Estimating Prognosis of Patients with Kidney Cancer

Anita Robert

Thesis submitted to the University of Ottawa in partial fulfillment of

requirements for the MSc degree in Epidemiology with Specialization in Biostatistics

School of Epidemiology and Public Health,

Faculty of Medicine,

University of Ottawa

Preface

Anita Robert (AR) is the student. AR contributed to all aspects of the project including assisting with Canadian Kidney Cancer Information System (CKCis) cohort data preparation, preliminary and statistical research, development of the statistical analysis plan, statistical analysis and preparation of all individual chapters. Dr. Rodney Breau (RB) is the thesis supervisor. RB is a surgical oncologist at The Ottawa Hospital, Professor of Medicine (Urology) at the University of Ottawa and Scientist at the Ottawa Hospital Research Institute Clinical Epidemiology program. RB oversaw the development of the project and supervised the preparation of the thesis in addition to providing clinical expertise. Dr.Ranjeeta Mallick (RM) is the thesis co-supervisor. RM is an adjunct professor at the School of Epidemiology and Public Health and senior statistician in biostatistics at the Ottawa Hospital Research Institute. RM oversaw statistical analysis plan development, statistical analysis and thesis preparation. Dr.Daniel McIsaac (DM) served as a Thesis Advisory Committee (TAC) member. Dr McIsaac is a scientist at ICES and the Ottawa Hospital Research Institute, associate professor at the University of Ottawa, chair in Innovative Perioperative Care at the University of Ottawa and an anesthesiologist at the Ottawa Hospital. DM provided expertise on prognostic modelling and model assessment. Dr. Tim Ramsey also served as a Thesis Advisory Committee (TAC) member. TM is a senior scientist, scientific director and associate director at the Ottawa Hospital Research Institute as well as an associate professor at the University of Ottawa. TR provided insight on the biostatistical assessment of the effect of varying baseline hazard functions.

Ethics approval was obtained for this study. The Ottawa Health Science Network Research Ethics Board (OHSN-REB) approval (Protocol ID# 20220143-01H) is included in the Appendix.

Abstract

Kidney Cancer has numerous subtypes with Clear Cell Renal Cell Carcinoma (ccRCC) being the most common. Pre-existing prognostic models have not been validated in Canadian patients for recurrence free survival (RFS) and other outcomes. We conducted four studies: 1) externally validated pre-existing RCC prognostic models; 2) assessed the impact of baseline hazard function miscalibration on model assessment; 3) created new models and risk groups for RFS in non-metastatic ccRCC patients; 4) compared new risk groups to existing Canadian guidelines and created new imaging schedules. Pre-existing model performance varied considerably with some models performing well. The effect of baseline hazard function miscalibration varied across distribution shapes but the calibration slope was useful in relatively ranking prognostic model performance. The CKCis prognostic model and risk groups performed better than the existing CUA risk groups. Based on CKCis risk groups fewer scans are recommended in low-risk patients and more scans are recommended in higher risk patients. External validation of the CKCis model is required to assess clinical utility in different populations.

Acknowledgements

I am very thankful for all the support and guidance I received while writing my thesis. I would like to acknowledge the valuable contributions of the following people to this thesis.

My parents and siblings for all their love, support, and encouragement.

My supervisor Dr.Rodney Breau and co-supervisor Dr.Ranjeeta Mallick for your mentorship and guidance throughout this process. I am very honoured to have learned from and have worked with you both.

Thesis Advisory Committee member Dr.Daniel McIsaac for your expertise and guidance on various matters related to prognostic modelling.

Dr.Tim Ramsey for your biostatistical expertise and guidance on baseline hazard function research.

Camilla Tajzler, Ranjena Maloni and all the staff working with the Canadian Kidney Cancer Information System for their collaboration.

Table of Contents

Chapter 1: Introduction

1.1 Rationale

1.1.1 Renal cell carcinoma

Renal cell carcinoma (RCC) is a malignancy which originates in the kidney parenchyma and accounts for approximately 85% of renal tumours (1). RCC has multiple different histologic subtypes caused different genetic mutations. Clear cell renal cell carcinoma (ccRCC), is the most common (~75%) RCC subtype and is most often caused by mutations in the tumour suppressor-von Hipple-Lindau gene (2). The second most common (~15%) RCC histologic subtype is papillary renal cell carcinoma (pRCC) and the third most common (~5%) is chromophobe renal cell carcinoma (chRCC). Approximately 10% of patients with RCC are symptomatic at clinical presentation. The most common symptoms are hematuria, flank pain, and a palpable abdominal mass (2).

1.1.2 Incidence and risk factors for renal cell carcinoma

In Canada, there were approximately 7500 new RCC diagnoses and 1950 RCC-related deaths in 2020 (3). In Canada (Province of Quebec not included), the national kidney cancer incidence rate increased from 14.2 per 100,000 population in 2009 to 17.5 per 100,000 population in 2018 (4). As such, the 5-year duration prevalence rate has also increased between 2006 and 2015 from 39.0 per 100,000 to 56.3 per 100,000 (4). RCC incidence has been associated with lifestyle related risk factors such as smoking, increased body mass index, alcohol consumption, and possibly a lack of physical activity (1). Medical conditions associated with increased incidence of RCC include hypertension, chronic kidney disease, and diabetes meillitus. Environmental exposure to trichloroethylene and aristolochic acid may also be risk factors for RCC(1).

1.1.3 Renal cell carcinoma diagnosis

RCC tumours are diagnosed during evaluation of signs/symptoms (eg hematuria) or incidentally (eg. at the time of imaging for an unrelated issue, such as appendicitis). A renal tumour is identified during imaging tests such as ultrasound, computed tomography (CT), or magnetic resonance imaging (MRI). Given the high likelihood that a renal tumour is RCC, these tumours are often treated without biopsy. In some patients, pre-treatment percutaneous biopsy is performed to establish the histologic diagnosis and help guide therapy.

1.1.4 Treatment of non-metastatic renal cell carcinoma

Approximately 85% of patients with RCC are diagnosed prior to radiographic evidence of metastases (i.e. cancer spread to other areas of the body). Surgical resection is the most common and effective treatment for these patients (5). Other, non-surgical management options include active surveillance (for small

tumours that may be indolent), percutaneous thermal ablation (for tumours less than 3 cm), and radiation (for patients who are too unwell for surgery).

Surgical resection (partial nephrectomy or radical nephrectomy) allows for histologic diagnosis of the tumour and may be curative. Partial nephrectomy is selective removal of the tumour and preservation of the surrounding kidney. Radical nephrectomy is removal of the entire tumour bearing kidney and surrounding tissues (6).

1.1.5 Prognosis after renal cell carcinoma treatment

Unfortunately, some patients will have RCC recurrence after surgery and many patients with recurrence will die from their disease. Recurrence free survival is defined as the time between kidney cancer surgery and last follow-up or the first incidence of kidney cancer recurrence post-surgery. Cancer-specific survival is defined as the time between kidney cancer surgery and last follow-up, or death due to kidney cancer. Overall survival is defined as the time between kidney cancer surgery and last follow-up or death due to any cause. The risk of recurrence at 2 years after surgery is 3.7% to 39.3% (7). Recurrence free survival ranges from 12.7%-98.7% at 5-years post-surgery in Clear Cell RCC patients (8). Clearly, the range of reported outcomes are broad and are associated with clinical characteristics such as tumour stage and patient age (7,9-10).

1.1.6 Follow-up of patients after surgery

The most recent Canadian clinical practice guidelines on the follow up of patients with non-metastatic kidney cancer was published by the Canadian Urological Association in 2018. They used evidence on risk of recurrence and other clinical outcomes to devise patient follow-up schedules for medical history/physical examination, blood tests, chest x-rays and abdominal CT, MRI or ultrasound scans (5). These follow-up schedules are stratified by pathological T and pathological N stage (5). Based on these guidelines: 'Low Risk' patients are those with pT1 stage cancer, 'Intermediate Risk' patients have pT2 stage cancer, 'High Risk' patients have pT3 or pT4 stage cancer and 'Very High Risk' patients have pN1 stage and any pathological T stage cancer (5). Other than tumour stage, the CUA guidelines to not consider other prognostic factors such as tumour grade.

1.1.7 Prognostic models

Prognostic modelling is defined as statistical modelling in which at least two laboratory based or clinical parameters are combined to estimate prognosis for outcomes of interest in either entire patient populations or for individual patients (11).

1.1.8 Benefits of prognostic models

The main benefit of prognostic models is to inform patients of their risk and to use that information to optimize treatment or surveillance decisions. (11). Beyond individual patient care, prognostic models have a wide range of applications such as informing public health policy, comparative effectiveness research, health services research, and assessing baseline risk in diagnostic studies and clinical trials (12).

1.1.9 Potential problems with prognostic model research

Most prognostic models are developed in a specific sample of patients and these models may not perform as well in a broader population of patients. Indeed, prognostic models should be externally validated prior to clinical use (11). Common prognostic model pitfalls are inadequate sample size and the poor data quality, issues that are often exposed during external validation (11-13). There are also reporting issues in prognostic model research that do not allow for validation. Many prognostic model publications do not include important information to allow external validation. A common example of a reporting omission is the baseline hazard function. The baseline hazard function provides the underlying hazard of an event occurring over time, or the underlying survival distribution (14). Failure to consider the baseline hazard function of the population in which a prognostic model was developed may affect prognostic model assessment as the baseline hazard function of the population in which the model is assessed is imputed without considering the original population in which the model was developed (14). Another significant consideration is that prognostic models must be continuously evaluated over time as treatment efficacy can change and the model itself becomes a "victim of success" (15). In other words, a prognostic model may be externally assessed on the population of interest and found to have excellent prognostic performance, the model then informs patient treatment decisions, the underlying distribution of clinical events in the population changes as clinical decisions are impacted by the prognostic model and the model no longer performs as well in the population of interest (15). Therefore, prognostic models must be continuously assessed over time with new prognostic models created to assure accurate prognosis assessment among contemporary patients (15).

1.1.10 Prognostic models for patients with RCC

Over the last two decades, numerous prognostic models for kidney cancer recurrence free survival, cancer specific survival, and overall survival have been created using laboratory and clinical data. Various prognostic models have been developed for these clinical outcomes for all patients and patient subgroups based on tumour histology or tumour stage. In addition to tumour features such as pathological T stage, pathological N stage, nuclear grade, tumour necrosis, imaging size, sarcomatoid features and positive margins, demographic variables such as age at surgery, race, and sex are also present in some prognostic

models. Clinical features such as Charlson comorbidity score, Eastern Cooperative Oncology Group (ECOG) score, American Society of Anesthesiology (ASA) score, diabetes and hypertension have also been used as predictors in pre-existing models. The full list of prognostic models for kidney cancer recurrence, cancer specific survival, and overall survival is provided in the next chapter. Some models for kidney cancer recurrence have been externally validated in various populations (16). These studies had examined the discriminative and calibration accuracy in different populations (16).

While prognostic models exist, there is a need to further assess model performance. Many existing models have not been validated, the validation may not apply to contemporary patients, or the validation was performed in single institutions. In addition, many prognostic models fail to specify the baseline hazard function and external validation studies fail to consider the impact of differences in the baseline hazard function between the populations in which the model was developed and validated.

1.1.11 Importance of this thesis

In addition to assessing existing models, it is possible that new prognostic models will improve clinical care by more accurately informing patients of their prognosis and allowing for more appropriate follow up imaging to detect recurrence. Importantly, establishing the most accurate models will allow for better selection of patients for adjuvant treatments or inclusion into clinical trials. Thus, the aim of this thesis is four-fold: 1. Identify and externally validate existing prognostic models in a large Canadian cohort, 2. measure the impact of baseline hazard function misspecification on calibration and discrimination, 3. create new prognostic models based using data from Canadian kidney cancer patients, and 4. create new patient follow-up schedules based on newly-defined patient risk groups.

1.2 Objectives

Overall, in this thesis we aim to examine and improve the assessment of kidney cancer patients for risk of recurrence and death using data from Canadian kidney cancer patients. Based on existing clinical evidence, the following research objectives were defined for this thesis:

Objective 1: Externally Assess the Performance of Existing Kidney Cancer Prognostic Models in Canadian Patients

In Canadian kidney cancer patients, how do pre-existing prognostic models for kidney cancer recurrence free survival, cancer specific survival and overall survival perform in terms of discrimination and calibration?

Objective 2: Assess the Impact of Baseline Hazard Function Miscalibration on Assessment of Discrimination and Calibration

As different clinical study populations have different underlying baseline hazard functions for the event of interest, how does baseline hazard function miscalibration affect the assessment of calibration and discrimination in the external assessment of prognostic models? Of the different techniques used to address this issue, does the ideal method of baseline hazard function adjustment vary based on the type and level of deviation?

Objective 3: Create New Prognostic Models for Kidney Cancer Recurrence Free Survival and Stratify Patients into Patient Risk Groups

Can a new prognostic model for kidney cancer recurrence free survival be created based on data from Canadian kidney cancer patients? How do these new models perform in terms of calibration and discrimination compared to pre-existing prognostic models? Based on the new prognostic model created, can patients be divided into risk groups?

Objective 4: Recommend Optimal Follow-up Schedules Based on Risk of Recurrence.

How does the optimal follow-up schedule vary based on patient risk groups and recurrence free survival?

1.3 Chapter Descriptions

The research objectives discussed will be elaborated on as described below:

- Chapters 2 provides evidence regarding pre-existing prognostic models, introduces the patient population of interest and explains theoretical concepts related to prognostic model validation and development. Methods used in data analysis and prognostic model creation are also explained here. Chapter 7 is an overall discussion section.
- Chapter 3 elaborates on the first research objective and holds the first manuscript. This manuscript contains analysis on the performance of pre-existing prognostic models on kidney cancer recurrence free survival, cancer specific survival and overall survival in Canadian kidney cancer patients.
- Chapter 4 elaborates on the second research objective and holds the second manuscript. This manuscript contains results from a simulation study examining the effects of baseline hazard function on the assessment of calibration and discrimination in external validation studies. This study also compared various methods of incorporating baseline hazard function related

information against not adjusting for the original baseline hazard function. The final part of this study examined whether the selection of the best performing clinical model varied based on the shape of the baseline hazard function assumed if researchers decide to impute a given type of distribution.

- Chapter 5 elaborates on the third research objective and holds the third manuscript. This manuscript contains the new prognostic model for kidney cancer recurrence free survival. This chapter also covers the creation of new risk groups.
- Chapter 6 elaborates on the final research objective and holds the final manuscript. This manuscript contains details on new patient risk groups based on the best performing prognostic model for kidney cancer recurrence in addition to the revised patient follow-up schedules based on patient risk groups.

1.4 References

1. Scelo G, Larose TL. Epidemiology and Risk Factors for Kidney Cancer. J Clin Oncol. 2018 Oct 29;36(36):JCO2018791905. doi: 10.1200/JCO.2018.79.1905.
2. Petejova N, Martinek A. Renal cell carcinoma: Review of etiology, pathophysiology and risk factors. Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub. 2016 Jun;160(2):183-94. doi: 10.5507/bp.2015.050.
3. Brenner DR, Weir HK, Demers AA, Ellison LF, Louzado C, Shaw A, Turner D, Woods RR, Smith LM; Canadian Cancer Statistics Advisory Committee. Projected estimates of cancer in Canada in 2020. CMAJ. 2020 Mar 2;192(9):E199-E205. doi: 10.1503/cmaj.191292. Epub 2020 Mar 2. PMID: 32122974; PMCID: PMC7055947.
4. Statistics Canada. Table 13-10-0111-01 Number and rates of new cases of primary cancer, by cancer type, age group and sex. DOI: https://doi.org/10.25318/1310011101-eng
5. Kassouf W, Monteiro LL, Drachenberg DE, Fairey AS, Finelli A, Kapoor A, Lattouf JB, Leveridge MJ, Power NE, Pouliot F, Rendon RA, Sabbagh R, So AI, Tanguay S, Breau RH. Canadian Urological Association guideline for followup of patients after treatment of non-metastatic renal cell carcinoma. Can Urol Assoc J. 2018 Aug;12(8):231-238. doi: 10.5489/cuaj.5462.
6. Krabbe LM, Bagrodia A, Margulis V, Wood CG. Surgical management of renal cell carcinoma. Semin Intervent Radiol. 2014 Mar;31(1):27-32. doi: 10.1055/s-0033-1363840.
7. Dragomir A, Aprikian A, Kapoor A, Finelli A, Pouliot F, Rendon R, Black PC, Moore R, Breau RH, Kawakami J, Drachenberg D, Lattouf JB, Tanguay S. Follow-up imaging after nephrectomy for cancer in Canada: urologists' compliance with guidelines. An observational study. CMAJ Open. 2017 Dec 11;5(4):E834-E841. doi: 10.9778/cmajo.20170005.
8. Leibovich BC, Blute ML, Cheville JC, Lohse CM, Frank I, Kwon ED, Weaver AL, Parker AS, Zincke H. Prediction of progression after radical nephrectomy for patients with clear cell renal cell carcinoma: a stratification tool for prospective clinical trials. Cancer. 2003 Apr 1;97(7):1663-71. doi: 10.1002/cncr.11234.
9. Chin AI, Lam JS, Figlin RA, Belldegrun AS. Surveillance strategies for renal cell carcinoma patients following nephrectomy. Rev Urol. 2006 Winter;8(1):1-7.

10. De P, Otterstatter MC, Semenciw R, Ellison LF, Marrett LD, Dryer D. Trends in incidence, mortality, and survival for kidney cancer in Canada, 1986-2007. Cancer Causes Control. 2014 Oct;25(10):1271-81. doi: 10.1007/s10552-014-0427-x.
11. Vogenberg FR. Predictive and prognostic models: implications for healthcare decision-making in a modern recession. Am Health Drug Benefits. 2009 Sep;2(6):218-222.
12. Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A, Briggs A, Udumyan R, Moons KG, Steyerberg EW, Roberts I, Schroter S, Altman DG, Riley RD; PROGRESS Group. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. BMJ. 2013 Feb 5;346:e5595. doi: 10.1136/bmj.e5595.
13. Riley RD, Hayden JA, Steyerberg EW, Moons KG, Abrams K, Kyzas PA, Malats N, Briggs A, Schroter S, Altman DG, Hemingway H; PROGRESS Group. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. PLoS Med. 2013;10(2):e1001380. doi: 10.1371/journal.pmed.1001380.
14. Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. BMC Med Res Methodol. 2013 Mar 6;13:33. doi: 10.1186/1471-2288-13-33.
15. Lenert MC, Matheny ME, Walsh CG. Prognostic models will be victims of their own success, unless…. J Am Med Inform Assoc. 2019 Dec 1;26(12):1645-1650. doi: 10.1093/jamia/ocz145.
16. Usher-Smith JA, Li L, Roberts L, Harrison H, Rossi SH, Sharp SJ, Coupland C, Hippisley-Cox J, Griffin SJ, Klatte T, Stewart GD. Risk models for recurrence and survival after kidney cancer: a systematic review. BJU Int. 2021 Dec 16. doi: 10.1111/bju.15673.

Chapter 2: Background on Prognostic Factors and Prognostic Models

2.1 The Importance of Early Detection of Renal Cell Carcinoma Recurrence

After surgical resection, clear cell renal cell carcinoma recurrence can theoretically occur anywhere, the most common sites are in the lungs (70% of all recurrences), lymph nodes (45%), bones (32%), liver (18%) and adrenal gland (10%) (1). Computed Tomography (CT) scans are used most often for abdominal imaging and CT is much more sensitive than chest x-ray to detect metastases in the thorax. Magnetic Resonance Imaging (MRI) scans can be also used as a substitute for CT or when greater soft tissue detail is required, or when CT is not feasible (2). As most recurrences occur within three years of treatment, most intense screening for recurrence occurs within this time (2). This may often by supplemented with bone evaluations and brain scans on a case-by-case basis (2). Although guidelines vary in terms of specific time-points where imaging is required, all guidelines are consistent regarding two points: the recommendation of more frequent imaging in higher risk patients (i.e. more aggressive or advanced stage disease) and the use of stage as a predictor of risk (2).

Early detection of recurrence may be important as it allows for consideration of salvage treatments when the recurrent disease burden is lowest (3). For example, thermal ablation, radiation, or metastasectomy may be used to treat local disease recurrence or oligometastatic recurrence (2).

2.2 Prognostic Factors Potentially Associated with Kidney Cancer Clinical Outcomes

Prognostic models for kidney cancer recurrence free survival, cancer specific survival and overall survival have numerous predictors as outlined here.

Table 1: Relationship Between Predictors in Pre-Existing Prognostic Models and Recurrence Free Survival, Cancer Specific Survival and Overall Survival

Predictor	**Definition**	**Association with Recurrence Free Survival**	**Association with Cancer Specific Survival**	**Association with Overall Survival**
Demographic Predictors				
Age at Surgery	This refers to the age of the patient at time of surgery.	Older age has been associated with a non-significant decrease in recurrence risk (4).	Being between 42-91 yrs old has been associated with increased cancer specific mortality risk (RR=2.3, p<0.0001) compared to those aged 10-42 yrs (5).	Each additional decade in age was associated with increased risk for death from any cause (HR=1.48, 1.31-1.67) (4).

Sex	This refers to whether a patient's sex is male or female.	Female sex has been associated with lower risk of kidney cancer recurrence (HR=0.64, 95%CI:0.479-0.829) (6).	Female sex has been associated with lower cancer specific mortality (HR=0.70, 95%CI:0.61-0.80) (7).	Female sex has been associated with lower risk of death from any cause (HR=0.80, 95%CI:0.73-0.88) (7).
Race/Ethnicity	Race is often described as a group of people with common history, nationality, or geographic distribution. Ethnicity is often a self-described belonging to an ethnic group.	Race was not found to be a significant predictor of the risk of recurrence (8).	In terms of race/ethnicity, Black and White race/ethnicities were found to have higher kidney cancer specific mortality risk than patients with other race/ethnicity (9).	In terms of race/ethnicity, Black race/ethnicity was associated with higher mortality compared to White race/ethnicity (HR=1.28, 95%CI:1.00-1.64) (10).
Tumour Related Predictors				
Pathologic Tumour Size	Tumour size is usually defined as the maximum tumour diameter from the surgical specimen (6).	Increased tumour size has been associated with increased recurrence risk (HR=1.02, 95%CI:1.01-1.02) (6).	Increased tumour size is associated with increased risk of cancer specific death (11).	Increased tumour size is associated with increased risk of death from any cause (12).
Pathologic Tumour (pT) stage	This refers to the T stage component of the American Joint Committee on Cancer's TNM staging system based on the surgical specimen (13). T1 and T2 tumours are restricted to the kidney only with T1 tumours having size ≤7 cm and T2 tumours having size >7cm (14). T3 tumours are those invading perinephric tissues and/or major veins and T4 tumours extend further than the Gerota's fascia (13).	pT3-4 stage kidney cancer compared to T1-2 stage cancer has been associated with a greater risk of recurrence (HR=7.8, 95%CI:5.3-11.3) (8).	pT2 (HR=3.69, 95%CI:2.50-5.46), pT3 (HR=7.76, 95%CI:5.72-10.53) and pT4 (HR=22.69, 95%CI:12.06-42.69) patients have higher risk of cancer specific death compared to T1 stage patients (15).	pT2 (HR=1.97, 95%CI:1.43-2.71), pT3 (HR=4.43, 95%CI:3.42-5.49) and pT4 (HR=10.46, 95%CI:5.69-19.25) patients have higher risk of death from any cause compared to pT1 patients (15).
Pathologic Lymph Node (pN) stage	pNx indicates that lymph node involvement could not be assessed since nodes were not surgically resected. pN0 indicates no cancer in resected lymph nodes, and pN1 stage	pN1 is associated with increased risk of recurrence (HR=4.9, 95%CI:3.7-6.7) (16).	pN1 is associated with risk of cancer specific death (HR=8.4, 95%CI:5.2-13.6) (15).	pN1 is associated with risk of death from any cause (HR=6.5, 95%CI:4.3-9.9) (15).

	indicates the presence of cancer in lymph nodes (13).			
Nuclear Grade	Nuclear grade refers to the microscopic appearance of the cells. For clear cell carcinoma, grade 1 has nuclei which are not clearly visible at 400x magnification, grade 2 has slightly larger nuclei visible at 200x magnification, grade 3 has clearly visible nuclei at 100x magnification and grade 4 has large pleomorphic cells (16).	For ccRCC Grade 3-4 is associated with increased risk of recurrence compared to grade 1-2 (HR=1.02, 95%CI:1.10-2.37) (14).	For ccRCC Grade 3-4 is associated with increased risk of cancer specific death compared to grade 1-2 (HR=3.84, 95%CI:2.66-5.54) (15)	For ccRCC Grade 3-4 is associated with increased risk of death from any cause (HR=2.34, 95%CI:1.79-3.05) (15)
Tumour histologic subtypes	Based on genetic etiology and morphological features, different subtypes of kidney cancer exist. The most common are Clear Cell Carcinoma (ccRCC), Papillary Renal Cell Carcinoma (pRCC), and Chromophobe Renal Cell Carcinoma (chRCC) (17).	ccRCC has been associated with an increased risk for recurrence compared to chromophobe/papillary /oncocytoma (HR=2.20, 95%CI:1.24-3.91) (14).	ChRCC has been associated with decreased risk of cancer specific mortality compared to clear cell (RR=0.2, p=0.004) and papillary RCC (5).	Clear cell RCC has been associated with increased risk of death from any cause (HR=1.01, 95%CI:0.71-1.44) (14).
Necrosis	Tumour necrosis is defined as the non-ischemic homogenous accumulation of tumour cells that are dead and degraded (18).	Tumour necrosis has been associated with increase recurrence risk (p=0.0003) (19).	Tumour necrosis has been associated with increased cancer specific mortality risk (HR=2.70, 95%CI:1.89-3.86) (15).	Tumour necrosis has been associated with increased mortality risk (HR=1.82, 95%CI:1.32-2.53) (15).
Lymphovascular Invasion	Lymphovascular invasion is present when tumour cells are located in spaces lined by endothelium that lack muscular walls (20).	Lymphovascular invasion has been associated with increased kidney cancer recurrence (OR=3.1, 95%CI:2.2–4.2) (20).	Lymphovascular invasion has been associated with increased cancer specific mortality (HR=3.6, 95%CI:2.3-5.3) (15).	Lymphovascular invasion has been associated with increased all-cause mortality (HR=2.66, 95%CI:1.88-3.78) (15).
Sarcomatoid Features	Sarcomatoid features are defined as tumour cells with giant cell and spindle cell features with large variation in morphology (21).	While sarcomatoid features is associated with increase risk of kidney cancer recurrence, this increased risk was not statistically significant (22).	Sarcomatoid features is associated with increased risk of cancer specific mortality (HR=5.6,	Sarcomatoid features is associated with increased risk of all-cause mortality (HR=3.8,

			95%CI:3.17-9.78) (15).	95%CI:2.22-6.47) (15).
Perinephric and or Renal Sinus Fat Invasion	Renal sinus fat invasion occurs when tumour cells are in direct contact with fat cells or renal sinus stroma (23). Perinephric fat invasion occurs when tumour cells are in direct contact with fat cells or perinephric stroma (23). The presence of both or either are key features of pT3a stage tumours (23).	Neither were significantly associate with risk of kidney cancer recurrence (24).	Perinephric and renal sinus fat invasion was associated with increased risk for cancer specific mortality (HR=1.19, 95%CI: 1.01-1.40) (25).	Perinephric and renal sinus fat invasion was associated with increased risk for all-cause mortality (HR=1.21, 95%CI:1.06-1.38) (25).
Tumour Thrombus	Tumour thrombus is classified into levels based on extension of the tumour into the venous circulatory system: level 0 is defined as limited extension to the renal vein, level 1 is defined as extension <2cm from the renal vein ostium, level 2 is defined as extension between >2cm from the renal vein ostium and the hepatic vein, level 3 is defined as extension between the hepatic vein and the diaphragm and level 4 is defined as extension of the tumour into the right atrium (26). The presence of tumour thrombus is the key feature of pT3b/c stage tumour.	When compared to level 3-4 thrombus, the level of tumour thrombus does not significantly affect recurrence free survival (26).	When compared to level 3-4 thrombus, the level of tumour thrombus does not significantly affect cancer specific survival (26).	When compared to level 3-4 thrombus, the level of tumour thrombus does not significantly affect overall survival (26).
Extension Beyond Kidney	This refers to whether the tumour extends beyond the Gerota's fascia of the kidney into adjacent tissues, such as the liver, spleen, intestine, diaphragm, or skeletal muscle. This is the key feature of pT4 stage tumour.	Extension beyond the kidney is associated with increased risk for kidney cancer recurrence (HR=1.86, 95%CI:1.27-2.70) (11).	Extension beyond the kidney may be a factor affecting cancer specific survival in Clear Cell RCC patients(11).	Extension beyond the kidney has historically resulted in higher death rates when compared to localized kidney cancer (27).
Positive Margin	Positive margins are ascertained by examining inked tissue specimens used in the pathological evaluation of kidney cancer with the presence of aggregated tumour cells at the inked margin indicating positive surgical margins (28).	Positive margins are not associated with recurrence free survival (28).	Positive margins are not associated with cancer specific survival (28).	Positive margins are not associated with overall survival (28).

Clinical Predictors				
Surgical extent	This refers to whether a patient was treated with partial or radical nephrectomy. Partial nephrectomy indicates that only the portion of the kidney bearing the tumour is resected. Radical nephrectomy indicates that the entire kidney and surrounding tissue is resected.	Partial nephrectomy has been associated with lower kidney cancer recurrence risk compared to radical nephrectomy (HR=0.37, 95%CI: 0.25-0.55) (8).	Partial nephrectomy has been associated with lower risk of cancer specific mortality (HR=0.17, 95%CI: 0.12-0.25) (29).	Partial nephrectomy has been associated with lower risk of death from any cause (HR=0.46, 95%CI:0.38-0.55) (29).
Clinical Presentation (i.e. symptoms)	Symptoms at clinical presentation. Incidental tumours are defined as those found without any symptoms (i.e. those found when examining other medical conditions) (6). Examples of local signs/symptoms include hematuria and flank pain; examples of systemic symptoms include weight loss and fatigue (6).	Local symptoms (HR=1.98, 95%CI:1.45-2.64) and systemic symptoms (HR=4.10, 95%CI:2.93-5.69) have been shown to increase the risk of kidney cancer recurrence (6).	While local symptoms did not significantly increase the risk of cancer specific death, systemic symptoms increased the risk of cancer specific death (RR=2.6, p=0.001) (30).	Local or systemic symptoms were associated with an increased risk of death from any cause (HR=1.57, 95%CI:1.21-2.03) (14).
ECOG Score	ECOG score is a clinical measure of function and can be classified as follows: 0 indicated complete functioning without restrictions, 1 indicates functioning with mild physical restrictions, 2 indicates the ability to complete all self-care related tasks but unable to work, 3 indicates limited capacity for self-care, 4 indicates complete disability and 5 indicates being deceased (31).	ECOG scores >0 increase the risk of recurrence but this increase is not statistically significant (32).	Increased ECOG scores result in increased risk of cancer specific mortality (HR=2.3, 95%CI:1.5-3.5) (33).	Increased ECOG scores result in increased risk of death from any cause (HR=2.1, 95%CI:1.5-2.7) (33).
Diabetes Mellitus	This refers to whether a patient has diabetes mellitus.	Diabetes has been associated with increased kidney cancer recurrence risk (HR=1.95, 95%CI: 1.23-3.09) (8).	Diabetes has been associated with a non-significant increase in cancer specific mortality risk (HR=1.13, 95%CI:0.78-1.62) (34).	Diabetes has been associated with a non-significant increase in the risk of all-cause mortality (HR=1.14,0.94-1.38) (34).
Hypertension	This refers to whether a patient has hypertension.	Hypertension has been associated with increased kidney	Hypertension has been associated with increased	Hypertension has been associated with increased

		cancer recurrence risk (HR=1.8, 95%CI: 1.2-2.7) (8).	cancer specific mortality (HR=1.5, 95%CI:1.1-2.0) (15).	all-cause mortality (HR=1.4, 95%CI:1.1-1.8) (15).
ASA Score	The American Society of Anesthesiologists (ASA) score is a scoring system used to classify patient status and has the following categories:1=normal healthy patient, 2=patient has mild systemic disease, 3=patient has non-life threatening severe systemic disease, 4=patient has life threatening severe systemic disease, 5=patient that is not expected to survive without operation and 6=patient that is brain dead and being considered as an organ donor (15).	Higher ASA scores were associated with increased risk of kidney cancer recurrence (HR=1.6, 95%CI: 1.2-2.1) (8).	ASA scores of 3 and 4 were associated with higher risk of kidney cancer specific mortality compared to ASA scores of 1 and 2 (HR=3.2, 95%CI:2.0-5.3) (15).	ASA scores of 3 and 4 were associated with higher risk of mortality compared to ASA scores of 1 and 2 (HR=4.2, 95%CI: 2.9-5.9) (15).
Age adjusted Charlson Co-morbidity Score	The Charlson co-morbidity score is a measure used to define patient health status. In this score, each co-morbidity is given a point value based on its contribution to a patient's mortality risk. In the age adjusted Charlson co-morbidity score, each decade over the age of 40 is given one point up to a maximum of 4 points and this is added to the patient's total points based on co-morbidity (35).	The age adjusted Charlson co-morbidity score is not associated with recurrence risk (36).	Age adjusted Charlson co-morbidity scores above 6 were associated with an increased cancer specific mortality risk (HR=9.4, 95%CI:2.6-33.7) (36).	Age adjusted Charlson co-morbidity scores of 4-5 (HR=2.7, 95%CI:1.1-6.4) and ≥6 (HR=12.8, 95%CI:5.1-32.1) were associated with increased mortality risk (36).
Adrenalectomy	Adrenalectomy is the surgical removal of the adrenal gland (37).	Adrenalectomy is associated with increased risk of kidney cancer recurrence (HR=1.2, 95%CI:1.0-1.7) (38).	Adrenalectomy is not associated with cancer specific survival (HR=1.1, 95%CI: 0.95-1.22) (39).	Adrenalectomy is not associated with overall survival (HR=1.09, 95%CI: 0.86-1.38) (38).

