the covariates, but it allows an unspecified form for the underlying survivor function.(26, 27)
We utilized OLT data to estimate parameters for the declining exponential method of life expectancy prediction. We used the values of $e_x$ (life expectancy at age $x$) to estimate mortality rates specific to each age of interest. We also used the values of $p_x$ (the probability that person aged exactly $x$ will survive to exact age $x+1$) to estimate the baseline survival function by the cumulative multiplication of the $p_x$ values, for each age $x$. Details of how these procedures were employed in the development of the life expectancy prediction models are included in the results section (Chapter 4).

3.3.3 Comparing Accuracy of Life Expectancy Predictions

To test the predictive accuracy of a life expectancy prediction model we compared the model-predicted life expectancy values to actual (observed) lifespan from the time of diagnoses with prostate cancer to death (or until censoring) for all members of the sub-cohort population. We chose to decompose predictive accuracy by testing two common components; discrimination and calibration. All models were compared based on these two criteria to chose a ‘best’ or ‘most accurate’ LE prediction model.

Discrimination

Discrimination measures a prediction model’s ability to separate patients with different responses. A model that correctly places each patient in the response class to which he truly belongs would be said to have perfect discrimination. The method chosen to measure the discriminatory ability of the models is the concordance index (c-index), as proposed by Harrell et al. For survival data, the c-index indicates the proportion of all usable pairs of subjects in which the predictions and outcomes are concordant. If the predicted survival time is larger for the patient who actually lived longer, the predictions for that pair are said to be concordant with the actual outcomes (otherwise the pair is said to be discordant). A pair of patients is considered usable if both patients have an observed survival time, or when one has the outcome and a shorter survival
time than the censored survival time of the other subject. (29) Given the set of all pairs of actual survival times \{X_1, X_2, ..., X_n\} and all model-predicted survival times \{Y_1, Y_2, ..., Y_n\}:

a concordant pair is defined as \(\pi_c = (X_i < X_j \text{ and } Y_i < Y_j) \text{ or } (X_i > X_j \text{ and } Y_i > Y_j)\)

a tied pair is defined as \(\pi_{tie} = (X_i = X_j \text{ and } Y_i < Y_j) \text{ or } (X_i = X_j \text{ and } Y_i > Y_j)\)

a discordant pair is defined as \(\pi_d = (X_i < X_j \text{ and } Y_i > Y_j) \text{ or } (X_i > X_j \text{ and } Y_i < Y_j)\)

To calculate the concordance index we assign 1 to each concordant pair, 0.5 to each tied pair, and 0 to each discordant pair. For all possible pairs we can define:

\[ c_{ij} = \begin{cases} 1 & \text{if the pair is concordant} \\ 0.5 & \text{if the pair is tied} \\ 0 & \text{if the pair is discordant} \end{cases} \] (3)

Denote by \(U\) the set of all usable pairs if subjects (i,j), then the c-index is can be estimated by equation 4.

\[ C = \frac{1}{Q} \sum_{(i,j) \in U} c_{ij} \] (4),

where \(Q\) is the number of all comparisons made (i.e. total number of usable pairs).

Calibration

Calibration refers to the extent of bias in the model’s predictions. Specifically, it describes how closely the predicted probabilities agree numerically with the actual outcomes. (29)

Calibration was measured using calibration plots. A calibration plot has predictions on the x axis, and the observed outcome on the y axis. For the current study we plotted mean predicted 10-year survival probabilities obtained from proposed models (for the sub-cohort population) grouped by quintiles, against actual Kaplan-Meier (K-M) survival estimates, with 95% confidence intervals around the actual survival at 10-years. A line of identity or reference at 45° was plotted to help for
orientation. Perfect predictions would lie on this line of identity, indicating that the model’s predictions align exactly with the observed outcomes.

3.3.4 Clinical Usefulness of ‘Most Accurate’ Model

We decided *a priori* to further assess the clinical usefulness of the ‘most accurate’ LE prediction model by exploring two further measures of accuracy. First, the accuracy of predictions across three clinically meaningful ‘time-windows’ of patient life expectancy was assessed. Secondly, we determined the optimal negative predictive value (NPV) associated with different ‘cut-offs’ of model-predicted probability of 10-year survival. These tests were intended to help clinicians assess the usefulness of the life expectancy model’s predictions as an aid in either the screening or treatment decision for potential curative treatment candidates.

3.3.5 Life Expectancy Reference Tables

In order to achieve our objective of developing a ‘product’ or ‘tool’ that was practical and easy-to-use in a clinical setting we created a table of our ‘most accurate’ model’s predicted life expectancy values. The table was formatted to include patient age (down the rows) and CIRS-Gpros (across the columns). Life expectancy values were calculated for all combinations of age and CIRS-Gpros. This resulted in an easy to use table whereby a clinician can easily find a patient’s estimated life expectancy by locating the row associated with the patient’s age and the column associated with the patient’s CIRS-Gpros score and reading the life expectancy value where that row and column meet.
3.3.6 Supplementary Logistic Regression Analysis

As an alternative to the outcome of remaining life years (produced by the life expectancy prediction models developed in this study) we also developed a logistic regression prediction model that focuses on 10-year survival probability only. Logistic regression is a useful alternative to linear regression (and Cox PH regression in the case of survival analysis) when the dependent variable is binary (i.e. includes only two outcomes). (32, 33)

The outcome of interest for this model was the probability of surviving 10-years from the date of diagnosis with prostate cancer. As detailed in the rationale and literature review 10-years of survival is a common cut-off used by clinicians in the treatment decision-making process. Therefore the binary dependent variable (Y) in this analysis was either died within 10 years of diagnosis (Y=0) or survived at least 10 years after diagnosis (Y=1). The logistic model was fit for the sub-cohort population using the LOGISTIC procedure in SAS (version 9.2, SAS Institute, Inc., Cary, NC). We included age and CIRS-G_pros as covariates in this analysis, in an attempt to increase the likelihood of accurately predicting 10-year survival probability. The results of this analysis allowed us to produce a table of predicted probabilities, similar to the life table style table of life expectancy values produced as the end ‘product’ or ‘tool’ in the main analysis.

3.4 References


48


Chapter 4: Results

4.1 Overview

The results chapter includes descriptive statistics for the target, case-cohort, and sub-cohort study populations. Patient specific characteristics relevant to life expectancy prediction model development include age, CIRS-Gpros score, and follow-up time. We also include the type of treatment received and several characteristics of the patient’s disease severity including PSA, T-stage, and Gleason score. Then, the development of each life expectancy calculation model is detailed, including the structure of each model, the methods, and the results of the analysis to derive parameter estimates for each model. We compare the apparent predictive accuracy of each model and choose the ‘most accurate’ among them. Next, results of further analysis of the clinical usefulness of the ‘most accurate’ model are presented. We also present the predictions of the ‘most accurate’ model in the form of life expectancy reference tables. Lastly, we include details of the methods and results related to the supplementary analysis using logistic regression modeling.

4.2 Description of Study Populations

4.2.1 Target Population

The characteristics of the updated target population are summarized in Table 4.1. The mean age of this population was 66.1 years (SD=6.2). The age distribution within the Target Population is slightly skewed to the right (older ages). The mean follow-up time from date of diagnosis to either death or censoring in the target population was 10.9 years, with a standard deviation of 4.2 years. This follow-up time was accumulated over a study induction period of Jan 1, 1990 to December 31, 1998, with follow-up occurring until December 31, 2008. The majority of men received either external beam radiotherapy or radical prostatectomy as their initial curative treatment type.
Table 4.1 Baseline characteristics of Target population.

<table>
<thead>
<tr>
<th>Age Distribution, n (%)</th>
<th>Target Population n= 17,431</th>
</tr>
</thead>
<tbody>
<tr>
<td>50 – 54</td>
<td>777 (4.5)</td>
</tr>
<tr>
<td>55 – 59</td>
<td>1949 (11.2)</td>
</tr>
<tr>
<td>60 - 64</td>
<td>3744 (21.5)</td>
</tr>
<tr>
<td>65-69</td>
<td>5478 (31.4)</td>
</tr>
<tr>
<td>70-74</td>
<td>4141 (23.8)</td>
</tr>
<tr>
<td>75-79</td>
<td>1342 (7.7)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Treatment Type, n (%)</th>
<th>Target Population n= 17,431</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lymph node dissection</td>
<td>1401 (8.0)</td>
</tr>
<tr>
<td>Surgery</td>
<td>6780 (38.9)</td>
</tr>
<tr>
<td>Radiotherapy</td>
<td>8835 (50.7)</td>
</tr>
<tr>
<td>Surgery + RT</td>
<td>415 (2.4)</td>
</tr>
</tbody>
</table>

4.2.2 Sub-cohort and Case-Cohort Populations

The characteristics of both the sub-cohort and case groups are summarized in Table 4.2 (below). We analyzed the characteristics of patients selected to the sub-cohort separately from the cases within the case-cohort study population. There was some overlap between the sub-cohort and cases (n=251) because of the sampling approach used in this design.

Sub-cohort group

Among members of the sub-cohort population the mean age was 66.7 years, with a standard deviation of 6.2 years. The age distribution was skewed to the right (older ages), this can be seen by the frequency of patients in each of six 5-year age groups presented in Table 4.2. Patients between the ages of 65 to 74 comprise more than 50% of the entire case-cohort study population. As with the Target population, the average follow-up in this population was 10.7 years (SD=4.1 years).

The sub-cohort had a mean CIRS-G_{pros} score of 4.1 (SD=2.5) and a median of 4.0 (note that all CIRS-G_{pros} scores of 1 and above have 2 added to them, as detailed in the methods
The histogram presented in Figure 4.1 (below) provides more details on the CIRS-G_{pros} distribution for the sub-cohort and cases. It demonstrates that very few sub-cohort patients had CIRS-G_{pros} scores greater than 6, and more than half had scores 4 or less. The majority of patients suffer from single organ system co-morbidities that are not considered severe. Ninety percent (90%) of sub-cohort members have severity scores of 1 or 2 for five of the seven single organ systems included in the CIRS-G_{pros} variable. Sub-cohort members have the most severe co-morbidities for cardiovascular disease (17% with severity scores of 3 or 4) and respiratory disease (35% with severity scores of 3 or 4).

The baseline disease severity characteristics of this population indicate the population was diagnosed with predominantly localized forms of prostate cancer. The majority of patients (71.8%) have PSA scores of less than 20. They predominantly have stage T1c, T2a, or T2b tumours, and most (85.1%) have Gleason scores of 7 or less. For over half (59.1%) of sub-cohort members, radiotherapy was indicated as their initial curative treatment modality. A large group (37.5%) also received radical prostatectomy, while very small numbers of patients received lymph node dissection only or a planned combination of radiotherapy and prostatectomy.

**Case Group**

The case group had a shorter mean follow-up time of 4.1 years (SD=2.2) because they were sampled as of December 31, 1999. The case group had a mean age of 68.1 years (SD=6.1). The average CIRS-G_{pros} score among the cases was 5.1 (SD=2.9). The histogram in Figure 4.1 demonstrates that the CIRS-G_{pros} distribution for the cases contains a larger percentage of scores 6 and higher. However, almost all cases in the case-cohort population also suffer from a combination of mild co-morbidities from different organ systems. Ninety percent (90%) of cases also have single organ system scores of 1 or 2 for five of the seven organ systems. Cardiovascular and respiratory diseases are experienced in greatest severity by the cases, with 25% and 43% of patients having scores of 3 or 4 for these two single organ systems respectively.
The cases have baseline disease severity characteristics that are more severe than the sub-cohort group, but still indicate mainly early stage disease. The majority of cases (72.6%) received radiotherapy as their initial curative treatment modality. Radical prostatectomy was the initial treatment for 23.3%.
### Table 4.2 Baseline Characteristics of sub-cohort and cases separately.\(^1\)

<table>
<thead>
<tr>
<th></th>
<th>Sub-cohort (n=1,596)</th>
<th>Cases (n=1,068)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age Distribution n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 – 54</td>
<td>63 (4.0)</td>
<td>33 (3.1)</td>
</tr>
<tr>
<td>55 – 59</td>
<td>160 (10.0)</td>
<td>69 (6.5)</td>
</tr>
<tr>
<td>60 – 64</td>
<td>302 (18.9)</td>
<td>174 (16.3)</td>
</tr>
<tr>
<td>65 – 69</td>
<td>507 (31.8)</td>
<td>315 (29.5)</td>
</tr>
<tr>
<td>70-74</td>
<td>417 (26.1)</td>
<td>331 (31.0)</td>
</tr>
<tr>
<td>75-79</td>
<td>147 (9.2)</td>
<td>146 (13.7)</td>
</tr>
<tr>
<td><strong>T stage, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tx</td>
<td>21 (1.3)</td>
<td>19 (1.8)</td>
</tr>
<tr>
<td>T1a</td>
<td>17 (1.1)</td>
<td>9 (0.8)</td>
</tr>
<tr>
<td>T1b</td>
<td>93 (5.8)</td>
<td>89 (8.3)</td>
</tr>
<tr>
<td>T1c</td>
<td>372 (23.3)</td>
<td>147 (13.8)</td>
</tr>
<tr>
<td>T2a</td>
<td>463 (29.0)</td>
<td>240 (22.5)</td>
</tr>
<tr>
<td>T2b</td>
<td>508 (31.8)</td>
<td>411 (38.5)</td>
</tr>
<tr>
<td>T3a</td>
<td>66 (4.1)</td>
<td>77 (7.2)</td>
</tr>
<tr>
<td>T3b</td>
<td>46 (2.9)</td>
<td>52 (4.9)</td>
</tr>
<tr>
<td>T4</td>
<td>11 (0.7)</td>
<td>24 (2.3)</td>
</tr>
<tr>
<td><strong>Pre-treatment PSA, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>148 (9.3)</td>
<td>140 (13.1)</td>
</tr>
<tr>
<td>&lt; 4</td>
<td>160 (10.0)</td>
<td>112 (10.5)</td>
</tr>
<tr>
<td>4-10</td>
<td>553 (34.6)</td>
<td>237 (22.2)</td>
</tr>
<tr>
<td>10-20</td>
<td>433 (27.1)</td>
<td>223 (20.9)</td>
</tr>
<tr>
<td>&gt;=20</td>
<td>302 (18.9)</td>
<td>356 (33.3)</td>
</tr>
<tr>
<td><strong>Gleason Score, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>56 (3.5)</td>
<td>49 (4.6)</td>
</tr>
<tr>
<td>&lt; 6</td>
<td>572 (36.0)</td>
<td>316 (30.0)</td>
</tr>
<tr>
<td>6</td>
<td>447 (28.1)</td>
<td>247 (23.4)</td>
</tr>
<tr>
<td>7</td>
<td>344 (21.7)</td>
<td>221 (21.0)</td>
</tr>
<tr>
<td>&gt;7</td>
<td>170 (10.7)</td>
<td>222 (21.0)</td>
</tr>
<tr>
<td><strong>Treatment Type, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lymph node dissection</td>
<td>43 (2.7)</td>
<td>41 (3.8)</td>
</tr>
<tr>
<td>Surgery</td>
<td>598 (37.5)</td>
<td>249 (23.3)</td>
</tr>
<tr>
<td>Radiotherapy</td>
<td>945 (59.1)</td>
<td>775 (72.6)</td>
</tr>
<tr>
<td>Surgery + Radiotherapy</td>
<td>11 (0.7)</td>
<td>3 (0.3)</td>
</tr>
</tbody>
</table>

\(^1\) Note there are 251 cases who overlap between the sub-cohort and case groups.
4.3 Life Expectancy Prediction

In this section we propose several life expectancy prediction models and obtain life expectancy estimates for each model. The first three models we considered are based on the Cox PH model and differ in the data used to estimate parameters and other small adjustments made in an attempt to predict survival experience with greater accuracy. Two subsequent are based on declining exponential method and utilize a combination of Ontario life table data and Cox PH modeling to estimate parameters.
4.3.1 Model #1 – Sub-Cohort Model

We considered the sub-cohort population data exclusively to develop this model. The advantage of using the sub-cohort population was that it provided us with both the most recent vital status information (not available in the case-cohort population) and the detailed CIRS-Gpros information (not available in the target population). Since the sub-cohort population is a randomly selected (albeit geographically stratified) sample of the target population, we felt that the baseline hazard and hazard ratios derived from the analysis of this population would be relevant to estimate the survival experience for our population of interest – potential candidates for curative treatment.

The proposed model for the sub-cohort data is a Cox's PH model, and the survival function of the patients in the sub-cohort data is modeled by equation (5) below.

\[
S(t) = S_0(t)^{\exp(\beta_1(AGE-\bar{AGE})+(\beta_2(CIRS-CIRS)))} \quad (5)
\]

where \( S_0(t) \) is the baseline survival function, which reflects the survival experience at the mean values of age (\( \bar{AGE} = 66.7 \)) and CIRS-Gpros (\( \bar{CIRS} = 4.1 \)) for members of the sub-cohort population, and \( \beta_1 \) and \( \beta_2 \) are the effects of the continuous form of age and CIRS-Gpros on the survival experience of the patients.

The SAS procedure PHREG is used to estimate the parameters in model (5) and the estimated covariate effects are shown in Table 4.3, Row 1 (below). The regression coefficient for age, while controlling for the confounding effect of CIRS-Gpros, was 0.082 (p<0.0001). This coefficient corresponds to a hazard ratio of 1.09 (95% CI 1.07, 1.10). It implies that for each 1-year increase in age, the risk of all-cause mortality increase 1.09 times, while holding the effect of CIRS-Gpros constant. The beta coefficient for CIRS-Gpros, controlling for the confounding effect of Age, was 0.127 (p<0.0001). This results in an all-cause mortality risk increase of 1.14 times, per 1-unit increase in CIRS-G score.
Table 4.3 Results of the development of inputs derived for use in the Case-cohort Target, Simplified Case-cohort Target, and Sub-cohort LE prediction models.

<table>
<thead>
<tr>
<th></th>
<th>AGE</th>
<th>CIRS-Gₚᵣₒₓ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>SE</td>
</tr>
<tr>
<td>Sub-Cohort¹</td>
<td>0.082</td>
<td>0.007</td>
</tr>
<tr>
<td>Target²</td>
<td>0.079</td>
<td>0.002</td>
</tr>
<tr>
<td>Case-Cohort³</td>
<td>0.020</td>
<td>0.006</td>
</tr>
</tbody>
</table>

1. From Cox PH Regression Analyses (CIRS-Gₚᵣₒₓ and Age included as covariates) of Sub-Cohort (N=1,596).
2. From Cox PH Regression Analysis (Age included as covariate) of Target Population (N=17,431).
3. From Cox PH Regression Analyses (CIRS-Gₚᵣₒₓ and Age included as covariates) of Case-Cohort (N=2,416) (Langholz Case-Cohort Survival Analysis Methodology used).

Based on the fitted model (5), the LE of a patient given age (AGEᵢ) and CIRS-Gₚᵣₒₓ (CIRSᵢ) can be calculated by the method described in section 3.3.2.

Predictive Accuracy

To evaluate the performance of the Sub-Cohort model in LE estimation, we compared the predicted LEs of patients in the sub-cohort with their actual survival times using the c-index, as described in section 3.3.3. The resulting c-index was 0.65. This means that the Sub-Cohort model correctly orders the time to death, for any usable pair of patients in the sub-cohort population 65% of the time. A value of 0.5 indicates no predictive discrimination, beyond that obtained by chance, and a value of 1.0 indicates perfect discrimination. Therefore this model indicates fairly modest discriminative ability.

The calibration plot for the model (5)’s predictions of 10-year survival for members of the sub-cohort population is presented in Figure 4.2a. A calibration plot is meant to illustrate the degree of correspondence agreement between model-predicted and actual 10-year survival probabilities (i.e. the extent of bias in the model’s performance). This figure demonstrates good
calibration at all quintiles of predicted 10-year survival probability. The line connecting the mean-predicted values of each quintile lies close to the line of identity (45° line), and the 95% CI for observed 10-year survival also contain this line. The model appears to perform best for the highest and lowest quintiles of model-predicted 10-year survival, as predictions lie very close to the line of identity. The bias present in the Sub-Cohort model’s predictions demonstrate no obvious trend; the model-predicted values appear to slightly over-estimate actual survival for some quintiles, but slightly under-estimate actual survival for others.

Figure 4.2a Calibration Plot for Sub-Cohort LE Prediction Model
The Impact of Co-morbid Illness status on Predictive Accuracy

The objective of this study was to improve on the accuracy of LE prediction by considering a measure of co-morbid illness status, in addition to age, in the assessment of remaining lifespan. To assess the impact of formally considering the effect of CIRS-$G_{pros}$ we considered a secondary Sub-Cohort prediction model (the ‘Age Only’ model), that is model (5) with the effect of CIRS-$G_{pros}$ eliminated. Therefore this secondary model only considers the baseline survival function and hazard due to age in the Cox PH analysis of the sub-cohort population. We tested this ‘Age Only’ model by comparing the predicted LEs of patients in the sub-cohort with their actual survival times using the c-index and calibration plot.

The c-index that resulted from predicting the life expectancy of members of the sub-cohort population using this model was 0.60. This indicates a relative improvement of approximately 5% in discriminative ability when the impact of CIRS-$G_{pros}$ is formally considered in the prediction model. The calibration plot, Figure 4.2b, demonstrates the performance of the ‘Age Only’ model is worse than the original Sub-Cohort model. At the mid-range of 10-year survival probabilities (0.70 < Pr(T>10) < 0.85) calibration appears similar to that of the full Sub-Cohort prediction model. However, for all other probabilities, the ‘Age Only’ model calibrates worse than the Sub-Cohort model.
4.3.2 Model #2 – Case-Cohort Target LE Prediction Model

This PH model combines a baseline survival estimate and age effect coefficient in a Cox's PH model for the target population with a co-morbidity effect coefficient in a Cox's PH model for the case-cohort study population. The model is presented in equation (6) below

\[ S(t) = (S_0(t)^\alpha \exp((\beta_1 \cdot \text{AGE} - \bar{\text{AGE}}) + (\beta_2 \cdot \text{CIRS} - \bar{\text{CIRS}})) \] (6),

where \( S_0(t) \) and \( \beta_1 \) are from the Cox model \( S(t) = S_0(t)^{\exp(\beta_1 \cdot \text{AGE})} \) for the target population, \( \beta_2 \) is from the Cox PH model \( S(t) = S_0(t)^{\exp(\beta_0 \cdot \text{AGE} + (\beta_2 \cdot \text{CIRS} - \bar{\text{CIRS}}))} \) for the case-cohort data, and \( \alpha \) is an adjustment to the baseline survival function \( S_0(t) \), and age coefficient \( \beta_1 \) from the Cox PH model for the target population due to the missing CIRS-G_{pros} information. The values for \( \bar{\text{AGE}} = 66.1 \) is
the mean value of age from the target population (the data used to derive the age effect $\beta_1$), and $CIRS = 4.1$ is the mean CIRS-$G_{\text{pros}}$ value for the case-cohort population (the data used to derive the co-morbidity effect $\beta_2$).

The estimated $\beta_1$, an estimate of the effect of age on the hazard of death due to any cause from the Cox PH model for the target population model is presented in Table 4.3, Row 2 (above). The age effect coefficient was $\beta_1 = 0.079$. The corresponding hazard ratio is $1.08$ (95% CI: 1.078, 1.087), meaning that for each 1-year increase in age a patient is approximately 1.1 times more likely to die. We hypothesized that because of its very large population size the target population would provide more stable and robust estimates of $S_0(t)$ and $\beta_1$, compared to the other data sources available to us.

When we fit the model, we also performed an analysis of the fit of the continuous vs categorical form of the age variable (described in detail in Appendix III) and concluded that the continuous form fit the data well.

The second Cox PH regression model utilized the case-cohort population to estimate $\beta_2$ the effect of co-morbid illnesses on death due to any cause after age is controlled. A similar Cox’s PH model was previously used by Dr. Groome.(14) to estimate the effects of co-morbidity and age on survival when only deaths not caused by prostate cancer are considered as the event of interest, while we considered all deaths due to any causes. Since the data are from a case-cohort study, it was necessary to incorporate some adjustments to the standard partial likelihood method to fit the Cox proportional hazards model. Of particular concern was dealing appropriately with the weighting of cases. The subjects selected to this population were not a random sample because the cases were over-sampled and were chosen based on their outcome status. Therefore the number of events in the study is much higher than would have occurred in a cohort study.

Prentice recognized this design issue and proposed to account for it by using a pseudolikelihood instead of partial likelihood to estimate the hazard ratios.(2) Using this method, a
case that occurs outside of the sub-cohort is not considered to be at risk until just prior to failure and is not included in these risk sets until just prior to failure. (3) Prentice, Self and Prentice, and Barlow have proposed weighting schemes for the sub-cohort and different methods to handle non-sub-cohort cases at failure. (2, 4, 5) Each method provides a way to calculate a robust standard error to adjust for the case-cohort sampling. (6) We incorporated the methods of proposed by Therneau and Li that provides a straightforward way to calculate point estimates and variances for Cox’s proportional hazard regression for case-cohort studies using the method of Self and Prentice. (29)

A second concern due to the case-cohort design is proper estimation of the variance given the stratified nature of the study sample. Recent work by Langholz and Jiao (7) addresses the stratification issue and Dr. Groome has consulted with Professors Bryan Langholz, Sven Ove Samuelsen and Yingwei (Paul) Peng to solve the case-sampling issue. (1) We were able to incorporate this approach by including a SAS macro developed by Langholz (8) in our analysis. These methods provided us with stable parameter estimates, robust variance estimates, and hazard ratios associated with the effect of each covariate included in the model. Table 4.3, Row 3, shows the results for the Cox PH model derived from the case-cohort population to quantify the effect of co-morbid illness status on death due to any cause. The CIRS-G_pros coefficient was $\beta_2 = 0.111$. The corresponding hazard ratio is 1.12 (95% CI: 1.09, 1.14), meaning that for each 1-unit increase in CIRS-G_pros score a patient's likelihood of dying increases by 12%.

We considered incorporating age-group specific CIRS effects in the model by stratifying by age group. The data demonstrated significant variation in CIRS-G_pros scores across 5-year age groups. The mean and median CIRS-G_pros scores varied between 2 and 5 depending on age, with the lowest scores appearing at the youngest ages and the highest scores for ages 65-74. However, the size of individual age groups at younger ages was small and would have resulted in an unstable estimate of CIRS effect. One of the strengths of the Dr. Groome’s work to calculate the CIRS effect is the stability of the estimates produced. Therefore we decided not to use age-group specific
CIRS effects, as we felt it would take away from one of the main strengths of the current study. This decision does assume that the CIRS effect is constant across age groups.

The inclusion of $\alpha$ in the Case-Cohort Target model equation (6) was done to account for bias that may result from unmeasured confounding due to co-morbid illness. Ideally, we would have been able to estimate the age effect and CIRS-$G_{pros}$ effect by building a Cox-PH model from the target population with both variables included as covariates. We were unable to do this because CIRS-$G_{pros}$ information was not available for the entire target population. Appendix V explains how to estimate this adjustment ($\alpha=0.78$). This adjustment was not necessary in model (5) because both age and CIRS-$G_{pros}$ are available in patients in the sub-cohort population.

We anchored the age and CIRS-$G_{pros}$ estimates at the mean age ($\text{AGE} = 66.1$) of the target population and mean CIRS-$G_{pros}$ ($\text{CIRS} = 4.1$) in the case-cohort population. These values differ slightly from those used in the Sub-Cohort LE prediction model because we used the target population to derive the age parameter ($\beta_1$) and case-cohort population to derive the co-morbidity parameter ($\beta_2$) in this model.

Predictive Accuracy

The overall c-index for comparing model (6) LE estimates to the actual survival times of the patients in the sub-cohort population was 0.64. We interpreted this result to mean that the Case-Cohort Target model correctly orders the time to death, for any usable pair of patients in the sub-cohort population 64% of the time. This value is 0.01 smaller than the c-index value obtained for the Sub-Cohort LE prediction model (c=0.65), and therefore the two models would be considered more or less equal in terms discriminative ability.

Figure 4.3 shows the calibration curve for the Case-Cohort Target model. The calibration plot for this model illustrates quite poor accuracy. The model performs fairly well for lowest quintile of predicted probabilities, however, at higher probabilities the 95% confidence interval

63
(around actual K-M 10-year survival probability) does not even cross the line of identity (dashed line). A perfectly calibrated model would lie directly on that line. It would be expected that the model would at least demonstrate 95% confidence intervals that include the reference line. The predicted probabilities appear to consistently underestimate the actual survival probabilities.

Figure 4.3 Calibration Plot for Case-Cohort Target LE Prediction Model
4.3.3 Model #3 – Simplified Case-Cohort Target Model

The third model we developed, the Simplified Case-Cohort Target model, attempted to improve both the discriminative ability and calibration of the Case-Cohort Target model by removing some of the complexities and assumptions associated with it. Therefore the adjustments (\( \alpha=0.78 \)) made to the baseline survival function \( S_0(t) \) and age effect parameter (\( \beta_1 \)) in the Case-cohort Target model were removed from this model equation, presented in equation (7), below. As such this equation is identical to that for the Sub-Cohort LE prediction model, presented in equation (5). But, the parameters including in this model are the same as those estimated for the Case-Cohort Target model.

\[
S(t) = S_0(t)\exp(\beta_1(AGE_i-\bar{AGE}))+\beta_2(CIRS_i-\bar{CIRS}))
\]  

(7)

Predictive Accuracy

To evaluate the performance of the Simplified Case-Cohort Target model’s predictive ability, we compared the model (7) predicted LEs of all patients in the sub-cohort with their actual survival times, using the c-index. The resulting c-index was 0.63, which indicates the Simplified Case-Cohort model actually performed slightly worse than the original Case-Cohort Target model. Similar to the both of the previous LE predictions models, this model demonstrates modest discriminative ability. However, of the three models the Simplified Case-Cohort Target model demonstrates the poorest discriminative ability.

The calibration for the Simplified Case-cohort Target LE prediction model is shown in Figure 4.4a. This model demonstrates fairly good predictive accuracy. Compared to the two previous models, it performs better than the Case-Cohort Target model, but slightly worse than the Sub-Cohort model. At lower predicted probabilities of 10-year survival this model does not perform well, as the line of identity does not fall within the 95% CI surrounding the observed 10-
year survival. However, at all other quintiles of 10-year survival probability the model performs well.

**Figure 4.4a Calibration Plot for Simplified Case-Cohort Target LE Prediction Model**

![Calibration Plot](image)

The Impact of Co-morbid Illness status on Predictive Accuracy

In order to quantify the effect of including CIRS-G_{pros} in the calculation of life expectancy for both the Case-Cohort Target model and Simplified Case-Cohort Target model, we developed a secondary (‘Age Only’) life expectancy prediction model that included only the parameters estimated from the Cox PH modeling of the target population. Therefore this model includes the estimate of baseline survival function, \( S_0(t) \), and age effect parameter (\( \beta_i \)) included in both the Case-Cohort Target and Simplified Case-Cohort Target LE prediction models.

Testing this model on the sub-cohort population resulted in a c-index of 0.61. Both the Case-Cohort Target model (C=0.64) and the Simplified Case-cohort Target model (C=0.63)
improve on the predictions of this ‘Age Only’ model. Therefore it appears that by including CIRS-
G$_{pros}$ as a measure of co-morbid illness status we improve predictions based on the c-index by 0.02
to 0.03.

The calibration plot for the Age Only model (Figure 4.4b, below) demonstrates poor
predictive accuracy. Although it can be difficult to quantify and compare the level of calibration
between the three models, it does appear that the Simplified Case-cohort Target model makes some
improvement over ‘Age Only’ model, while the Case-cohort Target model has similarly poor
calibration to the ‘Age Only’ model.

**Figure 4.4b Calibration Plot for 'Age Only' Target LE Prediction Model**
4.3.4 Comparing the Three Cox PH Type LE Prediction Models

To determine overall predictive accuracy both discriminative ability and calibration should be considered.(9) A model that demonstrates good discrimination does not automatically imply good calibration and vice versa. However, it has been argued that more weight should be given to discriminative ability, when comparing two prediction models.(9, 10)

The Sub-Cohort LE prediction model appears to be the ‘best’ or ‘most accurate’ of the three models presented thus far in this study, which all utilize Cox PH regression modeling. The Sub-Cohort LE prediction models performs best in terms of discriminative ability, with a c-index of 0.65, compared to 0.64 and 0.63 for the Case-Cohort Target and Simplified Case-Cohort Target model respectively. The Sub-Cohort model also appears to demonstrate the best model calibration. As demonstrated in Figure 4.2a, the calibration for this model is very good, as the mean-predicted 10-year survival probabilities for all quintiles lie close to the line of identity, and all 95% CI intervals around actual K-M survival also contain the line of identity. Generally performance appears worse for both of the Case-Cohort Target model and Simplified Case-Cohort Target model (Figures 4.3 and 4.4a). Specifically the 95% CIs did not contain the line of identity for at least one quintile, for both models.

4.3.5 Assessing Proportional Hazards Assumption

Even though the Sub-Cohort LE prediction model performed the ‘best’ of the first three models developed, the discriminative ability of this model is still somewhat lower than we would have expected. We decided to investigate the proportional hazards (PH) assumption employed in this prediction model.

The assumption of proportional hazards, which is made when we perform any Cox PH regression analysis, is that the hazard function for an individual (i.e., patient in the analysis) depends on the values of the covariates and the value of the baseline hazard. Given two patients
with particular values for the covariates, the ratio of the estimated hazards over time will be constant - hence the name of the method: the proportional hazard model.\(^{(11, 12)}\)

We first plotted the natural log of the cumulative hazard function or log (-log (Kaplan-Meier survival)) for deciles of the age variable against survival time in Figure 4.5a. Several of the K-M curves appear to cross for times up to approximately 8 years, but are relatively parallel at times thereafter. If the proportional hazards assumption holds true for the data, the K-M curves remain relatively parallel (i.e. do not cross). Therefore this plot suggests that the proportional hazards assumption may be violated for the age variable in the Sub-Cohort model.
Figure 4.5a Plot of Log (Cumulative Hazard) vs Survival Time, by Deciles of Age

Figure 4.5b Score Process Plot for Cox PH modeling of Sub-Cohort Data
We performed a more stringent test of the PH assumption for analysis of age by including the ASSESS statement in the PHREG procedure of SAS used to develop the model inputs. This statement performs the graphical and numerical methods proposed by Lin, et al(13) for assessing the adequacy of the Cox PH regression model.