Abbreviations: HR=Hazard Ratio; CI=Confidence Interval; RR=Risk Ratio; pT=pathological T stage; pN=pathological N stage; T=T stage; ccRCC=Clear Cell Carcinoma; pRCC=Papillary Renal Cell Carcinoma; chRCC=Chromophobe Renal Cell Carcinoma; RCC=Renal Cell Carcinoma

The age adjusted Charlson Co-morbidity score could be calculated using the following with one point added for each co-morbidity present:

Table 2: Age Adjusted Charlson Co-morbidity Score Calculation (35)

Number of Points	Co-morbidity
1	Myocardial Infarction, congestive heart failure, peripheral vascular disease, cerebral vascular disease, dementia, Chronic Obstructive Pulmonary Disease (COPD), connective tissue disease, ulcers, mild liver disease, diabetes
2	Hemiplegia, moderate/severe renal disease, diabetes with end-organ disease, any tumour, leukemia/lymphoma
3	Moderate/severe liver disease
6	Metastatic solid tumour, Acquired Immunodeficiency Syndrome (AIDS)
1	1 point for each decade above 40 years of age up to a maximum of 4 points for ages 70 and above

2.3 Importance of Prognostic Models

In addition to improving patient counselling, the implementation of prognostic models also have other benefits. Prognostic models inform public health policy through modelling disease burden in the population, allowing for the assessment of primary and secondary prevention healthcare activities (40). Comparative effective and health services research also evaluates healthcare interventions through changes in prognosis over time or examining prognostic distributions pre- vs. post-intervention (40). While new technology allows for the identification of new disease markers, clinicians must examine if prognosis changes due to the implementation of these new techniques (40). Prognostic models also model outcome distributions across clinical trial arms, assisting with sample size calculation (40). The use of prognostic models also detects disease mechanisms through the examination of prognostic outcomes in rare diseases (40). Additional information to understand diagnosis and improve clinical management may also be obtained from prognostic models (40). Finally, understanding the distribution of prognosis can contribute to the definition of disease characteristics (40). Given the wide range of benefits of using prognostic models, it is important to study prognostic models in great detail prior to use to improve patient outcomes.

2.4 Pre-Existing Prognostic Models

Over the past 20 years, numerous prognostic models have been developed to estimate kidney cancer recurrence free survival, cancer specific survival, and overall survival. Prognostic models have been created for all renal cell carcinoma (RCC) patients and specific patient subgroups (e.g. specific to RCC

histologic subtypes or low/high stage) (41). The pre-existing prognostic models included in this thesis are summarized below (Table 3).

Table 3: Kidney Cancer Recurrence, Cancer Specific Survival, Overall Survival Prognostic Models

Model	Patient Population Used to Generate Model	Prognostic Factors in Model; C-statistic Reported; Parameter Selection Method
Cancer Recurrence Free Survival Prognostic Models		
All Histologic RCC Subtypes		
Yaycioglu et al, 2001 (42)	N=296; Open nephrectomy patients excluding: bilateral synchronous disease or metastases, von Hippel-Lindau disease, lymphadenopathy, tumour that extends further than Gerota's fascia nor vena cava involved further than diaphragm	Clinical presentation, size largest tumour; c-statistic not available; parameter selection method not given
Cindolo et al, 2003 (43)	N=660; Surgical patients excluding: bilateral synchronous/metachronous disease, von Hippel-Lindau disease, tumour that extends further than Gerota's fascia, lymphadenopathy nor metastases	Clinical presentation, size largest tumour; c-statistic not available; forward selection used to select parameters
Zisman et al, 2002 (44)	N=814; Surgical patients for unilateral kidney cancer.	Pathological T stage, Nuclear Grade, ECOG, pathological N stage; c-statistic not available; parameter selection method not given
Kattan et al, 2001 (45)	N=601; Partial or radical nephrectomy patients for unilateral tumours. Excluding: bilateral synchronous disease, metastases, lymph node positive disease	Clinical Presentation, Histology, pathological T stage, size largest tumour; c=0.74; parameter selection method not given
Yacioglu et al, 2013 (6)	N=1889; Surgical patients excluding: von Hippel-Lindau disease or bilateral tumours	Age at surgery, gender, clinical presentation, imaging size, clinical T stage, radiological lymph nodes (clinical N stage); c-statistic=0.74; used all parameters significant in univariate analysis to create multivariate models
van der Mijn et al, 2019 (8)	N=873 Surgically treated (radical or partial nephrectomy) patients with non-metastatic kidney cancer patients aged ≥18 years who underwent partial or radical nephrectomy	Gender, extent, ASA score, pathological T stage, nuclear grade, WHO kidney cancer histology classification, diabetes mellitus, hypertension; c-statistic not reported; used all parameters significant in univariate analysis to create multivariate models
Passalacqua et al 2014 (46)	N=310; Surgically treated (radical or partial nephrectomy) patients with aged <75 years	Age, pathological T stage, nuclear grade, pathological nodal status (pathological N stage); c-statistic not reported; parameter selection method not given
Clear Cell Renal Cell Carcinoma Patients		
Leibovich et al, 2003 (16)	N=1671; Radical nephrectomy patients for clear cell renal cell carcinoma excluding: bilateral synchronous tumours, hereditary cancers, von Hippel-Lindau and tuberous sclerosis syndromes, Wilms tumour, metastases, age <18 years	Pathological T stage, pathological N stage, nuclear grade, necrosis, size largest tumour; c=0.819; stepwise selection used to select parameters

Sorbellini et al, 2005 (47)	N=701; Nephrectomy patients for clear cell renal cell carcinoma excluding: pT3c/pT4 cancers, metastases, von Hippel Lindau disease, hereditary papillary renal cell carcinoma, bilateral tumours	Size largest tumour, pathological T stage, nuclear grade, necrosis, lymphovascular invasion, clinical presentation; c=0.82; parameter selection method not given
Leibovich et al, 2018-Clear Cell Renal Cell Carcinoma (cRCC) (11)	N=2726; Clear cell renal cell carcinoma. Excluding metastatic patients	Clinical presentation, nuclear grade, necrosis, sarcomatoid, size largest tumour, perinephric or renal sinus fat invasion, tumour thrombus, extension beyond kidney, nodal involvement; c=0.83 for ccRCC, forward selection and backward selection (
Papillary Renal Cell Carcinoma Patients		
Klatte et al, 2019 (48)	N=556; Sporadic non-metastatic, unilateral kidney cancer patients with papillary renal cell carcinoma who underwent partial or radical nephrectomy	Size largest tumour, pathological T stage, pathological N stage, nuclear grade, venous tumour thrombus; c-statistic not reported; backward selection used to select parameters
Leibovich et al, 2018-Papillary Renal Cell Carcinoma (pRCC) (11)	N=607; Non-metastatic pRCC	Nuclear grade, perinephric or renal sinus fat invasion, tumour thrombus; c=0.77; parameter selection method not given
Chromophobe Renal Cell Carcinoma Patients		
Leibovich et al, 2018-Chromophobe Renal Cell Carcinoma (chRCC) (11)	N=222; Non-metastatic chRCC	Nuclear grade, perinephric or renal sinus fat invasion, tumour thrombus; c=0.78 for pRCC; parameter selection method not given
High Risk Patients Only		
Correa et al, 2021 (49)	N=1943; Kidney cancer patients with pathological stage pT1b and nuclear grade G3/G4 or pathological stage pT2/3/4 or pathological nodal stage pN1 who underwent radical nephrectomy	Lymphovascular invasion,histological subtype, size largest tumour, nuclear grade, necrosis, regional lymph nodes (pathological N stage); c=0.68; include all clinically significant variables whose inclusion resulted in ≥1% increase in c-statistic
Low Risk Patients Only		
van der Mijn et al 2019 (8)	Non-metastatic kidney cancer patients aged ≥18 years who underwent partial or radical nephrectomy with pathological T1/T2 stage	Gender, extent, ASA score, pathological T stage, nuclear grade, WHO kidney cancer histology classification, diabetes mellitus, hypertension; c-statistic not reported; used all parameters significant in univariate analysis to create multivariate models
Cancer Specific Survival Prognostic Models		
All Patients		

Kutikov et al, 2010 (9)	N=30801; Kidney Cancer patients who underwent surgery aged ≥30 years and tumour size <20cm	Race, Gender, age at diagnosis, size largest tumour; c-statistic not available; parameter selection method not given
Kutikov et al, 2012 (50)	N=6665; First time kidney cancer patients Excluding nodal positive, age under 66 years, uncommon histological subtypes, and tumours >20cm	Race, Gender, size largest tumour, Charlson Comorbidity Score, age at diagnosis; c-statistic not available; parameter selection method not given
Zisman et al, 2002 (44)	N=814; Patients who underwent surgery for unilateral kidney cancer	Pathological T stage, Nuclear Grade, ECOG, pathological N stage; c-statistic not available; parameter selection method not given
Clear Cell Renal Cell Carcinoma Patients		
Frank et al, 2002 (51)	N=1801; ccRCC patients who underwent radical nephrectomy excluding: bilateral synchronous tumors, herediary von Hippel-Lindau or tuberous sclerosis syndromes, Wilms tumor and those aged ≤18 years at surgery	Pathological T stage, pathological N stage, pathological M stage, size largest tumour, nuclear grade, necrosis; c-statistic =0.841; stepwise selection used to create multivariate models
Leibovich et al, 2018-Clear Cell Renal Cell Carcinoma (cRCC) (11)	N=2726; Non-metastatic ccRCC	Age at surgery, ECOG score, adrenalectomy, positive margins, clinical presentation, nuclear grade, necrosis, sarcomatoid, size largest tumour, perinephric or renal sinus fat invasion, tumour thrombus, nodal involvement; c=0.86 for ccRCC, forward selection and backward selection (Clear Cell Renal Cell Carcinoma only) used to select parameters
Papillary Renal Cell Carcinoma Patients		
Leibovich et al, 2018-Papillary Renal Cell Carcinoma (pRCC) (11)	N=607; Non-metastatic pRCC	Nuclear grade, perinephric or renal sinus fat invasion, tumour thrombus; c=0.83 for pRCC; parameter selection method not given
Overall Survival Prognostic Models		
All Patients		
Zisman et al, 2001 (12)	N=661; Renal cell carcinoma patients who underwent radical or partial nephrectomy	ECOG score, nuclear grade, size largest tumour; c-statistic not available; forward selection used to select parameters
Zisman et al, 2002 (44)	N=814; Patients who underwent surgery for unilateral kidney cancer	Pathological T stage, Nuclear Grade, ECOG, pathological N stage; c-statistic not available; parameter selection method not given
Passalacqua et al 2014 (46)	N=310; Kidney cancer patients aged <75 years who underwent partial or radical nephrectomy	Age, pathological T stage, nuclear grade, pathological nodal status (pathological N stage); c-statistic not reported; parameter selection method not given
High Risk Patients		

Correa et al, 2021 (49)	N=1943; Kidney cancer patients with pathological stage pT1b and nuclear grade G3/G4 or pathological stage pT2/3/4 or pathological nodal stage pN1 who underwent radical nephrectomy	Age at surgery, histological subtype, size largest tumour, nuclear grade, necrosis, regional lymph nodes (pathological N stage); c=0.69; include all clinically significant variables whose inclusion resulted in ≥1% increase in c-statistic

Abbreviations:N=number of patients; c=c-statistic; G=grade;pT=pathological T stage; pN=pathological N stage; ECOG=Eastern Cooperative Oncology Group, ASA=American Society of Anesthesiologists

2.5 Methods Used to Assess Accuracy of Prediction Models

Minimum sample sizes are required for validation of a prognostic model and empirical procedures have been developed to estimate the minimum required sample size for external validation of prognostic models (52). To start, the following parameters must be specified: time point, linear predictor value, survival function, survival time distribution, censoring time distribution and targeted standard error (i.e. 0.05) (52). Second, an arbitrary sample size is suggested and a dataset with the suggested sample size is created (52). Possible linear predictor values for the sample size are simulated and the values of a function using the baseline hazard function and linear predictor values are subsequently generated for each observation in the previously simulated dataset (52). Survival times are generated from the parameters specified in the first step adjusting for the linear predictor (52). The censoring distribution is generated and the maximum follow-up time is specified (52). Simulated observations are generated based on the previous steps and the estimated calibration slope and standard error are calculated (52). All steps between the second and seventh step are repeated and the estimates are saved (52). The mean calibration slope and mean standard error are calculated and compared against the targeted standard error; sufficient sample size is achieved when the mean standard error is approximately equal to the targeted standard error (52). If this is not achieved, this process is repeated using a different sample size (52).

If the sample size is adequate, common methods used to assess the accuracy of a prognostic model include c-statistics, calibration curves, Brier Scores, and decision curve analysis.

2.5.1 Discrimination: C-statistic and C-index

Harrell's c-statistic uses Kendall's tau for censored survival data and is a rank-based method relying on ordered pairs; this method also relies on convergence to a measure of association obtained when the censoring distribution is studied (53). The c-statistic is a measure of discriminative ability of a model and is quantified using the area under the receiver operating curve, which is generated when the false positive rate is plotted against the true positive rate (54). A c-statistic of 1.0 signifies perfect discrimination as it

means that the individual with higher risk of the event is identified among all possible pairs of individuals; a c-statistic of 0.5 indicates no discriminative ability (54). As such, a c-statistic ≥0.7 often indicates some clinical utility to discriminate. The c-index gives discriminative ability over time and is the proportion of pairs where the higher risk individual is identified correctly when a subset of all individuals with follow-up times from baseline until a pre-defined time point is generated (55). However, the c-statistic has limitations such as dependence on the censoring distribution (55), dependence on the distribution of event time and non-linear relationship between c-statistic and the number of errors when identifying the individual with higher risk of the event (56).

2.5.2 Calibration Curve

A calibration curve is used to assess predictive performance of prognostic models by plotting the predicted and observed probability of event-free survival in groups of patients (57). In calibration curves for survival analysis, Kaplan-Meier analysis is used to obtain observed event free survival, plotted on the y-axis, and the predicted probability is obtained from the fitted Cox Proportional Hazards model (57). Kaplan-Meier survival analysis has the following fundamental assumptions: censoring probability is not dependent on outcome; survival probabilities are consistent at the time of enrollment and the events actually occurred at the observed specific time-point (58). The inverse survival probability, probability that an individual is event-free at a specific time point may be given as $P(Inverse\ Survival) = \frac{Number\ at\ risk - Number\ of\ Events}{Number\ at\ risk}$ (58). Smoothed calibration curves generated using hazard-based regression were used to visually assess calibration as comparative analyses of various calibration methods showed that this method had the best performance under circumstances of model misspecification (59).

2.5.3 Brier Score

The Brier Score is a measure of the usefulness of clinical models by simultaneously measuring discrimination and calibration (60,61). Brier scores range from 0 to 1 with values closer to 0 indicating better discrimination and calibration (60). The Brier Score is calculated from taking the average over all observations of the squared difference between the event status at a given time point (0 or 1) and the predicted event-free survival probability weighted for the censoring probability (61). The Brier Score is prone to limitations such as being dependent on the prevalence of the event and its usefulness in assessing clinical models questioned as models with the lowest Brier Scores may not always have the best discrimination or calibration (60). As such, it is important to include other assessments of calibration and discrimination in addition to Brier Scores.

2.5.4 Decision Curve Analysis

Decision curve analysis can be applied to consider the clinical consequences of using a prognostic model (62). The net benefit compares the benefits and harms of prognostic approaches compared to assuming no one will recur or every patient will recur (62). After defining the probability threshold of patient risk, the number of true and false positives are obtained (63). The net benefit is computed using the given formula: $Net\ Benefit = \frac{True\ Positives}{N} - \frac{False\ Positives}{N} \times \frac{p_t}{1-p_t}$, where N=total sample size and p_t=threshold probability (62). This calculation is repeated for the entire range of threshold probability and this procedure is repeated for all prognostic models under consideration (in addition to the default states of assuming all patients or no patients will recur) (62). An alternative formulation of net benefit depends on sensitivity, prevalence, specificity and odds of the event (w) at the threshold probability: $Net\ Benefit = sensitivity\ \times prevalence - (1 - specificity) \times (1 - prevalence) \times w$ (62).

For decision curves of prognostic models, the net benefit is framed as relatively how many more 'true predictions' occur compared to 'false predictions' (63). When comparing two prognostic models for use, the "better" prognostic model is defined as the model yielding higher net benefits over a range of clinically useful threshold probabilities (63). For example, for consideration of an effective adjuvant therapy in kidney cancer (which may have toxicity) recurrence model thresholds that are too low or too high, are not clinically useful. We would assume the clinically useful thresholds for most patients with kidney cancer would be between 20% and 30%.

There are two main limitations to decision curve analysis (64). First, the assumption of predicted probability of events and threshold probabilities being independent ignores confounding factors that may affect both the predicted probability and threshold probability (i.e. age can affect the probability that disease recurrence may occur and the threshold probability that a patient prefers) (64). In addition to this, decision curve analysis has limited analytic options to quantify variance (tools such as confidence intervals are under development) (64).

2.6 Cox Proportional Hazards Model

For the purposes of this thesis, we are interested in prognostic models that use a time-to-event outcome (recurrence).

The Cox Proportional Hazards model is a semi-parametric multivariate model used for survival analysis, examining the relationship between various parameters and event-free survival without assuming any given shape of the baseline hazard function (65). This model is specified as $\ln h(t) = \ln h_0(t) + b_1 x_1 + \cdots + b_p x_p$ where $h(t)$=hazard at time t, h_0=baseline hazard (defined as the hazard when all covariates

are equal to 0), $x_1 \ldots x_p$ refer to covariates associated with $b_1 \ldots b_p$ are coefficients (65). The proportional hazards assumption is the main assumption when creating Cox Proportional Hazards models and requires that the relative hazard of an event is constant over time (66). Common methods used to assess Cox Proportional Hazards assumption include the visual assessment of Schoenfeld residual plots (65) and examining logarithmic scale cumulative hazard plots with non-overlap or considerable separation of curves indicating that the assumption is not met (67). Hazard ratios compare the relative hazard of one level of a predictor relative to a baseline level to provide a comparison of relative risk with hazard ratios above one indicating increased risk and below one indicating decreased risk (68). Important considerations in Cox Proportional Hazards modelling are: time dependent variables pose complications as the Cox Proportional Hazard model requires that all variables are specified at the start of survival time, there must also be a sufficient number of events for reliability or the number of parameters that can be included will be limited, there must be sufficient sample size for statistical power and if censoring rates are different in different groups compared, estimates will be biased (68).

2.7 Prognostic Model Variable Selection Methods

Various techniques are used for parameter selection in the development of prognostic models. A summary of methods used in this thesis is shown in the table below. While one technique is typically used in model development, we used each of the techniques below in this thesis to gain experience and compare outcomes using various approaches.

Table 4: Parameter Selection Methods for Prognostic Model Building

Method	Description	Advantages	Disadvantages
Forward Selection (69)	• This method sequentially adds individual variables to the model. First, start with a model without variables, iteratively test addition of each variable and add the most significant variable (i.e. variable with largest test statistic or lowest p-value). This process repeats until all the significant variables are added to the model.	• This model is less prone to problems arising from collinearity.	• During the process of adding variables, an existing variable in the model may become insignificant. It is not possible to remove the insignificant variable.
Backward Selection (69)	• Starting with all model parameters, the least significant parameter (i.e. variable with smallest test statistic or highest p-value) is	• The predictive ability of variables when combined with other variables can be examined using this method.	• Once a variable is dropped from the model, it is not possible to add the variable to the model despite the possibility that the variable

	sequentially dropped with the model refitted at each step to account for existing variables. This process is repeated until only significant variables are present in the model.		can be significant if added at a later step.
Stepwise Selection (69)	• This method has the properties of forward and backward selection. This method can start with forward selection to add variables and then perform backward selection to remove variables from the model. This pattern repeats until all significant variables are included. Significance tests are required at both steps.	• This method is easy to use and allows for the examination of the effects of different groupings of variables that may not be intuitive to the researcher.	• This method is prone to issues due to collinearity and bias in regression coefficients.
Elastic Net Penalization (70,71)	• This technique is a modification of LASSO based parameter selection using two linear regression-based penalties. The second penalty-imposed groups intercorrelated variables together, resulting in the retention or deletion of parameters as a group.	• This technique is more effective when there are more predictors than observations.	• Due to parameter grouping, there is the potential to include too many parameters.
Resampling Based Variable Selection (72)	• Repeated bootstrap sampling with replacement is performed to select parameters.	• The stability provided by bootstrap estimates helps control the false discovery rate.	• Further study is needed to determine the most appropriate sampling rate.

2.8 Nomogram Creation

A common method used to build easy to use nomograms for the prediction of outcomes in cancer patients weighs parameters based on their relative contribution to the predicted risk of an outcome (as opposed to their statistical significance) (73). Once the Cox Proportional Hazards model with all parameters has been fitted, the β coefficients for all categorical is retained and the β coefficients for quantitative variables are recalculated by multiplying the β value from the model by the range (difference between maximum and minimum) of the variable values (73). The maximum of the β coefficients (β_{max}) for categorical variables and the recalculated β values of quantitative variables is obtained (73). The number of points for each variable and category is calculated by using the following equation: $Points = 100 \times \frac{\beta}{\beta_{max}}$ with

the number of points rounded to the nearest whole number for categorical variables (73). Quantitative variables must be readjusted as per the following formula to obtain the number of points for each unit of the variable above the variable's minimum value: $Points = \frac{Points\ from\ relative\ weighting\ of\ coefficients}{(maximum\ variable\ value - minimum\ variable\ value)}$ (73).

2.9 Minimum Sample Size Required for Model Development

Adequate sample size is important for validation of existing models. Similarly, it is also important to assess adequacy of sample size for a new prediction model. To estimate sample size, parameters such as c-statistic, Cox-Snell's R^2 statistic, log-likelihood of the final model, the number of events, population size and total follow-up time can be used (74). Complete mathematical and statistical formulae have been previously described (74). Once the number of parameters has been specified, values for such as the Cox-Snell's R^2 are calculated using the c-statistic of the model and the number of events (74). Using the formulae specified, the sample size required to achieve a global shrinkage factor of 0.9 is determined (74). Based on the adjusted Cox-Snell's R^2 value calculated in the second step, the shrinkage factor is calculated; if the calculated value is less than 0.9, no further calculations are required (74). Based on the event rate and follow-up time, the overall risk interval size is calculated using the exponential model to ensure that the interval is sufficiently narrow (74). The sample size meeting all these criteria is used as the minimum required sample size for model development (74). If it is not possible to meet the sample size required, the expected shrinkage can be decreased; but this could increase model overfitting (74).

2.10 Model performance

In addition to c-statistics/c-index over time, calibration plots and decision curve analysis, calibration slopes can be used for internal validation of prognostic models.

2.10.1 Calibration Slopes

It has been previously mentioned that in the absence of information on the baseline hazard function, the Cox Proportional Hazard model is fitted, as is, and the calibration slope is given by the value of the coefficient β (75). The calibration slope obtained by performing regression analysis using the predicted event-free survival and observed event-free survival is an indicator of model performance with values that are close to 1 suggesting excellent calibration. Calibration slopes that are extremely low suggest poor calibration and high potential for model overfit (76). Calibration slopes are a simple way to assesses calibration (because it is quantitative), but in general, calibration plots are preferred because it is possible to obtain calibration slopes near 1, even when accuracy is low (i.e. linear calibration curves with

consistently higher or lower observed event rates compared to the predicted event rate) (77). Calibration plots also provide ranges of predicted risk where the model is more (or less) predictive of outcome.

2.10.2 Internal Validation: Bootstrap Sampling vs. k-fold Validation

Bootstrapping uses sampling with replacement to estimate the values of parameters using the mean and confidence intervals of the parameter values for all selected samples (78). This is often the preferred method for internal validation (78). While thousands of bootstrap samples can be used for the most conservative parameter estimates, 100-500 samples have been shown to be sufficient to generate computationally robust estimates of parameter values (79) and the most frequently used sampling rate is 0.632 (i.e. a sample size that is 63.2% of the total sample size) (80). Key limitations of bootstrap sampling include: processing time required, bootstrap estimates varying based on individuals included in bootstrap samples and the dependence of bootstrap parameter estimate on the parameter estimate in each individual bootstrap value (81).

An alternate method of internal validation is k-fold validation. K-fold validation involves the division of the total sample into k sub-samples with one sample used as the training set and the others used as the testing datasets (82). K different estimates of parameter values are generated, and these are pooled to create an overall estimate (82). Typically, k=10 is used (82) and this method has been identified as the best performing method in addition to bootstrap-based methods (83). Limitations of this method include: the need to evaluate the given algorithm k times, training sets and test sets not being independent as both originate from the same dataset, and larger variances in estimates generated using this method (83).

2.10.3 Handling Missing Data: Multiple Imputation Sensitivity Analysis

Missing data can be described using three categories: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR) (84). MCAR assumes that all cases have equal probability of having missing data (e.g. equipment used to weigh kidney cancer patients is broken, which results in no patients having observations for weight) (84). MAR assumes that missingness is based on previously observed information (e.g. if there is a survey on ethnicity and it has been previously observed that Caucasian people are more likely to complete such surveys compared to non-Caucasians, missing data distributions may vary between race/ethnicity) (84). MNAR assumes that missingness is based on unobserved data (e.g. if there is a survey on quality of life and those with severe renal dysfunction are less likely to complete this survey) (84).

Multiple Imputation imputes data based on given information for other variables (84). While the consideration of variability due to sampling and imputation provides more accurate estimates of

variability, this method has significant potential for error due to model misspecification (84). This method has three main steps: selection of variables for multiple imputation, creating multiple imputed datasets with missing data imputed from distributions given observed data, and calculating the desired measure of association and using Rubin's rules to combine estimates of the measure of association (84). Mathematically, multiple imputation imputes missing data based on the known values of missing variable, complete outcome, and other variables (85). Chained equations are used to arbitrarily fill missing values and calculate values for other variables until convergence is achieved with the t-distribution using degrees of freedom calculated from Rubin's rules used to obtained for confidence intervals (85). For survival data, the imputation model created is based on a logit function that imputes missing values based on the distribution of the event indicator and baseline hazard function (85).

The substantiative model compatible fully conditional specification (SMC-FCS) has been identified as the best method for multiple imputation due to imputation models and study analysis models being compatible in addition to being adapted for time-varying effects (86). The full mathematical derivation has been described elsewhere with repeated iterations being required until convergence (86). Convergence is determined by visually examining plots of partially observed variable by iteration number or the evaluation of inter-chain and intra-chain variability (87). SMC-FCS has the following mathematical properties: covariate models and joint models are mutually compatible, each iteration has the joint model factorized as independent reparameterizations and each iteration has a prior distribution (87). Meeting these conditions, normal linear regression being used for covariate models, using all non-informative priors, having equivalent imputation as the Bayesian joint model and correctly specifying models provide consistent estimators of the missing value (87).

2.11 Creation of New Patient Risk Groups

Many different methods have been developed to divide patients into risk groups such as arbitrarily grouping risk scores into groups (16), dividing patients into risk groups based on whether the probability of an event is equivalent to a pre-determined value (42) and dividing patients into groups with an approximately equal number of patients in each group (i.e. quintiles) (88). Patients may also be ranked in terms of percentile based on risk score with specified cut-offs such as 16^{th} percentile, 50^{th} percentile and 84^{th} percentile being used to specify risk groups (89).

2.11.1 Techniques for Comparing Risk Group Classification Systems

The effectiveness of risk group classification systems can be compared through various measures. Common measures include overall c-statistics and the predicted number vs. observed number of

recurrences (90). Risk group classification systems with larger overall c-statistics, are considered more effective (90). The predicted and observed number of events based on the predicted event risk and the observed event risk multiplied by the risk group size can also be compared with smaller differences indicating more predictive accuracy; predicted event risk is determined based on Cox Proportional Hazards modelling and observed event risk is determined based on Kaplan-Meier survival modelling (90).

2.12 Background Information on Canadian Kidney Cancer Information System (CKCis) Cohort

Canadian Kidney Cancer Information System (CKCis) is a prospective cohort with over 10,000 patients at 14 academic tertiary care centres across six provinces (British Columbia, Alberta, Manitoba, Ontario, Quebec and Nova Scotia) (91). CKCis was established in 2009 and initiated data collection on January 1, 2011 (91). CKCis has been shown to capture 20.7% of all Canadian kidney cancer patients with provincial coverage ranging from 12.7% to 51.4% of all kidney cancer patients diagnosed within specific provinces (91). While CKCis site specific differences in surgical quality indicators and outcomes exists (92),this cohort has been validated against external data sources such as the Canadian Cancer Registry and has been shown to be representative of the Canadian population across demographic and pathological features (91).

2.13 Conclusion

In conclusion, a number of prognostic models have been developed for patients with RCC. In the following chapters we will assess the performance of these models in Canadian kidney cancer patients and the development of new prognostic models using Canadian kidney cancer patients.

2.14 References

1. Dudani S, de Velasco G, Wells JC, Gan CL, Donskov F, Porta C, Fraccon A, Pasini F, Lee JL, Hansen A, Bjarnason GA, Beuselinck B, Pal SK, Yuasa T, Kroeger N, Kanesvaran R, Reaume MN, Canil C, Choueiri TK, Heng DYC. Evaluation of Clear Cell, Papillary, and Chromophobe Renal Cell Carcinoma Metastasis Sites and Association With Survival. JAMA Netw Open. 2021 Jan 4;4(1):e2021869. doi: 10.1001/jamanetworkopen.2020.21869.
2. Patel U, Sokhi H. Imaging in the follow-up of renal cell carcinoma. AJR Am J Roentgenol. 2012 Jun;198(6):1266-76. doi: 10.2214/AJR.11.8381.
3. Israel O, Kuten A. Early detection of cancer recurrence: 18F-FDG PET/CT can make a difference in diagnosis and patient care. J Nucl Med. 2007 Jan;48 Suppl 1:28S-35S.
4. Bazzi WM, Sjoberg DD, Feuerstein MA, Maschino A, Verma S, Bernstein M, O'Brien MF, Jang T, Lowrance W, Motzer RJ, Russo P. Long-term survival rates after resection for locally advanced kidney cancer: Memorial Sloan Kettering Cancer Center 1989 to 2012 experience. J Urol. 2015 Jun;193(6):1911-1916. doi: 10.1016/j.juro.2014.12.022.
5. Karakiewicz PI, Trinh QD, de la Taille A, Abbou CC, Salomon L, Tostain J, Cindolo L, Artibani W, Ficarra V, Patard JJ. ECOG performance status 0 or 1 and symptom classification do not improve the ability

to predict renal cell carcinoma-specific survival. Eur J Cancer. 2007 Apr;43(6):1023-1029. doi: 10.1016/j.ejca.2007.01.020.
6. Yaycioglu O, Eskicorapci S, Karabulut E, Soyupak B, Gogus C, Divrik T, Turkeri L, Yazici S, Ozen H; Society of Urooncology Study Group for Kidney Cancer Prognosis. A preoperative prognostic model predicting recurrence-free survival for patients with kidney cancer. Jpn J Clin Oncol. 2013 Jan;43(1):63-68. doi: 10.1093/jjco/hys192.
7. May M, Aziz A, Zigeuner R, Chromecki T, Cindolo L, Schips L, De Cobelli O, Rocco B, De Nunzio C, Tubaro A, Coman I, Truss M, Dalpiaz O, Hoschke B, Gilfrich C, Feciche B, Stoltze A, Fenske F, Fritsche HM, Figenshau RS, Madison K, Sánchez-Chapado M, Martin Mdel C, Salzano L, Lotrecchiano G, Joniau S, Waidelich R, Stief C, Brookman-May S; Members of the CORONA project the Young Academic Urologists Renal Cancer Group. Gender differences in clinicopathological features and survival in surgically treated patients with renal cell carcinoma: an analysis of the multicenter CORONA database. World J Urol. 2013 Oct;31(5):1073-1080. doi: 10.1007/s00345-013-1071-x.
8. van der Mijn JC, Al Hussein Al Awamlh B, Islam Khan A, Posada-Calderon L, Oromendia C, Fainberg J, Alshak M, Elahjji R, Pierce H, Taylor B, Gudas LJ, Nanus DM, Molina AM, Del Pizzo J, Scherr DS. Validation of risk factors for recurrence of renal cell carcinoma: Results from a large single-institution series. PLoS One. 2019 Dec 9;14(12):e0226285. doi: 10.1371/journal.pone.0226285.
9. Kutikov A, Egleston BL, Wong YN, Uzzo RG. Evaluating overall survival and competing risks of death in patients with localized renal cell carcinoma using a comprehensive nomogram. J Clin Oncol. 2010 Jan 10;28(2):311-317. doi: 10.1200/JCO.2009.22.4816.
10. Schwartz K, Ruterbusch JJ, Colt JS, Miller DC, Chow WH, Purdue MP. Racial disparities in overall survival among renal cell carcinoma patients with young age and small tumors. Cancer Med. 2016 Feb;5(2):200-208. doi: 10.1002/cam4.578.
11. Leibovich BC, Lohse CM, Cheville JC, Zaid HB, Boorjian SA, Frank I, Thompson RH, Parker WP. Predicting Oncologic Outcomes in Renal Cell Carcinoma After Surgery. Eur Urol. 2018 May;73(5):772-780. doi: 10.1016/j.eururo.2018.01.005.
12. Zisman A, Pantuck AJ, Dorey F, Said JW, Shvarts O, Quintana D, Gitlitz BJ, deKernion JB, Figlin RA, Belldegrun AS. Improved prognostication of renal cell carcinoma using an integrated staging system. J Clin Oncol. 2001 Mar 15;19(6):1649-1657. doi: 10.1200/JCO.2001.19.6.1649.
13. Swami U, Nussenzveig RH, Haaland B, Agarwal N. Revisiting AJCC TNM staging for renal cell carcinoma: quest for improvement. Ann Transl Med. 2019 Mar;7(Suppl 1):S18. doi: 10.21037/atm.2019.01.50.
14. Bazzi WM, Sjoberg DD, Feuerstein MA, Maschino A, Verma S, Bernstein M, O'Brien MF, Jang T, Lowrance W, Motzer RJ, Russo P. Long-term survival rates after resection for locally advanced kidney cancer: Memorial Sloan Kettering Cancer Center 1989 to 2012 experience. J Urol. 2015 Jun;193(6):1911-1916. doi: 10.1016/j.juro.2014.12.022.
15. Kim SH, Park B, Hwang EC, Hong SH, Jeong CW, Kwak C, Byun SS, Chung J. Retrospective Multicenter Long-Term Follow-up Analysis of Prognostic Risk Factors for Recurrence-Free, Metastasis-Free, Cancer-Specific, and Overall Survival After Curative Nephrectomy in Non-metastatic Renal Cell Carcinoma. Front Oncol. 2019 Sep 4;9:859. doi: 10.3389/fonc.2019.00859.
16. Leibovich BC, Blute ML, Cheville JC, Lohse CM, Frank I, Kwon ED, Weaver AL, Parker AS, Zincke H. Prediction of progression after radical nephrectomy for patients with clear cell renal cell carcinoma: a stratification tool for prospective clinical trials. Cancer. 2003 Apr 1;97(7):1663-1671. doi: 10.1002/cncr.11234.