The results of this analysis are displayed in Figure 4.5b (above). This plot demonstrates that the age variable is in violation of the PH assumption. The solid line on this plot represents the observed score process. When the proportional hazards assumption is upheld the observed path will lie somewhere in the middle of the simulated paths (the other lines plotted). It is obvious in this case that the observed is atypical of the simulations and therefore the PH assumption is violated. Furthermore the p-value in the lower right-hand corner of this figure indicates that the probability that the absolute maximum from any of the 1000 simulations exceeds that of the observed standardized score process is very low (0.0250). This result supports that of Figure 4.5a, indicating non-proportional hazards for the age variable in this model. This suggests that the form of the hazard of death due to age specified by the Cox PH model may not adequately fit the data. This may explain the modest discriminative ability (c-index) results obtained and suggests that Cox PH analysis may not be appropriate to estimate the hazard of death due to age in this population.

We also assessed the PH assumption for the CIRS-\(G_{\text{pros}}\) variable by first creating Figure 4.6 (below), a plot of the log (-log(K-M survival)) curves for four categories of CIRS-\(G_{\text{pros}}\) versus follow-up time. The curves representing the four of CIRS-\(G_{\text{pros}}\) displayed in this plot remain relatively parallel, and do not cross at any time. This demonstrates that the proportional hazards assumption holds for CIRS-\(G_{\text{pros}}\).
Further Exploration of Sub-Cohort Model properties

Another issue with the predictive accuracy results for the Sub-Cohort model that we wanted to investigate was the discrepancy between the two components of predictive accuracy; discrimination and calibration. The Sub-Cohort model demonstrated modest discrimination, yet good calibration. As indicated previously, these two tests of predictive accuracy should not be considered as standalone tests of overall accuracy, nor does a good result for one necessarily imply an equally good result for the other.

We grouped patients who died during study follow-up according to the quintile of predicted 10-years survival probability they were assigned to for the calibration analysis (Figure 4.2a, above). For each of these quintiles we calculated a c-index. This analysis allowed us to
explore the discriminative ability of the different subsets of the sub-cohort population grouped by quintiles of survival probability. Table 4.4 (below) includes c-index values for each of these groups.

<table>
<thead>
<tr>
<th>Quintile</th>
<th>Range of Predicted 10-Year Survival Probabilities</th>
<th>c-index</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>&lt;0.55</td>
<td>0.57</td>
</tr>
<tr>
<td>2</td>
<td>0.55-0.64</td>
<td>0.53</td>
</tr>
<tr>
<td>3</td>
<td>0.65-0.74</td>
<td>0.52</td>
</tr>
<tr>
<td>4</td>
<td>0.75-0.82</td>
<td>0.52</td>
</tr>
<tr>
<td>5</td>
<td>0.83-0.96</td>
<td>0.42</td>
</tr>
</tbody>
</table>

The calibration plot for the Sub-Cohort LE prediction model (Figure 4.2a) shows that the first quintile calibrates very well with observed survival probabilities – with mean predicted 10-year survival probability (0.44) and observed 10-year survival probability (0.43) being almost identical, and therefore lying on the line of identity. However, when we look at the individual discriminative ability for members of this quintile the c-index is only 0.57. This analysis confirms the inconsistency between calibration and discriminative ability of the Sub-Cohort LE prediction model.

4.3.6 Model #4 – OLT DEALE LE Prediction Model

Due to the potential violation of the PH assumption for age in the sub-cohort data, we wanted to consider a method that did not depend on the PH assumption for the age effect. This model gets it’s name because it utilizes the Ontario Life Table (OLT) data and borrows from the structure of the declining exponential approximation of life expectancy (DEALE), both described
previously in the methods section.\((14, 15)\) By utilizing the OLT data to accommodate the age effect we were able to avoid the use of the PH assumption and hoped to increase predictive accuracy. This model is presented in equation (8), below.

\[ LE = \frac{1}{(\mu_{\text{baseline/age}} + \mu_{\text{prostate}} + \mu_{\text{CIRS-G_pros}})} \] (8),

where \( LE \) is the predicted life expectancy, \( \mu_{\text{baseline/age}} = \frac{1}{e_x} \), and \( e_x \) is the life expectancy value at age \( x \) from the OLT data, \( \mu_{\text{prostate}} \) represents the additional hazard of death experienced by potential candidates for curative treatment of prostate cancer in the absence of treatment, and \( \mu_{\text{CIRS-G_pros}} \) is an excess mortality rate due to the co-morbidity.

Although \( \mu_{\text{baseline/age}} \) varies depending on age, at each age it is a constant rate that represents the baseline or population mortality rate. Because it was derived from data on all Ontario males (see OLT methodology in methods section) we feel it is a representative approximation of the average man in Ontario. Because the \( e_x \) values are age-specific (i.e. different for each age \( x \)), \( \mu_{\text{baseline/age}} \) also represents the hazard of death due to age.

This baseline estimate is derived from the general population of Ontario males, and therefore includes the average excess mortality experienced by this population due to any given disease, including prostate cancer. Since our objective was to identify the survival experience of potential candidates for curative treatment of prostate cancer, we also factored in the yearly excess probability of death (i.e. mortality rate), due to prostate cancer among curable prostate cancer patients. This value (\( \mu_{\text{prostate}} = 0.004 \)) was derived previously by Dr. Groome’s team using results from Holmberg et al’s Swedish randomized trial.\((16)\) It was calculated as the difference in mortality experienced by the watchful waiting group, as presented in the results of that study, versus that of the general male Swedish population of the same age. We compared this estimate to age specific excess mortality of the target population (who were treated) and found it to be of a similar magnitude (see Appendix V). This quantity represents the additional hazard of death.
experienced by potential candidates for curative treatment of prostate cancer, in the absence of treatment. Since this estimate was derived from a population identified during the same era as the OLT tables and our sub-cohort population, and also shares similar characteristics to our sub-cohort population we felt it represented the excess mortality due to prostate cancer for this model.

For the consideration of co-morbid illness status we utilized the coefficient derived from the Cox PH regression analysis for the Sub-Cohort population, as detailed in the description of the Sub-Cohort LE prediction model (above). As described in our assessment of the PH assumption for the sub-cohort data (above) the CIRS-G\textsubscript{pros} variable did not violate the proportional hazards assumption, therefore we were confident that this estimate would provide an accurate approximation of the hazard of death due to co-morbid illness in this prediction model.

We converted the CIRS-G\textsubscript{pros} coefficient (\( \beta_2 \)) to an excess mortality rate to fit the form of inputs required by this model (i.e constant mortality rates). For each age (from 50-79) we first calculated the relative risk \( RR = \exp(CIRS_i \cdot \beta_2) \) for each value of CIRS-G\textsubscript{pros} (\( CIRS_i \)). We then standardized these relative risk scores by dividing each by the relative risk score at the mean CIRS-G\textsubscript{pros} score (\( \bar{CIRS} \)) for each age: \( RR_{std} = (\exp(CIRS_i \cdot \beta_2))/\exp(\bar{CIRS} \cdot \beta_2) \). To calculate the excess mortality we then multiplied the age-specific constant mortality rate by the constant survival rate due to CIRS-G\textsubscript{pros}:

\[
\mu_{comorbidity} = (\mu_{baseline/age} \cdot (RR_{std} - 1))
\]

**Predictive Accuracy**

As with the previous three models we tested the predictive accuracy of our OLT DEALE predictions, based on model (8), by calculating a predicted life expectancy value for each member of the sub-cohort population and comparing it to the patients actual LE. The c-index that resulted from this analysis was 0.66. This indicates that given a randomly selected pair of usable patients from the sub-cohort population, the one with the better outcome (i.e. longer survival) would be identified 66% of the time. This measure of discriminative ability is a slight improvement over
the ‘best’ of the previous three LE prediction models, the Sub-Cohort model ($C=0.65$). However, the calibration plot for the OLT DEALE model, Figure 4.7a, demonstrates very poor calibration. The 95% confidence intervals surrounding actual K-M survival was very far from the line of identity. These results indicate that the model-predicted 10-year survival probabilities systematically and drastically underestimate the observed 10-year survival probabilities for the sub-cohort population. This may imply that the hazard of death due to co-morbid illness experienced by the sub-cohort population likely does not follow the form implied by the OLT DEALE model.

**Figure 4.7a Calibration Plot for OLT DEALE LE Prediction Model**

The Impact of Co-morbid Illness status on Predictive Accuracy

As with the three previous prediction models we created a secondary OLT DEALE model that eliminated the consideration of CIRS-$G_{pes}$ from the models predictions. This model
had c-index of 0.63, and calibration plot presented in Figure 4.7b. The c-index for this OLT DEALE ‘Age Only’ model indicates that when we incorporate the impact of CIRS-Gpros into the model, discriminative ability improves by 0.03 (from 0.63 to 0.66) on the c-index scale from 0.50 to 1.0. Interestingly, the calibration plot for this ‘Age Only’ model performs better than the full OLT DEALE LE prediction model.

![Figure 4.7b Calibration Plot for OLT DEALE 'Age Only' LE Prediction Model](image)

### 4.3.7 Model #5: Non-Constant OLT Model

The LE estimation in the OLT DEALE model relies on a constant hazard assumption for a given age, which may be inappropriate and lead to possible inaccuracies in LE. Therefore, we considered to use the OLT data to derive the non-constant baseline survival function at a given
age and calculate the LE after the baseline survival function is adjusted by the excess prostate cancer mortality and hazard due to CIRS-G_{pros} (β₂) as in the OLT DEALE model. The model is given in equation (9).

\[ S(t) = S_0(t)^{\exp(\beta_2(CIRS - \overline{CIRS}))} \]  

(9),

where \( S_0(t) \) is the baseline survival function calculated by creating a survival function specific to each age \( x \), from the \( p_x \) column of the OLT, \( \beta_2 \) is obtained from the Cox PH model

\[ S(t) = S_{00}(t)^{\exp(\beta_{01} \cdot AGE + \beta_2 \cdot \text{CIRS}_{pros})} \]  

for the sub-cohort data, and \( \overline{\text{CIRS}} = 4.1 \) is the mean CIRS-G_{pros} score from the sub-cohort data.

The baseline survival function \( S_0(t) \) is age-specific and non-constant because it is created by the cumulative multiplication of the \( p_x \) values, for each age \( x \) (starting at age \( x \) for all values of \( p_x \) up to \( p_{104} \) (the last age included in the OLT)). This results in an age-specific survival function, for each age \( x \) (in the OLT). We adjusted the set of survival probabilities \{ \Pr(T > t_1), ..., \Pr(T > t_n) \} \) of each age-specific survival function by the constant excess mortality due to prostate cancer \( (\mu_{prostate} = 0.004) \) by subtracting 0.004 from each survival probability. Therefore in estimating LE for members of the sub-cohort population the baseline survival function \( S_0(t) \) is equal to the age-specific survival function \( S_{age_x}(t) \), that corresponds to the patient’s age \( (AGE_i) \).

We developed a macro to calculate patient specific survival functions by considering the patient’s age and CIRS-G_{pros} score. This macro operated a bit differently from the macro used to calculate the survival function \( S(t) \) in the Case-cohort Target, Simplified Case-cohort Target, and Sub-Cohort models, because the baseline survival function varied depending on age in the latest model (as described above). However, the methods to calculate predicted LE from the patient specific survival function were the same as introduced for the previous three models that followed the form of the Cox PH regression equation.
Predictive Accuracy

The c-index for the Non-Constant OLT model that resulted from comparing the model (9) predicted LEs to actual LEs for members of the sub-cohort population was 0.66. This is the same value that resulted from the OLT DEALE LE prediction model and may be interpreted as; given two randomly chosen patients the one with the better outcome would be identified 66% of the time by this prediction model. As with all other c-index values obtained in this work, this model demonstrates a fairly modest level of discriminative ability.

The calibration plot, Figure 4.8a, demonstrates poor calibration, even though it is better than the calibration of OLT DEALE LE prediction model. For the first two quintiles (10-year survival probabilities <0.60) the calibration is worst, because the 95% CI surrounding the observed values does not contain the line of identity. The model performs slightly better at higher probabilities, but still does not illustrate good calibration.
The Impact of Co-morbid Illness status on Predictive Accuracy

When we remove the CIRS-G$_{pros}$ variable from this Non-constant OLT model to assess its impact on predictive accuracy, the discriminative ability falls to a c-index of 0.63. Therefore, including CIRS-G$_{pros}$ as a predictor in this model increases the PA by 0.03, from 0.63 to 0.66. The calibration plot for the Non-constant OLT ‘Age Only’ model is presented in Figure 4.8b. It demonstrates similar performance to the model containing CIRS-G$_{pros}$.
Figure 4.8b Calibration Plot for Non-Constant OLT 'Age Only' LE Prediction Model
4.3.8 Determining the ‘Best’ LE Prediction Model

In total we developed five distinct LE prediction models in this study, with the aim of developing a prediction model that demonstrated a high level of predictive accuracy and improved on predictions made based on the consideration of age alone. The first step we took was to develop a model based on the Cox PH regression model. A model of this type is semi-parametric, and therefore requires no assumptions about the form of the baseline hazard, which we thought would be an advantage in estimating the survival experience for potential candidates of curative prostate cancer treatment. We started with a model that estimated all parameters using just one data source. Then we moved on to develop models that utilized several different data sources we thought would provide more stable and robust estimates of the model parameters, and therefore increase predictive accuracy. When it became evident that the proportional hazards assumption may be violated in the estimation of some parameters we developed a model that allowed us to avoid this assumption for the age parameter, by estimating age-specific constant mortality rates from Ontario Life Table data. With concerns about possible inaccuracies due to the constant mortality rate we developed a model using OLT data with methods that provided a non-constant survival function and still avoid the use of Cox PH modeling for age.

Although neither calibration nor discrimination of a model should be considered independently as a measure of the overall predictive performance of a model, when considered in combination, they can be a reasonable measure of overall performance.(9) The highest level of discrimination was obtained by both the OLT DEALE and Non-Constant OLT LE prediction models, as both demonstrated a c-index= 0.66. Of these two models, the Non-Constant OLT model demonstrated better calibration. However, this model did not perform well on this measure, as the calibration curve did not lie close to the line of identity, and at the lowest quintile the 95% CI for observed survival did not contain this line of identity. Although we were not able to quantify the degree of difference in calibration between all five models, the calibration plot for
the Sub-Cohort model clearly demonstrated much better performance than the Non-Constant OLT model.

The Sub-Cohort model also demonstrated comparable discriminative ability (c-index=0.65) to the OLT DEALE and Non-Constant OLT models. With the Sub-Cohort model’s discriminative ability being only slightly worse than the Non-Constant OLT models, and the Sub-Cohort model performing a great deal better than the Non-Constant OLT model in terms of calibration, we concluded that the Sub-Cohort LE prediction model was the ‘best’ model developed in this study for predicting life expectancy in this context.

Furthermore, in consideration of our objective to improve methods of LE prediction based on age alone, we consider the ‘Age Only’ OLT DEALE model and compare it to the Sub-Cohort model. The ‘Age Only’ OLT DEALE model best represents the most common form of LE prediction currently used in clinical settings – population-based life tables – because it involves only one small adjustment to the life expectancy estimates presented in the OLT (see section 4.3.6 and Appendix V for details on this adjustment). The Sub-Cohort model clearly demonstrates better PA than the ‘Age Only’ OLT DEALE model. The c-index is 0.02 higher (C=0.65 and 0.63 respectively) and it also calibrates far better (Figure 4.2a vs Figure 4.7b). This improvement in PA exceeds that demonstrated by the Non-Constant OLT model when compared to the ‘Age Only’ OLT DEALE model, because the degree of improvement in calibration is not as significant in the latter comparison. We felt these comparisons helped to confirm our decision that the Sub-Cohort model is the ‘best’ LE prediction model of those developed in this study.

This conclusion is not necessarily the ‘correct’ decision as there was a certain degree of imprecision in our comparisons of these models. Beyond the difficulties in quantifying the degree of difference in calibration and a meaningful difference in discrimination, comparing these models was also difficult because they represent different forms of model validation. The Sub-Cohort model is strictly an internal validation process because the model parameters were
optimized for the same population used to validate its performance (the sub-cohort population). The Non-Constant OLT model (and the OLT DEALE model) represent a semi-external validation because the Ontario Life Tables, 1995-1997, Males (i.e. an external data source) were used to optimize some model parameters. The sub-cohort population was used to optimize the CIRS-Gprot effect in this model, so its validation cannot be considered truly external. Because the Sub-Cohort models calibration plot represents a strictly internal validation we would expect it to perform well, whereas we may not judge the Non-Constant OLT model on the same terms because of the differing validation process.

However, without a truly external dataset available to test all models, we concern ourselves with the prediction of the sub-cohort population. At this level the Sub-Cohort model performs ‘best’ for the reasons outlined above. Although the Sub-Cohort model is our ‘best’ model we recommend that it, at minimum, undergo a process of external validation before it can be recommended for clinical use.

4.4 Clinical Usefulness

In addition to performance measures of discrimination and calibration, we assessed the clinical usefulness of the most accurate life-expectancy prediction model (the Sub-Cohort model). We wanted to determine if the model would be potentially beneficial in clinical practice, to guide screening and treatment decision-making among men considered to be potential candidates for curative treatment of prostate cancer.

Level of agreement across clinically meaningful ‘time-windows’

As discussed in the introduction, clinicians often use a life expectancy estimate of 10-years as a cut-off to determine treatment candidacy. There is often a ‘grey-zone’ around this 10-year rule when the decision whether or not to treat is unclear. To assess the clinical usefulness of
the Sub-Cohort model, predicted life expectancy values and observed survival times for members of the sub-cohort were grouped into three ‘time-windows’ 1) life expectancy less than 7 years, where the decision not to treat is usually clear, 2) life expectancy between 7 to 13 years, the ‘grey-zone’ in which the decision whether or not to treat is unclear, and 3) life expectancy greater than 13 years, where the decision to treat is usually clear.

We then grouped the Sub-cohort model’s LE predictions and observed survival times for the sub-cohort population into the categories outlined to create the contingency table shown below in Table 4.5a. The results of Fleiss’ Kappa statistic indicated an agreement of just 0.09 (95% CI: 0.06, 0.12), which is considered to be a very poor level of agreement, and indicates that the Sub-Cohort model predictions do not agree well with observed survival times, when categorized into the three ‘time-windows’ of life expectancy. Specifically, the predictions disproportionately over-estimate patient life expectancy, incorrectly categorizing many patients into the greater than 13 years of life expectancy group.

Table 4.5a Contingency Table for Sub-Cohort Model Predicted vs Observed Life Expectancy, treat decision window of 7-13 years.

<table>
<thead>
<tr>
<th>Sub-Cohort Population - Observed Survival Time (years)</th>
<th>&lt;7yrs</th>
<th>7-13yrs</th>
<th>&gt;13yrs</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;7yrs</td>
<td>7</td>
<td>3</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>7-13yrs</td>
<td>172</td>
<td>369</td>
<td>136</td>
<td>677</td>
</tr>
<tr>
<td>&gt;13yrs</td>
<td>98</td>
<td>400</td>
<td>225</td>
<td>723</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>277</strong></td>
<td><strong>772</strong></td>
<td><strong>361</strong></td>
<td><strong>1420</strong></td>
</tr>
</tbody>
</table>

\[ k = 0.09 \text{ (95\% CI: 0.06, 0.12)} \]

We created a second contingency table (Table 4.5b) that used the 10-year cut-off as the lower end and 15 years as the upper end of the treatment decision ‘window’. Using these cut-off values the results improved slightly: Fleiss’ Kappa statistic was 0.13 (95% CI: 0.09,0.16).
However, this improvement does not change the interpretation of the results; the model predictions consistently and disproportionately over-estimate the observe life expectancy.

Table 4.5b Contingency Table for Sub-Cohort Model Predicted vs Observed Life Expectancy, treat decision window of 10-15 years.

<table>
<thead>
<tr>
<th>Sub-Cohort Population - Observed Survival Time (years)</th>
<th>&lt; 10yrs</th>
<th>10-15yrs</th>
<th>&gt;15yrs</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;10yrs</td>
<td>98</td>
<td>64</td>
<td>2</td>
<td>164</td>
</tr>
<tr>
<td>10-15yrs</td>
<td>228</td>
<td>448</td>
<td>73</td>
<td>749</td>
</tr>
<tr>
<td>&gt;15yrs</td>
<td>76</td>
<td>305</td>
<td>66</td>
<td>447</td>
</tr>
<tr>
<td>Total</td>
<td>402</td>
<td>817</td>
<td>141</td>
<td>1420</td>
</tr>
</tbody>
</table>

\[ k = 0.13 \text{ (95\% CI: 0.09, 0.16)} \]

The obvious over-estimation of LE demonstrated by the contingency tables are consistent with the findings in the scatterplot (Figure 4.9, below) of predicted vs actual outcomes (for all patients who died during study follow-up). The plot demonstrates a clear trend, as Sub-Cohort model LE predictions systematically overestimate observed survival times. Almost all data-points lie above the line of identity (dashed line at 45°), meaning that the predicted value for each patient is higher than the observed value (since the predicted values are plotted along the y-axis, and observed along the x-axis). This is illustrative of a systematic error in our model’s predictions. A model that performs well would have all data-points clustered around the line of identity. Another important property of the Sub-Cohort model that is illustrated by Figure 4.9 is that the model’s LE estimates are bounded at the upper limit at 19-years - the maximum observed follow-up time used to estimate parameters in the model.
To assess the impact on the level of agreement due to the formal consideration of co-morbid illness status we created contingency tables using the OLT predicted life expectancy. For each member of the sub-cohort population we assigned predicted life expectancy as the value of $e_x$ from the OLT specific to their age. We compared these predictions to observed survival times to create the same two contingency tables and Fleiss’ Kappa statistics developed for the Sub-cohort model predictions.

The results, presented in Table 4.6a-b, indicate poorer agreement than the Sub-cohort model predictions. For the 7-13 year ‘time window’ we are unable to calculate a Fleiss’ Kappa
(k) because the OLT predictions did not categorize any of the sub-cohort members to a life expectancy of less than 7 years. For the 10-15 year ‘time window’ the Fleiss’s kappa statistic is 0.06 (95% CI: 0.03, 0.09). Therefore, although the Sub-cohort model’s LE predictions result in a poor level of agreement with actual sub-cohort population LE, it does perform better than OLT LE predictions that do not formally consider the impact of co-morbid illness status.

Table 4.6a Contingency Table of Ontario Life Tables Predicted vs Observed Life Expectancy, treatment decision window of 7-13 years.

<table>
<thead>
<tr>
<th>Sub-Cohort Population - Observed Survival Time (years)</th>
<th>&lt; 7yrs</th>
<th>7-13yrs</th>
<th>&gt;13yrs</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;7yrs</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7-13yrs</td>
<td>185</td>
<td>326</td>
<td>162</td>
<td>673</td>
</tr>
<tr>
<td>&gt;13yrs</td>
<td>130</td>
<td>451</td>
<td>342</td>
<td>923</td>
</tr>
<tr>
<td>Total</td>
<td>315</td>
<td>777</td>
<td>504</td>
<td>1596</td>
</tr>
</tbody>
</table>

\(^k=0.06 (95\% CI: 0.03, 0.09)\)

Table 4.6b Contingency Table of Ontario Life Tables Predicted vs Observed Life Expectancy, treatment decision window of 10-15 years.

<table>
<thead>
<tr>
<th>Sub-Cohort Population - Observed Survival Time (years)</th>
<th>&lt; 10yrs</th>
<th>10-15yrs</th>
<th>&gt;15yrs</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;10yrs</td>
<td>55</td>
<td>37</td>
<td>5</td>
<td>97</td>
</tr>
<tr>
<td>10-15yrs</td>
<td>324</td>
<td>422</td>
<td>114</td>
<td>860</td>
</tr>
<tr>
<td>&gt;15yrs</td>
<td>156</td>
<td>389</td>
<td>114</td>
<td>659</td>
</tr>
<tr>
<td>Total</td>
<td>402</td>
<td>817</td>
<td>141</td>
<td>1596</td>
</tr>
</tbody>
</table>

\(^k=0.06 (95\% CI: 0.03, 0.09)\)
Probability of 10-year survival and Optimal Negative Predictive Value

We defined several ‘cutoffs’ of Sub-Cohort predicted probability of 10-year survival, that would be useful in assisting clinicians with the identification of patients with insufficient LE to be considered candidates for screening or curative treatment. At each probability cut-off we define several quantities: 1) the number of patients predicted to have a 10-year survival probability below the cut-off value 2) True Negatives (TN) – patients who were correctly identified by the prediction model with a LE less than 10 years, 3) False Negatives (FN) – patients who were predicted to fall below the cut-off but actually survived longer than 10 years, 4) True Positives (TP) – patients who were correctly identified as surviving longer than 10 years, and 5) Negative Predictive Value (NPV) = TN/# of patients with predicted 10-year survival probability below cut-off. The NPV is useful in quantifying the frequency (%) with which the model will correctly predict that a patient will actually survive less than 10 years.

The results of this analysis, presented in Table 4.7a, demonstrate the accuracy of the Sub-Cohort model’s predictions in the context of identifying patients who likely will not survive long enough to be considered candidates for screening or curative treatment. For instance, the clinician may recommend against curative treatment for patients with a 10-year survival probability of less than 50%. At this survival probability the Sub-Cohort model has an NPV of 61.8%. This means that the Sub-cohort model correctly identifies individuals with a 10-year survival probability of 0.50 or less, 62% of the time.
Table 4.7a Sub-cohort Model-Predicted Probability Cutoffs for 10-Year Survival After Diagnoses, among sub-cohort population members (n=1596).

<table>
<thead>
<tr>
<th>Predicted Probability of 10-year Survival(%)</th>
<th>Patients Below Cutoff (of Total)</th>
<th>True Negatives*</th>
<th>False Negatives+</th>
<th>True Positives++</th>
<th>Negative Predictive Value (%)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
<td>n</td>
</tr>
<tr>
<td>40</td>
<td>83</td>
<td>5.2</td>
<td>60</td>
<td>3.8</td>
<td>23</td>
</tr>
<tr>
<td>50</td>
<td><strong>227</strong></td>
<td><strong>14.2</strong></td>
<td><strong>140</strong></td>
<td><strong>8.8</strong></td>
<td><strong>87</strong></td>
</tr>
<tr>
<td>60</td>
<td>474</td>
<td>29.7</td>
<td>243</td>
<td>15.2</td>
<td>231</td>
</tr>
<tr>
<td>70</td>
<td>807</td>
<td>50.6</td>
<td>362</td>
<td>22.7</td>
<td>445</td>
</tr>
</tbody>
</table>

* Patients below cutoff who were correctly identified with an LE less than 10 years.
+ Patients below cutoff who nonetheless survived beyond 10 years.
++ Patients correctly identified by the prediction model to survive beyond 10 years.
$ NPV = True Negatives/#Patients Predicted Below cutoff

Impact of Co-morbid Illness Status on NPV Performance

Again, we wanted to assess the impact of formally including co-morbid illness status in the prediction of patient life expectancy for this outcome. We used the OLT data to derive age-specific survival functions, by the same methods described in the development of Non-constant OLT prediction model. Table 4.7b presents the performance of the OLT data at various cuts off 10-year survival probability for the sub-cohort population. The OLT data performs worse than the Sub-cohort model at all cut-off values. This result, again, supports the formal consideration of co-morbid illness status in predicting survival and life expectancy among men with clinically localized prostate cancer.
Table 4.7b Ontario Life Tables-Predicted Probability Cutoffs for 10-Year Survival After Diagnoses, among sub-cohort population members (n=1596).

<table>
<thead>
<tr>
<th>Predicted Probability of 10-year Survival(%)</th>
<th>Patients Below Cutoff (of Total)</th>
<th>True Negatives*</th>
<th>False Negatives+</th>
<th>True Positives++</th>
<th>Negative Predictive Value (%)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>147  9.2</td>
<td>85  5.3</td>
<td>62  3.9</td>
<td>1019 63.8</td>
<td>57.8</td>
</tr>
<tr>
<td>50</td>
<td>293 18.4</td>
<td>149 9.3</td>
<td>144 9.0</td>
<td>937 58.7</td>
<td>50.9</td>
</tr>
<tr>
<td>60</td>
<td>564 35.3</td>
<td>263 16.5</td>
<td>301 18.9</td>
<td>780 48.9</td>
<td>46.6</td>
</tr>
<tr>
<td>70</td>
<td>967 60.6</td>
<td>389 24.4</td>
<td>578 36.2</td>
<td>503 31.5</td>
<td>40.2</td>
</tr>
</tbody>
</table>

*Patients below cutoff who were correctly identified with an LE less than 10 years.
+ Patients below cutoff who nonetheless survived beyond 10 years.
++ Patients correctly identified by the prediction model to survive beyond 10 years.
$ NPV = True Negatives/#Patients Predicted Below cutoff

4.5 Table of Predicted Life Expectancy Values derived from the Sub-Cohort Model

The tables of life expectancy values presented in Table 4.8a and 4.8b (below) are the ‘end product’ of the current study. Table 4.8a contains the Sub-Cohort model predicted, post-treatment life expectancy values for potential screening or curative treatment candidates. Table 4.8b contains Sub-Cohort model predicted pre-treatment (adjusted) life expectancy values. The values differ based on the inclusion of a treatment effect adjustment - calculated to remove the survival benefit of treatment received by the sub-cohort population - included in the predictions for Table 4.8b (see Appendix VII for details). Therefore when considered together, the two tables provide a range of LE’s depending on the effect of curative treatment. The tables can be used as an aid in either the screening or treatment decision. As indicated previously, the rule of thumb (10-years of life expectancy) most often used when incorporating the consideration of life expectancy in the treatment decision-making process can also be applied to either the PSA testing or biopsy decision. Therefore, this tool has application to all time periods of pre-treatment decision-making.
In order to calculate the patient specific life expectancy one simply needs to follow across the row corresponding to the patient’s age and down the column corresponding to the patients CIRS-G pros score, until they meet. The grey area in the middle of the table corresponds to the ‘grey-zone’ defined during the development of the contingency tables (Tables 4.5-4.6), as the ‘time-window’ of life expectancy when the decision whether or not to screen or treat with curative intent is unclear (7-13 years life expectancy). The ‘unshaded’ area refers to the life expectancy values for which the decision to treat is often clear (i.e. when life expectancy is greater that 13 years), while the ‘lightly’ shaded area refers to the set of age and CIRS-G pros score, and corresponding life expectancy values for which the decision to not treat is usually clear (i.e. when life expectancy is less than 7 years). The life expectancy values that are in bold, underlined, and italicized text refer to the LE values that are closest to the Ontario Life Table LE value for each age in the table. If there is no model-predicted LE value within 1 year of the OLT LE value for a given age than no value is highlighted for that age. The shaded area in the top left corner of each table that does not contain LE values corresponds to the age and CIRS-G pros combinations for which the Sub-Cohort model could not predict LE. As noted previously, the Sub-Cohort model’s predictions are bound at the upper limit, by the maximum observed follow-up used to estimate parameters in the model. As such we were unable to estimate the LE for the age and CIRS-G pros combinations that indicated a LE that would exceed 19 years. Furthermore, as illustrated in Tables 4.5a-b and Figure 4.9, the LE predictions estimated by the Sub-Cohort model are optimistic on average, i.e., they overestimate life expectancy, and we caution that external validation is needed before they can be used for clinical decision making.
Table 4.8a Life Expectancy Reference Table of Sub-Cohort Model predicted, post-treatment LE Estimates\textsuperscript{1,2}

<table>
<thead>
<tr>
<th>AGE</th>
<th>CIRS-G\textsubscript{pros}</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>50</td>
<td></td>
</tr>
<tr>
<td>51</td>
<td></td>
</tr>
<tr>
<td>52</td>
<td></td>
</tr>
<tr>
<td>53</td>
<td></td>
</tr>
<tr>
<td>54</td>
<td></td>
</tr>
<tr>
<td>55</td>
<td></td>
</tr>
<tr>
<td>56</td>
<td></td>
</tr>
<tr>
<td>57</td>
<td></td>
</tr>
<tr>
<td>58</td>
<td></td>
</tr>
<tr>
<td>59</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td></td>
</tr>
<tr>
<td>61</td>
<td></td>
</tr>
<tr>
<td>62</td>
<td></td>
</tr>
<tr>
<td>63</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td></td>
</tr>
<tr>
<td>65</td>
<td></td>
</tr>
<tr>
<td>66</td>
<td></td>
</tr>
<tr>
<td>67</td>
<td></td>
</tr>
<tr>
<td>68</td>
<td></td>
</tr>
<tr>
<td>69</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td></td>
</tr>
<tr>
<td>71</td>
<td></td>
</tr>
<tr>
<td>72</td>
<td></td>
</tr>
<tr>
<td>73</td>
<td></td>
</tr>
<tr>
<td>74</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td></td>
</tr>
<tr>
<td>76</td>
<td></td>
</tr>
<tr>
<td>77</td>
<td></td>
</tr>
<tr>
<td>78</td>
<td></td>
</tr>
<tr>
<td>79</td>
<td></td>
</tr>
</tbody>
</table>

1. due to the poor PA and systematic over-estimation of LE demonstrated by the Sub-Cohort model we caution against the use of this table in clinical settings.
2. LE values in bold, underlined, and italicized text represent the model-predicted LE value closest to (if within 1 year) the OLT LE value for that age.
### Figure 4.8b Life Expectancy Reference Table of Sub-Cohort Model predicted, pre-treatment (adjusted) LE Estimates$^{1,2}$