17. Petejova N, Martinek A. Renal cell carcinoma: Review of etiology, pathophysiology and risk factors. Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub. 2016 Jun;160(2):183-194. doi: 10.5507/bp.2015.050.
18. Sengupta S, Lohse CM, Leibovich BC, Frank I, Thompson RH, Webster WS, Zincke H, Blute ML, Cheville JC, Kwon ED. Histologic coagulative tumor necrosis as a prognostic indicator of renal cell carcinoma aggressiveness. Cancer. 2005 Aug 1;104(3):511-520. doi: 10.1002/cncr.21206.
19. Ito K, Seguchi K, Shimazaki H, Takahashi E, Tasaki S, Kuroda K, Sato A, Asakuma J, Horiguchi A, Asano T. Tumor necrosis is a strong predictor for recurrence in patients with pathological T1a renal cell carcinoma. Oncol Lett. 2015 Jan;9(1):125-130. doi: 10.3892/ol.2014.2670.
20. Brookman-May S, May M, Shariat SF, Xylinas E, Stief C, Zigeuner R, Chromecki T, Burger M, Wieland WF, Cindolo L, Schips L, De Cobelli O, Rocco B, De Nunzio C, Feciche B, Truss M, Gilfrich C, Pahernik S, Hohenfellner M, Zastrow S, Wirth MP, Novara G, Carini M, Minervini A, Simeone C, Antonelli A, Mirone V, Longo N, Simonato A, Carmignani G, Ficarra V; Members of the CORONA project and the SATURN project. Features associated with recurrence beyond 5 years after nephrectomy and nephron-sparing surgery for renal cell carcinoma: development and internal validation of a risk model (PRELANE score) to predict late recurrence based on a large multicenter database (CORONA/SATURN Project). Eur Urol. 2013 Sep;64(3):472-477. doi: 10.1016/j.eururo.2012.06.030.
21. Blum KA, Gupta S, Tickoo SK, Chan TA, Russo P, Motzer RJ, Karam JA, Hakimi AA. Sarcomatoid renal cell carcinoma: biology, natural history and management. Nat Rev Urol. 2020 Dec;17(12):659-678. doi: 10.1038/s41585-020-00382-9.
22. Merrill MM, Wood CG, Tannir NM, Slack RS, Babaian KN, Jonasch E, Pagliaro LC, Compton Z, Tamboli P, Sircar K, Pisters LL, Matin SF, Karam JA. Clinically nonmetastatic renal cell carcinoma with sarcomatoid dedifferentiation: Natural history and outcomes after surgical resection with curative intent. Urol Oncol. 2015 Apr;33(4):166.e21-9. doi: 10.1016/j.urolonc.2014.11.021.
23. Thompson RH, Leibovich BC, Cheville JC, Webster WS, Lohse CM, Kwon ED, Frank I, Zincke H, Blute ML. Is renal sinus fat invasion the same as perinephric fat invasion for pT3a renal cell carcinoma? J Urol. 2005 Oct;174(4 Pt 1):1218-21. doi: 10.1097/01.ju.0000173942.19990.40.
24. Shimizu T, Miyake M, Hori S, Iida K, Ichikawa K, Sakamoto K, Miyamoto T, Nakai Y, Inoue T, Anai S, Tanaka N, Fujimoto K. Clinical Significance of Tumor Size, Pathological Invasion Sites Including Urinary Collecting System and Clinically Detected Renal Vein Thrombus as Predictors for Recurrence in pT3a Localized Renal Cell Carcinoma. Diagnostics (Basel). 2020 Mar 12;10(3):154. doi: 10.3390/diagnostics10030154.
25. Wang Z, Yu K, Zhu Y, Feng C, Liu C, Liu S, Wang J, Zeng X. Multiple Patterns of Perirenal Fat Invasion Are Associated With a Poorer Prognosis Compared With Isolated Invasion: A Proposal for a Revision of T3aN0M0 TNM Staging System. Front Oncol. 2020 Mar 11;10:336. doi: 10.3389/fonc.2020.00336.
26. Shiff B, Breau RH, Mallick R, Pouliot F, So A, Tanguay S, Kapoor A, Lattouf JB, Lavallée L, Fairey A, Finelli A, Bhindi B, Kawakami J, Rendon R, Bansal RK. Prognostic significance of extent of venous tumor thrombus in patients with non-metastatic renal cell carcinoma: Results from a Canadian multi-institutional collaborative. Urol Oncol. 2021 Dec;39(12):836.e19-836.e27. doi: 10.1016/j.urolonc.2021.08.016.
27. Gandaglia G, Ravi P, Abdollah F, Abd-El-Barr AE, Becker A, Popa I, Briganti A, Karakiewicz PI, Trinh QD, Jewett MA, Sun M. Contemporary incidence and mortality rates of kidney cancer in the United States. Can Urol Assoc J. 2014 Jul;8(7-8):247-252. doi: 10.5489/cuaj.1760.
28. Bensalah K, Pantuck AJ, Rioux-Leclercq N, Thuret R, Montorsi F, Karakiewicz PI, Mottet N, Zini L, Bertini R, Salomon L, Villers A, Soulie M, Bellec L, Rischmann P, De la Taille A, Avakian R, Crepel M,

Ferriere JM, Bernhard JC, Dujardin T, Pouliot F, Rigaud J, Pfister C, Albouy B, Guy L, Joniau S, van Poppel H, Lebret T, Culty T, Saint F, Zisman A, Raz O, Lang H, Spie R, Wille A, Roigas J, Aguilera A, Rambeaud B, Martinez Piñeiro L, Nativ O, Farfara R, Richard F, Roupret M, Doehn C, Bastian PJ, Muller SC, Tostain J, Belldegrun AS, Patard JJ. Positive surgical margin appears to have negligible impact on survival of renal cell carcinomas treated by nephron-sparing surgery. Eur Urol. 2010 Mar;57(3):466-471. doi: 10.1016/j.eururo.2009.03.048.

29. May M, Aziz A, Zigeuner R, Chromecki T, Cindolo L, Schips L, De Cobelli O, Rocco B, De Nunzio C, Tubaro A, Coman I, Truss M, Dalpiaz O, Hoschke B, Gilfrich C, Feciche B, Stoltze A, Fenske F, Fritsche HM, Figenshau RS, Madison K, Sánchez-Chapado M, Martin Mdel C, Salzano L, Lotrecchiano G, Joniau S, Waidelich R, Stief C, Brookman-May S; Members of the CORONA project the Young Academic Urologists Renal Cancer Group. Gender differences in clinicopathological features and survival in surgically treated patients with renal cell carcinoma: an analysis of the multicenter CORONA database. World J Urol. 2013 Oct;31(5):1073-1080. doi: 10.1007/s00345-013-1071-x.
30. Karakiewicz PI, Trinh QD, Bhojani N, Bensalah K, Salomon L, de la Taille A, Tostain J, Cindolo L, Altieri V, Ficarra V, Schips L, Zigeuner R, Mulders PF, Valeri A, Descotes JL, Mejean A, Patard JJ. Renal cell carcinoma with nodal metastases in the absence of distant metastatic disease: prognostic indicators of disease-specific survival. Eur Urol. 2007 Jun;51(6):1616-1624. doi: 10.1016/j.eururo.2006.12.015.
31. Sok M, Zavrl M, Greif B, Srpčič M. Objective assessment of WHO/ECOG performance status. Support Care Cancer. 2019 Oct;27(10):3793-3798. doi: 10.1007/s00520-018-4597-z.
32. Haas NB, Manola J, Uzzo RG, Flaherty KT, Wood CG, Kane C, Jewett M, Dutcher JP, Atkins MB, Pins M, Wilding G, Cella D, Wagner L, Matin S, Kuzel TM, Sexton WJ, Wong YN, Choueiri TK, Pili R, Puzanov I, Kohli M, Stadler W, Carducci M, Coomes R, DiPaola RS. Adjuvant sunitinib or sorafenib for high-risk, non-metastatic renal-cell carcinoma (ECOG-ACRIN E2805): a double-blind, placebo-controlled, randomised, phase 3 trial. Lancet. 2016 May 14;387(10032):2008-2016. doi: 10.1016/S0140-6736(16)00559-6.
33. Kroeger N, Klatte T, Birkhäuser FD, Rampersaud EN, Seligson DB, Zomorodian N, Kabbinavar FF, Belldegrun AS, Pantuck AJ. Smoking negatively impacts renal cell carcinoma overall and cancer-specific survival. Cancer. 2012 Apr 1;118(7):1795-1802. doi: 10.1002/cncr.26453.
34. Nayan M, Jalali S, Kapoor A, Finelli A, So A, Rendon R, Breau RH, Lavallee LT, Tanguay S, Heng D, Kawakami J, Basappa NS, Bjarnason G, Pouliot F, Hamilton RJ. Diabetes and kidney cancer survival in patients undergoing nephrectomy: A Canadian multi-center, propensity score analysis. Urol Oncol. 2019 Sep;37(9):576.e11-576.e16. doi: 10.1016/j.urolonc.2019.06.006.
35. Wu CC, Hsu TW, Chang CM, Yu CH, Lee CC. Age-adjusted Charlson comorbidity index scores as predictor of survival in colorectal cancer patients who underwent surgical resection and chemoradiation. Medicine (Baltimore). 2015 Jan;94(2):e431. doi: 10.1097/MD.0000000000000431.
36. Kang HW, Kim SM, Kim WT, Yun SJ, Lee SC, Kim WJ, Hwang EC, Kang SH, Hong SH, Chung J, Kwon TG, Kim HH, Kwak C, Byun SS, Kim YJ; KORCC (KOrean Renal Cell Carcinoma) Group. The age-adjusted Charlson comorbidity index as a predictor of overall survival of surgically treated non-metastatic clear cell renal cell carcinoma. J Cancer Res Clin Oncol. 2020 Jan;146(1):187-196. doi: 10.1007/s00432-019-03042-7.
37. Vazquez BJ, Richards ML, Lohse CM, Thompson GB, Farley DR, Grant CS, Huebner M, Moreno J. Adrenalectomy improves outcomes of selected patients with metastatic carcinoma. World J Surg. 2012 Jun;36(6):1400-1405. doi: 10.1007/s00268-012-1506-3.
38. Lavi A, Breau RH, Mallick R, Kapoor A, Finelli A, So A, Pouliot F, Tanguay S, Lavallée LT, Rendon R, Fairey A, Drachenberg DE, Lattouf JB, Maloni R, Power NE. Adrenalectomy During Radical

Nephrectomy- Incidence and Oncologic Outcomes From the Canadian Kidney Cancer Information System (CKCis) -A Modern Era, Nationwide, Multicenter Cohort. Urology. 2021 Nov;157:168-173. doi: 10.1016/j.urology.2021.05.053.
39. Weight CJ, Kim SP, Lohse CM, Cheville JC, Thompson RH, Boorjian SA, Leibovich BC. Routine adrenalectomy in patients with locally advanced renal cell cancer does not offer oncologic benefit and places a significant portion of patients at risk for an asynchronous metastasis in a solitary adrenal gland. Eur Urol. 2011 Sep;60(3):458-464. doi: 10.1016/j.eururo.2011.04.022.
40. Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A, Briggs A, Udumyan R, Moons KG, Steyerberg EW, Roberts I, Schroter S, Altman DG, Riley RD; PROGRESS Group. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. BMJ. 2013 Feb 5;346:e5595. doi: 10.1136/bmj.e5595.
41. Sun M, Vetterlein M, Harshman LC, Chang SL, Choueiri TK, Trinh QD. Risk Assessment in Small Renal Masses: A Review Article. Urol Clin North Am. 2017 May;44(2):189-202. doi: 10.1016/j.ucl.2016.12.005.
42. Yaycioglu O, Roberts WW, Chan T, Epstein JI, Marshall FF, Kavoussi LR. Prognostic assessment of nonmetastatic renal cell carcinoma: a clinically based model. Urology. 2001 Aug;58(2):141-145. doi: 10.1016/s0090-4295(01)01207-9.
43. Cindolo L, de la Taille A, Messina G, Romis L, Abbou CC, Altieri V, Rodriguez A, Patard JJ. A preoperative clinical prognostic model for non-metastatic renal cell carcinoma. BJU Int. 2003 Dec;92(9):901-905. doi: 10.1111/j.1464-410x.2003.04505.x.
44. Zisman A, Pantuck AJ, Wieder J, Chao DH, Dorey F, Said JW, deKernion JB, Figlin RA, Belldegrun AS. Risk group assessment and clinical outcome algorithm to predict the natural history of patients with surgically resected renal cell carcinoma. J Clin Oncol. 2002 Dec 1;20(23):4559-4566. doi: 10.1200/JCO.2002.05.111.
45. Kattan MW, Reuter V, Motzer RJ, Katz J, Russo P. A postoperative prognostic nomogram for renal cell carcinoma. J Urol. 2001 Jul;166(1):63-67.
46. Passalacqua R, Caminiti C, Buti S, Porta C, Camisa R, Braglia L, Tomasello G, Vaglio A, Labianca R, Rondini E, Sabbatini R, Nastasi G, Artioli F, Prati A, Potenzoni M, Pezzuolo D, Oliva E, Alberici F, Buzio C; POLAR-01 Trial Investigators. Adjuvant low-dose interleukin-2 (IL-2) plus interferon-α (IFN-α) in operable renal cell carcinoma (RCC): a phase III, randomized, multicentre trial of the Italian Oncology Group for Clinical Research (GOIRC). J Immunother. 2014 Nov-Dec;37(9):440-447. doi: 10.1097/CJI.0000000000000055.
47. Sorbellini M, Kattan MW, Snyder ME, Reuter V, Motzer R, Goetzl M, McKiernan J, Russo P. A postoperative prognostic nomogram predicting recurrence for patients with conventional clear cell renal cell carcinoma. J Urol. 2005 Jan;173(1):48-51. doi: 10.1097/01.ju.0000148261.19532.2c.
48. Klatte T, Gallagher KM, Afferi L, Volpe A, Kroeger N, Ribback S, McNeill A, Riddick ACP, Armitage JN, 'Aho TF, Eisen T, Fife K, Bex A, Pantuck AJ, Stewart GD. The VENUSS prognostic model to predict disease recurrence following surgery for non-metastatic papillary renal cell carcinoma: development and evaluation using the ASSURE prospective clinical trial cohort. BMC Med. 2019 Oct 3;17(1):182. doi: 10.1186/s12916-019-1419-1.
49. Correa AF, Jegede OA, Haas NB, Flaherty KT, Pins MR, Adeniran A, Messing EM, Manola J, Wood CG, Kane CJ, Jewett MAS, Dutcher JP, DiPaola RS, Carducci MA, Uzzo RG. Predicting Disease Recurrence, Early Progression, and Overall Survival Following Surgical Resection for High-risk Localized and Locally Advanced Renal Cell Carcinoma. Eur Urol. 2021 Jul;80(1):20-31. doi: 10.1016/j.eururo.2021.02.025.

50. Kutikov A, Egleston BL, Canter D, Smaldone MC, Wong YN, Uzzo RG. Competing risks of death in patients with localized renal cell carcinoma: a comorbidity based model. J Urol. 2012 Dec;188(6):2077-2083. doi: 10.1016/j.juro.2012.07.100.
51. Frank I, Blute ML, Cheville JC, Lohse CM, Weaver AL, Zincke H. An outcome prediction model for patients with clear cell renal cell carcinoma treated with radical nephrectomy based on tumor stage, size, grade and necrosis: the SSIGN score. J Urol. 2002 Dec;168(6):2395-2400. doi: 10.1097/01.ju.0000035885.91935.d5.
52. Riley RD, Collins GS, Ensor J, Archer L, Booth S, Mozumder SI, Rutherford MJ, van Smeden M, Lambert PC, Snell KIE. Minimum sample size calculations for external validation of a clinical prediction model with a time-to-event outcome. Stat Med. 2021 Dec 16. doi: 10.1002/sim.9275.
53. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. Stat Med. 2011 May 10;30(10):1105-1117. doi: 10.1002/sim.4154.
54. Caetano SJ, Sonpavde G, Pond GR. C-statistic: A brief explanation of its construction, interpretation and limitations. Eur J Cancer. 2018 Feb;90:130-132. doi: 10.1016/j.ejca.2017.10.027.
55. Correa AF, Jegede O, Haas NB, Flaherty KT, Pins MR, Messing EM, Manola J, Wood CG, Kane CJ, Jewett MAS, Dutcher JP, DiPaola RS, Carducci MA, Uzzo RG. Predicting Renal Cancer Recurrence: Defining Limitations of Existing Prognostic Models With Prospective Trial-Based Validation. J Clin Oncol. 2019 Aug 10;37(23):2062-2071. doi: 10.1200/JCO.19.00107.
56. Longato E, Vettoretti M, Di Camillo B. A practical perspective on the concordance index for the evaluation and selection of prognostic time-to-event models. J Biomed Inform. 2020 Aug;108:103496. doi: 10.1016/j.jbi.2020.103496.
57. Steyerberg EW. Clinical Prediction Models A Practical Approach to Development, Validation, and Updating. 2nd ed. Cham (Switzerland):Springer Nature Switzerland;c2019.558p.
58. Koletsi D, Pandis N. Survival analysis, part 2: Kaplan-Meier method and the log-rank test. Am J Orthod Dentofacial Orthop. 2017 Oct;152(4):569-571. doi: 10.1016/j.ajodo.2017.07.008.
59. Austin PC, Harrell FE Jr, van Klaveren D. Graphical calibration curves and the integrated calibration index (ICI) for survival models. Stat Med. 2020 Sep 20;39(21):2714-2742. doi: 10.1002/sim.8570.
60. Assel M, Sjoberg DD, Vickers AJ. The Brier score does not evaluate the clinical utility of diagnostic tests or prediction models. Diagn Progn Res. 2017 Dec 2;1:19. doi: 10.1186/s41512-017-0020-3.
61. Kronek LP, Reddy A. Logical analysis of survival data: prognostic survival models by detecting high-degree interactions in right-censored data. Bioinformatics. 2008 Aug 15;24(16):i248-253. doi: 10.1093/bioinformatics/btn265.
62. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. BMJ. 2016 Jan 25;352:i6. doi: 10.1136/bmj.i6.
63. Zhang Z, Rousson V, Lee WC, Ferdynus C, Chen M, Qian X, Guo Y; written on behalf of AME Big-Data Clinical Trial Collaborative Group. Decision curve analysis: a technical note. Ann Transl Med. 2018 Aug;6(15):308. doi: 10.21037/atm.2018.07.02.
64. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making. 2006 Nov-Dec;26(6):565-574. doi: 10.1177/0272989X06295361.
65. Bewick V, Cheek L, Ball J. Statistics review 12: survival analysis. Crit Care. 2004 Oct;8(5):389-394. doi: 10.1186/cc2955.
66. Kuitunen I, Ponkilainen VT, Uimonen MM, Eskelinen A, Reito A. Testing the proportional hazards assumption in cox regression and dealing with possible non-proportionality in total joint

arthroplasty research: methodological perspectives and review. BMC Musculoskelet Disord. 2021 May 28;22(1):489. doi: 10.1186/s12891-021-04379-2.
67. Lee M, Han J. Statistical methods and models in the analysis of time to event data. Ann Transl Med. 2020 Feb;8(4):73. doi: 10.21037/atm.2019.12.66.
68. Zwiener I, Blettner M, Hommel G. Survival analysis: part 15 of a series on evaluation of scientific publications. Dtsch Arztebl Int. 2011 Mar;108(10):163-169. doi: 10.3238/arztebl.2010.0163.
69. Chowdhury MZI, Turin TC. Variable selection strategies and its importance in clinical prediction modelling. Fam Med Community Health. 2020 Feb 16;8(1):e000262. doi: 10.1136/fmch-2019-000262.
70. Sanchez-Pinto LN, Venable LR, Fahrenbach J, Churpek MM. Comparison of variable selection methods for clinical predictive modeling. Int J Med Inform. 2018 Aug;116:10-17. doi: 10.1016/j.ijmedinf.2018.05.006.
71. Waldmann P, Mészáros G, Gredler B, Fuerst C, Sölkner J. Evaluation of the lasso and the elastic net in genome-wide association studies. Front Genet. 2013 Dec 4;4:270. doi: 10.3389/fgene.2013.00270.
72. Sauerbrei W, Perperoglou A, Schmid M, Abrahamowicz M, Becher H, Binder H, Dunkler D, Harrell FE Jr, Royston P, Heinze G; for TG2 of the STRATOS initiative. State of the art in selection of variables and functional forms in multivariable analysis-outstanding issues. Diagn Progn Res. 2020 Apr 2;4:3. doi: 10.1186/s41512-020-00074-3.
73. Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. J Clin Oncol. 2008 Mar 10;26(8):1364-1370. doi: 10.1200/JCO.2007.12.9791.
74. Riley RD, Snell KI, Ensor J, Burke DL, Harrell FE Jr, Moons KG, Collins GS. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. Stat Med. 2019 Mar 30;38(7):1276-1296. doi: 10.1002/sim.7992.
75. Crowson CS, Atkinson EJ, Therneau TM. Assessing calibration of prognostic risk scores. Stat Methods Med Res. 2016 Aug;25(4):1692-1706. doi: 10.1177/0962280213497434.
76. Rahman MS, Ambler G, Choodari-Oskooei B, Omar RZ. Review and evaluation of performance measures for survival prediction models in external validation settings. BMC Med Res Methodol. 2017 Apr 18;17(1):60. doi: 10.1186/s12874-017-0336-2.
77. Stevens RJ, Poppe KK. Validation of clinical prediction models: what does the "calibration slope" really measure? J Clin Epidemiol. 2020 Feb;118:93-99. doi: 10.1016/j.jclinepi.2019.09.016.
78. Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol. 2016 Jan;69:245-247. doi: 10.1016/j.jclinepi.2015.04.005.
79. Pattengale ND, Alipour M, Bininda-Emonds OR, Moret BM, Stamatakis A. How many bootstrap replicates are necessary? J Comput Biol. 2010 Mar;17(3):337-354. doi: 10.1089/cmb.2009.0179.
80. Sauerbrei W, Perperoglou A, Schmid M, Abrahamowicz M, Becher H, Binder H, Dunkler D, Harrell FE Jr, Royston P, Heinze G; for TG2 of the STRATOS initiative. State of the art in selection of variables and functional forms in multivariable analysis-outstanding issues. Diagn Progn Res. 2020 Apr 2;4:3. doi: 10.1186/s41512-020-00074-3.
81. Young GA. Bootstrap: More than a Stab in the Dark? Statist. Sci. 1994 Aug; 9 (3):382 – 395. doi: 10.1214/ss/1177010383
82. Koul A, Becchio C, Cavallo A. Cross-Validation Approaches for Replicability in Psychology. Front Psychol. 2018 Jul 2;9:1117. doi: 10.3389/fpsyg.2018.01117.

83. Borra S, Di Ciaccio. Measuring the prediction error. A comparison of cross-validation, bootstrap and covariance penalty methods. Comput. Stat. Data Anal. 2010 Dec 1;54(12):2976-2989.doi: 10.1016/j.csda.2010.03.004
84. Pedersen AB, Mikkelsen EM, Cronin-Fenton D, Kristensen NR, Pham TM, Pedersen L, Petersen I. Missing data and multiple imputation in clinical epidemiological research. Clin Epidemiol. 2017 Mar 15;9:157-166. doi: 10.2147/CLEP.S129785.
85. White IR, Royston P. Imputing missing covariate values for the Cox model. Stat Med. 2009 Jul 10;28(15):1982-1998. doi: 10.1002/sim.3618.
86. Carroll OU, Morris TP, Keogh RH. How are missing data in covariates handled in observational time-to-event studies in oncology? A systematic review. BMC Med Res Methodol. 2020 May 29;20(1):134. doi: 10.1186/s12874-020-01018-7.
87. Bartlett JW, Seaman SR, White IR, Carpenter JR; Alzheimer's Disease Neuroimaging Initiative*. Multiple imputation of covariates by fully conditional specification: Accommodating the substantive model. Stat Methods Med Res. 2015 Aug;24(4):462-487. doi: 10.1177/0962280214521348.
88. Aujesky D, Roy PM, Le Manach CP, Verschuren F, Meyer G, Obrosky DS, Stone RA, Cornuz J, Fine MJ. Validation of a model to predict adverse outcomes in patients with pulmonary embolism. Eur Heart J. 2006 Feb;27(4):476-481. doi: 10.1093/eurheartj/ehi588.
89. Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. BMC Med Res Methodol. 2013 Mar 6;13:33. doi: 10.1186/1471-2288-13-33.
90. Heng DY, Xie W, Regan MM, Harshman LC, Bjarnason GA, Vaishampayan UN, Mackenzie M, Wood L, Donskov F, Tan MH, Rha SY, Agarwal N, Kollmannsberger C, Rini BI, Choueiri TK. External validation and comparison with other models of the International Metastatic Renal-Cell Carcinoma Database Consortium prognostic model: a population-based study. Lancet Oncol. 2013 Feb;14(2):141-8. doi: 10.1016/S1470-2045(12)70559-4.
91. Tajzler C, Tanguay S, Mallick R, Ahrens B, Ly TL, Breau RH, Basappa NS, Kapoor A, Heng DYC, Pouliot F, Finelli A, Lavallée LT, So AI, Drachenberg DE, Soulières D, Bjarnason GA, Richard PO, Maloni R, Power NE, Haan M, Wood LA. Determining generalizability of the Canadian Kidney Cancer information system (CKCis) to the entire Canadian kidney cancer population. Can Urol Assoc J. 2020 Oct;14(10):E499-E506. doi: 10.5489/cuaj.6716.
92. Lawson KA, Saarela O, Liu Z, Lavallée LT, Breau RH, Wood L, Jewett MAS, Kapoor A, Tanguay S, Moore RB, Rendon R, Pouliot F, Black PC, Kawakami J, Drachenberg D, Finelli A. Benchmarking quality for renal cancer surgery: Canadian Kidney Cancer information system (CKCis) perspective. Can Urol Assoc J. 2017 Aug;11(8):232-237. doi: 10.5489/cuaj.4397.

Chapter 3: Validation of Prognostic Models for Renal Cancer Recurrence, Renal Cancer Specific Survival and Overall Survival in Canadian Patients

3.1 Abstract

Introduction Given the wide range of recurrence risk after kidney cancer surgery, an accurate prediction tool is useful to counsel patients, assess eligibility for adjuvant treatments, and ensure appropriate follow-up surveillance. The purpose of this analysis is to validate existing prognostic models for kidney cancer recurrence, cancer specific death, and death from any cause.

Methods Using data from the Canadian Kidney Cancer Information System (CKCis), 15 models for kidney cancer recurrence, six for cancer specific survival, and four for overall survival were assessed. Discrimination was measured using c-statistics and c-index over time. Brier Scores and Calibration plots were constructed at 2-, 3- and 5-years post-surgery. Decision curve analysis at 5-years post-surgery was also performed. The accuracy of prognostic models was compared to pathological T stage alone.

Results There was a wide range in model performance for discrimination and calibration of recurrence free survival, cancer specific survival, and overall survival. For recurrence free survival, c-statistics ranged from 0.61 to 0.83, depending on whether the model was derived, and applied, to patients of all risk/histology, specific risk groups, or to specific histological subtypes. Cancer specific survival models had c-statistics ranging from 0.59 to 0.88 and overall survival models had c-statistics ranging from 0.59 to 0.75. C-index over time for all three outcomes was highest from 1 to 3 years post-surgery. The best performing models yielded a higher net benefit compared to T-stage alone for 5-year recurrence risk thresholds in a clinically useful range for adjuvant therapy (0-60% recurrence risk).

Conclusions Significant differences in model performance were observed. Some models provide good discrimination and calibration when applied to a large cohort of Canadian patients.

3.2 Introduction

Renal cell carcinoma (RCC) is common in Canada with an estimated 7500 individuals newly diagnosed and 1950 RCC-related deaths in 2020 (1). Approximately 80% of patients are diagnosed with clinically localized disease (2) and most are treated with surgery (radical or partial nephrectomy) (3). Within 5 years of surgery, the risk of recurrence is between 8% and 60%, cancer specific survival ranges from 55% to 91%, and overall survival ranges from 44% to 84% (3). Given the wide range in risk, more specific estimates of prognosis are important for patient counseling and to guide clinical decisions.

Prognostic models include variables that predict a disease event, and often have clinical, histologic, and laboratory components (4). Prognostic models have an important role in personalized medicine as

intensity of treatment or surveillance can be proportional to the individual risk (4). Ideally, prognostic models would result in accurate prediction of the future and allow patients at highest risk of disease recurrence to receive more intensive care, and patients at lowest risk of recurrence to avoid unnecessary testing and treatments (4). Nomograms are often derived from prognostic models and risk group categories created to facilitate clinical adoption. Several kidney cancer prognostic models have been developed and it is important to assess if these models are valid for contemporary patients.

In the past 20 years, at least 15 models have been published to assess risk of recurrence, six to predict cancer-related death, and four for death from any cause (3,5-19). These models are usually derived from single institution databases (3,5-19). RCC histologic subtype is an important consideration as factors such as tumour grade, are prognostic for clear cell or papillary RCC subtypes, but not others (5-19). Some prognostic models have been derived for specific histologic subtypes and specific risk groups (high-stage or low-stage) and some have been externally validated (20,21). To our knowledge, no validation studies have been conducted in Canadian RCC patients. To examine the predictive accuracy and usefulness of these prognostic models in Canadian patients, we used a large prospective cohort of Canadian RCC patients.

3.3 Patients and Methods

3.3.1 Study Population

The Canadian Kidney Cancer Information System (CKCis) is a prospective cohort of individuals undergoing RCC treatment at 14 academic tertiary care institutions in Canada since January 1, 2011. Patients were included if treated with a radical or partial nephrectomy for RCC. Patients were included if they had local adenopathy, but patients with distant adenopathy or metastases were excluded. Additional exclusion criteria for this study were: age <18 years at surgery, treatment other than partial or radical nephrectomy (i.e surveillance or thermal ablation), presence of multiple tumours, a previously treated RCC, and presence of congenital RCC syndromes (e.g. von Hipple-Lindau disease). Figure 1 outlines the development of the patient cohort used in prognostic model development.

3.3.2 Data Collection

Trained data abstractors at each institution collect and enter patient baseline demographics and clinical information. Treatment and pathological information are captured from the medical record. Medical data entry is then performed at least annually for patients without recurrence after surgery, and at least every 3 months for patients with metastases. Last follow up is the last RCC clinical encounter (often at 5 years post treatment if no recurrence is detected). Patients in the database were designated as “lost-to-follow-

up" if they did not attend their scheduled cancer follow up or moved to another geographic location and could not be contacted. For external validation, all follow-up to June 30, 2021 was used.

3.3.3 Description of Prognostic Models

After searching relevant medical literature, 15 prognostic models for recurrence free survival, 6 models for cancer specific survival and 4 models for overall survival were identified (Tables 1, 2, and 3). While some models were derived from patients with all histologic subtypes (eg. clear cell, papillary, or chromophobe) or patient risk (eg high and low risk), others were specific to certain sub-groups of disease. Selection of models for this validation study was based on the model being constructed specifically for prognosis of patients without clinical metastases at the time of kidney cancer surgery (i.e. prognostic models that included patients with metastases were excluded). Models were excluded if they included genomic and laboratory markers not routinely collected in clinical practice.

3.3.4 Outcome Variables

The fundamental definitions of recurrence and cancer-specific death were consistent across studies. Post-surgical recurrence is defined as any RCC recurrence, including local recurrence in the resection bed or distant metastases as identified by imaging as part of post-surgical surveillance. Cancer specific death is defined as death directly related to kidney cancer as reported to the participating CKCis site. All cause mortality is defined as deaths due to any cause as reported to the participating CKCis site.

3.3.5 Predictor Variables

Different models had a different definitions or groupings of variables. For example, histologic grade 1 and 2 were grouped together as "low grade" in some models and analyzed separately in others. Similarly, clinical variables such as "symptoms" had various definitions between models. For each model validation, CKCis data was transformed to match the groupings and definitions of the model being assessed. For each model assessment we used all patients that had complete baseline information. Therefore, the cohort sizes were different, depending on the model being assessed.

3.3.6 Assessment of Models

Model accuracy was assessed using c-statistics, c-index at various time points, calibration plots, and Brier scores. Each prognostic model was assessed for the outcome from which it was derived. When nomograms were provided, the nomogram was used as the sole predictor of outcome in the Cox Proportional Hazards model. In the absence of a nomogram, prognostic indexes were derived from the Cox Proportional Hazards model using the sum of predictor coefficients. As a referent comparator, we derived a model using pathologic tumour stage as the only predictor for each outcome.

Discrimination was evaluated using the overall c-statistic and the c-index at time points (1-year, 2-year, etc). Given pairs of observations, the c-statistic is the proportion of pairs where the predicted higher risk for the outcome is correctly identified (22). A c-statistic of 0.5 indicates no discriminative ability, 1.0 indicates perfect discriminative ability and c-statistic ≥0.7 indicates potential clinical utility (22). The c-index is a measure of discrimination at various follow-up time points (23). The same benchmarks for no discriminative ability, perfect discriminative ability, and potential clinical utility apply for c-index (23).

The Brier Score is a composite measure of discrimination and calibration (24). The Brier score is the averaged-squared-difference between the predicted survival status and the observed survival status at a given period of follow-up (eg at 5 years post-surgery). The Brier score could range between 0 and 1, with values closer to 0 indicating better accuracy (24).

Calibration and discrimination were assessed by inspection of calibration plots using LOESS (locally weighted scatterplot smoothing). LOESS smoothing combines linear least squares regression and nonlinear regression in modeling for complex relationships in data (25). Predicted probability of outcomes were computed using the Cox Proportional Hazards model with the prognostic index as the sole parameter and observed outcomes were derived using the Kaplan Meier non-parametric survival analysis (26). For prognostic models where authors did not provide a specified model and only grouped patients into risk groups based on deciles (or patient groups with at least 30 patients where decile-based groupings are not adequate due to small sample size), or had categorical prognostic indices, grouped bar graphs of predicted versus observed risk were created in lieu of calibration plots. Patients were grouped based on predicted probability with the mean predicted probability used for each group. Brier scores and calibration plots were calculated and created for 2-years post-surgery, 3-years post-surgery, and 5-years post-surgery.

Decision Curve Analysis is an analytic method for evaluating prognostic model performance by calculating the net benefit to patients against various thresholds of patient risk (27). The range in threshold probabilities at which the net benefit is greater than or equal to zero represent the range of thresholds where patients are expected to benefit from the model (27). If the model were used to select patients for adjuvant therapy (where toxicity would be expected), we estimate the clinically meaningful model threshold for most patients to be between 20% and 40% for recurrence and between 5% and 15% for death.