<table>
<thead>
<tr>
<th>AGE</th>
<th>0</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>19</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>17</td>
<td>16</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>19</td>
<td>19</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>17</td>
<td>16</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>19</td>
<td>19</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>17</td>
<td>16</td>
<td>15</td>
<td>15</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>19</td>
<td>19</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>17</td>
<td>16</td>
<td>15</td>
<td>15</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>54</td>
<td>19</td>
<td>19</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>17</td>
<td>16</td>
<td>15</td>
<td>15</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>55</td>
<td>19</td>
<td>19</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>17</td>
<td>16</td>
<td>15</td>
<td>14</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>56</td>
<td>19</td>
<td>19</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>17</td>
<td>16</td>
<td>15</td>
<td>14</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>57</td>
<td>18</td>
<td>18</td>
<td>17</td>
<td>16</td>
<td>15</td>
<td>14</td>
<td>13</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>18</td>
<td>18</td>
<td>17</td>
<td>16</td>
<td>15</td>
<td>14</td>
<td>13</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>59</td>
<td>18</td>
<td>18</td>
<td>16</td>
<td>15</td>
<td>14</td>
<td>13</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>19</td>
<td>18</td>
<td>17</td>
<td>16</td>
<td>15</td>
<td>14</td>
<td>13</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>61</td>
<td>19</td>
<td>17</td>
<td>16</td>
<td>15</td>
<td>14</td>
<td>13</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>62</td>
<td>18</td>
<td>17</td>
<td>16</td>
<td>15</td>
<td>14</td>
<td>13</td>
<td>12</td>
<td>12</td>
<td>10</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>63</td>
<td>18</td>
<td>16</td>
<td>15</td>
<td>14</td>
<td>13</td>
<td>12</td>
<td>11</td>
<td>11</td>
<td>10</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>64</td>
<td>18</td>
<td>15</td>
<td>14</td>
<td>14</td>
<td>13</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>65</td>
<td>18</td>
<td>14</td>
<td>14</td>
<td>13</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>66</td>
<td>17</td>
<td>14</td>
<td>14</td>
<td>13</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>67</td>
<td>16</td>
<td>14</td>
<td>13</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>68</td>
<td>16</td>
<td>13</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>69</td>
<td>15</td>
<td>15</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>70</td>
<td>14</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>71</td>
<td>14</td>
<td>11</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>72</td>
<td>13</td>
<td>11</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>73</td>
<td>12</td>
<td>10</td>
<td>11</td>
<td>10</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>74</td>
<td>12</td>
<td>10</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>75</td>
<td>12</td>
<td>10</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>76</td>
<td>11</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>77</td>
<td>11</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>78</td>
<td>10</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>79</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

1. due to the poor PA and systematic over-estimation of LE demonstrated by the Sub-Cohort model we caution against the use of this table in clinical settings.
2. LE values in bold, underlined, and italicized text represent the model-predicted LE value closest to (if within 1 year) the OLT LE value for that age.
4.6 Supplementary Analysis – Logistic Regression Prediction Model

With concerns regarding violation of the PH assumption in the Sub-cohort model and the lack of accurate prediction by models that limited the use of Cox PH analysis we decided to build a prediction model for the probability of surviving 10-years only using logistic regression based on the sub-cohort population. This model is represented by equation 9, below.

\[
\Pr(Y = 1) = \exp(\beta_0 + (\beta_1 \cdot AGE) + (\beta_2 \cdot CIRS - G_{pros}))
\]

where \( Y \) is the binary outcome of surviving 10 years (1=yes, 0=no), \( \beta_0 \) is the ‘intercept’ that represents the probability of 10-year survival when the value of all covariates is 0, \( \beta_1 \) and \( \beta_2 \) are the effects on the probability of the outcome due to age and CIRS-Gpros respectively for the sub-cohort data.

Since the logistic regression only considers 10-year survival, it involves less stringent assumptions than the Cox PH model and the prediction of 10-year survival would be more reliable than the PH model based predictions discussed above.

The results of the logistic regression analysis are presented in Table 4.9. The odds ratio for age indicates that for each 1-year increase in age (from 50 to 79) the odds of surviving 10-years decreases by 9%. The impact of CIRS-Gpros is even greater than that of age; for each 1-point increase on the CIRS-Gpros scale a patient’s likelihood of surviving 10-years decreases by 18%.

Table 4.9 Logistic Regression Model, Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>( \beta )</th>
<th>SE ( \beta )</th>
<th>( p)-value</th>
<th>( e^\beta ) (odds ratio)</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept (( \beta_0 ))</td>
<td>7.7993</td>
<td>0.7139</td>
<td>&lt;.0001</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Age (( \beta_1 ))</td>
<td>-0.0911</td>
<td>0.0104</td>
<td>&lt;.0001</td>
<td>0.913</td>
<td>(0.895,0.932)</td>
</tr>
<tr>
<td>CIRS-Gpros (( \beta_2 ))</td>
<td>-0.2013</td>
<td>0.0236</td>
<td>&lt;.0001</td>
<td>0.818</td>
<td>(0.781,0.856)</td>
</tr>
</tbody>
</table>

The result of central interest to us is the predictive ability of this model. The c-index that results from this model is 0.71. This statistic is the concordance index, like that used for the
measure of discriminative ability in our LE prediction models. In logistic regression this statistic is calculated by considering all possible pairs formed by taking one observation from each of the outcome groups (i.e. Y=1 for survival >= 10-years, Y=0 for survival < 10-years). Each pair is then compared to the model-predicted probability of 10-year survival. The comparison yields either a concordant pair (predicted outcome is higher for the subject that survived > 10-years), or discordant (predicted outcome is higher for the subject that survived < 10-years), or tied (predicted outcome is equal for both subjects). The c-index is then calculated as demonstrated in the methods chapter, section 3.3.3).

Therefore a C=0.71 means that given two patients (one who has survived 10 years, (i.e. the binary outcome variable Y=1), and one who has not, Y=0) the model will correctly predict the one with the better survival probability 71% of the time. This is very similar to the measure of discrimination that we calculated for the life expectancy prediction models in our main analysis. We cannot make direct comparisons between this result and the previous results because in the logistic model we are predicting the 10-year survival (a binary outcome, either yes=1 or no=0), where as in our main analysis we are predicting life expectancy values.

The logistic regression prediction model also performed well in terms of calibration. As shown in Figure 4.10, below, the model-predicted 10-year survival probabilities calibrate very well to the observed survival of the sub-cohort population for probabilities between 0.60 and 0.80. The performance in these ranges was better than we observed for any of our life expectancy prediction models. For probabilities higher than 0.80 and lower than 0.60 the model did not perform quite as well.
The Impact of Co-morbid Illness status on Predictive Accuracy

As has been the case throughout this work we were interested in the impact on predictive accuracy of including the CIRS-G$_{pros}$ variable in our analysis as a formal consideration of co-morbid illness status. To do so we built a logistic regression model using the same methods detailed for our Logistic regression model, but excluded the CIRS-G$_{pros}$ variable from the prediction of our outcome in the logistic regression model. This model resulted in a c-index of 0.66. Including CIRS-G$_{pros}$ in the prediction of 10-year survival probability improved the discriminative ability by 0.05 (from 0.66 to 0.71).
Logistic Regression Table of Predicted Probabilities

We created a table of predicted probabilities of surviving 10-years, based on the logistic regression model. The estimates presented above in Table 4.9 for the intercept ($\beta_0$), the change in slope due to age ($\beta_1$) and the change in slope due to CIRS-Gpros ($\beta_2$) were entered into this equation. Then the probability of surviving 10-years for all combinations of the two predictor variables included in the model: age (from 50-79) and CIRS-Gpros (0, 3-12), were calculated to form Table 4.10.

This table can help in the treatment decision-making process if the clinician has pre-specified probabilities of 10-year survival to determine screening or treatment candidacy. Although this may not be as conventional as using remaining lifetime (i.e. 10 years) it is simply another way to think about the decision surrounding treatment that may be more attractive to some patients and clinicians. As with Tables 4.8-b we caution against the use of this table in any clinical setting, without external validation.
Table 4.10 Look-Up Table of Predicted Probabilities of 10-year Survival

<table>
<thead>
<tr>
<th>AGE</th>
<th>0</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.96</td>
<td>0.93</td>
<td>0.92</td>
<td>0.90</td>
<td>0.88</td>
<td>0.86</td>
<td>0.84</td>
<td>0.81</td>
<td>0.77</td>
<td>0.74</td>
<td>0.70</td>
</tr>
<tr>
<td>51</td>
<td>0.96</td>
<td>0.93</td>
<td>0.91</td>
<td>0.90</td>
<td>0.87</td>
<td>0.85</td>
<td>0.82</td>
<td>0.79</td>
<td>0.76</td>
<td>0.72</td>
<td>0.68</td>
</tr>
<tr>
<td>52</td>
<td>0.96</td>
<td>0.92</td>
<td>0.91</td>
<td>0.89</td>
<td>0.86</td>
<td>0.84</td>
<td>0.81</td>
<td>0.78</td>
<td>0.74</td>
<td>0.70</td>
<td>0.66</td>
</tr>
<tr>
<td>53</td>
<td>0.95</td>
<td>0.91</td>
<td>0.90</td>
<td>0.88</td>
<td>0.85</td>
<td>0.83</td>
<td>0.80</td>
<td>0.76</td>
<td>0.72</td>
<td>0.68</td>
<td>0.64</td>
</tr>
<tr>
<td>54</td>
<td>0.95</td>
<td>0.91</td>
<td>0.89</td>
<td>0.87</td>
<td>0.84</td>
<td>0.81</td>
<td>0.78</td>
<td>0.74</td>
<td>0.70</td>
<td>0.66</td>
<td>0.61</td>
</tr>
<tr>
<td>55</td>
<td>0.94</td>
<td>0.90</td>
<td>0.88</td>
<td>0.86</td>
<td>0.83</td>
<td>0.80</td>
<td>0.76</td>
<td>0.73</td>
<td>0.68</td>
<td>0.64</td>
<td>0.59</td>
</tr>
<tr>
<td>56</td>
<td>0.94</td>
<td>0.89</td>
<td>0.87</td>
<td>0.84</td>
<td>0.82</td>
<td>0.78</td>
<td>0.75</td>
<td>0.71</td>
<td>0.66</td>
<td>0.62</td>
<td>0.57</td>
</tr>
<tr>
<td>57</td>
<td>0.93</td>
<td>0.88</td>
<td>0.86</td>
<td>0.83</td>
<td>0.80</td>
<td>0.77</td>
<td>0.73</td>
<td>0.69</td>
<td>0.64</td>
<td>0.60</td>
<td>0.55</td>
</tr>
<tr>
<td>58</td>
<td>0.93</td>
<td>0.87</td>
<td>0.85</td>
<td>0.82</td>
<td>0.79</td>
<td>0.75</td>
<td>0.71</td>
<td>0.67</td>
<td>0.62</td>
<td>0.57</td>
<td>0.52</td>
</tr>
<tr>
<td>59</td>
<td>0.92</td>
<td>0.86</td>
<td>0.83</td>
<td>0.81</td>
<td>0.77</td>
<td>0.73</td>
<td>0.69</td>
<td>0.65</td>
<td>0.60</td>
<td>0.55</td>
<td>0.50</td>
</tr>
<tr>
<td>60</td>
<td>0.91</td>
<td>0.85</td>
<td>0.82</td>
<td>0.79</td>
<td>0.76</td>
<td>0.72</td>
<td>0.67</td>
<td>0.63</td>
<td>0.58</td>
<td>0.53</td>
<td>0.48</td>
</tr>
<tr>
<td>61</td>
<td>0.90</td>
<td>0.84</td>
<td>0.81</td>
<td>0.77</td>
<td>0.74</td>
<td>0.70</td>
<td>0.65</td>
<td>0.61</td>
<td>0.56</td>
<td>0.51</td>
<td>0.46</td>
</tr>
<tr>
<td>62</td>
<td>0.90</td>
<td>0.82</td>
<td>0.79</td>
<td>0.76</td>
<td>0.72</td>
<td>0.68</td>
<td>0.63</td>
<td>0.58</td>
<td>0.53</td>
<td>0.48</td>
<td>0.43</td>
</tr>
<tr>
<td>63</td>
<td>0.89</td>
<td>0.81</td>
<td>0.78</td>
<td>0.74</td>
<td>0.70</td>
<td>0.66</td>
<td>0.61</td>
<td>0.56</td>
<td>0.51</td>
<td>0.46</td>
<td>0.41</td>
</tr>
<tr>
<td>64</td>
<td>0.88</td>
<td>0.80</td>
<td>0.76</td>
<td>0.72</td>
<td>0.68</td>
<td>0.64</td>
<td>0.59</td>
<td>0.54</td>
<td>0.49</td>
<td>0.44</td>
<td>0.39</td>
</tr>
<tr>
<td>65</td>
<td>0.87</td>
<td>0.78</td>
<td>0.75</td>
<td>0.71</td>
<td>0.66</td>
<td>0.62</td>
<td>0.57</td>
<td>0.52</td>
<td>0.47</td>
<td>0.42</td>
<td>0.37</td>
</tr>
<tr>
<td>66</td>
<td>0.86</td>
<td>0.77</td>
<td>0.73</td>
<td>0.69</td>
<td>0.64</td>
<td>0.59</td>
<td>0.54</td>
<td>0.49</td>
<td>0.44</td>
<td>0.39</td>
<td>0.35</td>
</tr>
<tr>
<td>67</td>
<td>0.84</td>
<td>0.75</td>
<td>0.71</td>
<td>0.67</td>
<td>0.62</td>
<td>0.57</td>
<td>0.52</td>
<td>0.47</td>
<td>0.42</td>
<td>0.37</td>
<td>0.33</td>
</tr>
<tr>
<td>68</td>
<td>0.83</td>
<td>0.73</td>
<td>0.69</td>
<td>0.65</td>
<td>0.60</td>
<td>0.55</td>
<td>0.50</td>
<td>0.45</td>
<td>0.40</td>
<td>0.35</td>
<td>0.31</td>
</tr>
<tr>
<td>69</td>
<td>0.82</td>
<td>0.71</td>
<td>0.67</td>
<td>0.62</td>
<td>0.58</td>
<td>0.53</td>
<td>0.48</td>
<td>0.43</td>
<td>0.38</td>
<td>0.33</td>
<td>0.29</td>
</tr>
<tr>
<td>70</td>
<td>0.81</td>
<td>0.69</td>
<td>0.65</td>
<td>0.60</td>
<td>0.55</td>
<td>0.50</td>
<td>0.45</td>
<td>0.40</td>
<td>0.36</td>
<td>0.31</td>
<td>0.27</td>
</tr>
<tr>
<td>71</td>
<td>0.79</td>
<td>0.67</td>
<td>0.63</td>
<td>0.58</td>
<td>0.53</td>
<td>0.48</td>
<td>0.43</td>
<td>0.38</td>
<td>0.34</td>
<td>0.29</td>
<td>0.25</td>
</tr>
<tr>
<td>72</td>
<td>0.78</td>
<td>0.65</td>
<td>0.61</td>
<td>0.56</td>
<td>0.51</td>
<td>0.46</td>
<td>0.41</td>
<td>0.36</td>
<td>0.32</td>
<td>0.27</td>
<td>0.24</td>
</tr>
<tr>
<td>73</td>
<td>0.76</td>
<td>0.63</td>
<td>0.59</td>
<td>0.54</td>
<td>0.49</td>
<td>0.44</td>
<td>0.39</td>
<td>0.34</td>
<td>0.30</td>
<td>0.26</td>
<td>0.22</td>
</tr>
<tr>
<td>74</td>
<td>0.74</td>
<td>0.61</td>
<td>0.56</td>
<td>0.51</td>
<td>0.46</td>
<td>0.41</td>
<td>0.37</td>
<td>0.32</td>
<td>0.28</td>
<td>0.24</td>
<td>0.20</td>
</tr>
<tr>
<td>75</td>
<td>0.72</td>
<td>0.59</td>
<td>0.54</td>
<td>0.49</td>
<td>0.44</td>
<td>0.39</td>
<td>0.34</td>
<td>0.30</td>
<td>0.26</td>
<td>0.22</td>
<td>0.19</td>
</tr>
<tr>
<td>76</td>
<td>0.71</td>
<td>0.57</td>
<td>0.52</td>
<td>0.47</td>
<td>0.42</td>
<td>0.37</td>
<td>0.32</td>
<td>0.28</td>
<td>0.24</td>
<td>0.21</td>
<td>0.18</td>
</tr>
<tr>
<td>77</td>
<td>0.69</td>
<td>0.55</td>
<td>0.49</td>
<td>0.44</td>
<td>0.40</td>
<td>0.35</td>
<td>0.30</td>
<td>0.26</td>
<td>0.23</td>
<td>0.19</td>
<td>0.16</td>
</tr>
<tr>
<td>78</td>
<td>0.67</td>
<td>0.52</td>
<td>0.47</td>
<td>0.42</td>
<td>0.37</td>
<td>0.33</td>
<td>0.29</td>
<td>0.25</td>
<td>0.21</td>
<td>0.18</td>
<td>0.15</td>
</tr>
<tr>
<td>79</td>
<td>0.65</td>
<td>0.50</td>
<td>0.45</td>
<td>0.40</td>
<td>0.35</td>
<td>0.31</td>
<td>0.27</td>
<td>0.23</td>
<td>0.20</td>
<td>0.17</td>
<td>0.14</td>
</tr>
</tbody>
</table>

1. We caution against the use of this table in clinical settings at this time.
4.7 References


5.1 Overview

The use of a life expectancy prediction tool is a proposed method to increase life expectancy prediction accuracy, and aid in both screening and treatment decision making for potential curative treatment candidates. These tools can help refine traditional methods of life expectancy estimation by formally considering patient and disease specific characteristics. For the current study we considered the impact of co-morbid illness status, in addition to age, on patient survival while exploring the best method to predict life expectancy in a population of men with clinically localized prostate cancer. We found that including co-morbid illness status in our prediction models consistently improved predictive accuracy, regardless of the methods used to develop each model.