Statistical analyses were performed using SAS Enterprise Guide 8.2 and Microsoft Excel 2016 with decision curve analysis performed in R 4.0.2.

3.4 Results

From January 1, 2011, 5321 CKCis patients met the criteria of solitary localized renal tumours treated with surgery (partial or radical nephrectomy). Table 4 presents clinical and pathological characteristics of the CKCis cohort. 606 (11.4%) of the cohort were lost-to-follow up at 2-years post-surgery, 937 (17.6%) were lost-to-follow up at 3-years post-surgery and 972 (18.3%) were lost-to-follow up at 5-years post-surgery. There were 5321 patients with any follow-up for recurrence free survival, 5234 patients with any follow-up for cancer specific survival and 5321 patients with any follow-up for overall survival that will be included in the subsequent analyses for external prognostic model validation. The discrepancy in numbers between cancer specific survival and recurrence/overall survival is that 87 patients died during follow up but the cause of death was unknown.

3.4.1 Recurrence Free Survival

Overall, at last follow-up, 1049 patients experienced kidney cancer recurrence during follow up (18.7% at 2-years, 23.6% at 3-years and 31.0% at 5-years post-surgery) (Figure 2). Among histologic subtypes, more recurrences were observed in clear-cell renal cell carcinoma (ccRCC; 33.9% 5yr recurrence) compared to papillary renal cell carcinoma (pRCC; 22.4% 5yr recurrence) and chromophobe renal cell carcinoma (chRCC; 15.9% 5yr recurrence) (Figure 3).

Recurrence in All Histological Subtypes and Risk Groups

Models for recurrence are described in Table 2. Most of the prognostic models for recurrence are based on pathological features or a combination of clinical and pathological features. The exceptions are Yaycioglu et al 2001, Cindolo et al 2003 and Yaycioglu et al 2013, which are entirely based on clinical features available prior to surgery. C-statistics and Brier scores for each model are summarized in Table 5.

For all RCC histology and RCC risk, Yaycioglu et al 2013 had the highest overall c-statistic (Table 5). All models except for the Yaycioglu model, had lower c-statistic than T-stage alone. All models had similar or inferior c-index over time compared to pT-stage alone (Figure 4). Kattan et al 2001 had the lowest Brier Score at 2-, 3- and 5-years post-surgery.

While the models were reasonably calibrated at 2-, 3- and 5-years post-surgery, these models often over- and under-estimated recurrence risk at different ranges.

Decision Curve Analysis is presented in Figure 8. Yaycioglu et al 2013 and Zisman et al 2002 had the largest range of recurrence risk thresholds for clinical utility (Table 6) although in all models, there was no net-benefit in the upper thresholds of clinical utility for adjuvant therapy (eg. 25% to 40%).

Recurrence Models Including Low-Risk Patients Only

One prognostic model was specific for patients defined as low risk (i.e. pathological T1 and T2 stage) (8). At all timepoints, the model had lower c-index compared to pathological T stage alone (Figure 4).

Recurrence Models Including High Risk Patients Only

One prognostic model was based on patients from a clinical trial (ASSURE trial) that only included high risk patients (i.e. inclusion into the trial was pathological stage pT1b and nuclear grade G3/G4 or pathological stage pT2/3/4 or pathological nodal stage pN1). This model had a higher c-index over time at all timepoints compared to pathological T stage (Figure 4). Calibration performance based on the calibration plots at 2-, 3- and 5-years indicated poor calibration to CKCis patients. On decision curve analysis, there was a net benefit to patients with 0%-33.33% risk of recurrence(Figure 8).

Recurrence Models Including ccRCC Patients Only

Of the 3 ccRCC models, all had higher c-statistics than pathological T stage alone. Prognostic models for ccRCC patients were reasonably well calibrated to our cohort (Figures 5-7). Sorbellini et al 2005 had consistently underestimated recurrence risk. Leibovich et al 2003 provided net benefits to the widest range of patient risk and included the range of clinical utility (Table 6, Figure 8).

Recurrence Models Including pRCC Patients Only

Two models were specific to pRCC patients. C-statistic, Brier scores, and C-index over time are presented in Table 5 and Figures 4. Calibration is presented in Figures 5 to 7 and both had net-benefit across the range in thresholds for adjuvant therapy (Figure 8).

Recurrence Models Including chRCC Patients Only

One prognostic model was specific for chRCC and had lower c-statistics than pathological T stage alone. The model was poorly calibrated to our cohort based on calibration plots at 2-,3- and 5-years. the chRCC model provided net benefit to patients with 0%-73% and 81.25%-100% risk of recurrence (Table 8).

3.4.2 Cancer Specific Survival

Overall, at last follow-up, 311 of patients experienced kidney cancer specific death during follow up (3.7% at 2-years, 5.8% at 3-years and 10.6% at 5-years post-surgery) (Figure 9). Among histologic subtypes, more cancer deaths were observed in patinets with pRCC (11.9% 5yr cancer death) compared to ccRCC (10.2% 5yr cancer death) and chRCC (6.4% 5yr cancer death) (Figure 10).

Cancer specific survival Including All Histological Subtypes and Risk Groups

Three models included all RCC histology. Kutikov et al 2010 had the highest overall c-statistic and highest c-index over time (Table 6). C-indices were lower than the referent model of pathological T stage alone and c-index was highest at 1-2 years post-surgery (Figure 11). It was also observed that Brier Scores increased over time from 2- and 3-years to 5-years post-surgery. Calibration plots show excellent agreement between predicted and observed cancer specific death risk in prognostic models for all patients at 2-years post-surgery (Figure 12). Calibration performance worsened over time for prognostic models for all patients. Based on decision curve analysis at 5-years post-surgery, Zisman et al 2002 had the widest range of thresholds for which the prognostic model is clinically useful in CKCis patients (Table 8, Figure 15).

Cancer specific survival Including Clear Cell RCC Patients Only

Frank et al 2002 and Leibovich et al 2018 models were created using ccRCC patients. Between the two models Leibovich et al 2018 had higher overall c-statistic and c-index over time. C-index over time was higher than the reference model using pathological T stage as the sole predictor and c-index was highest at 1-year post-surgery using the Leibovich et al 2018 and 7-8 years post-surgery using the Frank et al 2002 model. On calibration plots, Frank et al 2002, demonstrated an underestimation of patient risk of cancer related death at 3- and 5-years post-surgery. (Figures 13-14). The models had a wide range of thresholds with net benefit compared to pathological T stage as the sole predictor (Table 8, Figure 15).

Cancer specific survival Including Papillary RCC Patients Only

Leibovich et al 2018 was the only model for pRCC related death and had a lower c-index over time compared to the referent model with pathological T stage as the sole predictor. The model was poorly calibrated to our cohort regardless of time post-surgery. Compared to univariate models with pathological T stage, prognostic models for pRCC patients had lower range of threshold risk for clinical utility.

3.4.3 Overall Survival

Overall, at last follow-up, 494 patients experienced death during follow up (5.6% at 2-years, 9.2% at 3-years and 15.9% at 5-years post-surgery) (Figure 16). Among histologic subtypes, by 5-years post surgery, more deaths occurred in patients with pRCC (18.8%) compared to ccRCC (15.6%) and chRCC (8.2%) (Figure 17).

Overall survival Including All Histological Subtypes and Risk Groups

Three prognostic models were derived from all RCC patients. Discriminative performance of the models are presented in Table 7 and Figure 18. Calibration is presented in Figures 19 to 21. Based on decision curve analysis, all models provide net benefit in a clinical useful range of thresholds (Table 8, Figure 22).

Overall survival Including High Risk Patients Only

One prognostic model was derived from the ASSURE trial and assessed overall survival in high-risk patients only. This model had a higher overall C-statistic and c-index compared to pathological T stage only. In our cohort, Correa et al 2021 under-estimated risk of death at higher predicted probabilities at 2- and 3-years post-surgery but over-estimated risk of death at higher predicted probabilities at 5-years post-surgery. (Figures 19-21). The model was useful in patients with threshold risk of death up to 21.9% (Table 8, Figure 22).

3.5 Discussion

External assessments of prognostic models are essential to ensure generalizability in different populations and contemporary patients. As expected, accuracy of prognostic models in the CKCis cohort was generally lower compared to the internal validations in the original publications. The distributions of prognostic factors such as tumour stage also varied slightly between the CKCis cohort and the model development cohorts. This further highlights the need for external validation of prognostic models prior to applying models to new populations. We identified a wide range in model performance with some models performing well in Canadian patients. We identified prognostic models that provided clinically useful discrimination, calibration, and net-benefit (if used for selecting patients for adjuvant therapy).

Some models for recurrence have previously been assessed in other external validation studies(28). Discriminative accuracy was the most common external validation metric with assessment of calibration or clinical utility frequently omitted. Compared to previous external validations, we found Yaycioglu et al 2001 (c-statistic 0.63-0.70 on external validation vs. 0.74 in this CKCis external validation) and Cindolo et al 2003 (c-statistic 0.63-0.75 on external validation vs. 0.74 in this CKCis validation study), and Kattan et al

2001 (c-statistic 0.61-0.84 on previous external validation vs. 0.78 in this CKCis validation study) to have similar discriminative accuracy (21). Consistent with other external validations studies, Leibovich 2003 seems to be the most well calibrated (20,21,28). Our findings are consistent with previous external validations where models that included only pathological (post-surgical) information performed better than those using only clinical (pre-operative) information (i.e. Yaycioglu et al 2001, Cindolo et al 2003 and Yaycioglu et al 2013)(20). Similarly, models specific to histologic subtypes performed better than models grouping all RCC histology (20,21).

For cancer-specific survival, Frank et al 2002 provided better discrimination in CKCis patients (c-statistic:0.83) compared to most previous external validation studies (range in c-statistics:0.71-0.86) (21). For overall survival, Zisman et al 2002 had much better discriminative accuracy in other publications (c-statistic 0.76 to 0.86) compared to this analysis (c-statistic 0.59)(23). The performance of Zisman et al 2001 in CKCis patients (c-statistic:0.75) was similar to other publications (range in c-statistics:0.64-0.83) (21). In general, consistent with previous publications, these models underestimated all-cause mortality at the extremes of patient risk(21).

For all three outcomes examined in this validation study, highest discriminative ability over time is achieved within 2 years of surgery, which is consistent with results from a previous analysis of c-index that used patients enrolled in the ASSURE trial (20).

3.5.1 Limitations and Interpretation Considerations

Direct comparisons of models in this study should be made with caution, as the cohorts were different depending on availability of baseline information required for application of each model. In addition, the coefficients for the pT stage referent group were derived from the CKCis cohort, biasing for better performance of pT stage compared to externally derived models.

External assessments of prognostic models may be limited since all publications did not report baseline hazard. Therefore, it could not be determined whether miscalibrations in the model were due to miscalibration in the model parameters or from differences between the baseline hazard function within CKCis patients and patients used in the original development of these prognostic models (26,29). Indeed, without baseline hazard information our calibration assessments may be overestimated.

The CKCis cohort also presents limitations. Our cohort had insufficient sample size or follow-up events for rare histologic subtypes (i.e. chromophobe) and models that incorporate recently discovered biomarkers could not be assessed (30).

3.5.2 Strengths

CKCis is a large prospective cohort that is representative of the Canadian kidney cancer population, containing approximately 20% of all kidney cancer patients diagnosed in Canada (31). CKCis also includes contemporary patients who received surgery (eg aggressive partial nephrectomy) and systemic therapies (eg immune checkpoint inhibitors) more consistent with modern clinical practice. In our analysis we used multiple methods to examine accuracy, which allowed for a more thorough understanding of model performance, including calibration, which is likely the most important characteristic for translating to patient care.

3.6 Conclusion

In conclusion, there are many prognostic models for patients with non-metastatic kidney cancer. We have presented the performance of these models in a large contemporary cohort of Canadian patients. These findings provide clinicians and patients a better understanding of model performance when applying to patient care. Most models are superior to using tumour stage alone and we encourage use of model-based risk when designing clinical trials, counseling patients, and scheduling follow-up imaging.

3.7 Other Information

There is no supplementary information available and this study did not receive any external funding.

3.8 References

1. Brenner DR, Weir HK, Demers AA, Ellison LF, Louzado C, Shaw A, Turner D, Woods RR, Smith LM. Projected estimates of cancer in Canada in 2020.CMAJ. 2020 Mar 2;192(9): E199-205. DOI: 10.1503/cmaj.191292.
2. Capitanio U, Montorsi F. Renal cancer. Lancet. 2016 Feb 27;387(10021):894-906. doi: 10.1016/S0140-6736(15)00046-X.
3. Zisman A, Pantuck AJ, Wieder J, Chao DH, Dorey F, Said JW, deKernion JB, Figlin RA, Belldegrun AS. Risk group assessment and clinical outcome algorithm to predict the natural history of patients with surgically resected RCC. J Clin Oncol. 2002 Dec 1;20(23):4559-4566. doi: 10.1200/JCO.2002.05.111.
4. Vogenberg FR. Predictive and prognostic models: implications for healthcare decision-making in a modern recession. Am Health Drug Benefits. 2009 Sep;2(6):218-222.
5. Yaycioglu O, Roberts WW, Chan T, Epstein JI, Marshall FF, Kavoussi LR. Prognostic assessment of nonmetastatic RCC: a clinically based model. Urology. 2001 Aug;58(2):141-145. doi: 10.1016/s0090-4295(01)01207-9.
6. Cindolo L, de la Taille A, Messina G, Romis L, Abbou CC, Altieri V, Rodriguez A, Patard JJ. A preoperative clinical prognostic model for non-metastatic RCC. BJU Int. 2003 Dec;92(9):901-905. doi: 10.1111/j.1464-410x.2003.04505.x.
7. Kattan MW, Reuter V, Motzer RJ, Katz J, Russo P. A postoperative prognostic nomogram for RCC. J Urol. 2001 Jul;166(1):63-67.
8. Yaycioglu O, Eskicorapci S, Karabulut E, Soyupak B, Gogus C, Divrik T, Turkeri L, Yazici S, Ozen H; Society of Urooncology Study Group for Kidney Cancer Prognosis. A preoperative prognostic

model predicting recurrence-free survival for patients with kidney cancer. Jpn J Clin Oncol. 2013 Jan;43(1):63-68. doi: 10.1093/jjco/hys192.
9. van der Mijn JC, Al Hussein Al Awamlh B, Islam Khan A, Posada-Calderon L, Oromendia C, Fainberg J, Alshak M, Elahjji R, Pierce H, Taylor B, Gudas LJ, Nanus DM, Molina AM, Del Pizzo J, Scherr DS. Validation of risk factors for recurrence of RCC: Results from a large single-institution series. PLoS One. 2019 Dec 9;14(12):e0226285. doi: 10.1371/journal.pone.0226285.
10. Passalacqua R, Caminiti C, Buti S, Porta C, Camisa R, Braglia L, Tomasello G, Vaglio A, Labianca R, Rondini E, Sabbatini R, Nastasi G, Artioli F, Prati A, Potenzoni M, Pezzuolo D, Oliva E, Alberici F, Buzio C; POLAR-01 Trial Investigators. Adjuvant low-dose interleukin-2 (IL-2) plus interferon-α (IFN-α) in operable RCC (RCC): a phase III, randomized, multicentre trial of the Italian Oncology Group for Clinical Research (GOIRC). J Immunother. 2014 Nov-Dec;37(9):440-7. doi: 10.1097/CJI.0000000000000055.
11. Leibovich BC, Blute ML, Cheville JC, Lohse CM, Frank I, Kwon ED, Weaver AL, Parker AS, Zincke H. Prediction of progression after radical nephrectomy for patients with clear cell RCC: a stratification tool for prospective clinical trials. Cancer. 2003 Apr 1;97(7):1663-1671. doi: 10.1002/cncr.11234.
12. Sorbellini M, Kattan MW, Snyder ME, Reuter V, Motzer R, Goetzl M, McKiernan J, Russo P. A postoperative prognostic nomogram predicting recurrence for patients with conventional clear cell RCC. J Urol. 2005 Jan;173(1):48-51. doi: 10.1097/01.ju.0000148261.19532.2c.
13. Leibovich BC, Lohse CM, Cheville JC, Zaid HB, Boorjian SA, Frank I, Thompson RH, Parker WP. Predicting Oncologic Outcomes in RCC After Surgery. Eur Urol. 2018 May;73(5):772-780. doi: 10.1016/j.eururo.2018.01.005.
14. Klatte T, Gallagher KM, Afferi L, Volpe A, Kroeger N, Ribback S, McNeill A, Riddick ACP, Armitage JN, 'Aho TF, Eisen T, Fife K, Bex A, Pantuck AJ, Stewart GD. The VENUSS prognostic model to predict disease recurrence following surgery for non-metastatic papillary RCC: development and evaluation using the ASSURE prospective clinical trial cohort. BMC Med. 2019 Oct 3;17(1):182. doi: 10.1186/s12916-019-1419-1.
15. Correa AF, Jegede OA, Haas NB, Flaherty KT, Pins MR, Adeniran A, Messing EM, Manola J, Wood CG, Kane CJ, Jewett MAS, Dutcher JP, DiPaola RS, Carducci MA, Uzzo RG. Predicting Disease Recurrence, Early Progression, and Overall Survival Following Surgical Resection for High-risk Localized and Locally Advanced RCC. Eur Urol. 2021 Jul;80(1):20-31. doi: 10.1016/j.eururo.2021.02.025.
16. Kutikov A, Egleston BL, Wong YN, Uzzo RG. Evaluating overall survival and competing risks of death in patients with localized RCC using a comprehensive nomogram. J Clin Oncol. 2010 Jan 10;28(2):311-317. doi: 10.1200/JCO.2009.22.4816.
17. Kutikov A, Egleston BL, Canter D, Smaldone MC, Wong YN, Uzzo RG. Competing risks of death in patients with localized RCC: a comorbidity based model. J Urol. 2012 Dec;188(6):2077-2083. doi: 10.1016/j.juro.2012.07.100.
18. Frank I, Blute ML, Cheville JC, Lohse CM, Weaver AL, Zincke H. An outcome prediction model for patients with clear cell RCC treated with radical nephrectomy based on tumor stage, size, grade and necrosis: the SSIGN score. J Urol. 2002 Dec;168(6):2395-2400. doi: 10.1097/01.ju.0000035885.91935.d5.

19. Zisman A, Pantuck AJ, Dorey F, Said JW, Shvarts O, Quintana D, Gitlitz BJ, deKernion JB, Figlin RA, Belldegrun AS. Improved prognostication of RCC using an integrated staging system. J Clin Oncol. 2001 Mar 15;19(6):1649-1657. doi: 10.1200/JCO.2001.19.6.1649.
20. Correa AF, Jegede O, Haas NB, Flaherty KT, Pins MR, Messing EM, Manola J, Wood CG, Kane CJ, Jewett MAS, Dutcher JP, DiPaola RS, Carducci MA, Uzzo RG. Predicting Renal Cancer Recurrence: Defining Limitations of Existing Prognostic Models With Prospective Trial-Based Validation. J Clin Oncol. 2019 Aug 10;37(23):2062-2071. doi: 10.1200/JCO.19.00107.
21. Usher-Smith JA, Li L, Roberts L, Harrison H, Rossi SH, Sharp SJ, Coupland C, Hippisley-Cox J, Griffin SJ, Klatte T, Stewart GD. Risk models for recurrence and survival after kidney cancer: a systematic review. BJU Int. 2021 Dec 16. doi: 10.1111/bju.15673.
22. Caetano SJ, Sonpavde G, Pond GR. C-statistic: A brief explanation of its construction, interpretation and limitations. Eur J Cancer. 2018 Feb;90:130-132. doi: 10.1016/j.ejca.2017.10.027.
23. Galfano A, Novara G, Iafrate M, Cavalleri S, Martignoni G, Gardiman M, D'Elia C, Patard JJ, Artibani W, Ficarra V. Mathematical models for prognostic prediction in patients with RCC. Urol Int. 2008;80(2):113-23. doi: 10.1159/000112599.
24. Assel M, Sjoberg DD, Vickers AJ. The Brier score does not evaluate the clinical utility of diagnostic tests or prediction models. Diagn Progn Res. 2017 Dec 2;1:19. doi: 10.1186/s41512-017-0020-3.
25. Bang H, Mazumdar M, Spence JD. Tutorial in biostatistics: Analyzing associations between total plasma homocysteine and B vitamins using optimal categorization and segmented regression. Neuroepidemiology. 2006;27(4):188-200. doi: 10.1159/000096149.
26. Steyerberg EW. Clinical Prediction Models A Practical Approach to Development, Validation, and Updating. 2nd ed. Cham (Switzerland):Springer Nature Switzerland;c2019.558p.
27. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. Diagn Progn Res. 2019 Oct 4;3:18. doi: 10.1186/s41512-019-0064-7.
28. Rosiello G, Larcher A, Fallara G, Giancristofaro C, Martini A, Re C, Cei F, Musso G, Tian Z, Karakiewicz PI, Mottrie A, Bertini R, Salonia A, Necchi A, Raggi D, Briganti A, Montorsi F, Capitanio U. Head-to-head comparison of all the prognostic models recommended by the European Association of Urology Guidelines to predict oncologic outcomes in patients with RCC. Urol Oncol. 2022 Jun;40(6):271.e19-271.e27. doi: 10.1016/j.urolonc.2021.12.010.
29. Crowson CS, Atkinson EJ, Therneau TM. Assessing calibration of prognostic risk scores. Stat Methods Med Res. 2016 Aug;25(4):1692-706. doi: 10.1177/0962280213497434.
30. Seles M, Posch F, Pichler GP, Gary T, Pummer K, Zigeuner R, Hutterer GC, Pichler M. Blood Platelet Volume Represents a Novel Prognostic Factor in Patients with Nonmetastatic Renal Cell Carcinoma and Improves the Predictive Ability of Established Prognostic Scores. J Urol. 2017 Dec;198(6):1247-1252. doi: 10.1016/j.juro.2017.07.036.
31. Tajzler, C., Tanguay, S., Mallick, R., Ahrens, B., Ly, T. L., Breau, R. H., Basappa, N. S., Kapoor, A., Heng, D. Y., Pouliot, F., Finelli, A., Lavallee, L. T., So, A. I., Drachenberg, D. E., Soulieres, D., Bjarnason, G., Richard, P. O., Maloni, R., Power, N. E., Haan, M., & Wood, L. A. Determining generalizability of the Canadian Kidney Cancer information system (CKCis) to the entire Canadian kidney cancer population. Can Urol Assoc J. 2020 Jun 16; 14(10). doi: 10.5489/cuaj.6716.

3.9 Figures and Tables

Table 1: Kidney Cancer Recurrence Models

Model	Patient Population Used to Generate Model	Prognostic Factors in Model	C-statistic Reported	Parameter Selection Method
All Histology RCC Patients				
Yaycioglu et al, 2001 (4)	N=296; Open nephrectomy patients without any: bilateral synchronous disease or metastases, von Hippel-Lindau disease, lymphadenopathy, tumour that extends further than Gerota's fascia nor vena cava involved further than diaphragm	Clinical presentation, size largest tumour	c-statistic not reported	parameter selection method not reported
Cindolo et al, 2003 (5)	N=660; Patients who underwent surgery for kidney without: bilateral synchronous/metachronous disease, von Hippel-Lindau disease, tumour that extends further than Gerota's fascia, lymphadenopathy nor metastases	Clinical presentation, size largest tumour	c-statistic not reported	forward selection
Zisman et al, 2002 (2)	N=814; Patients who underwent surgery for unilateral kidney cancer	Pathological T stage, Nuclear Grade, ECOG, pathological N stage, pathological M stage	c-statistic not reported	parameter selection method not reported
Kattan et al, 2001 (6)	N=601; Partial or radical nephrectomy patients for unilateral kidney disease except for those who had: bilateral synchronous disease, metastases, lymph node positive disease	Clinical Presentation, Histology, pathological T stage, size largest tumour	c=0.74	parameter selection method not reported
Yacioglu et al, 2013 (7)	N=1889; Kidney cancer patients who underwent surgery except for those with: von Hippel-Lindau disease or bilateral tumours	Age at surgery, gender, clinical presentation, imaging size, clinical T stage, radiological lymph nodes (clinical N stage)	c-statistic=0.75	used all parameters significant in univariate analysis to create multivariate models
van der Mijn et al, 2019 (8)	N=873 for all patients; Non-metastatic kidney cancer patients aged ≥18 years who underwent partial or radical nephrectomy	Gender, extent, ASA score, pathological T stage, nuclear grade, WHO kidney cancer histology classification, diabetes mellitus, hypertension	c-statistic not reported	used all parameters significant in univariate analysis to create multivariate models

Passalacqua et al 2014 (9)	N=310; Kidney cancer patients aged <75 years who underwent partial or radical nephrectomy	Age, pathological T stage, nuclear grade, pathological nodal status (pathological N stage)	c-statistic not reported	parameter selection method not reported
Clear Cell Histology Renal Cell Carcinoma Patients				
Leibovich et al, 2003 (10)	N=1671; Radical nephrectomy patients for clear cell renal cell carcinoma except those who had: bilateral synchronous tumours, hereditary cancers, von Hippel-Lindau and tuberous sclerosis syndromes, Wilms tumour, metastases, age <18 years	Pathological T stage, pathological N stage, nuclear grade, necrosis, size largest tumour	c=0.82	stepwise selection
Sorbellini et al, 2005 (11)	N=701; Nephrectomy patients for clear cell renal cell carcinoma except those who had: pT3c/pT4 cancers, metastases, von Hippel Lindau disease, hereditary papillary renal cell carcinoma, bilateral tumours	Size largest tumour, pathological T stage, nuclear grade, necrosis, lymphovascular invasion, clinical presentation	c=0.82	parameter selection method not reported
Leibovich et al, 2018 (12)	N=2726; overall for clear cell renal cell carcinoma (ccRCC)+ papillary renal cell carcinoma (papRCC)+ chromophobe renal cell carcinoma; Non-metastatic kidney cancer patients; for clear cell renal cell carcinoma only	Clinical presentation, nuclear grade, necrosis, sarcomatoid, size largest tumour, perinephric or renal sinus fat invasion, tumour thrombus, extension beyond kidney, nodal involvement	c=0.83	forward selection and backward selection
Papillary Histology Renal Cell Carcinoma Patients				
Klatte et al, 2019 (13)	N=556; Sporadic non-metastatic, unilateral kidney cancer patients with papillary renal cell carcinoma who underwent partial or radical nephrectomy	Size largest tumour, pathological T stage, pathological N stage, nuclear grade, venous tumour thrombus	c-statistic not reported	backward selection
Leibovich et al, 2018 (12)	N=607 overall for clear cell renal cell carcinoma (ccRCC)+ papillary renal cell carcinoma (papRCC)+ chromophobe renal cell carcinoma; Non-metastatic kidney cancer patients; for clear cell renal cell carcinoma only	Nuclear grade, perinephric or renal sinus fat invasion, tumour thrombus	c=0.77	parameter selection method not reported
Chromophobe Histology Renal Cell Carcinoma Patients				

Leibovich et al, 2018 (12)	N=222; overall for clear cell renal cell carcinoma (ccRCC)+ papillary renal cell carcinoma (papRCC)+ chromophobe renal cell carcinoma; Non-metastatic kidney cancer patients; for clear cell renal cell carcinoma only	Nuclear grade, perinephric or renal sinus fat invasion, tumour thrombus	c=0.78 for pRCC	parameter selection method not reported
All Histology High Risk Patients				
Correa et al, 2021 (14)	N=1943; Kidney cancer patients with pathological stage pT1b and nuclear grade G3/G4 or pathological stage pT2/3/4 or pathological nodal stage pN1 who underwent radical nephrectomy	Lymphovascular invasion,histological subtype, size largest tumour, nuclear grade, necrosis, regional lymph nodes (pathological N stage)	c=0.68	include all clinically significant variables whose inclusion resulted in ≥1% increase in c-statistic
All Histology Low Risk Patients				
van der Mijn et al 2019 (8)	Non-metastatic kidney cancer patients aged ≥18 years who underwent partial or radical nephrectomy with pathological T1/T2 stage	Gender, extent, ASA score, pathological T stage, nuclear grade, WHO kidney cancer histology classification, diabetes mellitus, hypertension	c-statistic not reported	used all parameters significant in univariate analysis to create multivariate models

Abbreviations:N=number of patients; c=c-statistic; G=grade;pT=pathological T stage; pN=pathological N stage; ECOG=Eastern Cooperative Oncology Group, ASA=American Society of Anesthesiologists; WHO=World Health Organization

Table 2: Kidney Cancer Specific Survival Models

Model	Patient Population Used to Generate Model	Prognostic Factors in Model	C-statistic Reported	Parameter Selection Method
All Histology RCC Patients				
Kutikov et al, 2010 (15)	N=30801; Kidney Cancer patients who underwent surgery aged ≥30 years and tumour size <20cm	Race, Gender, age at diagnosis, size largest tumour	c-statistic not reported	parameter selection method not reported
Kutikov et al, 2012 (16)	N=6665; First time kidney cancer patients with nodal negative disease excluding those who were under 66 years, had uncommon histological subtypes and tumours >20cm	Race, Gender, size largest tumour, Charlson Comorbidity Score, age at diagnosis	c-statistic not reported	parameter selection method not reported
Zisman et al, 2002 (2)	N=814; Patients who underwent surgery for unilateral kidney cancer	Pathological T stage, Nuclear Grade, ECOG, pathological N stage, pathological M stage	c-statistic not reported	parameter selection method not reported

Clear Cell Histology Renal Cell Carcinoma Patients				
Frank et al, 2002 (17)	N=1801; Clear cell renal carcinoma patients who underwent radical nephrectomy excluding: bilateral synchronous tumors, hereditary von Hippel-Lindau or tuberous sclerosis syndromes, Wilms tumor and those aged ≤18 years at surgery	Pathological T stage, pathological N stage, pathological M stage, size largest tumour, nuclear grade, necrosis	c=0.84	stepwise selection
Leibovich et al, 2018 (12)	N=2726; overall for clear cell renal cell carcinoma (ccRCC)+ papillary renal cell carcinoma (papRCC)+ chromophobe renal cell carcinoma; Non-metastatic kidney cancer patients; for clear cell renal cell carcinoma only	Age at surgery, ECOG score, adrenalectomy, positive margins, clinical presentation, nuclear grade, necrosis, sarcomatoid, size largest tumour, perinephric or renal sinus fat invasion, tumour thrombus, nodal involvement	c=0.86	forward selection and backward selection
Papillary Histology Renal Cell Carcinoma Patients				
Leibovich et al, 2018 (12)	N=607; overall for clear cell renal cell carcinoma (ccRCC)+ papillary renal cell carcinoma (papRCC)+ chromophobe renal cell carcinoma; Non-metastatic kidney cancer patients; for clear cell renal cell carcinoma only	Nuclear grade, perinephric or renal sinus fat invasion, tumour thrombus	c=0.83	parameter selection method not reported

Abbreviations:N=number of patients; c=c-statistic; G=grade;pT=pathological T stage; pN=pathological N stage; ECOG=Eastern Cooperative Oncology Group, ASA=American Society of Anesthesiologists

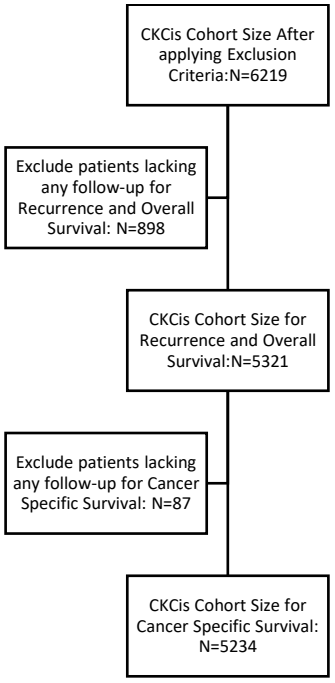
Table 3: Overall Survival Models

Model	Patient Population Used to Generate Model	Prognostic Factors in Model	C-statistic Reported	Parameter Selection Method
All Histology Patients				
Zisman et al, 2001 (18)	N=661; Renal cell carcinoma patients who underwent radical or partial nephrectomy	ECOG score, nuclear grade, size largest tumour	c-statistic not reported	forward selection
Zisman et al, 2002 (2)	N=814; Patients who underwent surgery for unilateral kidney cancer	Pathological T stage, Nuclear Grade, ECOG, pathological N stage, pathological M stage	c-statistic not reported	parameter selection method not reported

Passalacqua et al 2014 (9)	N=310; Kidney cancer patients aged <75 years who underwent partial or radical nephrectomy	Age, pathological T stage, nuclear grade, pathological nodal status (pathological N stage)	c-statistic not reported	parameter selection method not reported
All Histology High Risk Patients				
Correa et al, 2021 (14)	N=1943; Kidney cancer patients with pathological stage pT1b and nuclear grade G3/G4 or pathological stage pT2/3/4 or pathological nodal stage pN1 who underwent radical nephrectomy	Age at surgery, histological subtype, size largest tumour, nuclear grade, necrosis, regional lymph nodes (pathological N stage)	c=0.69	include all clinically significant variables whose inclusion resulted in ≥1% increase in c-statistic

Abbreviations:N=number of patients; c=c-statistic; G=grade;pT=pathological T stage; pN=pathological N stage; ECOG=Eastern Cooperative Oncology Group, ASA=American Society of Anesthesiologists

Figure 1: Summary of Patients Included for Analysis



Abbreviations: N=number of patients; CKCis=Canadian Kidney Cancer Information System

Table 4: Clinical and Pathological Features of CKCis Patients [1,2]

Feature	**N(%)**	**Number of Patients with Missing Data**
Age at Surgery (Years)	60.7±11.8	0
Gender (Male)	3502(65.8%)	0
Race		1885
White	2883(83.9%)	
Non-white	553(16.1%)	
CHMI score		190
2	647(12.6%)	
3	1020(19.9%)	
4	1188(23.2%)	
5	1051(20.5%)	
6	606(11.8%)	
7	322(6.3%)	
8	141(2.7%)	
9	83(1.6%)	
10	44(0.9%)	
11	20(0.4%)	
12	8(0.2%)	
13	1(0%)	
Hypertension (yes)	2662(51.9%)	190
Diabetes Mellitus (yes)	1115(21.7%)	190
ECOG score		2606
0	2092(77.1%)	
1	545(20.1%)	
2	62(2.3%)	
3	15(0.6%)	
4	1(0%)	
ASA Score		1542
1: Normal Healthy Patient	671(17.8%)	
2: Mild Systemic Disease	1705(45.1%)	
3: Severe Systemic Disease (that limits activity)	1232(32.6%)	
4: Severe Systemic Disease (life threatening)	167(4.4%)	
5: Moribound (not expected to survive)	2(0.1%)	
6: Declared brain-dead; organs are being removed for donor purposes	2(0.1%)	
Symptoms		0
No Symptoms	3705(69.6%)	
Local Symptoms	1305(24.5%)	
Systemic Symptoms	174(3.3%)	
Local and Systemic Symptoms	137(2.6%)	
Clinical T stage		0
T1a	2288(43%)	
T1b	1395(26.2%)	
T2a	496(9.3%)	
T2b	295(5.5%)	

T3a	649(12.2%)	
T3b	124(2.3%)	
T3c	29(0.5%)	
T4	45(0.8%)	
Clinical N stage		0
N0	5157(96.9%)	
N1	164(3.1%)	
Clinical Tumour Size (cm)	5.5±3.1	753
Extent		0
Partial	2659(50%)	
Radical	2662(50%)	
Adrenalectomy (yes)	672(12.6%)	0
Positive Surgical Margin (Yes)	391(7.7%)	213
Pathologic T stage		0
T1a	2246(42.2%)	
T1b	1063(20%)	
T2a	271(5.1%)	
T2b	127(2.4%)	
T3a	1425(26.8%)	
T3b	110(2.1%)	
T3c	31(0.6%)	
T4	48(0.9%)	
Pathologic N stage		0
N0	906(17%)	
N1	131(2.5%)	
NX	4284(80.5%)	
Pathologic M stage		0
M0	1624(30.5%)	
M1	10(0.2%)	
MX	3687(69.3%)	
Venous Tumour Thrombus		82
No	4422(84.4%)	
Level 0	574(11%)	
Level 1-4	243(4.6%)	
Perinephric Fat Invasion (yes)	608(12.1%)	285
Renal Sinus Fat Invasion (yes)	710(14.1%)	285
Extension Beyond Gerota's Fascia (yes)	22(0.4%)	285
Pathological Tumour Size (cm)	5.3±3.7	33
Histology		27
Conventional (clear cell) renal carcinoma	3817(72.1%)	
Chromophobe carcinoma	425(8%)	
Papillary renal cell carcinoma (type 1)	332(6.3%)	
Papillary renal cell carcinoma (type 2)	208(3.9%)	
Clear cell papillary renal cell carcinoma	173(3.3%)	
Papillary renal cell carcinoma (type not specified)	161(3%)	
Renal Cell Carcinoma unclassified	98(1.9%)	
Renal cell carcinoma	20(0.4%)	

Sarcomatoid carcinoma	18(0.3%)	
Mucinous tubular and spindle cell cancer	15(0.3%)	
Tubulocystic carcinoma	11(0.2%)	
Xp11 translocation	9(0.2%)	
Collecting duct carcinoma	6(0.1%)	
Renal medullary carcinoma	1(0%)	
Nuclear Grade		479
G1	325(6.7%)	
G2	2106(43.5%)	
G3	1825(37.7%)	
G4	586(12.1%)	
Necrosis (Yes)	1300(24.4%)	0
Sarcomatoid (Yes)	197(3.7%)	0
Lymphovascular Invasion (Yes)	198(3.9%)	285

Abbreviations:N=number of patients; ECOG=Eastern Cooperative Oncology Group, ASA=American Society of Anesthesiologists

[1]Data provided as % or as mean (SD).

[2]Analyses for cancer specific survival exclude 87 patients whose cause of death was unknown (N=5234) while analyses for recurrence free survival and overall survival include all patients with any follow-up (N=5231).

Table 5: External Validation of Recurrence free Survival Prognostic Models

Prognostic Model	Overall C-statistic	2-year Brier Score	3-year Brier Score	5-year Brier Score	N
All Histology and Risk Patients					
Yaycioglu et al, 2001	0.742	0.130	0.155	0.183	4568
Cindolo et al, 2003	0.738	0.132	0.157	0.185	4568
Zisman et al, 2002	0.606	0.115	0.129	0.143	2776
Kattan et al, 2001	0.779	0.058	0.085	0.126	4881
Yaycioglu et al, 2013	0.801	0.119	0.138	0.161	4542
Van der Mijn et al, 2019	0.771	0.117	0.135	0.160	3311
Passalacqua et al 2014	0.700	0.138	0.164	0.194	4842
Pathological T Stage	0.794	0.114	0.132	0.156	5321
Low Risk Patients (All Histology)					
Van der Mijn et at 2019	0.669	0.049	0.071	0.115	2304
Pathological T Stage	0.703	0.061	0.086	0.129	2304
High Risk Patients (All Histology)					
Correa et al 2021	0.709	0.194	0.206	0.207	2112
Pathological T Stage	0.642	0.203	0.216	0.216	2112
Clear Cell Renal Cell Carcinoma Patients					
Leibovich et al, 2003	0.826	0.115	0.133	0.145	3767
Sorbellini et al, 2005	0.790	0.127	0.148	0.169	3508

Leibovich et al, 2018-Clear Cell Renal Cell Carcinoma	0.809	0.124	0.142	0.153	3560
Pathological T Stage	0.782	0.128	0.146	0.168	3835
Papillary Renal Cell Carcinoma Patients					
Klatte et al, 2019	0.832	0.068	0.085	0.120	620
Leibovich et al, 2018-Papillary Renal Cell Carcinoma	0.759	0.082	0.108	0.144	597
Pathological T Stage	0.805	0.082	0.010	0.124	705
Chromophobe Renal Cell Carcinoma Patients					
Leibovich et al, 2018-Chromophobe Renal Cell Carcinoma	0.801	0.067	0.081	Maximum time is 4 years with event	390
Pathological T Stage	0.825	0.060	0.064	0.100	425

Abbreviation: N=number of patients

Table 6: External Validation of Cancer Specific Survival Prognostic Models

Prognostic Model	Overall C-statistic	2-year Brier Score	3-year Brier Score	5-year Brier Score	N
All Patients					
Kutikov et al, 2010	0.791	0.036	0.054	0.089	2866
Kutikov et al, 2012	0.744	0.047	0.062	0.112	996
Zisman et al, 2002	0.587	0.035	0.052	0.082	2722
Pathological T Stage	0.791	0.033	0.050	0.084	5234
Clear Cell Renal Cell Carcinoma Patients					
Frank et al, 2002	0.830	0.028	0.047	0.084	3682
Leibovich et al, 2018-Clear Cell Renal Cell Carcinoma	0.881	0.024	0.036	0.055	1796
Pathological T Stage	0.777	0.026	0.048	0.084	3766
Papillary Renal Cell Carcinoma Patients					
Leibovich et al, 2018-Papillary Renal Cell Carcinoma	0.750	0.035	0.054	0.081	574
Pathological T Stage	0.798	0.034	0.049	0.081	696

Abbreviation: N=number of patients

Table 7: External Validation of Overall Survival Prognostic Models

Prognostic Model	Overall C-statistic	2-year Brier Score	3-year Brier Score	5-year Brier Score	N
All Patients					
Zisman et al, 2001	0.752	0.045	0.072	0.116	2468

Zisman et al, 2002	0.587	0.052	0.079	0.127	2722
Passalacqua et al, 2014	0.720	0.050	0.080	0.125	4842
Pathological T Stage	0.731	0.051	0.080	0.125	5321
High Risk Patients (All Histology)					
Correa et al, 2021	0.710	0.078	0.115	0.166	2181
Pathological T Stage	0.625	0.083	0.126	0.177	2181

Abbreviation: N=number of patients

Table 8: Decision Curve Analysis Patient Risk Thresholds where Net Benefit Provided

Prognostic Model	**Patient Threshold Probability Range where Prognostic Model Yields Net Benefit**
Cancer Recurrence Free Survival Prediction Models	
All Patients	
Yaycioglu et al, 2001	0%-25% risk of recurrence
Cindolo et al, 2003	0%-24% risk of recurrence
Zisman et al, 2002	0%-31.25% risk of recurrence
Kattan et al, 2001	0%-13.5% risk of recurrence
Yaycioglu et al, 2013	0%-31.25% risk of recurrence
Van der Mijn et al, 2019	0%-25% risk of recurrence
Passalacqua et al, 2014	0%-18.75% risk of recurrence
Low Risk Patients	
Van der Mijn et al, 2019	0%-9.4% risk of recurrence
High Risk Patients	
Correa et al, 2021	0%-33.33% risk of recurrence
Clear Cell Renal Cell Carcinoma	
Leibovich et al, 2003	0%-56.25% risk of recurrence
Sorbellini et al, 2005	0%-31.25% risk of recurrence
Leibovich et al, 2018-Clear Cell Renal Cell Carcinoma	0%-36.5% risk of recurrence
Papillary Renal Cell Carcinoma	
Klatte et al, 2019	0%-68.75% risk of recurrence
Leibovich et al, 2018-Papillary Renal Cell Carcinoma	0%-49% risk of recurrence
Chromophobe Renal Cell Carcinoma	
Leibovich et al, 2018-Chromophobe Renal Cell Carcinoma	0%-73%,81.25%-100% risk of recurrence
Cancer Specific Survival Prediction Models	
All Patients	
Kutikov et al, 2010	0%-12.5% risk of cancer specific death
Kutikov et al, 2012	0%-12.5% risk of cancer specific death
Zisman et al, 2002	0%-14.0% risk of cancer specific death
Clear Cell Renal Cell Carcinoma	
Frank et al, 2002	0%-18.75% risk of cancer specific death
Leibovich et al, 2018-Clear Cell Renal Cell Carcinoma	0%-25% risk of cancer specific death
Papillary Renal Cell Carcinoma	

Leibovich et al, 2018-Papillary Renal Cell Carcinoma	0%-25% risk of cancer specific death
Overall Survival Prediction Models	
All Patients	
Zisman et al, 2001	0%-12.5% risk of death from any cause
Zisman et al, 2002	0%-15.5% risk of death from any cause
Passalacqua et al, 2014	0%-15.6% risk of death from any cause
High Risk Patients	
Correa et al, 2021	0%-21.9% risk of death from any cause

Figure 2: Recurrence Free Survival Over Time in All CKCis Patients

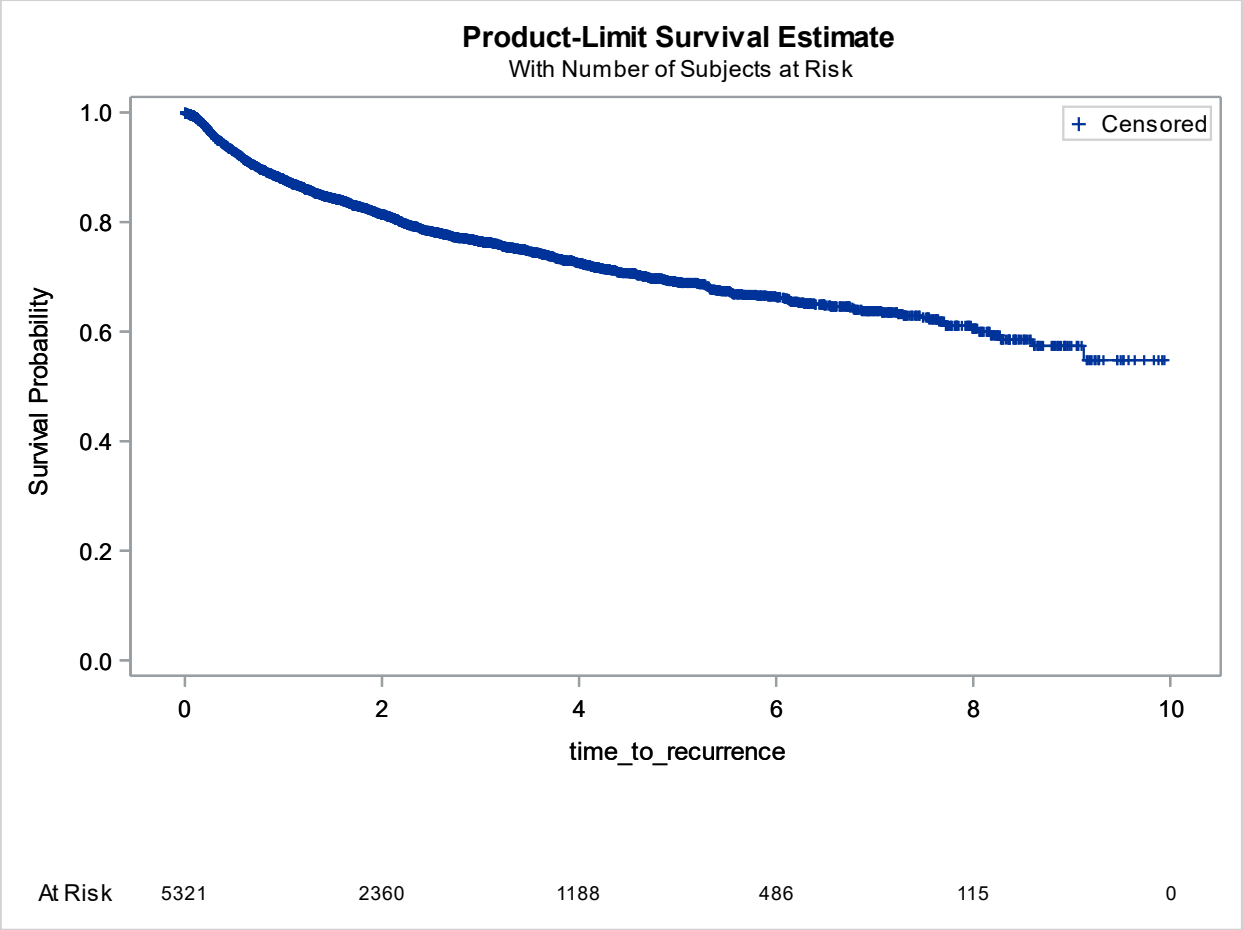
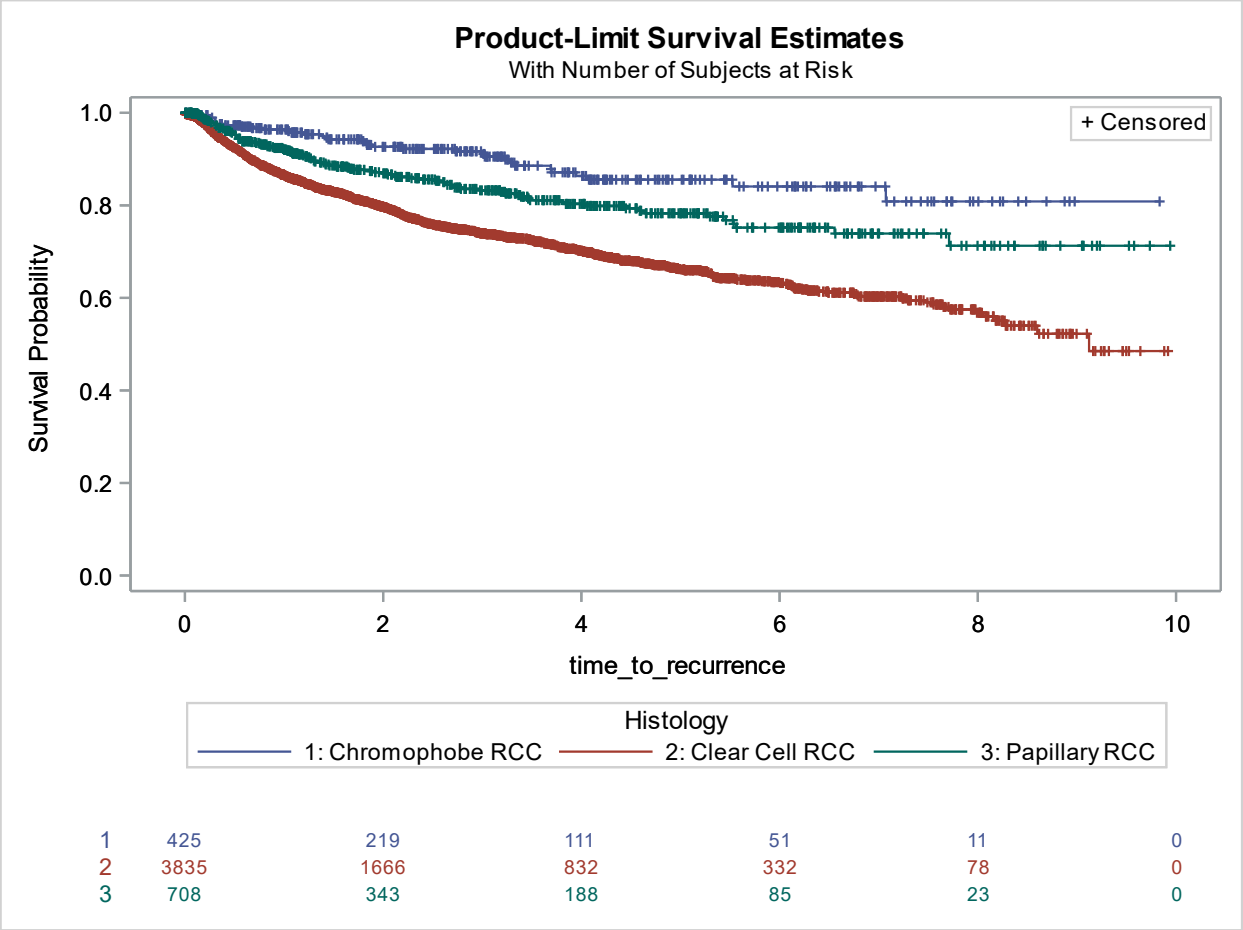
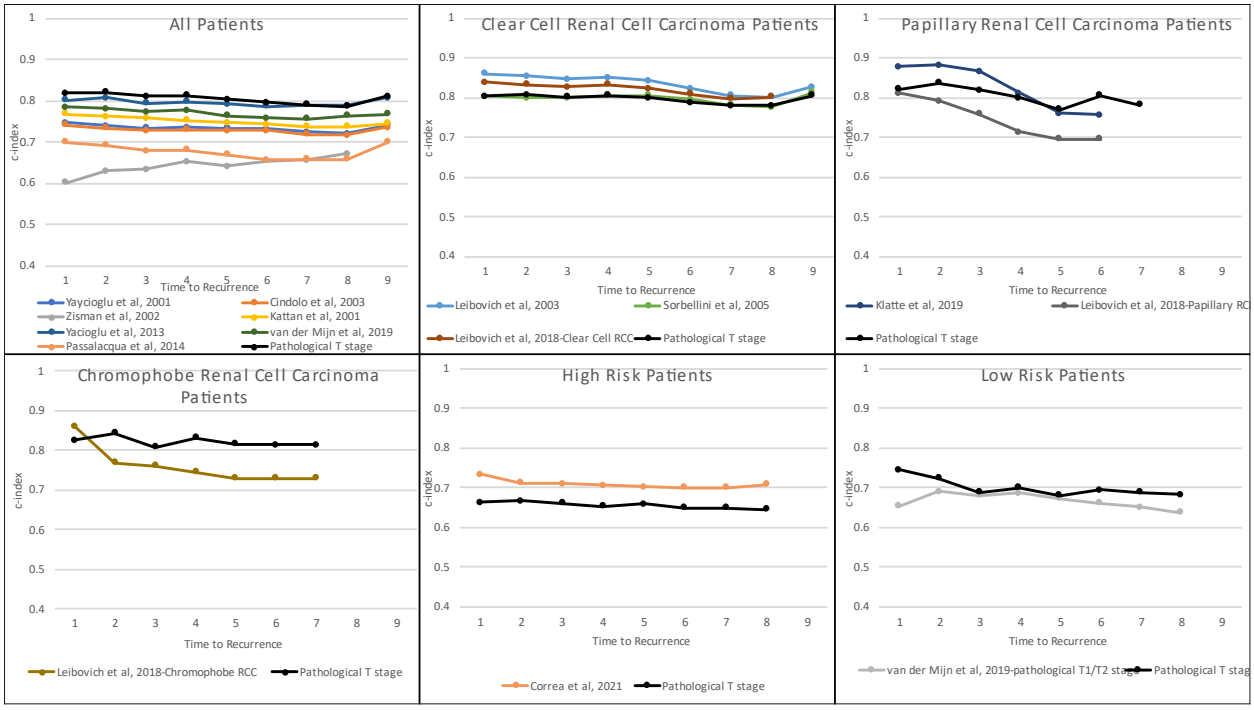


Figure 3: Recurrence Free Survival Over Time in CKCis Patients Stratified by Histological Subtype



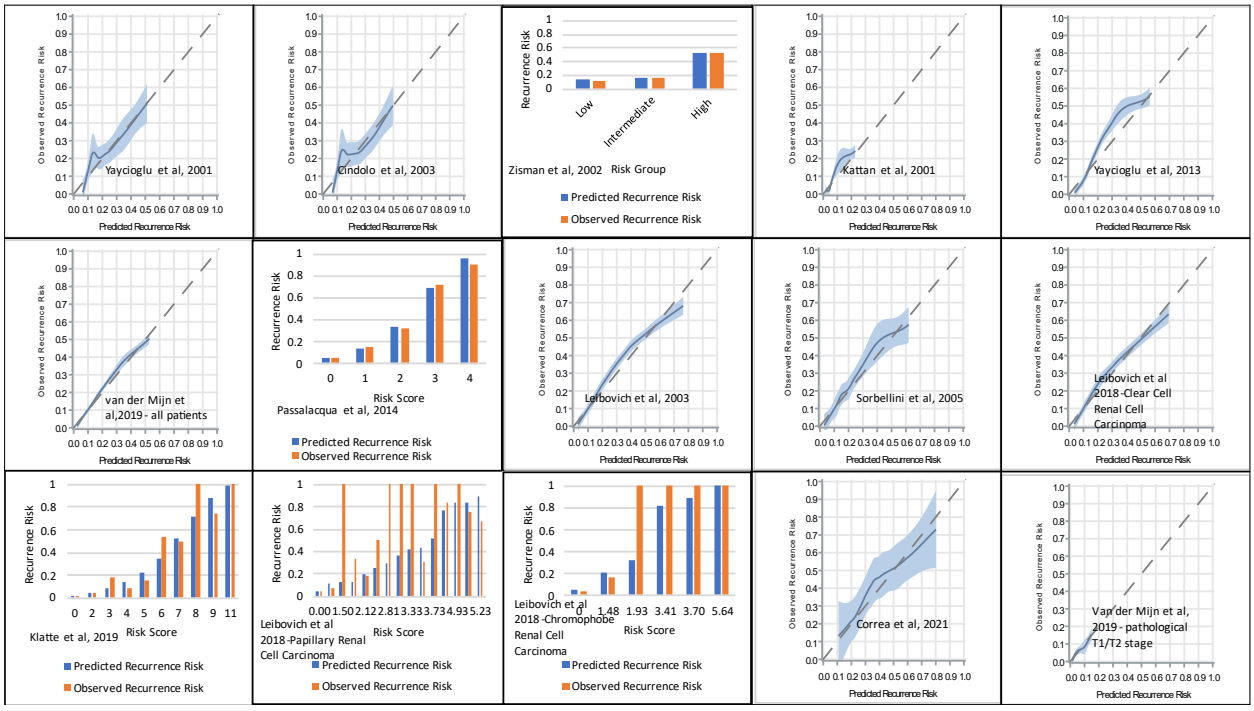
Abbreviation: RCC=Renal Cell Carcinoma

Figure 4: Estimated c-index Over Time for Recurrence Free Survival Prognostic Models



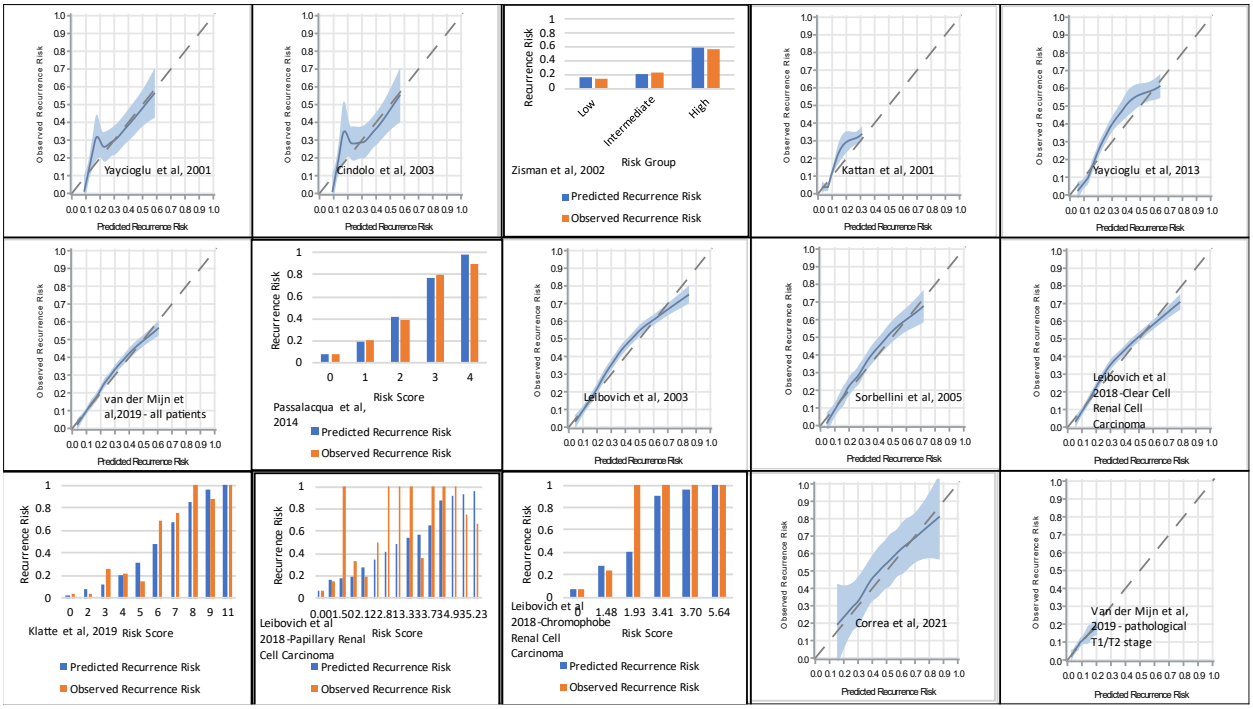
Abbreviation: RCC=Renal Cell Carcinoma

Figure 5: Calibration Curves for Recurrence Risk at 2-years post-surgery



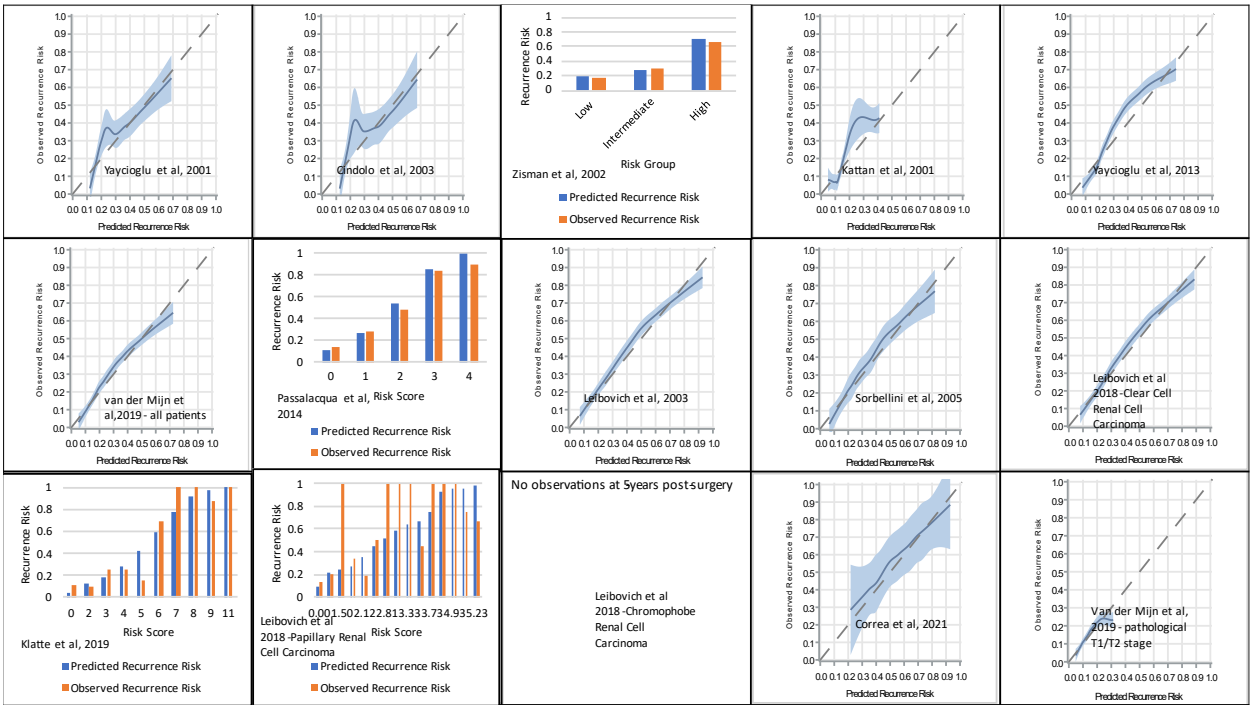
Abbreviation: RCC=Renal Cell Carcinoma

Figure 6: Calibration Curves for Recurrence Risk at 3-years post-surgery



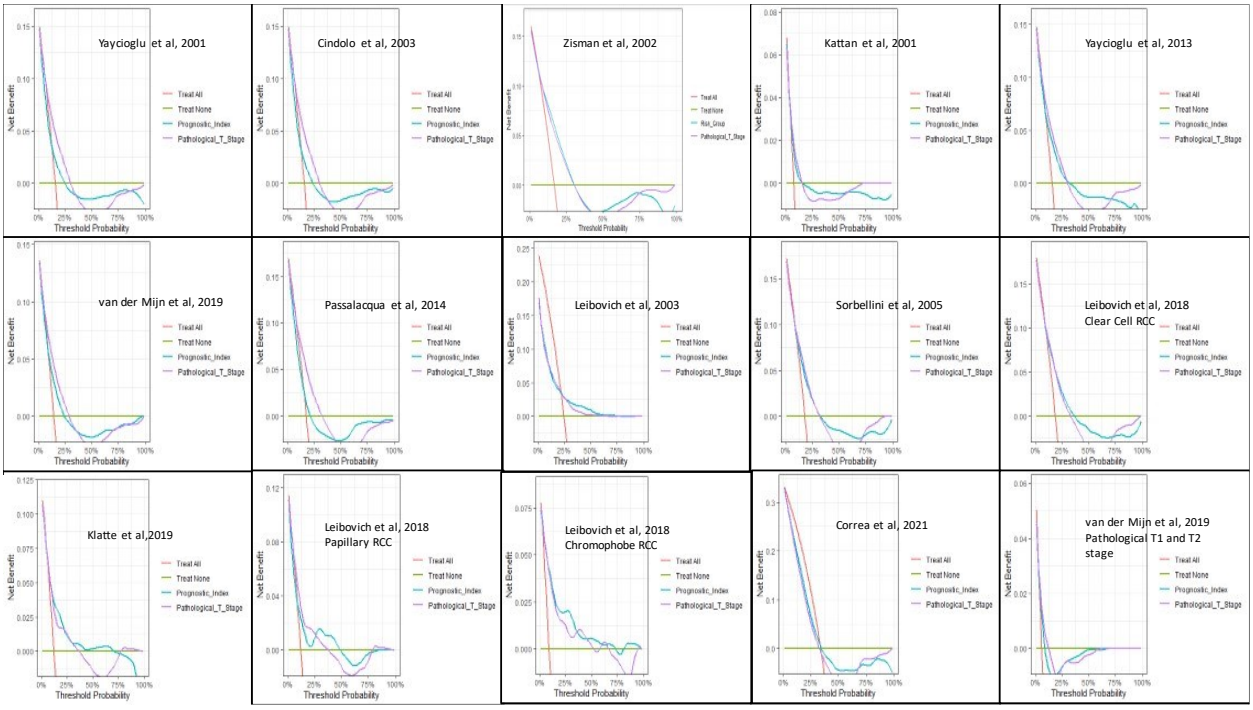
Abbreviation: RCC=Renal Cell Carcinoma

Figure 7: Calibration Curves for Recurrence Risk at 5-years post-surgery



Abbreviation: RCC=Renal Cell Carcinoma

Figure 8: Decision Curve Analysis for Recurrence Free Survival at 5-years post-Surgery



Abbreviation:RCC=Renal Cell Carcinoma

Figure 9: Cancer Specific Mortality in CKCis Patients **Figure 9:** Cancer Specific Mortality in CKCis Patients