The ‘best’ of our life expectancy prediction models demonstrated modest apparent predictive accuracy, but still exceeds the accuracy of models based on age alone, including those of population based life tables.(1) When compared to clinician’s intuitive estimates of LE for patients with localized prostate cancer our model appears to perform better.(2-4) When compared to several prostate cancer LE prediction tools identified in the current literature, our model does not perform quite as well. (5-7) There were several assumptions and limitations to each of the models we developed that may have hindered their predictive accuracy. The use of logistic regression modeling, which is without extensive assumptions, demonstrated good predictive accuracy. Although this model is limited to predicting a binary outcome, its results were most encouraging and provide an alternative method of approaching the screening and treatment decision-making process for potential curative treatment candidates.

In view of the fact that our most accurate model demonstrates at best modest predictive accuracy and clinical utility we feel more work is needed before our end product can be
recommend for use in clinical settings. An external validation of both the Sub-Cohort LE prediction model and our logistic regression model is required to truly understand the models’ utility as screening and treatment decision aids. Further exploration of the best method to predict LE in this setting is also suggested.

5.2 Interpretation of Results

Table 4.8b is one of three reference tables presented as the ‘end product’ of this work and provides life expectancy predictions, based on age and CIRS-G_{pros} score, to aid in the screening or treatment decision-making process for potential curative treatment candidates. These predictions, unlike those presented in Table 4.8a account for the survival benefit due to curative treatment, received by the population studied to estimate model parameters (details in Appendix VII). When the two tables are compared we see that adjusting for the treatment benefit resulted in predicted life expectancy decreasing by as much as 2 years depending on patient age and CIRS-G_{pros} score. If the clinician (or reader) feels this adjustment is too favorable, or not favorable enough, they may choose instead to alter the predictions presented in Table 4.8a by the magnitude (in months or years) they feel more appropriate.

When we view Table 4.8b, we can see the independent and combined impact of age and co-morbid illness status on patient predicted life expectancy. The age effect indicates a less than 1-year reduction in life expectancy per 1-year increase in age, while holding CIRS-G_{pros} score constant. For example, if we consider all patients with a CIRS-G_{pros} of 5; both 68 and 69 year old patients are expected to live 11 years. It is not until 70 years of age that these patients would experience a full year decrease in life expectancy. The effect of age demonstrated by our model is not unlike the per-year change in life expectancy we see in the common Life Tables like those produced by Statistics Canada. Although instinctively we may expect that for each 1-year increase in age, life expectancy would go down by at least one year, this is not the case. When a
person survives an additional year of life, their life expectancy does not decrease by a full 1-year from what it was 1-year prior, because they accrue the advantage of having survived that extra year of life.

The change in CIRS-\textit{G}_{\text{pros}} score from 0 to 3 (while holding age constant) results in the largest impact on life expectancy (between 1 and 3 years) at all ages. Thereafter, a 1-point change in CIRS-\textit{G}_{\text{pros}} score results in a loss of anywhere from 0 to 2 years of life expectancy. If we consider a man 65 years of age diagnosed with localized prostate cancer, our Sub-cohort model predicts that if the man were in otherwise perfect health (i.e. CIRS-\textit{G}_{\text{pros}} =0) he would have a survival advantage of 5 years, over a man of the same age with a CIRS-\textit{G}_{\text{pros}} score of 5. The survival advantage for the otherwise perfectly healthy man over a man of the same age in very poor health (i.e. CIRS-\textit{G}_{\text{pros}} =12) is 10 years. Although this appears to be a large loss of life expectancy, it does not have huge implications for treatment decisions at younger ages. A 55-year old man is only considered to be in the ‘grey-zone’ of uncertainty regarding treatment candidacy when they reach a CIRS-\textit{G}_{\text{pros}} score of 11. This is a very high and uncommon score among men diagnosed with localized prostate cancer, as only 4.8% of the sub-cohort population had CIRS-\textit{G}_{\text{pros}} scores of 11 or higher. If we look at the implications of our results for a 70-year old man, a different situation presents itself. Only the healthiest of 70-year old men (CIRS-\textit{G}_{\text{pros}} =0) would be considered clear candidates for curative treatment, with a predicted life expectancy of 14 years. This man would be expected to live 4-years longer than a man of the same age with a CIRS-\textit{G}_{\text{pros}} =5, and more than 8-years longer than a man of the same age with a CIRS-\textit{G}_{\text{pros}} score of 12. The results presented in Table 4.8b also imply that for a man age 70, only a patient with a CIRS-\textit{G}_{\text{pros}} of 11 or 12 would clearly not be a patient for curative treatment, because their predicted life expectancy is less than 7 years. Even at 70 years of age a CIRS-\textit{G}_{\text{pros}} score of 11 or 12 is still very uncommon (4.1% in the sub-cohort).

The results presented in Chapter 4, pertaining to the performance of the Sub-Cohort model indicated that the Sub-Cohort model’s LE predictions consistently over-estimate actual
patient survival. This over-estimation, and the model’s poor measures of PA suggest that the predicted LE values presented in the reference tables (Tables 4.8a-b) should be used with caution, in the context of both screening and treatment decision-making.

For younger patients, with high CIRS-Gpros scores the predictions may falsely indicate that the patient is a ‘clear’ curative treatment candidate and should proceed with the screening process or definitive treatment, depending on where they are in the spectrum of prostate cancer diagnosis and decision-making. This is illustrated in Tables 5.1 and 5.2 below. Although total numbers from the sub-cohort population are small for ages 50-59 (n=83) and even smaller for the combination of age 50-59 and CIRS-Gpros >=5 (n=47) the contingency tables indicate that among patients predicted to live long enough to receive screening or treatment (i.e. >=13 years) only 37% (30 of 82, Table 5.1) and 28% (13 of 46, Table 5.2) of patients actually lived >=13 years.

For patients in the middle age group (60-69 years old) Table 5.3 shows a similar trend in over-estimation by the Sub-Cohort model. For the patients with model predicted survival greater than 13 years only 32% (184 of 571) of patients actually survive that long. For model predicted survival times between 7-13 years, the model performs slightly better, as 52% (105 of 202) of the patients predicted to live between 7-13 years actually survived between 7-13 years.

<table>
<thead>
<tr>
<th>Sub-Cohort Population - Observed Survival Time (years)</th>
<th>Sub-Cohort Model - Predicted Life Expectancy (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;7yrs</td>
<td>7-13yrs</td>
</tr>
<tr>
<td>&lt;7yrs</td>
<td>0</td>
</tr>
<tr>
<td>7-13yrs</td>
<td>1</td>
</tr>
<tr>
<td>&gt;13yrs</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td>8</td>
</tr>
</tbody>
</table>
For older patients (70-79), the over-estimation of life expectancy, would have greatest
effect for men in good health (i.e., low CIRS-Gpros scores), because of the possibility of falsely
including them in either the ‘grey-zone’ or above (indicating they are a ‘clear’ curative treatment
candidates), when in fact they should be in the zone below (where delaying the screening process
or opting for conservative management is the ‘clear’ choice). Tables 5.4 and 5.5 demonstrate this
drawback of our model’s predictions. Focusing first on Table 5.4 (for men, 70 years or older) our
model over-estimates the LE of approximately 27% of patients (128 of 474) by predicting they
live 7-13 years, when they actually survived less than 7 years. It also incorrectly indicates that
treatment is the ‘clear’ chose (i.e. predicted LE >=13 years) 68% of the time, by incorrectly
including 45 of 66 patients in the >=13 year group.

When we restrict this analysis to just those individuals 70-79 years old who are also
healthiest (i.e. CIRS-Gpros <=3), the model incorrectly suggests screening or treatment as the
‘clear’ choice for an even greater proportion of patients. As indicated by Table 5.5, 73% (59 of
80) of patients the model predicts to have a life expectancy of >=13 years actually live less than
13 years.
Table 5.3 Contingency Table of Model-Predicted vs Observed Life Expectancy, Sub-Cohort population members with age 60-69 years.

<table>
<thead>
<tr>
<th>Sub-Cohort Model - Predicted Life Expectancy (years)</th>
<th>&lt; 7yrs</th>
<th>7-13yrs</th>
<th>&gt;13yrs</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;7yrs</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7-13yrs</td>
<td>43</td>
<td>105</td>
<td>54</td>
<td>202</td>
</tr>
<tr>
<td>&gt;13yrs</td>
<td>77</td>
<td>310</td>
<td>184</td>
<td>571</td>
</tr>
<tr>
<td>Total</td>
<td>120</td>
<td>415</td>
<td>238</td>
<td>773</td>
</tr>
</tbody>
</table>

Table 5.4 Contingency Table of Model-Predicted vs Observed Life Expectancy, Sub-Cohort population members with age 70-79 years.

<table>
<thead>
<tr>
<th>Sub-Cohort Model - Predicted Life Expectancy (years)</th>
<th>&lt; 7yrs</th>
<th>7-13yrs</th>
<th>&gt;13yrs</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;7yrs</td>
<td>7</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7-13yrs</td>
<td>128</td>
<td>264</td>
<td>82</td>
<td>474</td>
</tr>
<tr>
<td>&gt;13yrs</td>
<td>0</td>
<td>45</td>
<td>21</td>
<td>66</td>
</tr>
<tr>
<td>Total</td>
<td>135</td>
<td>312</td>
<td>103</td>
<td>564</td>
</tr>
</tbody>
</table>

Table 5.5 Contingency Table of Model-Predicted vs Observed Life Expectancy, Sub-Cohort population members with age 70-79 years and CIRS-Gpros <=3.

<table>
<thead>
<tr>
<th>Sub-Cohort Model - Predicted Life Expectancy (years)</th>
<th>&lt; 7yrs</th>
<th>7-13yrs</th>
<th>&gt;13yrs</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;7yrs</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7-13yrs</td>
<td>21</td>
<td>52</td>
<td>19</td>
<td>92</td>
</tr>
<tr>
<td>&gt;13yrs</td>
<td>14</td>
<td>45</td>
<td>21</td>
<td>80</td>
</tr>
<tr>
<td>Total</td>
<td>35</td>
<td>97</td>
<td>40</td>
<td>172</td>
</tr>
</tbody>
</table>
Because the Sub-Cohort LE prediction model is built on a survival function that is estimated from the sub-cohort populations observed survival times, the model’s predictions are bounded at the upper end by the maximum follow-up time observed in the sub-cohort (19 years). Therefore this element of the design of the Sub-Cohort model cannot explain the systematic over-estimation (error) between predicted and observed outcomes illustrated in these contingency tables and Figure 4.9 (scatterplot). This is perhaps better explained by violation of the PH assumption for the age effect in this model (discussed in further detail in section 5.4).

The model-predicted LE values refine the life expectancy estimates that may be commonly made by clinicians. The LE predictions in Tables 4.8a-b that are in bold, underlined, and italicized text represent the values that are closest to those found in the Ontario Life Tables, 1995-1997, males.(8) If there was no model-predicted LE value within 1-year of the OLT estimate for a given age, no value was highlighted on the table. These values help to illustrate how a comprehensive consideration of the patient’s co-morbid illness status can have a large impact on the patient’s predicted life expectancy and in some instances on either the screening or treatment decision.

When the post-treatment LE prediction values (Table 4.8a) are compared to the treatment adjusted LE prediction values (Table 4.8b) in the context of the OLT values, some differences are observed. The post-treatment values demonstrated a wider range of variability in terms of CIRS-G pros scores. For the post-treatment LE prediction closest to the OLT prediction, as age increases the corresponding CIRS-G pros score ranges from 0 (at lowest ages) to 5 (at the highest ages). While for the treatment adjusted predictions the CIRS-G pros score is not higher than 3 at any age.
5.3 Model Performance

As indicated in the results section, the Sub-Cohort LE prediction model demonstrated the ‘best’ overall performance: with a combination of discriminative ability demonstrated by a $c$-index of 0.65, and good calibration as illustrated by Figure 4.2a.

We demonstrated in the results section that each of the life expectancy prediction models we developed consistently provided more accurate predictions than a version of each model that excluded the consideration of CIRS-$G_{pros}$. Another way to determine if we were able to achieve our goal of developing a life-expectancy prediction model that is more accurate than calculations based on age alone was to calculate the overall $c$-index from the life expectancy estimates available in the Ontario Life Tables, males, 1995-1997 to predict the life expectancy of our sub-cohort population.(8) Life tables are the oldest and most widely available method for prediction of patient life expectancy during the treatment decision making process. They provide the average remaining life years based on age and gender considerations of a population.(9) The life expectancy values (represented by $e_x$) from the Ontario life table, males, 1995-1997, resulted in an overall $c$-index of 0.62. This measure of discrimination is 0.03 lower than that of our Sub-Cohort model.

Walz et al. demonstrated a similar predictive ability ($C=0.60$) of the Statistics Canada life tables for the prediction of a similar population of men with localized prostate cancer in Quebec.(1) So, all five of the prediction models developed in this study improved on the OLT’s predictions to some extent. However, none of the models outperform the OLT predictions by a great deal, which confirms the disappointing discriminative ability we were able to obtain.

It is worth noting here that there is some work involved in the calculation of a patient’s CIRS-G score that does add an extra step to the calculation of life expectancy. Calculating a patient’s CIRS-G involves reviewing their medical history to obtain information on other illnesses and morbidities suffered within the previous 5 years.(10, 11) However, it is often not practical for an attending clinician to perform this review during the patient consultation. It may
be more practical for this step to be performed prior to the patients visit by a member of the clinical staff. That said, the organ systems contained within the CIRS-G\textsubscript{pres} are those that have been shown to be prognostically important in this patient population and, therefore, could be used to guide the medical history taking for this patient population.\textsuperscript{(12)}

There is some evidence to suggest that clinicians intuitive estimates of prognostic outcomes may be biased and therefore inaccurate.\textsuperscript{(3, 13)} Several studies have focused on the accuracy of clinicians’ estimation of 10-year survival, because of its importance to the 10-year rule commonly adopted in treatment decision making.\textsuperscript{(2-4)} Walz et al assessed the accuracy of 19 clinicians predictions of LE for 50 patients treated with either radical prostatectomy or external beam radiotherapy. The clinicians were presented with patient age, co-morbidities, and Charlson co-morbidity index (CCI) scores, and asked to predict survival at 10-years. Overall predictive accuracy, as measured by the area under the receiver operative characteristics (ROC) curve was 0.68 (95\% CI:0.64-0.71) and individual accuracy rates ranged from 0.52 to 0.78. Since this study measured the accuracy of predicting a binary outcome (10-year survival, yes or no) it is hard to compare to the measures of accuracy obtained for the Sub-Cohort LE prediction model. However, the logistic regression we developed in our supplementary analysis demonstrates higher PA (C=0.71) than clinicians in the Walz study.

Koch et al, conducted a similar study, by comparing actuarial estimation of LE to the clinicians’ estimation of LE among a consecutive series of 261 men who underwent radical prostatectomy between 1995 to 1998. They found that clinicians accurately identified 51\% of patients who lived less than 10 years. These results demonstrate that clinicians’ estimates are barely better than the flip of a coin (pure chance).

However, Krahn et al, demonstrated that clinicians do predict the LE of curative prostate cancer treatment candidates with a high degree of accuracy. They asked 191 Canadian urologists and radiation oncologists to predict the LE for 18 patient scenarios: two prostate cancer scenarios, each with three ages and three levels of co-morbidity. They compared clinicians’ predictions to
‘true’ patient LE derived from a Markov model based decision-analytic model of life expectancy. Their results indicate that clinicians estimate LE greater than or less than 10 years with 82% accuracy. Furthermore they report that clinicians’ estimates were accurate within 1-year and 3-years of their Markov model predictions 31% and 67% of the time, respectively. Comparing these results directly to those produced by our Sub-Cohort LE prediction model may be difficult because the former are not based on actual outcomes and therefore are an overstatement of clinical accuracy. Nevertheless, our Sub-Cohort model demonstrated accuracy within 1-year and 3-years of actual survival 25% and 49% of the time, respectively.

Considering these results it is difficult to conclude whether clinicians are ‘poor’ raters of LE in our context, and whether our model’s predictions perform better than clinicians. At the very least our model is not likely to diminish clinicians’ estimates and in some instances it may significantly improve prediction accuracy and therefore lead to fewer errors in screening and treatment decision-making.

There are at least four other life expectancy prediction tools in the literature that were developed for the same objective as ours, using data from similar localized prostate cancer populations, and similar methods of analysis.(5-7,14) We compared the apparent predictive accuracy of the Sub-Cohort model to three of these prediction models, developed by Tewari, Cowen, and Walz, because they also considered co-morbid illness status and age as main predictors, used Cox PH regression analysis to estimate model parameters, and reported c-index as a measure of predictive accuracy.

Tewari, et al identified a cohort of 1,611 men diagnosed with clinically localized prostate cancer and 4,538 age, race and co-morbidity matched controls. They included patient age, race, co-morbidity status, PSA, Gleason grade, and treatment type as predictors for propensity score adjusted survival prediction models based on Kaplan-Meier and Cox PH methods. The model demonstrated an apparent discriminative ability of 63% (C=0.63). They validated their model internally using a 50/50 split-sample validation process that resulted in a c-index of 0.69. Cowen,
et al used a retrospective cohort of 506 men who were diagnosed and or treated for clinically localized prostate cancer between 1987 and 1989. They created a nomogram for deriving survival estimates based on 11 different patient demographic and disease severity variables using Cox PH regression modeling. These variables included patient age and Charlson co-morbidity index score, a common measure of co-morbid illness status. They performed a leave-one-out cross validation process among a subset of 133 men with complete data for all predictors to test the predictive accuracy of their model. The c-index for their nomogram’s predictions in this validation set was 0.73. Walz et al. studied a retrospective cohort of 9,131 men treated with radiotherapy or radical prostatectomy between 1989 and 2000, among whom deaths were due to causes other than prostate cancer. They used Cox PH regression models to create nomograms that predicted the probability of 10-year life expectancy. They included age, Charlson co-morbidity index score, and treatment type as predictors in their final model. Their model predictions were internally validated using a bootstrap re-sampling method. Their predictive model demonstrated very good discriminative ability, with a c-index of 0.83.

Although comparing different predictive models based on the c-index that results from the prediction of differing populations can be misleading and non-informative, all three of the studies we compared showed better discriminative ability than the best model from the current study. All three studies also performed some form of internal validation of their prediction tool, which helps support the validity of their findings.

One obvious difference between our LE prediction model(s) and those identified in the literature was the inclusion of more predictor variables in addition to age and co-morbid illnesses by the latter. These other predictors included indicators of disease severity such as PSA, Gleason score, and clinical stage, as well as treatment type, and other patient characteristics including smoking status and body mass index (note that although not accounted for independently, obesity and smoking status are considered in CIRS-G, see Appendix II). Our investigation did not consider disease severity indicators because, as indicated previously, in addition to being a
treatment decision aid, we wanted our tool to have utility as a screening decision aid. Some of the disease severity indicators included in other LE prediction tools appear to be strong predictors of LE. By not including these variables in our models we may have limited the potential predictive accuracy of our models. Given this potential limitation we feel that expecting predictive accuracy results comparable to these other models is not realistic.

There are also several popular website based prediction tools, often referred to as nomograms, that predict life expectancy for prostate cancer patients.(17, 18) Most notable is Kattan’s nomogram which has the ability to predict several outcomes including probability of disease metastasis and 5 and 10 year survival probabilities.(17). These nomograms are useful at both the pre-treatment and post-treatment decision making time-periods. For pre-treatment analysis the nomograms require up to 8 predictors of disease severity, while at post-treatment they require the consideration of up to 11 inputs. Cox PH regression analysis is used extensively on large cohorts of prostate cancer patients to estimate the parameters included in these nomograms.(17) They have been extensively validated and the PA varies between 0.52 – 0.81, depending on the outcome and type of validation performed.(13, 19-21) Kattan’s suite of prostate cancer nomograms also offers a simple life expectancy prediction tool, based on the patient’s current age and race. These estimates are based on the United States Life Tables, a set of population-based age and race specific tables of life expectancy estimates derived using similar methods to the life tables published by Statistics Canada.(22,8)

Another popular website based nomogram developed by Karakiewicz et al(18) predicts the probability of 10-year survival. The methods used to derive these predictions are based on the LE prediction tool developed by Walz et al(5). As described above, it includes the consideration of just two predictors; patient age and number of co-morbidities. This prediction tool demonstrated very discriminative ability with a c-index of 0.83.

There does not appear to be any prediction tools recommended for use at the time of biopsy or PSA screening. As outlined in the literature review section the decision of whether or
not to be screened for prostate cancer has become an increasingly important consideration for the entire population of men older than 50. The purpose of a LE prediction tool at the screening decision making stage should be to inform the patient and clinician of the patients’ remaining lifespan. Some of the current LE prediction tools and nomograms detailed above can be used to aid in the screening decision-making process, if they do not require the input of disease severity measures such as Gleason and PSA scores. We believe our LE prediction tool has a strong application at this stage because it includes the consideration of co-morbid illness status, which we’ve demonstrated provides an advantage over age only consideration and clinicians’ intuition.

The results of the logistic regression prediction model, developed as a supplementary analysis were encouraging for several reasons. First, this model provided a higher level of PA than we obtained in any of LE prediction models we developed in this work. However, the logistic regression model does have some limitations, when compared to the Cox PH model in this context. When the continuous form of the outcome is available, linear regression (or Cox PH regression in the case of time to event data) is often preferred, as it will often fit the data better.(23) Logistic regression cannot predict a continuous outcome. Since our original goal was to predict future life expectancy (a continuous outcome), logistic regression may be insufficient. However, the goal of predicting life expectancy was based on the need to advise clinicians and patients of the patient’s exact remaining lifespan in the context of the 10-year rule. The logistic regression model provides an alternative, the probability of survival at a fixed time point. By predicting the probability of survival at 10-years, based on the effects of just age and co-morbid illness status, the logistic model has application to both the screening and treatment decision-making processes. In fact it may provide a measure of the patient’s prognosis that is more meaningful to the patient or clinician and therefore is more useful in aiding their decisions.(24)

Besides demonstrating a high degree of predictive accuracy and providing an alternative interpretation of prognosis, the logistic regression model’s results were also encouraging because they demonstrated that the results we obtained from our other predictive models were reasonable.
The logistic regression model provided an opportunity to explore predictive accuracy for the available data without concerns for the error in prediction caused by various assumptions made in model development (see discussion of assumptions, below). The logistic regression model does not require the same assumptions as the Cox PH model with regards to the proportionality of hazards.

Since, this model did not improve predictive accuracy drastically, over our LE prediction models, it may be that the assumptions we made (and violated re: proportional hazard of the age effect) did not hinder our results a great deal. There is some expectation when using a measure of predictive accuracy with an upper limit (in this case 1.0) that a model should be judged by how closely it comes to that upper limit. However, the results of our logistic regression model, and the performance of several published LE prediction models, suggest that the upper limit may not be attainable in this context. Instead, it is possible that a higher degree of predictive accuracy than obtained in this work is not achievable.

5.4 Assumptions

The Case-cohort Target model’s predictions calibrated poorly to the observed survival of the sub-cohort, whereas the Simplified Case-cohort Target model’s predictions both calibrated well. Considering the sub-cohort data were used in the development of both of these models, it was surprising that the two models calibrated so differently. However, upon further investigation it appeared that this might have resulted in part because of the age effect adjustment (α) to account for unmeasured confounding due to co-morbid illnesses in the target population.

When we applied this adjustment we assumed that; 1) the survival experience estimated by Cox PH modeling of the target population was biased, and 2) the distribution of CIRS-Gₚₕₚ in the case-cohort study was representative of the CIRS-Gₚₕₚ distribution in the target population. First, we assumed an adjustment to the estimated survival experience in the Sub-Cohort model
was necessary because of potential confounding by CIRS-Gpros. As demonstrated in the analysis performed to calculate this adjustment (see Appendix V), in Cox PH analysis of the case-cohort population we observed a significant change in the age effect parameter depending on whether we controlled for the confounding effect of CIRS-Gpros. We assumed a similar change in age effect, depending on the consideration of CIRS-Gpros would exist in the target population. Therefore, we felt not adjusting for this potential confounding would leave of us with an estimate of the age effect that was bias, indicating a larger impact on death than the true impact.

Furthermore, we assumed that the methodology behind our calculation of the adjustment (see Appendix V) and its application to our Cox PH regression equation were free of errors and other false assumptions. When we applied the final adjustment $\alpha=0.78$, to the Case-Cohort Target model equation (see results section, equation 2) it lowers the independent age effect ($\beta_1$) and therefore results in age having less impact on future life expectancy than if the adjustment was not included. Applying the adjustment to the model also resulted in lowering the baseline hazard from that originally calculated based on the updated target population. The observed impact of the adjustment on both the age coefficient and baseline hazard are in the direction we expected, but we are uncertain if the size of the impact is reasonable.

Another a priori assumption made in the development of the of our model inputs was that the co-morbid illness effect seen in the case-cohort population is constant across time. This assumption had many implications for the current work. We were uncertain whether the effect seen in the case-cohort population, with an average follow-up of just 4.4 years, would be applicable to a model that was attempting to predict an outcome as much as 19 years (a maximum observed follow-up time) into the future. There was the potential that the co-morbidity effect changed significantly sometime after 4.4 years follow-up. If that were the case using the co-morbidity effect from the case-cohort study in our prediction models would result in at least some level of inaccuracy. Analysis done independent of this project by Dr. Groome found that the co-morbidity effect derived from the case-cohort population was not constant across time. The slope
of the effect estimate differed significantly from follow-up time less than 4 years, to follow-up greater than four years.(12) However, this finding was much less likely to be an issue in the Sub-Cohort model, because the CIRS-G<sub>pre</sub> effect was estimated from the sub-cohort data which contained much greater follow-up (up to 19 years) than the case-cohort data.

The results of our assessment of the proportional hazards assumption in the Sub-Cohort model suggest that use of Cox PH regression analysis to derive the estimate for the hazard due to age is not appropriate for the population studied. The violation of this assumption, the primary assumption required in Cox PH regression modeling, may undermine the validity of the entire model and may be responsible for its poor performance. When the proportional hazards assumption is violated it implies that the hazards are not proportional over time. The implications of fitting a model where the PH assumption is violated is that the parameter estimate for covariates with hazards that increase over time are overestimated, while for covariates with converging hazards the effect is underestimated.(25)

In this context it is likely that the parameter estimate for age ($\beta_1$) – for which the PH assumption was violated – is an underestimate. In turn this would imply that the Sub-Cohort LE prediction model overestimates LE (because the effect of age is underestimated). This error would have greater effect as age increases and is likely responsible, at least in part, for the consistent overestimation of LE made by our model.

Most researchers feel that when faced with data that demonstrates non-proportional hazards for a covariate of interest that their results are uninformative because of the biases of violating the assumption, and that in some cases this analysis should be abandoned. However, it has also been suggested that these concerns are unwarranted.(26) Allison has suggested that if for a particular covariate the PH assumption is violated it does not imply the coefficient is uninformative. Instead, he suggests “the coefficient represents a kind of ‘average’ effect over the period of observation”. Therefore, in many instances the coefficient may be sufficient and abandoning the analysis is unnecessary. In the context of the Sub-Cohort LE prediction model, if
we believe the age coefficient ($\beta_1$) represents an ‘average’ effect, we are likely limiting the predictive ability of the model, because of a lack of specificity of the effect estimate. But, the model as a whole should not be discredited because of the non-proportionality of age in the sub-cohort data.

There are also several methods available to perform Cox PH analysis in the presence of non-proportional hazards (when the violation of the assumption appears extreme) prior to abandoning the analysis. These include stratification by the covariate with non-proportional hazards, separate models for distinct time periods and the use of time-dependent covariate terms in the model. These are possible solutions that should be considered in future exploration of methods to predict LE in this context, if faced with non-proportional hazards.

5.5 Limitations

The current study has several important limitations. The lack of either internal or external validation for the predictions of our life expectancy models is one important limitation. For all prediction models the data used to test predictive accuracy (sub-cohort population) was also used, either in whole or in part, to train the data, i.e. develop the parameter estimates included in the models themselves. Because of these design limitations the evaluation of model performance in this study measures apparent predictive accuracy; the measure of accuracy derived when predictions are based on the same population used to estimate parameters included in the model. Apparent predictive accuracy usually leads to an optimistic estimate of performance, one that generally will not be improved upon by either internal or external validation. Therefore the predictive accuracy results we obtained are likely a best case scenario.

We had previously proposed to perform a leave-one-out cross validation process to identify a more valid measure of predictive accuracy then we obtained in this work. However, because the apparent performance our models were already so modest, validation
would not provide a more meaningful conclusion regarding the objectives we proposed for this study. Therefore, we decided that focusing on different methods to optimize parameter estimates to better fit the data was more important than performing a validation process.

There are other limitations with the methodology used to develop our models that may have led to their generally poor predictive accuracy. As mentioned previously, our models were not as complex as others identified in the literature. Including more variables in the Cox PH regression modeling as predictors of death, may have led to more accurate predictions. Another limitation related to the complexity of some of our models may be the lack of consideration of age specific CIRS-$G_{pros}$ effects. Stratifying our CIRS-$G_{pros}$ effect analysis by 5-year age groups could have resulted in stronger measures of the effect of co-morbidities. We did not have enough statistical power to perform these stratified analyses. A further limitation to our CIRS-$G_{pros}$ effect estimates may have resulted from performing this analysis with an outcome of all-cause mortality. Dr. Groome originally performed the co-morbid illness effect analysis within the case-cohort study with other cause death as the endpoint and demonstrated a stronger effect.\(^{(12)}\) For the patients in our study who died of prostate cancer, the co-morbid illness effect would likely be virtually zero. So, by including these patients in our analysis of the co-morbidity impact we may have diluted the strength of CIRS-$G_{pros}$ as a predictor of mortality.

A further limitation to our co-morbid illness effect estimates is that we did not consider the independent disease effects for the illnesses documented by the CIRS-$G_{pros}$ variable. In Dr. Groome’s work on the effect of co-morbid illnesses on mortality, in this population, very strong effects were demonstrated for some independent diseases/illnesses.\(^{(12)}\) Including the individual co-morbid illnesses that demonstrated the largest effect on mortality in our predictions models, as opposed to the cumulative effect of all illness (obtained using the CIRS-$G_{pros}$ scoring), may have increased the accuracy of life expectancy predictions.

A potential limitation of the current study is whether the survival experience in men diagnosed as many as 20 years ago (1990-1998) is relevant in men diagnosed today. The
characteristics and survival experience of men diagnosed with localized prostate cancer have changed since the advent of widespread PSA screening in the early 1990’s. In the period between the early 1990’s and 2000’s there was an increase in the average overall life expectancy in men ages 50 to 80 years diagnosed with prostate cancer. This increasing survival trend was greatest among 50 to 60 year olds, who gained 1 to 1.5 years of life expectancy. The gains in remaining life span decrease as age increases. Depending on the patient’s age and the magnitude of increase in life expectancy that may apply to our predictions there are implications on the management decision suggested by our model predictions. For many younger men (50 to 60 years), who confer the largest impact of the shifting trends across eras, the management decision would remain unchanged. However, at older ages, and higher CIRS-G scores, even a small increase in life expectancy could alter the management decision suggested by our Life Tables. Thus, if the change in life expectancy observed between the early 1990’s and 2000’s applies to our study population, some caution should be used in applying our results to men being diagnosed with the disease today.

Our results only apply to men 50 to 79 years of age. We believe the prediction of life expectancy as an aid in the screening or treatment decision-making process does not have application to patients outside of this age range. For patients younger than 50 the decision to screen and subsequently treat is straightforward, while in men older the 80 the decision not to screen or treat is often fairly straightforward. Additionally, the available data for the target population was made up of a relatively small proportion of individuals younger than 50 (n=232, 1.3%) or older than 79 (n=209, 0.5%), and we felt that the effects calculated using our data would not apply to very young and very old men diagnosed with localized prostate cancer.
5.6 Strengths

Our study has several strengths. It was a population-based study and, therefore, our results are free of potential case-selection bias often present in single institution studies. Also, because the study population was drawn from all of Ontario, and included a large number of subjects, our results are highly generalizable to a large geographic area. The use of the case-cohort design to study the effect of co-morbid illness status on mortality is also a strength of the current study, which was adopted from the parent study. We originally limited our follow-up to 10-years post diagnosis, to obtain a clear understanding of the role of co-morbidities during this important treatment decision-making time period. Because the outcome of interest is so rare, the case-cohort design is extremely efficient because it allowed us to ‘over-sample’ the deaths in a context that requires failure time analyses.(31) This efficiency produced precise and robust estimates of the effects of co-morbid illnesses indexed by the CIRS-Gpros.

Another strength of the current work was the exploration of several approaches to estimate the baseline hazard function in our LE prediction models. Originally, we did not feel comfortable with making an assumption that the survival experience of men with clinically localized prostate cancer followed a particular parametric form. Therefore, in three of our five models we used the Cox PH model to estimate the baseline survival experience. By definition this model makes no assumption about the form of the baseline hazard function, and therefore has the potential to remove inaccuracies in the prediction of life expectancy. We assumed, a priori that by using Cox PH survival analysis we would be able to produce a model with very good predictive accuracy, because we would be able to avoid the assumptions required by a parametric model. In two of the models we utilized a constant mortality rate that assumes the baseline hazard follows the form of an exponential function. Other life expectancy calculation tools have been criticized for assuming the hazard of death follows a particular parametric form. It has been demonstrated that assuming a particular form for the hazard leads to inaccuracies in many populations, including prostate cancer patients.(32, 33) Although it is difficult to conclude, based
on our model’s results, which method provides a better fit to the data in this context, the investigation of both methods was informative for the prediction of LE.

One of our underlying goals in the model development process was to find the best way to calculate life expectancy, while taking advantage of the extensive and high quality data available to us. We wanted to take advantage of the large number subjects in our calculations to improve the strength and stability of effect estimates. We also wanted to take advantage of the prostate cancer population-wide survival data available from the CCE Cancer database. Using Cox PH survival analysis for the model development allowed us to take advantage of the detailed and continually updated patient vital status information. It also allowed us to include the patient information for the variables we found to be important. We had large sample sizes (n=1,596 to 17,439) and no missing information in the calculations of our major effects: baseline hazard function, age effect, and CIRS effect, for all three models.