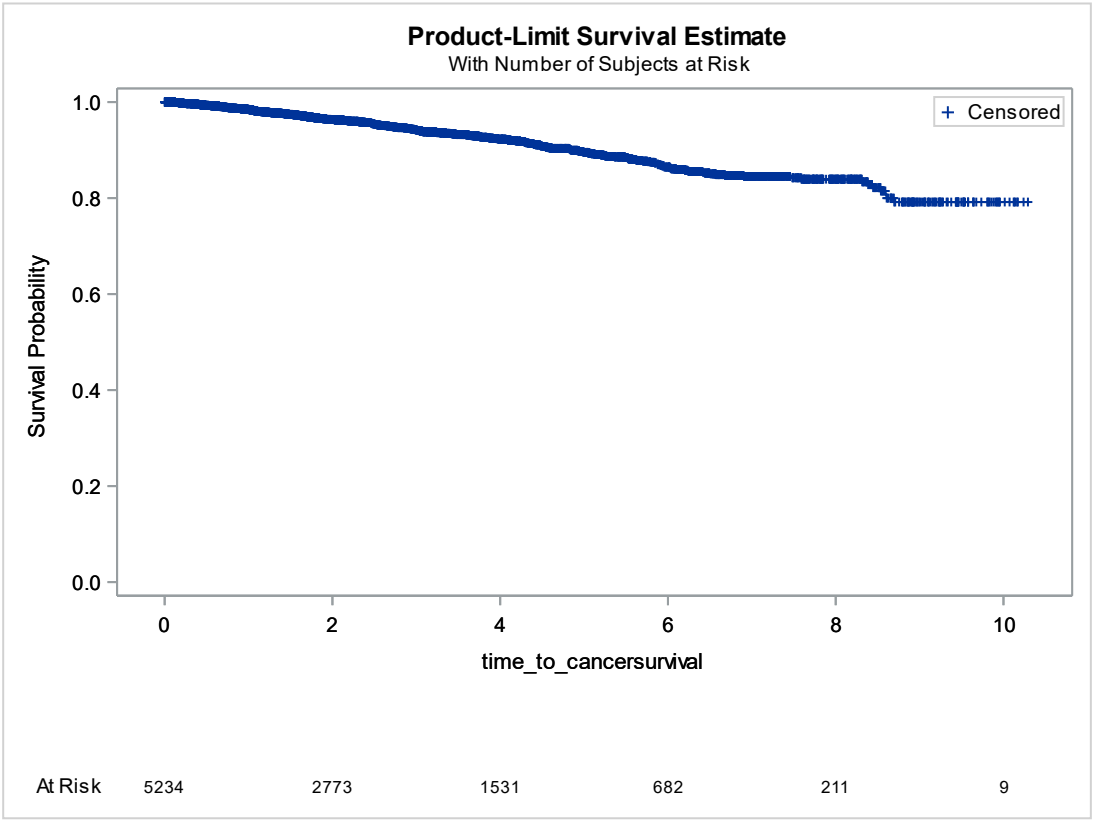
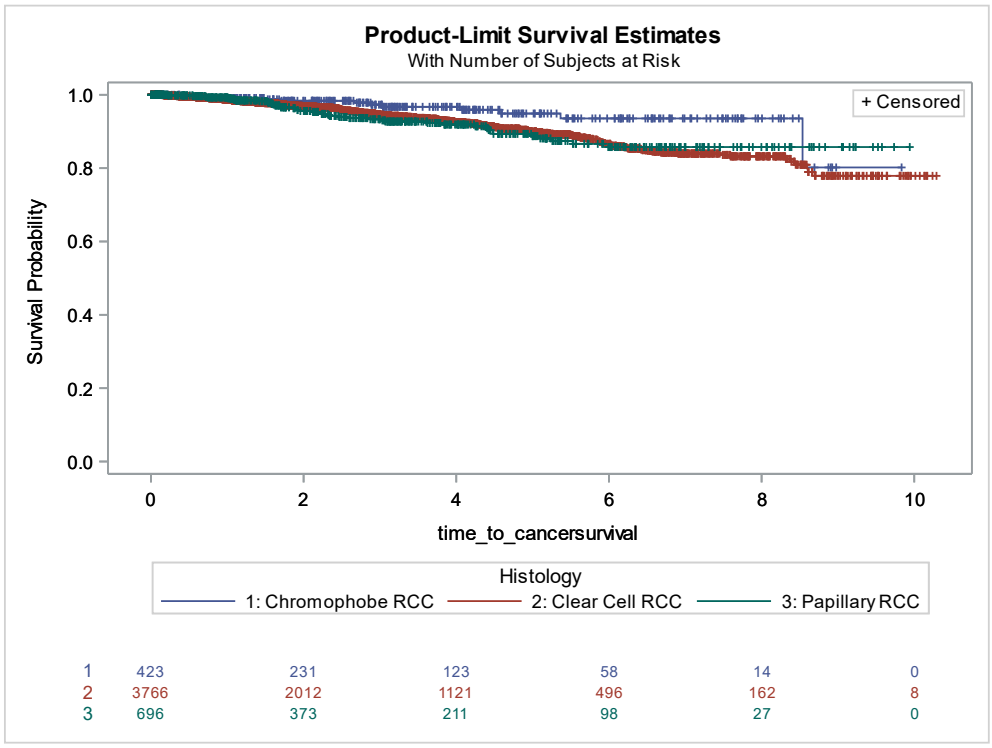
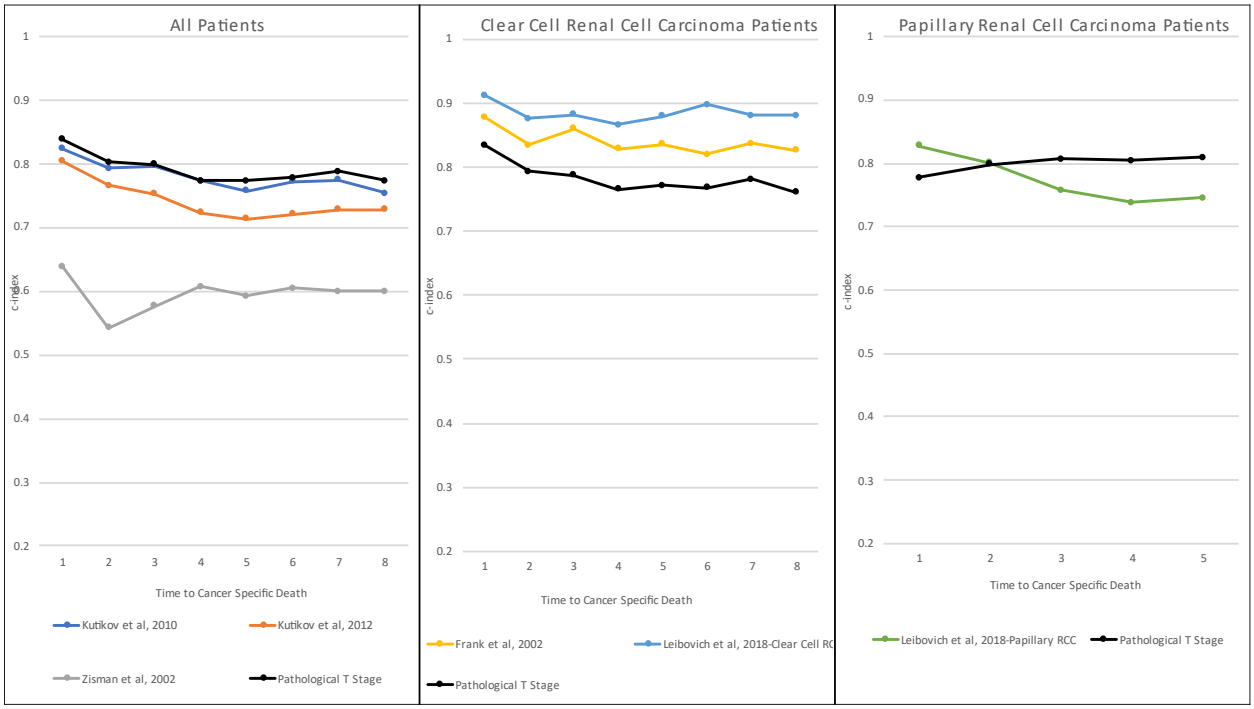


Figure 10: Cancer Specific Mortality in CKCis Patients by Histological Subtype



Abbreviation: RCC=Renal Cell Carcinoma

Figure 11: Estimated c-index Over Time for Cancer Specific Survival Prognostic Models



Abbreviation: RCC=Renal Cell Carcinoma

Figure 12: Calibration Curves for Cancer Specific Death Risk 2-years post-surgery

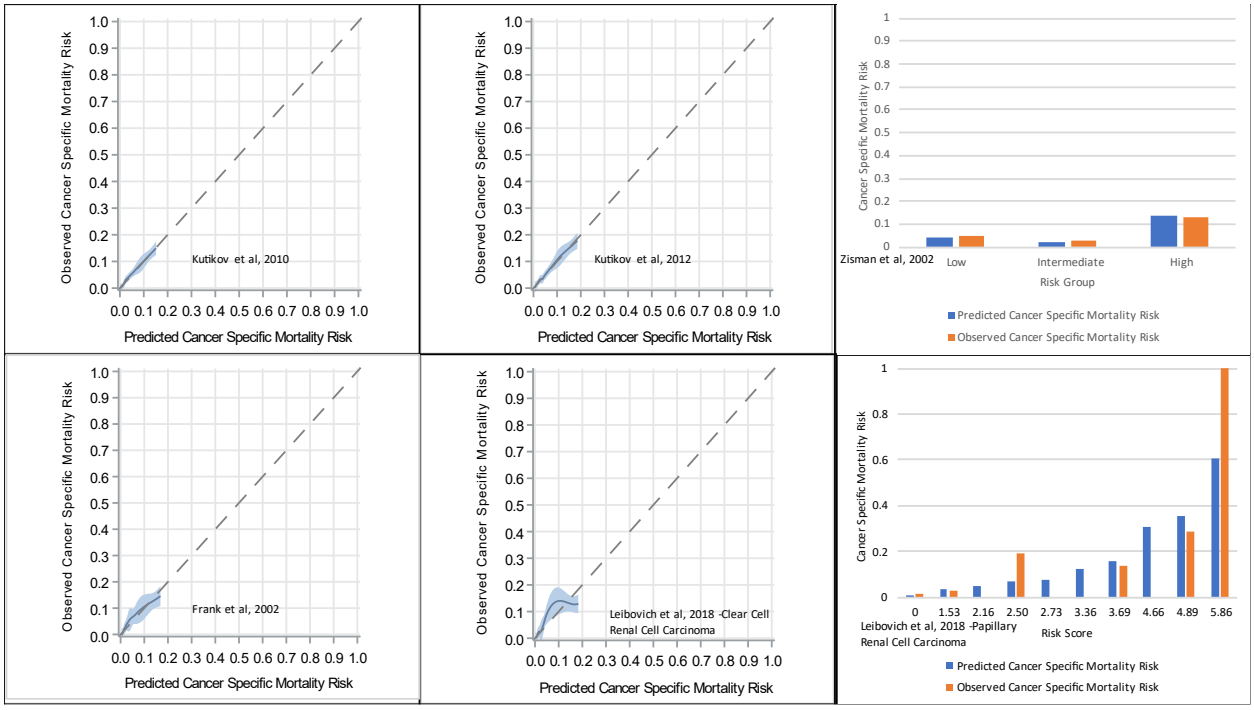


Figure 13: Calibration Curves for Cancer Specific Death Risk 3-years post-surgery

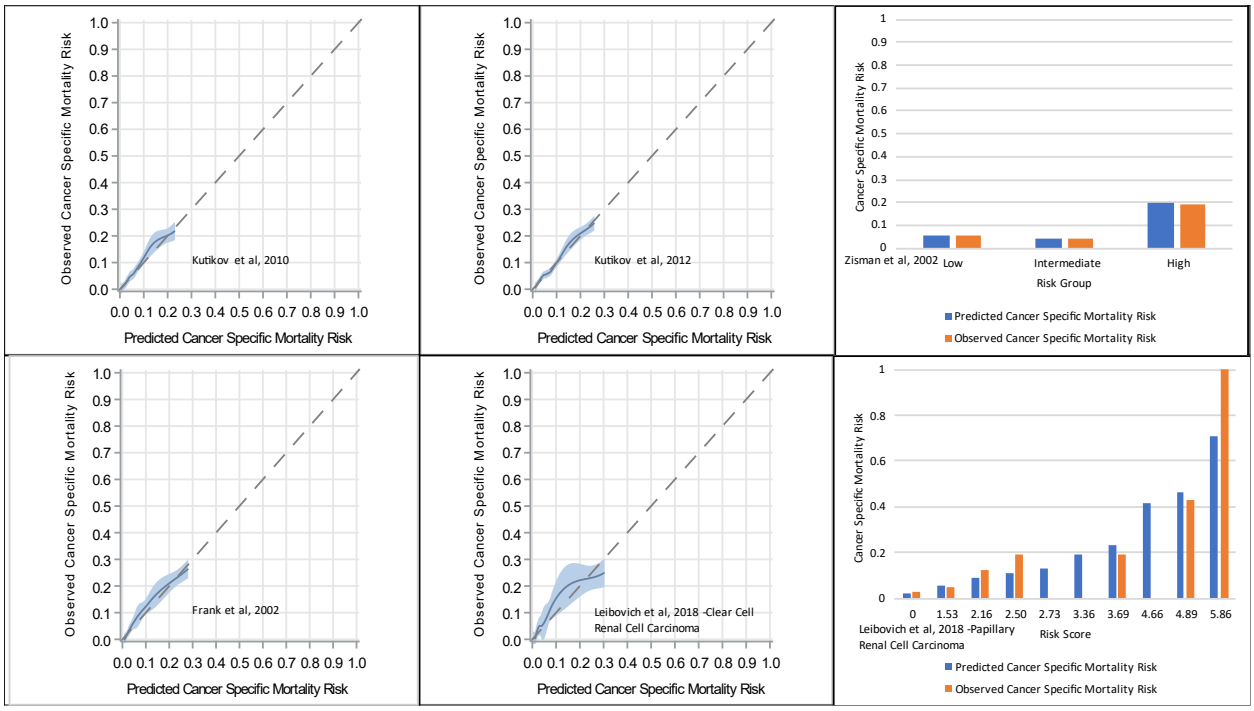


Figure 14: Calibration Curves for Cancer Specific Death Risk 5-years post-surgery

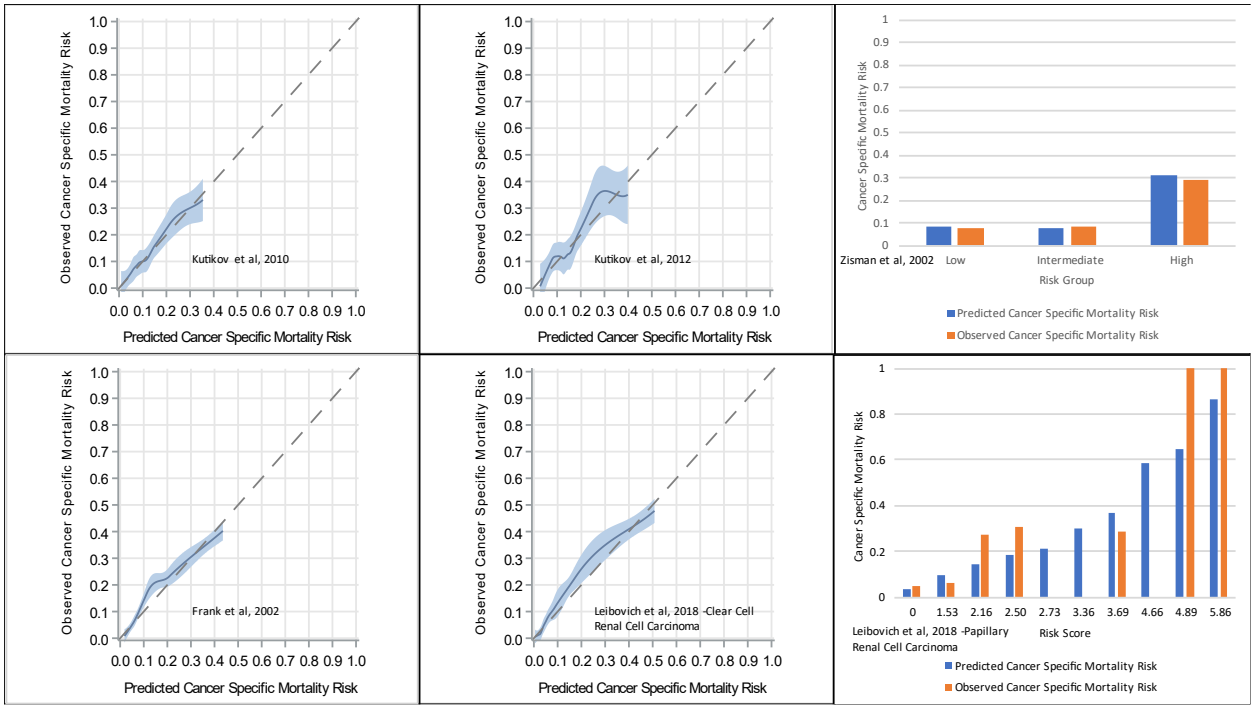
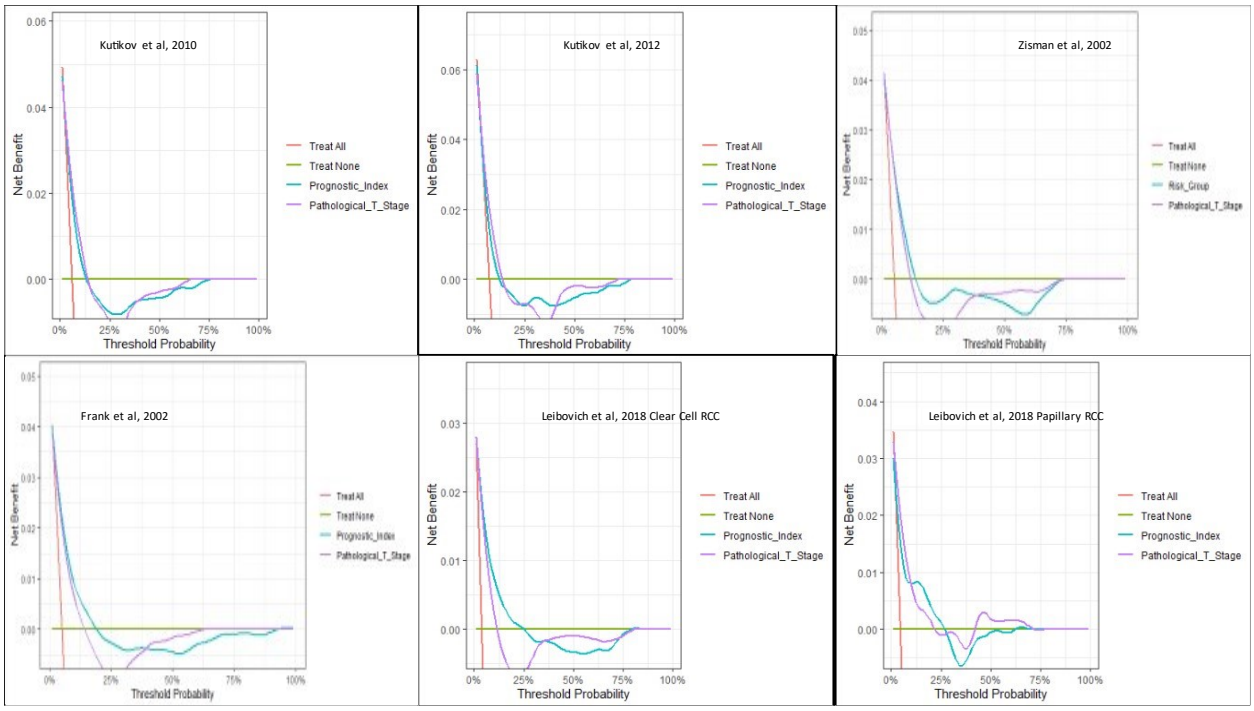


Figure 15: Decision Curve Analysis 5-years Post-Surgery Cancer Specific Survival



Abbreviation: RCC=Renal Cell Carcinoma

Figure 16: Overall Survival in CKCis Patients

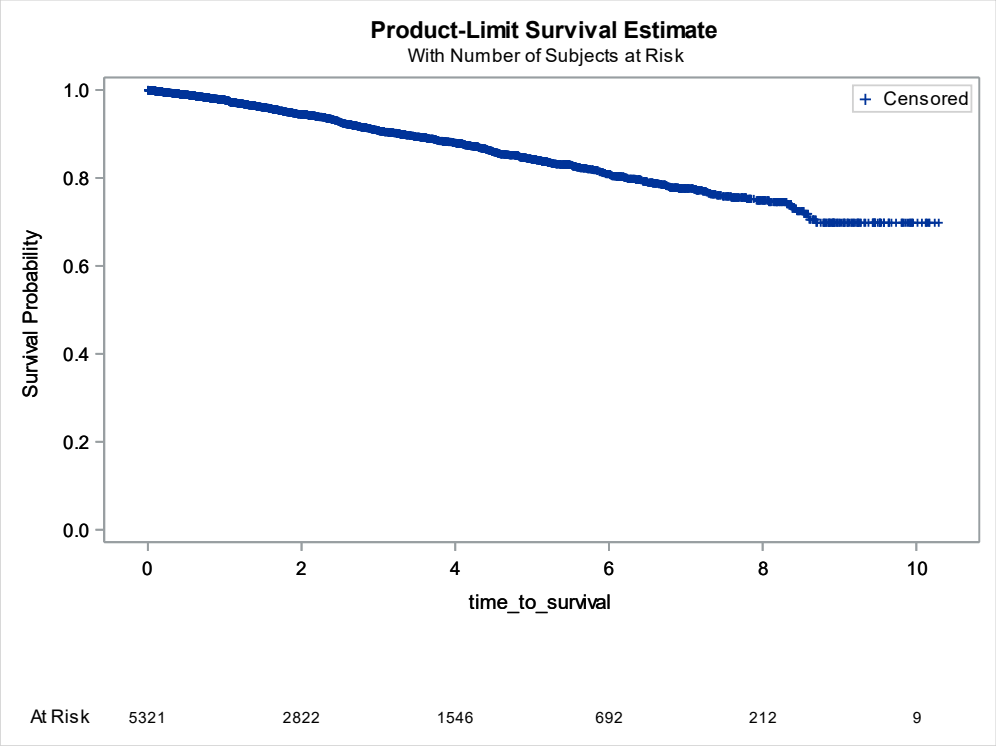
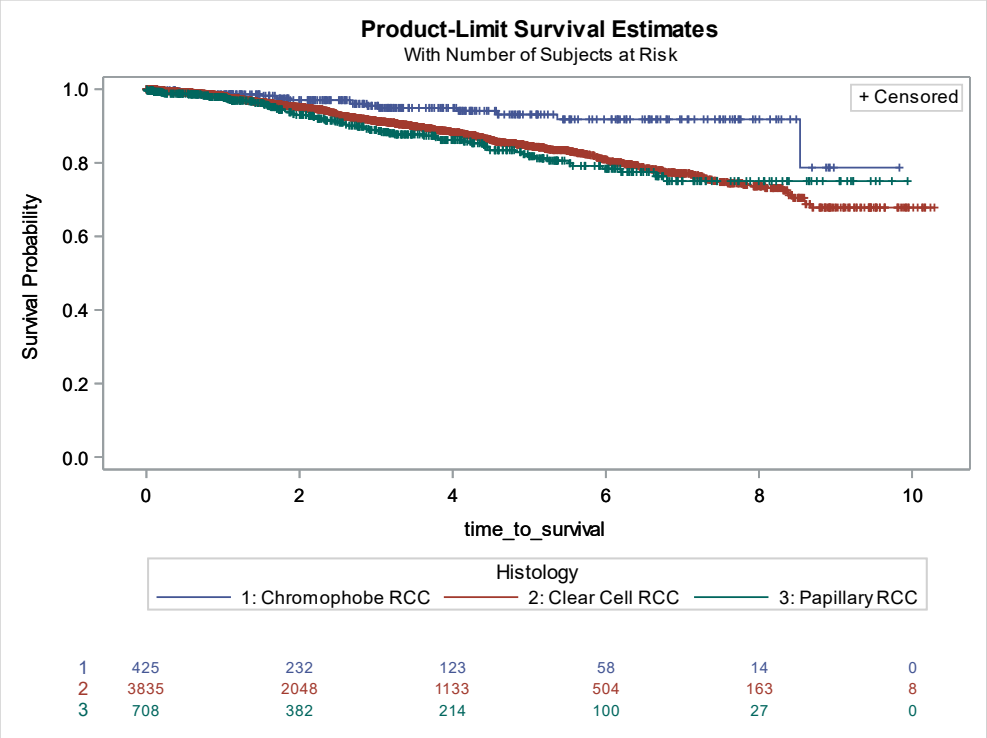


Figure 17: Overall Survival in CKCis Patients by Histological Subtype



Abbreviation: RCC=Renal Cell Carcinoma

Figure 18: Estimated c-index Over Time for Overall Survival Prognostic Model

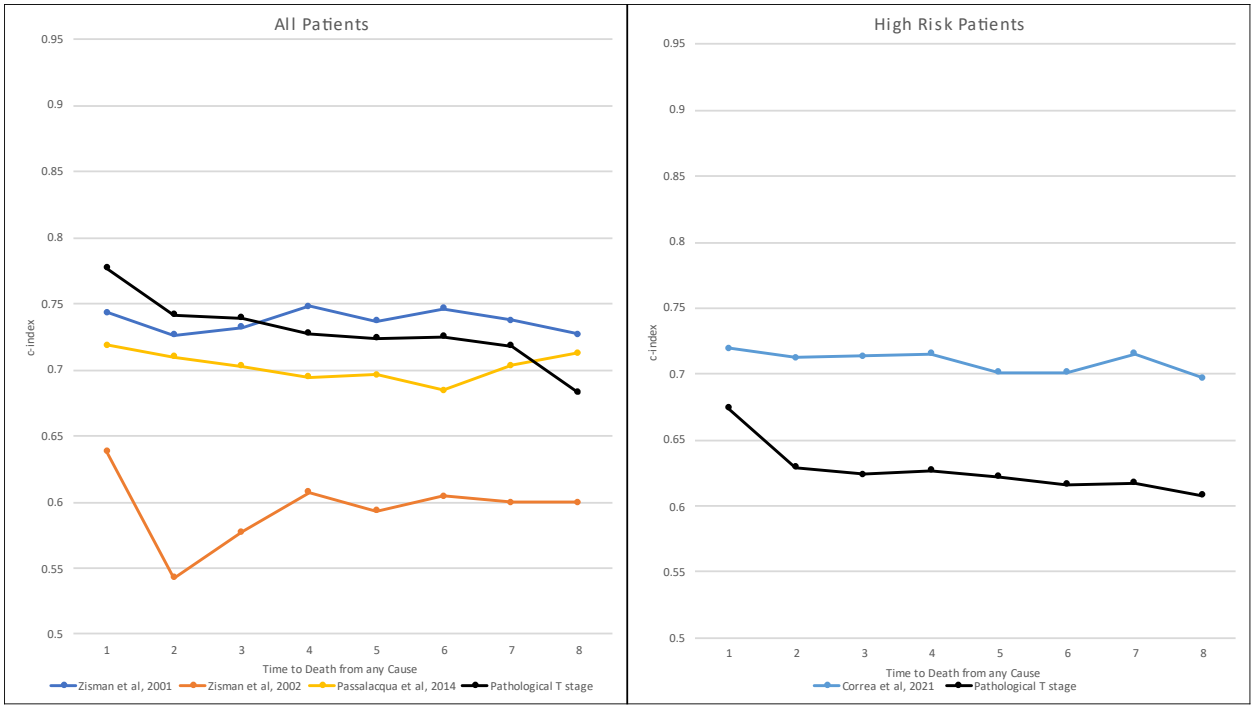


Figure 19: Calibration Curves for All-Cause Death Risk 2-years post-surgery

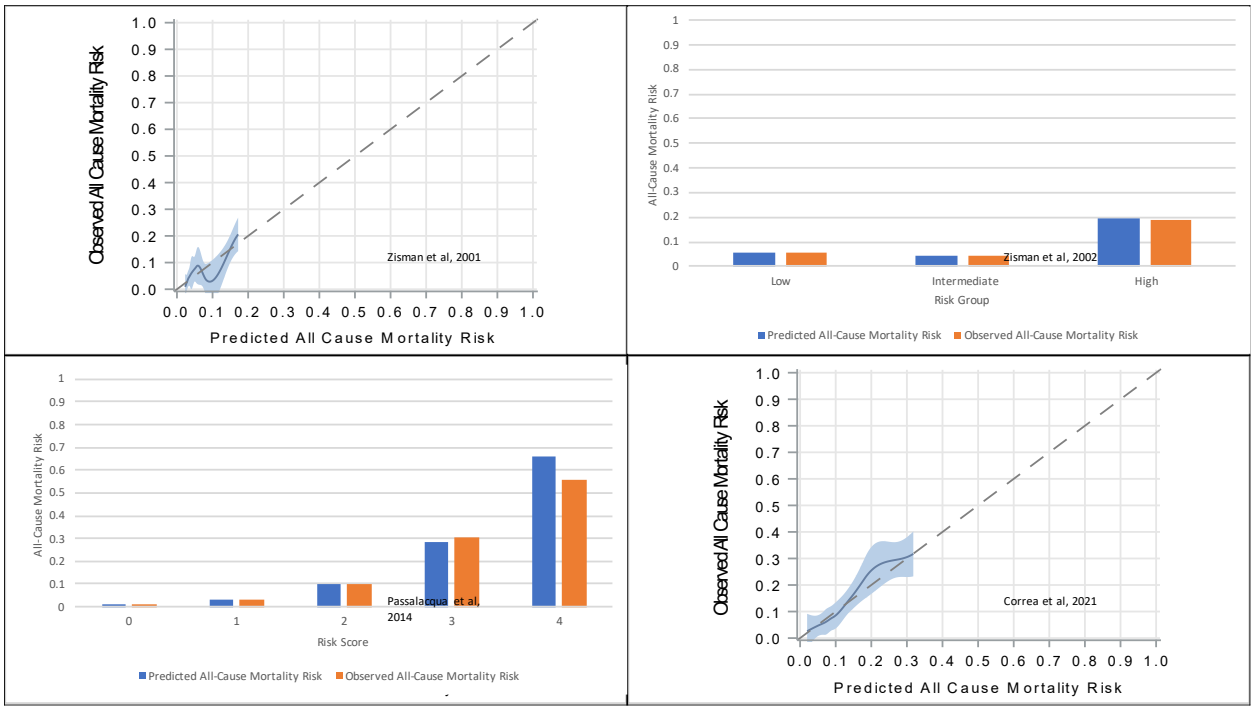


Figure 20: Calibration Curves for All-Cause Death Risk 3-years post-surgery

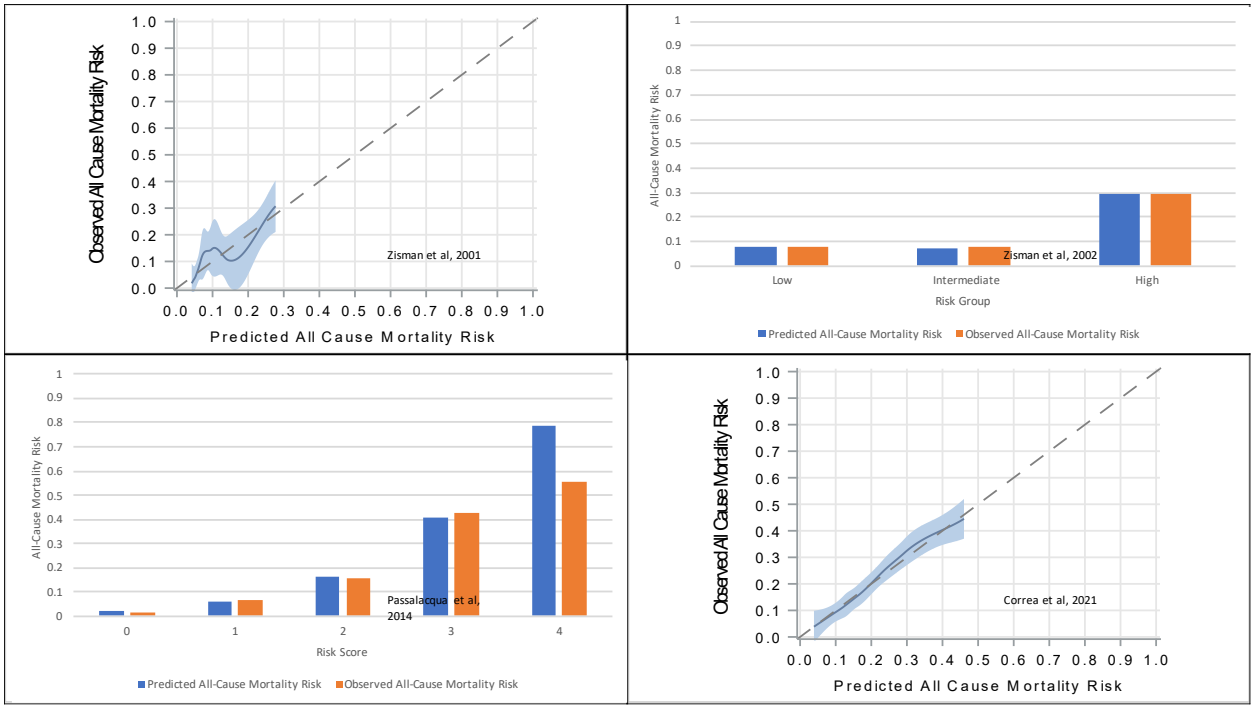


Figure 21: Calibration Curves for All-Cause Death Risk 5-years post-surgery

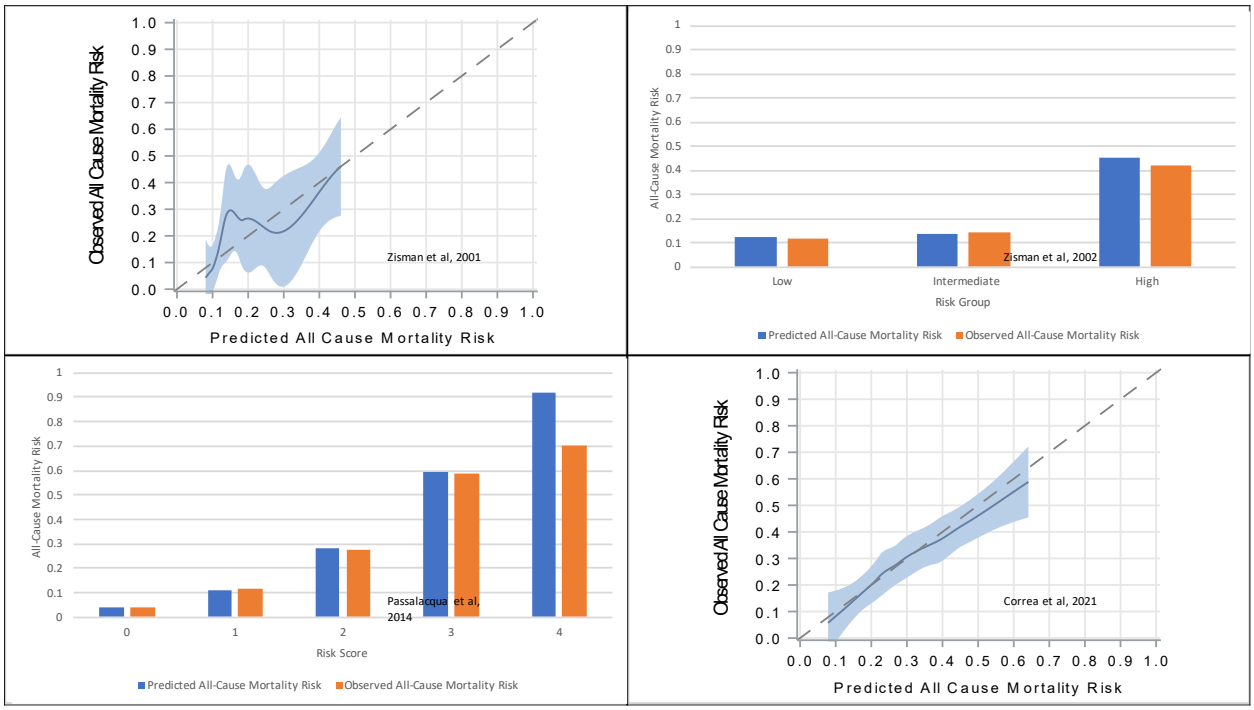
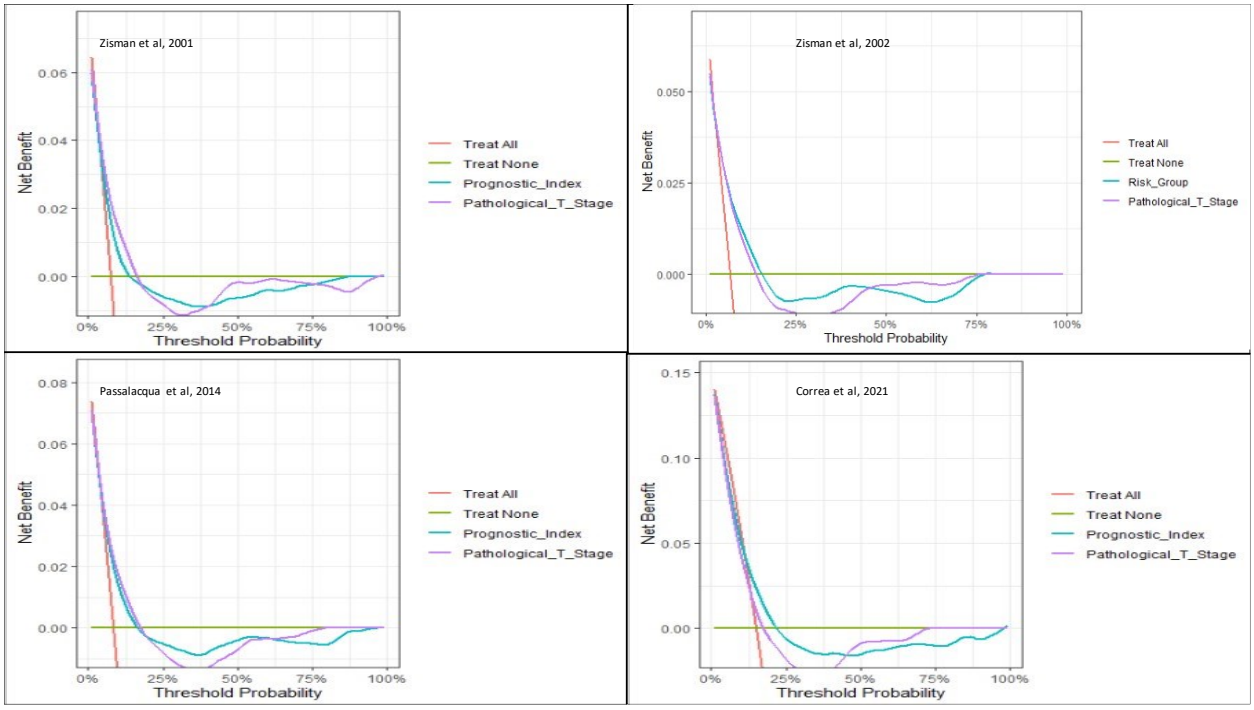


Figure 22: Decision Curve Analysis 5-years Post-Surgery All-Cause Mortality



Chapter 4: The Effect of Baseline Hazard Function on Assessment of Prognostic Model Performance

4.1 Abstract

Introduction On external validation, prognostic model inaccuracy can be due to issues with regression coefficients or baseline hazard function. The purpose of this analysis is to examine the effects of baseline hazard function variation on discrimination and calibration.