Furthermore, the detailed chart review process allowed us to collect the necessary information to calculate CIRS-Gpros scores for most patients in the case-cohort study population. We originally proposed contrasting predictions developed using the CIRS-Gpros to those using the chronic disease scale (CDS). However, we felt that a focus on modeling strategies using the psychometrically-stronger CIRS-Gpros was a better use of our time. The CIRS is a comprehensive co-morbid illness index that uses information typically recorded in patient charts. It has been shown to have content, criterion, concurrent and construct validity and reliability.(34, 35) Furthermore, the CIRS was previously piloted in this population and demonstrated substantial inter-rater reliability.(36)

A final strength of this study is that it adds to a sparse literature. The formal consideration of the impact of co-morbid illness status, in addition to age, in the prediction of life expectancy has not been thoroughly studied in this population. Albertsen et. al. developed a formula to predict patient life expectancy that included the impact of co-morbid illnesses over a decade
ago.(14) Beyond that work, and the recent work of Tewari, Cowen, and Walz (detailed above), there is very little available in the literature to help inform patients and clinicians on this issue.

5.7 Conclusions

The current studied was an exploration of methods to predict life expectancy for potential candidates of curative treatment in both the screening and treatment decision-making process. This exploration resulted in the development of five separate models for the prediction of life expectancy, based on the consideration of age and co-morbid illness status. All models demonstrated poor to fair predictive accuracy. Since the predictive accuracy of age alone is worse and other published life expectancy prediction tools provide similarly low levels of prediction, this may be as good as it gets. External validation of our sub-cohort and logistic models may be in order. The end result of this study are several look-up tables containing a) model-predicted life expectancy predictions, derived from the most accurate of the five life expectancy prediction models developed, and b) logistic regression model-predicted probabilities of 10-year survival. Considering the low level of accuracy and clinical utility demonstrated by these models’ predictions the tables are not ready for use in the clinic without further validation and exploration of other modeling approaches to improve accuracy.

5.8 References


Appendices

Appendix I - The Impact of Lymph Node Dissection on Treatment Candidacy

This investigation was meant to determine the impact on the case-cohort study population of including patients who had a lymph node dissection alone in the absence of bladder cancer in our definition of curative treatment candidates. We wanted to include this group because our objective was to develop a decision aid that would be applicable not just at the time of a treatment decision, but also at several other pre-treatment time periods including the time at which; 1) a PSA test decision is made, 2) a biopsy decision is made, and 3) a curative treatment decision is made, before lymph node dissection.

First we identified 7,921 patients from the initial group of curative treatment candidates (n=17,934) who received a lymph node dissection within seven months of diagnosis of their prostate cancer. Within this group it was determined that 1,443 of these patients received a lymph node dissection as the only treatment of their prostate cancer. We tried to determine how large this same group was within the case-cohort study population to determine if they (n=1443) should be excluded from the target population. We found that 279 patients within the case-cohort analysis received a lymph node dissection as the only treatment of their prostate cancer (approximately 10% of the entire study population). Among these patients there were only 52 who do not meet the inclusion criteria of being a potential candidate for curative treatment of prostate cancer. So, approximately 18% (52/279) of the individuals originally selected into the case-cohort study who had a lymph node dissection only were subsequently excluded because they were found to not be treatment candidates. If we then consider that 18% of the 1443 selected into the target population, potentially were not treatment candidates, we are dealing with roughly 268 patients. Because this 268 represent such a small fraction of the total target population (n=17934, before other exclusions) we chose not to alter the definition of curative candidates, and
retained the group of 1,443 who received a lymph node dissection only to better represent treatment candidates (rather than a group who ended up receiving treatment).
Appendix II – Cumulative Illness Rating Scale (CIRS)

Generic Categories:

0 = No problem
1 = Current mild problem or past significant problem
2 = Moderate disability or morbidity / requires first line therapy
3 = Severe / constant significant disability / ‘uncontrollable’ chronic problems
4 = Extremely severe / immediate treatment required / end organ failure / severe impairment function

Organ Systems Considered and Severity scores utilized in the revised scoring system specific to prostate cancer, CIRS-G$_{pros}$:

<table>
<thead>
<tr>
<th>Organ system</th>
<th>Severity Score</th>
<th>Revised Severity Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original CIRS-G</td>
<td>0   1   2   3   4</td>
<td></td>
</tr>
<tr>
<td>Cardiac</td>
<td>0   1   2   3   4</td>
<td></td>
</tr>
<tr>
<td>Vascular</td>
<td>0   1   3   4</td>
<td></td>
</tr>
<tr>
<td>Hematological</td>
<td>0   2   4</td>
<td></td>
</tr>
<tr>
<td>Respiratory</td>
<td>0   1   2   3   4</td>
<td></td>
</tr>
<tr>
<td>Eyes, ENT</td>
<td></td>
<td>Not included</td>
</tr>
<tr>
<td>Upper GI</td>
<td></td>
<td>Not included</td>
</tr>
<tr>
<td>Lower GI</td>
<td>0   3</td>
<td></td>
</tr>
<tr>
<td>Liver</td>
<td>0   4</td>
<td></td>
</tr>
<tr>
<td>Renal</td>
<td>0   1   2   4</td>
<td></td>
</tr>
<tr>
<td>Genitourinary</td>
<td></td>
<td>Not included</td>
</tr>
<tr>
<td>Musculoskeletal</td>
<td></td>
<td>Not included</td>
</tr>
<tr>
<td>Neurological</td>
<td>0   2   4</td>
<td></td>
</tr>
<tr>
<td>Endocrine/Metabolic</td>
<td>0   2   4</td>
<td></td>
</tr>
<tr>
<td>Psychiatric</td>
<td></td>
<td>Not included</td>
</tr>
</tbody>
</table>
## Example of CIRS-G Scoring Guide:

<table>
<thead>
<tr>
<th>CIRS - 0</th>
<th>Heart Disease (Cardiac)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No problem</td>
<td></td>
</tr>
</tbody>
</table>

| CIRS - 1 | Remote MI (> 5 yrs ago)/ occasional angina treated with prn meds/ mild atherosclerotic heart disease/ detectable murmurs that indicate valvular pathology without activity restriction (*more severely compromising valvular disease would require a higher rating, see instruction manual*) |

| CIRS - 2 | CHF compensated with daily meds/ daily anti-angina meds/ left ventricular hypertrophy/ ECG findings of atrial fibrillation/ right or left bundle branch block/ daily antiarrhythmic drugs/ moderate atherosclerotic heart disease/ placement of a pacemaker for an incidental finding of periods of bradycardia during a holter monitor |

| CIRS - 3 | Previous MI within 5 yrs/ abnormal stress test/ status post percutaneous coronary angioplasty or coronary artery bypass graft surgery/ CHF (intermediate severity)/ severe atherosclerotic heart disease/ bifascicular block/ placement of a pacemaker for cardiogenic syncope/ pericardial effusion/ pericarditis |

| CIRS - 4 | Marked activity restriction secondary to cardiac status (ie. unstable angina or intractable congestive heart failure)/ extremely severe atherosclerotic heart disease |

The original CIRS-G manual used for this study is available on the Internet. One source is: http://catalogue.iugm.qc.ca/GEIDEFile/CIRSG.PDF?Archive=102955792013&File=CIRSG_PD F
Appendix III: Analysis of Fit for Continuous vs Categorical Age variable in Target Population

We performed a graphical analysis of the fit several forms of the Age variable, to ensure the continuous form of the variable was correct and to avoid over-fitting the data in the Cox PH regression analysis. Over-fitting the data would result in overestimating the effect of age on death due to any cause and subsequently over-estimate the outcome – life expectancy.

First we performed a univariate investigation of the continuous form of the age variable, looking at the distribution of the variable and possible outliers. It was determined that the data were skewed slightly to the right, towards older ages between 70-79. We then compared the continuous form Age with a categorical form of age consisting of 6 categories to assess the fit graphically. The results of this analysis indicate poor fit, with the continuous form of age over-estimating the effect demonstrated by the categorical form of the variable.

Next we considered possible transformations of the age variable as an attempt to normalize the distribution of age and demonstrate a better fit to the categorical age groups. We calculate the square root (\(\sqrt{\text{age}}\)), log (\(\ln(\text{age})\)), and squared (\(\text{Age}^2\)) forms of the Age variable and performed univariate analyses on each of the three transformed variables. The distribution of \(\text{Age}^2\) was more normal than the other transformations and appeared to be slightly more normal than the continuous form of age. Therefore we explored the graphical fit of the age+\(\text{age}^2\) form of the age variable compared to the continuous age. This provided a slightly better fit to the effect of the categorical age groups as but still did not fit well – it is slightly outside the 95% CI toward older age categories.

Finally it was decided that we should investigate the fit of both the continuous form of age and \(\text{Age}^2\) to a categorical form of the variable containing more categories than initially analysed. We hypothesized that original categorical variable may not have been stable enough or sensitive enough to the change in age effect over time. This resulted in a much better fit of both
age and $Age + Age^2$. Although the $Age + Age^2$ transformation appeared to be a slightly better fit to the finer age categories, the original continuous form of age also fit well, and was a slight underestimate of the effect seen. In the end the continuous form of age was selected for the model because graphically it appeared to fit the data well, was more parsimonious than the $age + age^2$ transformation, and it slightly underestimated the effect of Age. The results of this analysis are presented in Figure 1a-b.

**Figure 1a Analysis of fit for Continuous vs Transformation vs 6-Categories of the Age variable, for Cox PH model building process.**
Figure 1b Analysis of fit for Continuous vs Transformation vs 10-Categories of the Age Variable, for Cox PH model building process.
Appendix IV – Macro to Calculate Life Expectancy from Survival Function

The macro performed the following steps for each patient in the sub-cohort population:

1) A dataset containing the hazard ratio part of the survival function equation:

\[ x = \exp((\beta_1 \cdot (AGE_i - \overline{AGE})) + (\beta_2 \cdot (CIRS_i - \overline{CIRS})) \] specific to each patient (based on age and CIRS-G_pro) was called into the macro.

2) The baseline survival function (previously calculated for each model) was raised to the power of the quantity \( x \) calculated in step 1, forming a set of probabilities of survival for each time interval observed for the baseline survival function, i.e. a patient specific survival function.

3) To calculate the median survival for each patient we used the patient specific macro-generated survival function. The macro read through the survival function probabilities to identify \( S(t)=0.50 \). If \( S(t)=0.50 \) existed among the survival probabilities then the time \( t \), corresponding to that probability was selected as the patient predicted life expectancy. In most cases the exact value \( S(t)=0.50 \) did not exist in the survival function, in which case the patient’s predicted life expectancy was calculated using the following macro generated steps:

   i. Identify the survival probability nearest to but less than 0.50 (\( S(a) \)) and nearest to but greater than 0.50 (\( S(b) \)).

   ii. Calculate the difference between the two survival times from step:

   \[ S(c)=S(b)-S(a) \]

   iii. Calculate the difference between the survival time greater than 0.50 in step i and 0.50: \( S(d)=S(b)-0.50 \)

   iv. Calculate the ratio of the quantities from step ii and iii: \( E=S(c)/S(d) \)

   v. Calculate the difference in survival times associated with the survival probabilities from step 1: \( F= t_b - t_a \)
vi. Multiply the quantity from step v by the ratio from step iv: $G = F \times E$

vii. Add the quantity from step vi to the survival time associated with the survival probability nearest to but greater than 0.50: $LE = S(b) + G$.

The quantity derived in step vii is the survival time at exactly $S(t) = 0.50$, and represents the predicted life expectancy.

Calculating life expectancy based on median survival was chosen as opposed to mean survival because the baseline survival curve of the target population contained no information for survival probabilities less than approximately $S(t) = 0.30$. This is simply a limitation of the data used to calculate the baseline survival function. Therefore calculating the mean survival by the area under the survival curve would have required extrapolating the curve beyond the observed data.
Appendix V: Adjustment for unmeasured CIRS-$G_{\text{pros}}$ in Target Population

In statistics if we have a Cox-PH model $H(t) = H_0(t) \cdot e^{(B_1 \cdot x_1 + B_2 \cdot x_2)}$, where $x_2$ is missing but is known to follow a certain distribution with parameter estimate $\alpha$, then the new model without $x_2$ can be calculated by the equation $H(t) = (H_0(t) \cdot e^{B_1 \cdot x_1})^\alpha$. This adjustment is applied to the model where an important potentially confounding covariate is missing as an attempt to account for the bias of the effect estimates of included covariates. In the literature there are several proposed methods to calculate such a parameter. (1-4) The method we chose is similar to that proposed by Hougaard et al. (4)

We utilized the case-cohort study population data to perform Cox-PH regression on two models:

i) The Adjusted model = Model A – included Age and CIRS-$G_{\text{pros}}$ as covariates, and

ii) The Unadjusted model = Model B – included only Age as covariate.

The methods to develop these two Cox PH regression models were analogous to that performed to generate the CIRS-$G_{\text{pros}}$ effect from the case-cohort population. Results are presented in Table 1, below.

<table>
<thead>
<tr>
<th>AGE</th>
<th>CIRS-$G_{\text{pros}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient</td>
<td>SE</td>
</tr>
<tr>
<td>Model A</td>
<td>0.020</td>
</tr>
<tr>
<td>Model B</td>
<td>0.026</td>
</tr>
</tbody>
</table>

4. From Cox PH Regression Analyses (CIRS-$G_{\text{pros}}$ and Age included as covariates) of Case-Cohort (N=2,416) (Langholz Case-Cohort Survival Analysis Methodology used).

5. From Cox PH Regression Analyses (Age included as covariates) of Case-Cohort (N=2,416) (Langholz Case-Cohort Survival Analysis Methodology used).

We calculated the ratio of the Age coefficient from Model A divided by the Age coefficient from Model B:

$$\alpha = \beta_{A(MODEL\ A)}/\beta_{A(MODEL\ B)} (11)$$
This quantity, $\alpha = 0.78$, represents the change in Age effect seen between a model that is adjusted for CIRS effect (Model A) and a model that is unadjusted (Model B). Since the effect comes from the case-cohort study population, which is sampled from the target population (and adjusted for sampling method), we felt it was reasonable to apply $\alpha$ to the target population.

This constant quantity is then applied to the Case-cohort Target model survival function equation. The adjustment is applied to the age effect by multiplication, and to the survival function, by raising each survival probability to the power of the adjustment, as illustrated in the model equation (6).
Appendix VI: Excess Mortality Due to Prostate Cancer

Excess rate due to prostate cancer = average yearly death rate in prostate cancer patients who are candidates for cure minus average yearly death rate for the general population.

Therefore, $\mu_{\text{prostate}} = 0.022 - 0.018 = 0.004$, where the baseline rate is from the Swedish Life Table data.(5)

<table>
<thead>
<tr>
<th></th>
<th>5-year survival</th>
<th>Average yearly rate</th>
<th>Baseline rate (pop)</th>
<th>Excess due to PC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Swedish Trial</td>
<td>0.897</td>
<td>0.022</td>
<td>0.018</td>
<td>0.004</td>
</tr>
<tr>
<td>Target population</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;55</td>
<td>0.916</td>
<td>0.018</td>
<td>0.004</td>
<td>0.014</td>
</tr>
<tr>
<td>55-59</td>
<td>0.927</td>
<td>0.015</td>
<td>0.007</td>
<td>0.008</td>
</tr>
<tr>
<td>60-64</td>
<td>0.904</td>
<td>0.020</td>
<td>0.012</td>
<td>0.009</td>
</tr>
<tr>
<td>65-69</td>
<td>0.881</td>
<td>0.025</td>
<td>0.019</td>
<td>0.007</td>
</tr>
<tr>
<td>70-74</td>
<td>0.849</td>
<td>0.033</td>
<td>0.029</td>
<td>0.003</td>
</tr>
<tr>
<td>75-79</td>
<td>0.768</td>
<td>0.053</td>
<td>0.050</td>
<td>0.003</td>
</tr>
<tr>
<td>&gt;79</td>
<td>0.635</td>
<td>0.091</td>
<td>0.089</td>
<td>0.002</td>
</tr>
</tbody>
</table>
Appendix VII: Treatment Effect Adjustment

We recognized that the baseline hazard function estimated from the target population may be low relative to a similar population of men who do not receive treatment, because of the benefit of receiving treatment with curative intent. We wanted to be able to account for this difference because our goal was to produce life expectancy estimates for men who have not yet received treatment. In order to account for the treatment effect in our baseline survival function we performed a literature review to identify high quality studies that provide an estimate of the mortality risk reduction due to curative treatment for a population of individuals (similar to our study population) who are treated with conservative management. Unfortunately we found very few published randomized controlled trials comparing the mortality experience of a treated vs watchful waiting group among men who were candidates for curative treatment.\(^{(6-8)}\) Among the studies we did identify Holmberg et al.’s randomized trial comparing radical prostatectomy to watchful waiting contained a study population most similar to ours in terms of demographics and the era in which the study was conducted. The published results section of this study presented an estimate of the mortality risk ratio for treated vs watchful waiting groups.

We calculated a constant offset (to be applied to The Sub-Cohort Model) by taking the inverse of the published risk ratio to quantify the risk increase for untreated men. Then computed the natural log of this quantity to produce a constant coefficient, \(\sigma_{Tx} = 0.301\), as shown in equation (12).

\[
\sigma_{Tx} = \ln(1/RR_{Holmberg}) \quad (12),
\]

where \(RR_{Holmberg} = 0.74\)

This value can then be added to hazard ratio part of the model (5) equation, to form the pre-treatment (adjusted) Sub-Cohort model, represented by equation (13)

\[
S(t) = S_0(t)^{e^{(\beta_x(t) + \beta_z(t) + \beta_2(t) \cdot \sigma_{Tx})}} \quad (13)
\]
Appendices References