Methods Baseline hazard functions for the development dataset were simulated from Weibull, Exponential, Log-Logistic and Log-Normal distributions. The baseline hazard functions were inflated and deflated by 5%, 10%, 15% and 20% and fitted to the covariate values and coefficient values of the original development dataset to create external validation datasets. The c-statistics and calibration slopes at 5-years for the external validation dataset were assessed. To determine the effect of misspecification in distribution type, this procedure was repeated pairwise on each of the original development dataset baseline hazard functions. For real-world application, five prognostic models were applied to a cohort of Canadian kidney cancer patients and the four distributions were assumed in parametric Cox models to examine whether the choice of the best prognostic model varied based on the functional form of the baseline hazard function.

Results Variation in the baseline hazard function resulted in minor changes in the c-statistic but large changes in calibration slope. Changes in calibration slope ranged up to 3.6% for the Weibull distribution, up to 2.0% for the Exponential distribution, up to 55.2% for the Log-Logistic distribution and up to 52.8% for the Log-Normal distribution. The three methods had different efficacies for each scenario examined and adjustment for the original baseline hazard function was often required. The specific method recommended also differed based on the distribution and particular change in the baseline hazard function. From the real-world application, it was determined that the ranking of prognostic models did not change based on the functional form assumed.

Conclusions The baseline hazard function of the development and external validation datasets may be different. We have shown that misspecification of the baseline hazard function may significantly impact assessment of calibration but are unlikely to impact discrimination or selection of the best performing model. Reporting of the baseline hazard function and distribution is important to facilitate external validation.

4.2 Introduction

Prognostic research is defined as the study of the relationship between the patient's state of health at a given time-point and future clinical outcomes such as disease recurrence, disease specific-death or death from any cause (1). Prognostic modelling has numerous purposes for clinical trials and is important for patient counseling and follow-up scheduling. After development, prognostic models should be externally validated prior to wide-spread adoption to ensure generalizability.

The Cox-Proportional Hazards Model, a regression model used in survival analysis considering the covariate values of individuals, can be specified as:$h(t, X) = h_0(t)e^{\beta_1 X}$ where $h_0(t)$ is the baseline hazard function, β_1=parameter coefficient, X=covariate (2). The baseline hazard function is the instantaneous probability that an event will happen at a given point in time, without adjustment for prognostic variables, and is specified as: $h_0(t) = \lim_{\Delta t \to 0} \frac{P(t<T<t+\Delta t | T \geq t)}{\Delta t}$ (2). Key properties of the hazard function are: $h_0(t) \geq 0$, $h_0(t)$ has no maximum and any shape is possible for $h_0(t)$ (2). The cumulative hazard function can be specified as: $H(t) = \int_0^t h_0(\tau) d\tau$ (2). Survival functions, $S(t)$, and hazard functions, $H(t)$, are related to each other and this relationship can be given as: $S(t) = e^{-H(t)}$ (2). Distributions for baseline hazard functions can also be different as outlined in Table 1.

Studies deriving prognostic models often report the model without providing information on the baseline hazard. When the baseline hazard is not available, it is estimated from the validation dataset using the Breslow estimator (3).

Parametric Cox models are an alternative to Cox Proportional Hazards models as these models assume a distribution for the baseline hazard function and have been shown to have less bias (4). If all prognostic models being compared had the same baseline hazard function distribution shape, a parametric cox model could be used.

Differences between the baseline hazard functions of the population in which the prognostic model was developed and the population in which the prognostic model is validated has been theorized to be a key cause of miscalibration in prognostic model validation (5).To address issues with baseline hazard during external validation various methods have been developed. When no baseline hazard function information is available, a Cox-Proportional Hazard model with the prognostic index is fitted to the external cohort (5). When baseline hazard function information is available the following Poisson regression model is fitted: $E[y] = f(\gamma_0 + \gamma_1 p)$ where $E[y]$ is the expected number of events based on outcome status y and γ_1 is the calibration slope (5). If the full original development dataset is available, a Poisson regression is fitted

using the original dataset and re-fitted to the new dataset to adjust for the baseline hazard function (5). However, if only one baseline hazard function value for a given timepoint (i.e. value of the baseline hazard function at 5 years of follow-up), baseline hazard function values are interpolated based on the timepoint for which the baseline hazard function is given and the overall time of follow-up (5).

Since the baseline hazard is often derived in the external dataset, it is crucial to examine the effect of baseline hazard function variation on discrimination and calibration. While there are methods to adjust for miscalibration of the baseline hazard function, it is also important to assess the efficacy of these methods across different distributions to ensure that model performance is assessed accurately. The first objective was to assess the effect of baseline hazard misspecification on discrimination and calibration for different distributions. Our second objective was to examine a real world application where data from a prospective cohort of Canadian kidney cancer patients were used to examine how variation of the baseline hazard function impacted model assessment and model comparisons.

4.3 Methods

The methods will be separated into two parts. The first is based on simulations examining the effect of baseline hazard function on discrimination and calibration. The simulations will also be used to assess the efficacy of different methods of estimating calibration slope when the baseline hazard function is known. The second is a real-world application to CKCis data where we examine if model calibration varies based on the selection of the baseline hazard function. All simulations and analysis were performed in RStudio (based on R 4.0.2) and Microsoft Excel 2016.

4.3.1 Simulation Study Examining Effect of Baseline Hazard Function Calibration

Baseline hazard functions for each distribution are specified as indicated in Table 2. 500 datasets with the baseline hazard function specified with 500 observations each, maximum time-to-event (i.e. time to recurrence) of 11 years and parameter coefficients from the prognostic model for recurrence free survival in chromophobe renal cell carcinoma patients such that $\boldsymbol{\beta} = [\ln(9.28) \quad \ln(4.38) \quad \ln(6.90)]$ (6) were simulated. The prognostic index was calculated using: $Prognostic\ Index = X_1 \ln(9.28) + X_2 \ln(4.38) + X_3 \ln(6.90)$ where X_1, X_2, X_3 are values of individual covariates. Using Cox Proportional Hazards models with the prognostic index as the sole predictor, the overall c-statistic for discrimination and calibration slope at 5-years were determined for each individual simulated dataset. From this, mean c-statistics and calibration slopes and 95% confidence intervals were calculated for the model development dataset.

New baseline hazard functions were specified with deviations from the original baseline hazard function as stated in Table 3 for Weibull distribution, Table 4 for Exponential distribution, Table 5 for Log-Logistic

distribution, Table 6 for Log-Normal distribution, and Table 7 for combinations of the different distributions studied. The following steps were followed for each of the 500 model development datasets derived in the previous part. The values of covariates X_1, X_2, X_3 were stored in a data frame and a new mis-specified dataset was simulated with the following parameters: 500 observations, maximum time-to-event of 11-years, covariate parameter coefficients of $\boldsymbol{\beta} = [\ln(9.28) \quad \ln(4.38) \quad \ln(6.90)]$ and the mis-specified baseline hazard function. The prognostic index was calculated using $Prognostic\ Index = X_1 \ln(9.28) + X_2 \ln(4.38) + X_3 \ln(6.90)$ where X_1, X_2, X_3 are values of individual covariates. Cox Proportional Hazards models with the prognostic index as the sole predictor were used to determine the overall model c-statistic in the new dataset with the mis-specified baseline hazard function.

Cox Proportional Hazards models were also used to determine the calibration slope at 5-years for the new dataset to simulate the effect of not adjusting for the original baseline hazard function. Poisson regression models using event status as the outcome such that: $E[y] = f(\gamma_0 + \gamma_1 p)$ where γ_1 is the calibration slope were fitted to show the effect of adjusting for the original baseline hazard function (5). The effect of adjustment for the original baseline hazard function given the original dataset was examined by fitting the original dataset's baseline hazard function to the new dataset using Poisson regression (5). The effect of adjusting for the original baseline hazard function given its value at a given time point (i.e. value of baseline hazard function at 5-years) were examined by interpolating the value of the baseline hazard function based on the given value for all other time points (5). R code used for both Poisson regression models have been previously published (5). From the 500 new datasets, the mean c-statistic, calibration slope at 5-years and 95% confidence intervals for both parameters were obtained. From this, recommendations were made regarding the 'best method' used to calculate the calibration slope at 5-years. The 'best method' is defined as the method that produces results closest to the original development calibration slope without over-estimating the calibration slope provided that the external validation calibration slope calculated does not understate the original development calibration slope by more than 0.1. If all three methods over-estimate the calibration slope, the 'best method' is the method that has the least deviation from the original calibration slope.

4.3.2 Real-World Application Using CKCis Data

Data used in the real-world analysis originated from Canadian Kidney Cancer Information System (CKCis), a Canadian prospective cohort of kidney cancer patients. Variables within the CKCis dataset were reparametrized as needed based on pre-existing models. The outcome was recurrence free survival at 5-

years post-surgery, with time to recurrence defined as the time between surgery and kidney cancer recurrence.

The following pre-existing prognostic models for recurrence free survival in all patients were examined: Kattan et al 2001 (7), Cindolo et al 2003 (8), Yaycioglu et al 2013 (9), Yaycioglu et al 2001 (10) and van der Mijn et al 2019 (11). Prognostic indexes were constructed as specified for all models except for van der Mijn et al 2019 where a surrogate prognostic index based on the addition of parameter coefficients derived from the multivariate hazard ratios provided was used as a point system was not specified (11). Parametric Cox models were fitted assuming the Weibull distribution for each prognostic model and the ranking of the prognostic models in terms of calibration slopes from best (closest to 1) to worst (furthest from 1) was examined. This was repeated assuming the exponential, log-logistic and log-normal distributions. The calibration slopes and the ranking of the prognostic models were examined to determine the effect of specifying different baseline hazard functions on external model assessment.

4.4 Results

4.4.1 Part 1: Simulation Study Examining Effect of Baseline Hazard Function Calibration

Weibull Distribution

The simulated Weibull distribution baseline survival functions are shown in Figure 1. Figures 2 and 3 demonstrate the effect of various rate and shape parameter miscalibrations on overall c-statistic and calibration slopes at 5-years. The first case examined is where there are changes in both the rate and shape parameters. The c-statistics calculated are slight overestimates where both parameters are overestimated, with the estimated c-statistic increasing with larger increases in both parameters and slight underestimates where both parameters are underestimated, with the estimated c-statistic decreasing with larger decreases in both parameters. Overall, estimating the calibration slope at 5-years without adjusting for the baseline hazard function provides the most liberal estimates, while adjustment of the baseline hazard function using the value of the original baseline function at 5-years is the most conservative. When both parameters are underestimated, larger decreases in both parameters are associated with larger underestimates of the calibration slope at 5-years. Hence, when both parameters are underestimated, calculating the calibration slope at 5-years without adjusting for the baseline hazard function is the best method. When both parameters are overestimated, estimates using both adjustment using the full dataset and without adjustment for the baseline hazard function fluctuate. Calibration slopes estimated by adjusting for the original baseline hazard function value at 5-years decrease between

0%-5% overestimate in both parameters and increase from 5%-20% overestimate. Thus, when both parameters are overestimated, adjustment using the full dataset is the best method.

The second case examined is where the rate parameter changed while the shape parameter remained constant. C-statistics, while still close the original value, were underestimated when the rate parameter was underestimated with larger underestimation associated with larger decreases in the rate parameter. When the rate parameter is overestimated, the C-statistic decreased between 0%-5% overestimate in rate and increased between 5%-20% overestimate in rate. The results in terms of the most liberal and conservative methods and trends in estimated 5-year calibration slope across increases/decreases in rate parameter were like the results from the first case. Thus, not adjusting for the baseline hazard function is the best method for underestimates in rate and adjustment using the full original dataset was the best method when the rate is overestimated, except at 20% overestimate in rate at which adjusting for the value of the original baseline hazard function at 5-years is the best method.

The final case studied using the Weibull distribution is where the shape parameter changed while the rate parameter is constant. When the shape parameter is underestimated, the c-statistic is also underestimated with the estimated 5-year calibration slope decreasing as the shape parameter becomes smaller. As the shape parameter increases, the c-statistic decreased from 0%-10% overestimation of the shape parameter and increased from 10%-20% overestimation of shape parameter. Like the two other cases, adjustment using the value of the original baseline hazard function at 5-years resulted in the most conservative estimates of 5-year calibration slopes. Underestimating the shape parameter resulted in constant calibration slopes for both estimates adjusting for the baseline hazard function using the original dataset and without adjusting for the baseline hazard function. As the calibration slopes derived without adjusting for the baseline hazard function were slight overestimates of the original calibration slope, it is recommended that the baseline hazard function be adjusted using the original dataset. Overestimating the shape parameter resulted in constant underestimates using both methods adjusting for the original dataset and without adjusting for the baseline hazard function. As the shape parameter increased the calibration slope decreased when computed adjusting for the value of the baseline hazard function at 5-years. Thus, it is recommended that the baseline hazard function is not adjusted for in this case.

Exponential Distribution

The simulated Exponential distribution baseline survival functions are shown in Figure 4. Figures 5 and 6 demonstrate the effect of various rate and scale parameter miscalibrations on overall c-statistic and calibration slopes at 5-years. The first case examined is where there were changes in both the rate and

scale parameters. Underestimates in both parameters were associated with underestimates in c-statistics with smaller c-statistics as both parameters decreased and overestimates in both parameters were associated with overestimates in c-statistics with larger c-statistics as both parameters increased. Overall, adjusting for the value of the original baseline hazard function at 5-years led to the most conservative results. For all three methods, smaller rates and scales were associated with smaller calibration slopes and larger rates and scales were associated with larger calibration slopes. Thus, not adjusting for the baseline hazard function is the best method. Not adjusting for the baseline hazard function caused slight overestimates of the calibration slopes when both rate and scale were overestimated. Adjusting for the value of the baseline hazard function at 5-years when both rate and scale were overestimated resulted in decreased estimated calibration slopes between 0%-5% overestimate in rate and scale and increased calibration slopes between 5%-20% overestimate in rate and scale. Therefore, adjusting for the baseline hazard function using the full original dataset was the best method when both rate and scale were overestimated by up to 15% and adjusting for the baseline hazard function using the original value at 5-years was the best method between 15-20% overestimates in rate and scale.

The second case studied was where the rate changed but the scale was fixed. Underestimating the rate resulted in fluctuating but constant c-statistics close to the original c-statistic and overestimating the rate resulted in overestimating the c-statistic. The calibration slopes at 5-years had similar results to the first case under the Exponential distribution when the rate was under- or over-estimated while the scale was fixed. Thus, when the rate was underestimated with the scale parameter fixed, the best method was to not adjust for the original baseline hazard function. When the rate was overestimated with the scale parameter fixed, adjustment using the original dataset was best for 0-10% increases in the rate and adjustment using the value of the original baseline hazard function at 5-years was best for 15%-20% increases in the rate.

The third case studied for the Exponential distribution was where there was a change in scale, but the rate was fixed. Underestimating the scale parameter resulted in constant but fluctuating c-statistics and overestimating the scale parameter resulted in overestimations of the c-statistic. For both underestimation and overestimation of the scale, the results were like the results obtained for the case where there were changes in shape while the rate is constant under the Weibull distribution. Therefore, adjustment for the original baseline hazard function using the full dataset is recommended when the scale is underestimated and not adjusting for the baseline hazard function is recommended when the scale is overestimated.

Log-Logistic Distribution

The simulated Log-Logistic distribution baseline survival functions are shown in Figure 7. Figures 8 and 9 demonstrate the effect of various rate and scale parameter miscalibrations on overall c-statistic and calibration slopes at 5-years. The first case examined was where both rate and shape changed. When both parameters were underestimated, the c-statistic increased from 0%-5% underestimate and decreased from 5%-20% underestimation of both rate and shape. Overestimating both parameters led to an increase in the c-statistic from 0%-5% overestimation in both parameters and a constant but fluctuating c-statistic between 0%-20% overestimations in both parameters. Regardless, the estimated c-statistic was overestimated in this case. All three methods overestimated the calibration slope with adjustment using the full original dataset being the most liberal and adjustment using the value of the original baseline hazard function at 5-years being the most conservative. Underestimating both parameters had led to increases in calibration slope between 0%-5% underestimation, with the calibration slope decreasing between 5%-20% underestimate using the adjustment for the full original dataset and value at 5-years methods and constant calibration slope when the baseline hazard function was not adjusted for. Overestimating both parameters led to an increase between 0%-5% deviation, with the calibration slope remaining constant when there was no adjustment for the baseline hazard function and adjustment using the full dataset between 5%-20% overestimations and the calibration slope calculated by adjusting for the value of the original baseline hazard function at 5-years between 5%-20% slowly increasing. Furthermore, it is recommended that the calibration slope be calculated by adjusting for the value of the original baseline hazard function at 5-years.

The second case examined was when the rate parameter changed but the shape parameter remained constant. C-statistics were overestimated across all studied deviations in shape parameter with increases in the c-statistic from 0% to ±5% and remained constant afterward with minor fluctuations. While all three methods overestimated the calibration slope at 5-years, not adjusting for the baseline hazard function provided the most liberal estimates of the calibration slope. When the rate is underestimated, all three methods had similar results with increases in the estimated calibration slope until 5% decrease in the rate and decreases in the estimated calibration slope afterward. When the rate is overestimated, all three methods increase until a 5% overstatement in rate. The calibration slope remained constant when there was no adjustment for the baseline hazard function and when there was adjustment using the full dataset; the estimated calibration slope increased when adjustment using the value of the original baseline hazard

function at 5-years was performed. Hence, it is recommended that adjustment using the value of the baseline hazard function at 5-years is performed.

The final case examined was when the shape parameter changed but the rate parameter remained constant. When examining the effects of deviations in the shape parameter on c-statistics, the results are like when the rate parameter changed while keeping the shape parameter constant. Like the second case, all three methods overestimate the calibration slopes at 5-years and not adjusting for the original baseline hazard function provided the most liberal results. Underestimates in the shape parameter led to increase in calibration slopes until 5% overestimate of the shape parameter in all methods, the estimated calibration slope remained constant for both methods except for adjustment for the value of the original baseline hazard function at 5-years, which increased as the shape parameter increased. Overestimates in the shape parameter led to increased calibration slope until 5% overestimate in shape and remained constant afterwards for all three methods. As such, adjustment for the value of the original baseline hazard function at 5-years provided the best estimates of the calibration slopes.

Log-Normal Distribution

The simulated Log-Normal distribution baseline survival functions are shown in Figure 10. Figures 11 and 12 demonstrate the effect of various rate and scale parameter miscalibrations on overall c-statistic and calibration slopes at 5-years. The first case studied is when mean and standard deviation both changed. When both parameters were underestimated, the c-statistic estimated decreased until 5% underestimate and increased afterward. Meanwhile, overestimates in both parameters resulted in decreased estimate of c-statistics. 5-year calibration slope estimates were most liberal when the baseline hazard function was not adjusted and most conservative when adjustment was performed using the original baseline hazard function at 5-years. Underestimates in both parameters are associated with increased calibration slope until 5% deviation and decrease afterwards when no adjustment is performed and when adjustment using the full original dataset is performed. The estimated 5-year calibration slope decreased as the understatement in both parameters increased. Overestimates in both parameters resulted in increased calibration slope until 5% when there was no adjustment and when there was adjustment for the full dataset. However, the calibration slopes remained constant when there was no adjustment and increased when there was adjustment using the full dataset. Yet, when the baseline hazard function was adjusted using the value of the original baseline hazard function at 5-years, the estimated calibration slope decreased until 5% change and increased afterward. For this case, adjustment using the full dataset is recommended.

The second case studied was where the mean changed but the standard deviation remained fixed. C-statistics were underestimated when the mean was underestimated with a decrease until 5% and increased afterward. When the mean was overestimated, the estimated c-statistic decreased until 5% and remained constant with minor fluctuations. Estimated calibration slopes at 5-years were overestimated using both methods except for the adjustment using the value of the original baseline hazard function at 5-years which was underestimated. Underestimation in the mean resulted in increased estimated calibration slopes both when the baseline hazard function was not adjusted and when it was adjusted using the full original dataset with decreased calibration slopes afterward. The estimated calibration slope when adjustment was performed using the value of the original baseline hazard function at 5-years decreased. Overestimation in the mean resulted in results like when both the mean and standard deviation changed. Thus, adjustment using the full dataset is recommended.

The final case study using the Log-Normal distribution was where the standard deviation changed but the mean remained constant. Estimates of the c-statistic follow trends analogous to those observed when both the mean and standard deviation changed. Not adjusting for the baseline hazard function continued to provide the most liberal estimates of calibration slope and adjustment using the value of the calibration slope at 5-years remained the most conservative in terms of calibration slopes at 5-years. Underestimation of the standard deviation yielded results like the first two cases under the Log-Normal distribution. However, overestimation of the standard deviation resulted in results like the first case except for adjustment using the value of the original baseline hazard function at 5-years where the calibration slope decreased until 5% and remained constant afterward. Overall, adjustment using the full dataset is recommended.

Changes in Types of Distribution

The effect of different distribution types in development datasets and external validation datasets were studied by comparing c-statistics and calibration slopes at 5-years among all three methods . Table 7 lists all distribution parameters used in these simulations. The overall results are given in Figure 13 with the original c-statistics and 5-year calibration slopes from the development datasets highlighted in grey and the best method for each pairwise comparison highlighted in yellow. While differences in the types of distributions resulted in slight under and overestimates in the c-statistics, the c-statistic is quite robust despite deviations in the baseline hazard function. Wider variation was observed in calibration slopes at 5-years with and without adjustment for the baseline hazard function using the value of the original baseline hazard function at 5-years. However, these results should be interpreted with caution as all shape

parameters were equal to 3 and all rate parameters were equal to 5 in these simulations for all distributions.

4.4.2 Part 2: Real-World Application Using CKCis Data

While the calibration slope at 5-years for each model is constant or varies slightly based on the distribution assumed, the ranking of the prognostic model remains constant regardless of the distribution used(Table 8).

4.5 Discussion

Many prognostic models are published without providing information on the baseline hazard or distribution. In this paper we show that the c-statistic is robust to changes in the baseline hazard function but calibration slope is not. While all distributions reported considerably different calibration slopes between the original and mis-specified baseline hazard functions, the differences were most profound in Log-Logistic and Log-Normal distributions. The calibration slope at 5-years estimates were most robust when the differences were between the type of distribution (i.e. development dataset had a baseline hazard function with a Weibull distribution and the external validation dataset had a baseline hazard function with an Exponential distribution). From the real-world application, it was evident that while small differences existed in the estimated calibration slope based on the assumed baseline hazard function, the ranking of the prognostic models using parametric Cox models remained constant. Hence, differences in baseline hazard function do not prevent the comparison of prognostic models using calibration-based techniques.

As limited research on the effect of baseline hazard function misspecification has been performed to date, it is difficult to directly compare the results of this study to prior research. However, previous research in interval censored data has shown that deviations in the baseline hazard function resulted in bias in inferential properties of the model, which suggested a significant effect (12). In case-control studies, differences in distributions have led to differences in estimates of the baseline hazard function (13). Differences in baseline hazard function estimates would then lead to differences in estimated survival probabilities, which ultimately would result in differences in the estimated calibration slope. As the Cox Proportional Hazards model imputes the baseline hazard function of observations in the validation dataset to calculate survival probabilities, it is plausible that not adjusting for the baseline hazard function would yield the most optimistic and liberal estimates of calibration slope at 5-years (2).

While misspecification of the baseline hazard function may significantly impact assessment of calibration, it is unlikely to impact discrimination or selection of the best performing model. Therefore, reporting of

the baseline hazard function is important for calibration, however, external validation without the baseline hazard function is still informative. Simulation studies similar to this one should be performed to identify the best method of estimating prognostic model calibration prior to external validation of the model.

4.5.1 Limitations

While this study is the first study examining the effects of misspecifications in baseline hazard functions on c-statistics and calibration slopes at 5-years in point censored survival data, there are important limitations that must be considered. As only one set of **β** parameters were considered, variation in effects across different combinations of **β** parameters was not examined. The comparison of different distribution types used the same rate and shape (scale for exponential distribution) in the distributions studied, which would mask potential differences in c-statistic and calibration slope that would be more visible had different rate and shape parameters been studied. This study also examined 5%-20% deviation in the baseline hazard functions across different parameters, which would ignore any possible change in trends at larger deviations (i.e. whether it is possible that estimates of c-statistic and calibration slope start to improve after a sufficiently large deviation). None of the models examined as part of the real-world application included information on the baseline hazard function. Thus, deviations between the baseline hazard functions assumed as part of the external assessment and the original baseline hazard function could not be quantified. As such, it could not be determined which of the four distributions assumed is the most accurate representation of the original baseline hazard function.

4.5.2 Strengths

Four commonly used distributions were examined and the impact of differences in the baseline hazard function were observed across a wide range of scenarios. Since the **β** parameters and the covariate values were held constant when fitting the mis-specified baseline hazard function, all deviations in calibration slopes at 5-years and c-statistics observed were mostly due to misspecification in the baseline hazard function (random error persists but this would be a small proportion of the variation). Three existing methods of calculating the calibration slope at 5-years were compared, allowing for the examination of each method under a wide range of situations to eventually select the best method. The real-world application examined differences in calibration slope at 5-years post-surgery observed across different distributions.

4.6 Conclusion

Overall, baseline hazard function misspecification effects calibration. Baseline hazard function misspecification may be an important source of model miscalibration and must be considered during external assessment of prognostic models. We recommend that authors provide information regarding the baseline hazard function when publishing prognostic models. Directions for future research include larger simulation studies using different sets of **β** parameters, wider range of deviations from the original baseline hazard function and more distribution types with the goal of developing overarching recommendations on how to adjust for baseline hazard function misspecification when externally validating prognostic models.

4.7 References

1. Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A, Briggs A, Udumyan R, Moons KG, Steyerberg EW, Roberts I, Schroter S, Altman DG, Riley RD; PROGRESS Group. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. BMJ. 2013 Feb 5;346:e5595. doi: 10.1136/bmj.e5595.
2. Emmert-Streib F, Dehmer M. Introduction to Survival Analysis in Practice. Mach. Learn. Knowl. Extr. 2019 Sep 8; 1(3):1013-1038. doi: 10.3390/make1030058.
3. Xia F, Ning J, Huang X. Empirical Comparison of the Breslow Estimator and the Kalbfleisch Prentice Estimator for Survival Functions. J Biom Biostat. 2018;9(2):392. doi: 10.4172/2155-6180.1000392.
4. Zare A, Mahmoodi M, Mohammad K, Zeraati H, Hosseini M, Naieni KH. Comparison between parametric and semi-parametric cox models in modeling transition rates of a multi-state model: application in patients with gastric cancer undergoing surgery at the Iran cancer institute. Asian Pac J Cancer Prev. 2014 Jan;14(11):6751-5. doi: 10.7314/apjcp.2013.14.11.6751.
5. Crowson CS, Atkinson EJ, Therneau TM. Assessing calibration of prognostic risk scores. Stat Methods Med Res. 2016 Aug;25(4):1692-1706. doi: 10.1177/0962280213497434.
6. Leibovich BC, Lohse CM, Cheville JC, Zaid HB, Boorjian SA, Frank I, Thompson RH, Parker WP. Predicting Oncologic Outcomes in Renal Cell Carcinoma After Surgery. Eur Urol. 2018 May;73(5):772-780. doi: 10.1016/j.eururo.2018.01.005.
7. Kattan MW, Reuter V, Motzer RJ, Katz J, Russo P. A postoperative prognostic nomogram for renal cell carcinoma. J Urol. 2001 Jul;166(1):63-67.
8. Cindolo L, de la Taille A, Messina G, Romis L, Abbou CC, Altieri V, Rodriguez A, Patard JJ. A preoperative clinical prognostic model for non-metastatic renal cell carcinoma. BJU Int. 2003 Dec;92(9):901-905. doi: 10.1111/j.1464-410x.2003.04505.x.
9. Yaycioglu O, Eskicorapci S, Karabulut E, Soyupak B, Gogus C, Divrik T, Turkeri L, Yazici S, Ozen H; Society of Urooncology Study Group for Kidney Cancer Prognosis. A preoperative prognostic model predicting recurrence-free survival for patients with kidney cancer. Jpn J Clin Oncol. 2013 Jan;43(1):63-68. doi: 10.1093/jjco/hys192.
10. Yaycioglu O, Roberts WW, Chan T, Epstein JI, Marshall FF, Kavoussi LR. Prognostic assessment of nonmetastatic renal cell carcinoma: a clinically based model. Urology. 2001 Aug;58(2):141-145. doi: 10.1016/s0090-4295(01)01207-9.

11. van der Mijn JC, Al Hussein Al Awamlh B, Islam Khan A, Posada-Calderon L, Oromendia C, Fainberg J, Alshak M, Elahjji R, Pierce H, Taylor B, Gudas LJ, Nanus DM, Molina AM, Del Pizzo J, Scherr DS. Validation of risk factors for recurrence of renal cell carcinoma: Results from a large single-institution series. PLoS One. 2019 Dec 9;14(12):e0226285. doi: 10.1371/journal.pone.0226285.
12. Withana Gamage PW, Chaudari M, McMahan CS, Kim EH, Kosorok MR. An extended proportional hazards model for interval-censored data subject to instantaneous failures. Lifetime Data Anal. 2020 Jan;26(1):158-182. doi: 10.1007/s10985-019-09467-z.
13. Hsu L, Gorfine M, Zucker DM. On Estimation of the Hazard Function from Population-based Case-Control Studies. J Am Stat Assoc. 2018;113(522):560-570. doi: 10.1080/01621459.2017.1356315.

4.8 Figures and Tables

Table 1: Hazard Functions and Parameters of Commonly Observed Distributions in Survival Analysis

Distribution	**Hazard Function**	**Parameters**
Weibull Distribution (2)	$h_0(t) = \lambda p(\lambda p)^{p-1}$	$\lambda > 0$ rate parameter, p >0 shape parameter
Exponential Distribution (2)	$h_0(t) = \lambda$	$\lambda > 0$ rate parameter
Log-Logistic Distribution (2)	$h_0(t) = \lambda\alpha(\lambda t)^{\alpha-1}(1 + (\lambda t)^{\alpha})^{-1}$	$\lambda > 0$ rate parameter, $\alpha > 0$ shape parameter
Log-Normal Distribution (2)	$h_0(t) = \frac{\alpha}{\sqrt{2\pi}t} e^{\left(\frac{-\alpha^2(\ln(\lambda t))^2}{2}\right)}\left(1 - \phi(\alpha \ln(\lambda t))\right)^{-1}$	Mean $\mu = -\ln\lambda$, standard deviation $\sigma = \alpha^{-1}$

Abbreviations: t =time; p=shape parameter; λ =rate parameter; α=shape parameter; μ=mean; σ=standard deviation

Table 2: Baseline Hazard Functions Specified in Simulations for Model Development Dataset[1]

Distribution	**Parameters**
Weibull Distribution	$\lambda = 5$ rate parameter, p=3 shape parameter
Exponential Distribution	$\lambda = 5$ rate parameter, p=3 scale parameter
Log-Logistic Distribution	$\lambda = 5$ rate parameter, p=3 shape parameter
Log-Normal Distribution	Mean $\mu = 5$, standard deviation $\sigma = 1.5$

Abbreviations: p=shape parameter; λ =rate parameter; μ=mean; σ=standard deviation

[1]R Studio requires the specification of the scale parameters in addition to the rate parameter to simulate the Exponential distribution

Table 3: Mis-specified Weibull Distributed Baseline Hazard Functions Specified in Simulations for the External Validation Dataset[1]

Scenario	**Baseline Hazard Function Parameters**
20% reduction in all parameters	$\lambda = 4$ rate parameter, p=4 shape parameter
15% reduction in all parameters	$\lambda = 4.25$ rate parameter, p=4.25 shape parameter

10% reduction in all parameters	$\lambda = 4.5$ rate parameter, p=4.5 shape parameter
5% reduction in all parameters	$\lambda = 4.75$ rate parameter, p=4.75 shape parameter
5% inflation in all parameters	$\lambda = 5.25$ rate parameter, p=5.25 shape parameter
10% inflation in all parameters	$\lambda = 5.5$ rate parameter, p=5.5 shape parameter
15% inflation in all parameters	$\lambda = 5.75$ rate parameter, p=5.75 shape parameter
20% inflation in all parameters	$\lambda = 6$ rate parameter, p=6 shape parameter
20% reduction in rate, all others held constant	$\lambda = 4$ rate parameter, p=5 shape parameter
15% reduction in rate, all others held constant	$\lambda = 4.25$ rate parameter, p=5 shape parameter
10% reduction in rate, all others held constant	$\lambda = 4.5$ rate parameter, p=5 shape parameter
5% reduction in rate, all others held constant	$\lambda = 4.75$ rate parameter, p=5 shape parameter
5% inflation in rate, all others held constant	$\lambda = 5.25$ rate parameter, p=5 shape parameter
10% inflation in rate, all others held constant	$\lambda = 5.5$ rate parameter, p=5 shape parameter
15% inflation in rate, all others held constant	$\lambda = 5.75$ rate parameter, p=5 shape parameter
20% inflation in rate, all others held constant	$\lambda = 6$ rate parameter, p=5 shape parameter
20% reduction in shape parameter, all others held constant	$\lambda = 5$ rate parameter, p=4 shape parameter
15% reduction in shape parameter, all others held constant	$\lambda = 5$ rate parameter, p=4 shape parameter
10% reduction in shape parameter, all others held constant	$\lambda = 5$ rate parameter, p=4.5 shape parameter
5% reduction in shape parameter, all others held constant	$\lambda = 5$ rate parameter, p=4.75 shape parameter
5% inflation in shape parameter, all others held constant	$\lambda = 5$ rate parameter, p=5.25 shape parameter
10% inflation in shape parameter, all others held constant	$\lambda = 5$ rate parameter, p=5.5 shape parameter
15% inflation in shape parameter, all others held constant	$\lambda = 5$ rate parameter, p=5.75 shape parameter
20% inflation in shape parameter, all others held constant	$\lambda = 5$ rate parameter, p=6 shape parameter

Abbreviations: p=shape parameter; λ=rate parameter

[1]R Studio requires the specification of the scale parameters in addition to the rate parameter to simulate the Exponential distribution

Table 4: Mis-specified Exponential Distributed Baseline Hazard Functions Specified in Simulations for the External Validation Dataset[1]

Scenario	**Baseline Hazard Function Parameters**
20% reduction in all parameters	$\lambda = 4$ rate parameter, p=4 scale parameter
15% reduction in all parameters	$\lambda = 4.25$ rate parameter, p=4.25 scale parameter
10% reduction in all parameters	$\lambda = 4.5$ rate parameter, p=4.5 scale parameter
5% reduction in all parameters	$\lambda = 4.75$ rate parameter, p=4.75 scale parameter
5% inflation in all parameters	$\lambda = 5.25$ rate parameter, p=5.25 scale parameter
10% inflation in all parameters	$\lambda = 5.5$ rate parameter, p=5.5 scale parameter
15% inflation in all parameters	$\lambda = 5.75$ rate parameter, p=5.75 scale parameter
20% inflation in all parameters	$\lambda = 6$ rate parameter, p=6 scale parameter
20% reduction in rate, all others held constant	$\lambda = 4$ rate parameter, p=5 scale parameter
15% reduction in rate, all others held constant	$\lambda = 4.25$ rate parameter, p=5 scale parameter
10% reduction in rate, all others held constant	$\lambda = 4.5$ rate parameter, p=5 scale parameter
5% reduction in rate, all others held constant	$\lambda = 4.75$ rate parameter, p=5 scale parameter
5% inflation in rate, all others held constant	$\lambda = 5.25$ rate parameter, p=5 scale parameter
10% inflation in rate, all others held constant	$\lambda = 5.5$ rate parameter, p=5 scale parameter
15% inflation in rate, all others held constant	$\lambda = 5.75$ rate parameter, p=5 scale parameter
20% inflation in rate, all others held constant	$\lambda = 6$ rate parameter, p=5 scale parameter
20% reduction in scale parameter, all others held constant	$\lambda = 5$ rate parameter, p=4 scale parameter

15% reduction in scale parameter, all others held constant	$\lambda = 5$ rate parameter, p=4 scale parameter
10% reduction in scale parameter, all others held constant	$\lambda = 5$ rate parameter, p=4.5 scale parameter
5% reduction in scale parameter, all others held constant	$\lambda = 5$ rate parameter, p=4.75 scale parameter
5% inflation in scale parameter, all others held constant	$\lambda = 5$ rate parameter, p=5.25 scale parameter
10% inflation in scale parameter, all others held constant	$\lambda = 5$ rate parameter, p=5.5 scale parameter
15% inflation in scale parameter, all others held constant	$\lambda = 5$ rate parameter, p=5.75 scale parameter
20% inflation in scale parameter, all others held constant	$\lambda = 5$ rate parameter, p=6 scale parameter

Abbreviations: p=shape parameter; λ=rate parameter

[1]R Studio requires the specification of the scale parameters in addition to the rate parameter to simulate the Exponential distribution

Table 5: Mis-specified Log-Logistic Distributed Baseline Hazard Functions Specified in Simulations for the External Validation Dataset[1]

Scenario	**Baseline Hazard Function Parameters**
20% reduction in all parameters	$\lambda = 4$ rate parameter, p=4 shape parameter
15% reduction in all parameters	$\lambda = 4.25$ rate parameter, p=4.25 shape parameter
10% reduction in all parameters	$\lambda = 4.5$ rate parameter, p=4.5 shape parameter
5% reduction in all parameters	$\lambda = 4.75$ rate parameter, p=4.75 shape parameter
5% inflation in all parameters	$\lambda = 5.25$ rate parameter, p=5.25 shape parameter
10% inflation in all parameters	$\lambda = 5.5$ rate parameter, p=5.5 shape parameter
15% inflation in all parameters	$\lambda = 5.75$ rate parameter, p=5.75 shape parameter
20% inflation in all parameters	$\lambda = 6$ rate parameter, p=6 shape parameter
20% reduction in rate, all others held constant	$\lambda = 4$ rate parameter, p=5 shape parameter
15% reduction in rate, all others held constant	$\lambda = 4.25$ rate parameter, p=5 shape parameter
10% reduction in rate, all others held constant	$\lambda = 4.5$ rate parameter, p=5 shape parameter
5% reduction in rate, all others held constant	$\lambda = 4.75$ rate parameter, p=5 shape parameter
5% inflation in rate, all others held constant	$\lambda = 5.25$ rate parameter, p=5 shape parameter
10% inflation in rate, all others held constant	$\lambda = 5.5$ rate parameter, p=5 shape parameter
15% inflation in rate, all others held constant	$\lambda = 5.75$ rate parameter, p=5 shape parameter
20% inflation in rate, all others held constant	$\lambda = 6$ rate parameter, p=5 shape parameter
20% reduction in shape parameter, all others held constant	$\lambda = 5$ rate parameter, p=4 shape parameter
15% reduction in shape parameter, all others held constant	$\lambda = 5$ rate parameter, p=4 shape parameter
10% reduction in shape parameter, all others held constant	$\lambda = 5$ rate parameter, p=4.5 shape parameter
5% reduction in shape parameter, all others held constant	$\lambda = 5$ rate parameter, p=4.75 shape parameter
5% inflation in shape parameter, all others held constant	$\lambda = 5$ rate parameter, p=5.25 shape parameter
10% inflation in shape parameter, all others held constant	$\lambda = 5$ rate parameter, p=5.5 shape parameter
15% inflation in shape parameter, all others held constant	$\lambda = 5$ rate parameter, p=5.75 shape parameter
20% inflation in shape parameter, all others held constant	$\lambda = 5$ rate parameter, p=6 shape parameter

Abbreviations: p=shape parameter; λ=rate parameter

[1]R Studio requires the specification of the scale parameters in addition to the rate parameter to simulate the Exponential distribution

Table 6: Mis-specified Log-Normal Distributed Baseline Hazard Functions Specified in Simulations for the External Validation Dataset

Scenario	**Baseline Hazard Function Parameters**
20% reduction in all parameters	Mean $\mu = 4$, standard deviation $\sigma = 1.2$

15% reduction in all parameters	Mean $\mu = 4.25$, standard deviation $\sigma = 1.275$
10% reduction in all parameters	Mean $\mu = 4.5$, standard deviation $\sigma = 1.35$
5% reduction in all parameters	Mean $\mu = 4.75$, standard deviation $\sigma = 1.425$
5% inflation in all parameters	Mean $\mu = 5.25$, standard deviation $\sigma = 1.575$
10% inflation in all parameters	Mean $\mu = 5.5$, standard deviation $\sigma = 1.65$
15% inflation in all parameters	Mean $\mu = 5.75$, standard deviation $\sigma = 1.725$
20% inflation in all parameters	Mean $\mu = 6$, standard deviation $\sigma = 1.8$
20% reduction in mean, standard deviation held constant	Mean $\mu = 4.$, standard deviation $\sigma = 1.5$
15% reduction in mean, standard deviation held constant	Mean $\mu = 4.25$, standard deviation $\sigma = 1.5$
10% reduction in mean, standard deviation held constant	Mean $\mu = 4.5$, standard deviation $\sigma = 1.5$
5% reduction in mean, standard deviation held constant	Mean $\mu = 4.75$, standard deviation $\sigma = 1.5$
5% inflation in mean, standard deviation held constant	Mean $\mu = 5.25$, standard deviation $\sigma = 1.5$
10% inflation in mean, standard deviation held constant	Mean $\mu = 5.5$, standard deviation $\sigma = 1.5$
15% inflation in mean, standard deviation held constant	Mean $\mu = 5.75$, standard deviation $\sigma = 1.5$
20% inflation in mean, standard deviation held constant	Mean $\mu = 6$, standard deviation $\sigma = 1.5$
20% reduction in standard deviation parameter, mean held constant	Mean $\mu = 5$, standard deviation $\sigma = 1.2$
15% reduction in standard deviation parameter, mean held constant	Mean $\mu = 5$, standard deviation $\sigma = 1.275$
10% reduction in standard deviation parameter, mean held constant	Mean $\mu = 5$, standard deviation $\sigma = 1.35$
5% reduction in standard deviation parameter, mean held constant	Mean $\mu = 5$, standard deviation $\sigma = 1.425$
5% inflation in standard deviation parameter, mean held constant	Mean $\mu = 5$, standard deviation $\sigma = 1.575$
10% inflation in standard deviation parameter, mean held constant	Mean $\mu = 5$, standard deviation $\sigma = 1.65$
15% inflation in standard deviation parameter, mean held constant	Mean $\mu = 5$, standard deviation $\sigma = 1.725$
20% inflation in standard deviation parameter, mean held constant	Mean $\mu = 5$, standard deviation $\sigma = 1.8$

Abbreviations: μ=mean; σ=standard deviation

Table 7: Baseline Hazard Functions Specified in Simulations Examining Effects of Different Distributions[1]

Original Baseline Hazard Function and Parameters	**External Validation Dataset Baseline Hazard Function Distribution**	**Parameters of Baseline Hazard Function of External Validation Dataset**
Weibull: $\lambda = 5$ rate parameter, p=3 shape parameter	Exponential	$\lambda = 5$ rate parameter, p=3 scale parameter
	Log-Logistic	$\lambda = 5$ rate parameter, p=3 shape parameter
	Log-Normal	Mean $\mu = 5$, standard deviation $\sigma = 1.5$
Exponential: $\lambda = 5$ rate parameter, p=3 scale parameter	Weibull	$\lambda = 5$ rate parameter, p=3 shape parameter
	Log-Logistic	$\lambda = 5$ rate parameter, p=3 shape parameter
	Log-Normal	Mean $\mu = 5$, standard deviation $\sigma = 1.5$
Log-Logistic: $\lambda = 5$ rate parameter, p=3 shape parameter	Weibull	$\lambda = 5$ rate parameter, p=3 shape parameter
	Exponential	$\lambda = 5$ rate parameter, p=3 scale parameter
	Log-Normal	Mean $\mu = 5$, standard deviation $\sigma = 1.5$
Log-Normal: Mean $\mu = 5$, standard deviation $\sigma = 1.5$	Weibull	$\lambda = 5$ rate parameter, p=3 shape parameter
	Exponential	$\lambda = 5$ rate parameter, p=3 scale parameter
	Log-Logistic	$\lambda = 5$ rate parameter, p=3 shape parameter

Abbreviations: p=shape parameter; λ=rate parameter; μ=mean; σ=standard deviation

[1]R Studio requires the specification of the scale parameter in addition to the rate parameter to simulate the Exponential distribution

Figure 1: Weibull Distribution Baseline Survival Functions Simulated

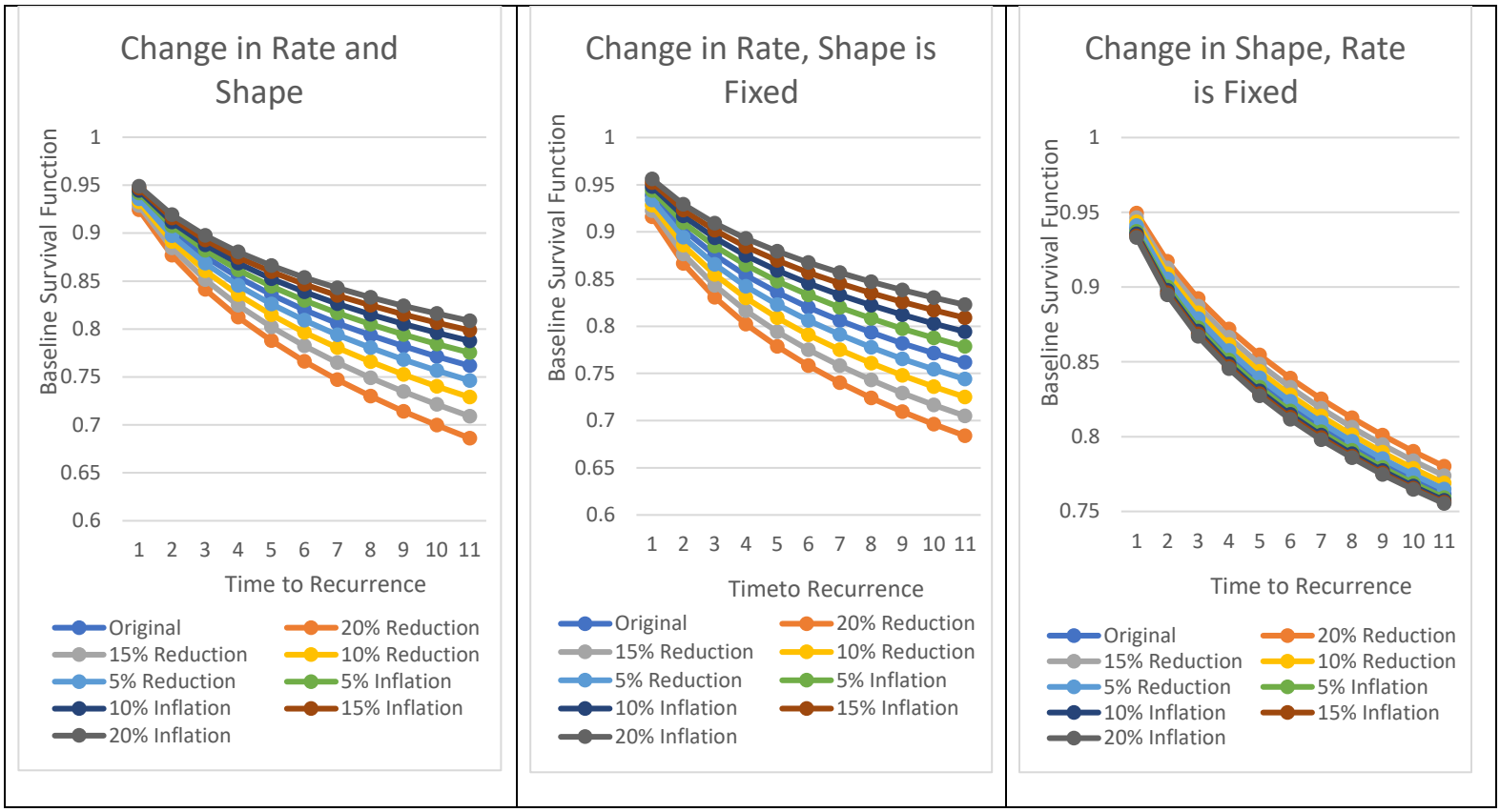


Figure 2: Weibull Distribution- Effect of Baseline Survival Function Mis-calibration on Overall C-Statistics

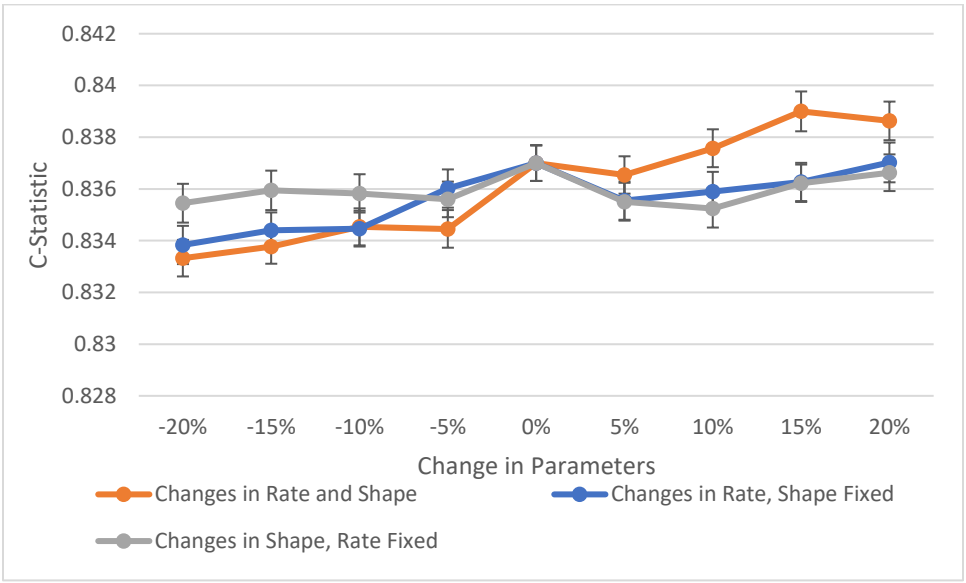


Figure 3: Weibull Distribution-Effect of Baseline Survival Function Mis-calibration on Calibration Slopes at 5-years

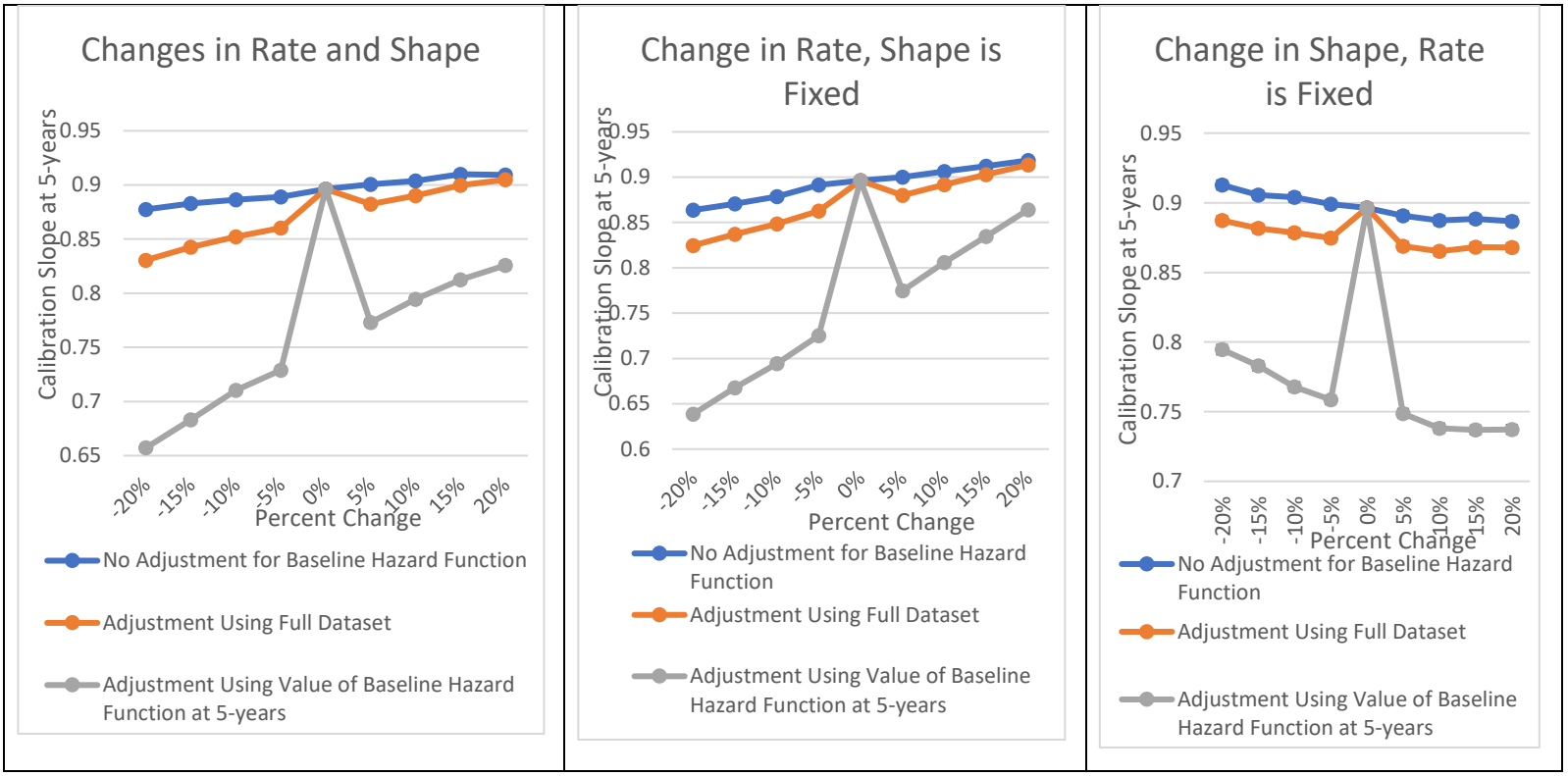


Figure 4: Exponential Distribution Baseline Survival Functions Simulated

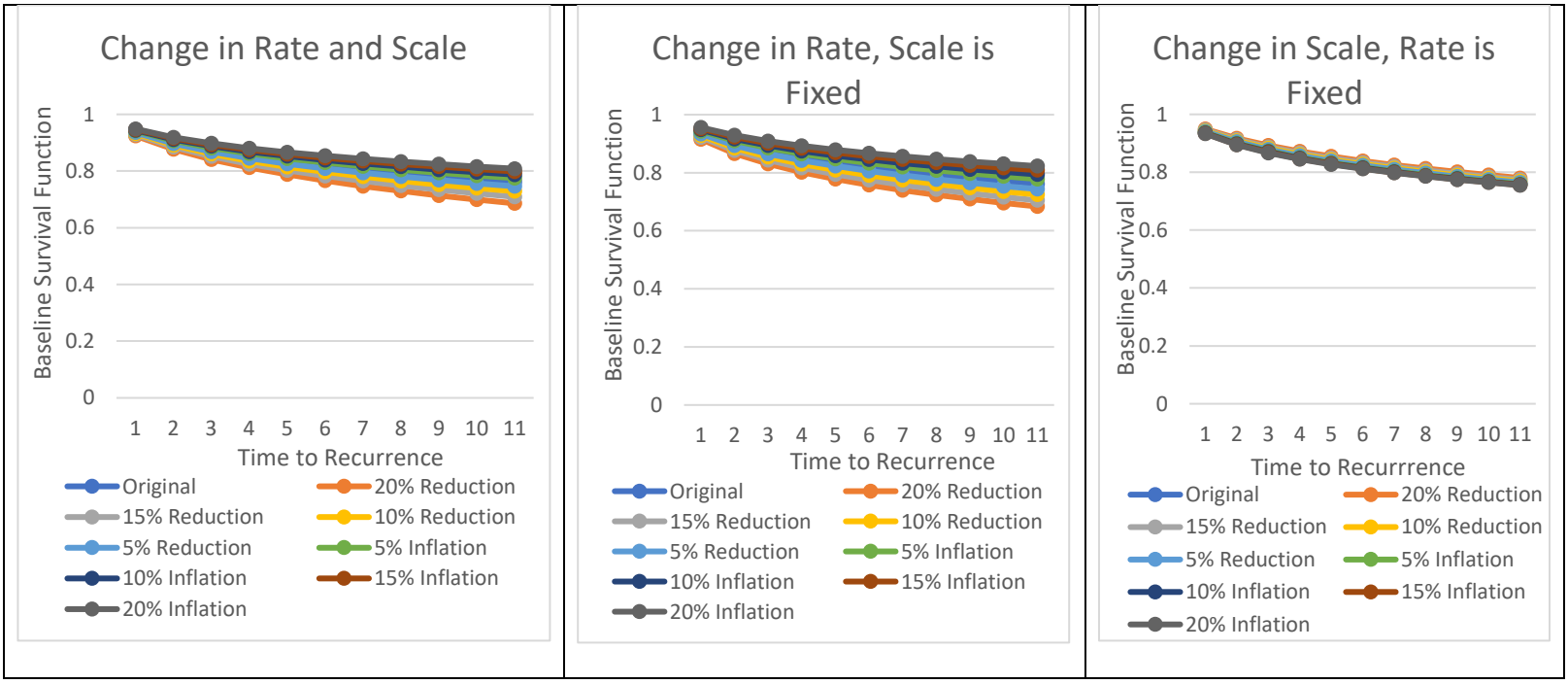


Figure 5: Exponential Distribution- Effect of Baseline Survival Function Mis-calibration on Overall C-Statistics

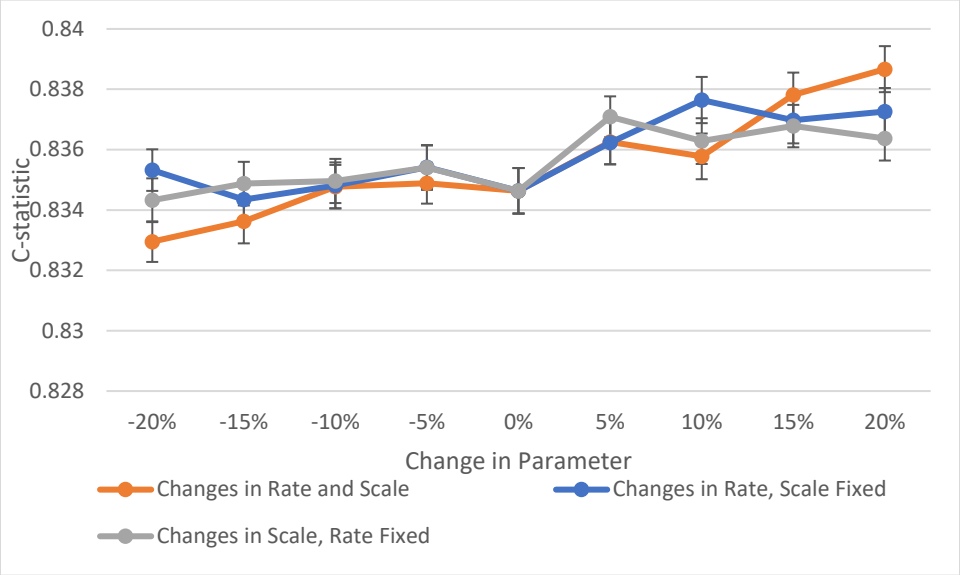


Figure 6: Exponential Distribution-Effect of Baseline Survival Function Mis-calibration on Calibration Slopes at 5-years

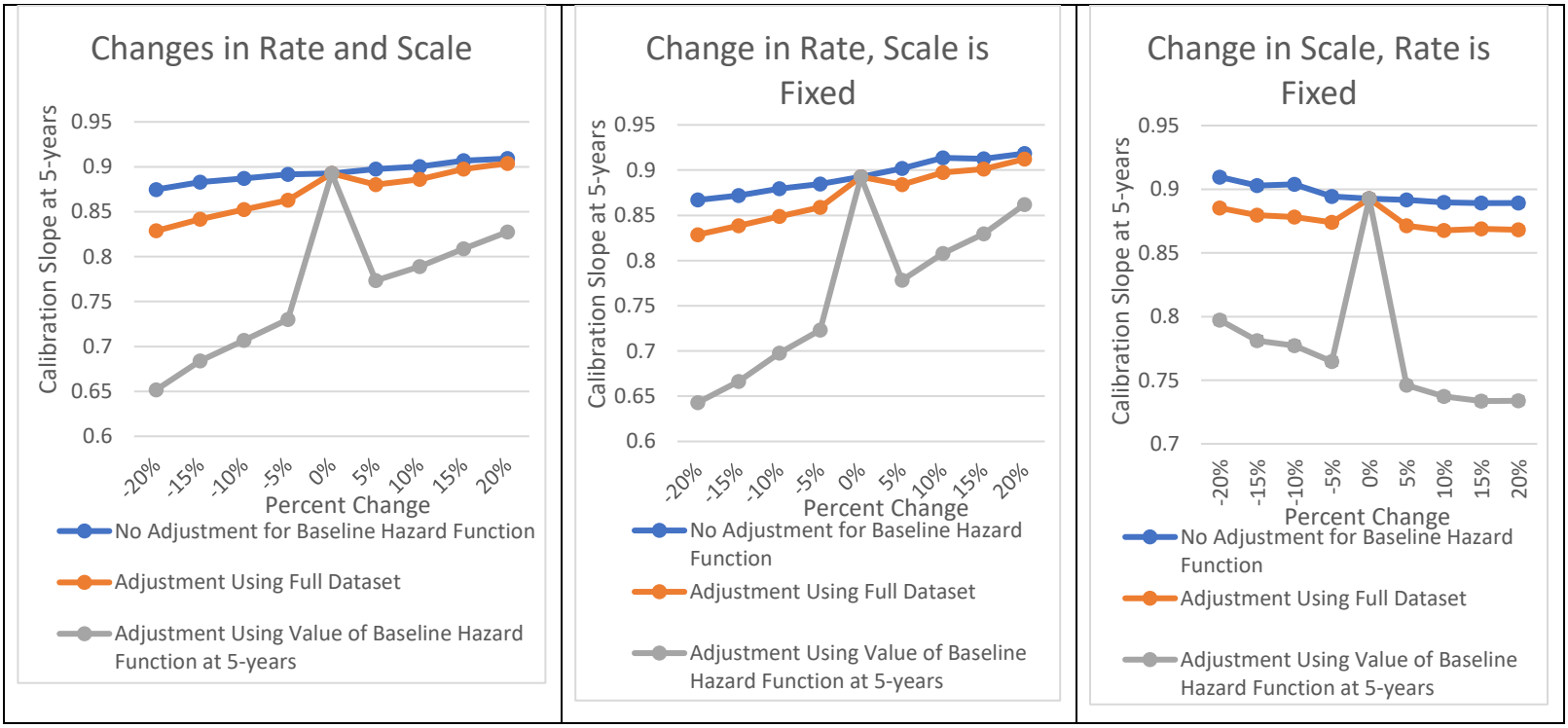


Figure 7: Log-Logistic Distribution Baseline Survival Functions Simulated

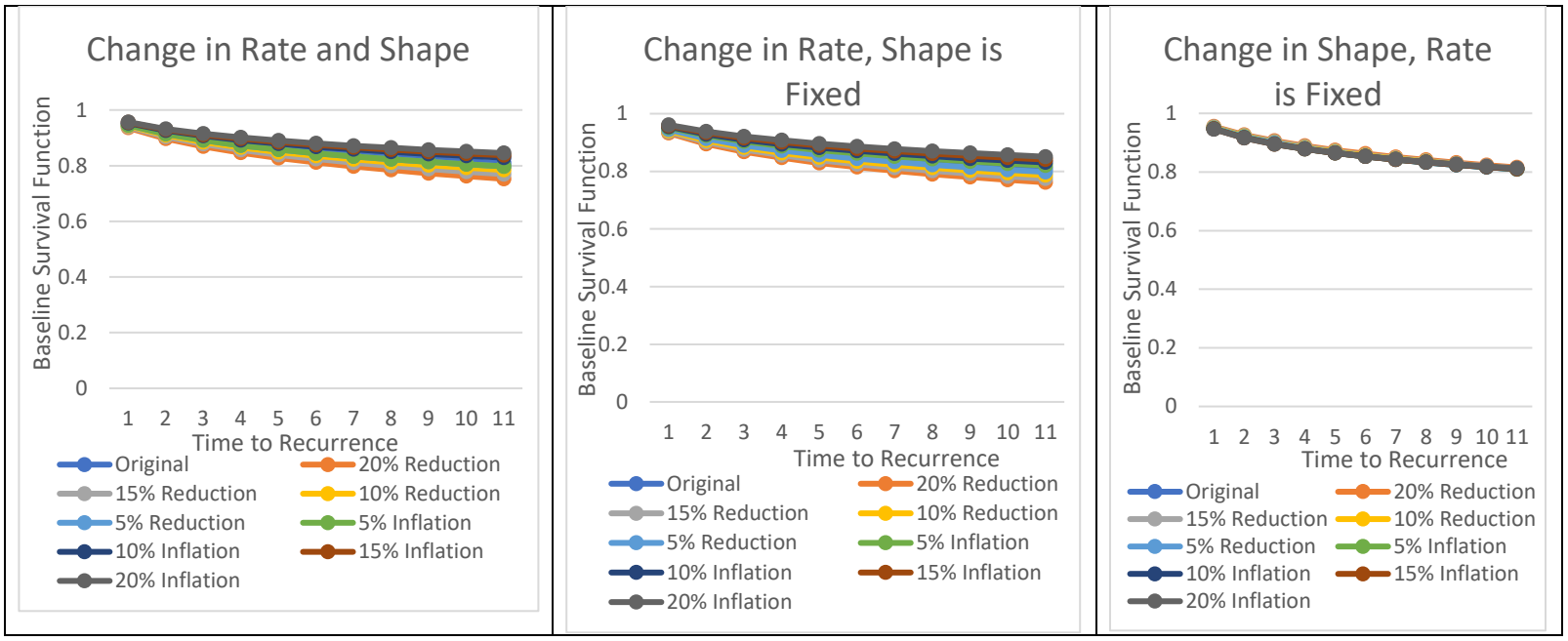


Figure 8: Log-Logistic Distribution- Effect of Baseline Survival Function Mis-calibration on Overall C-Statistics

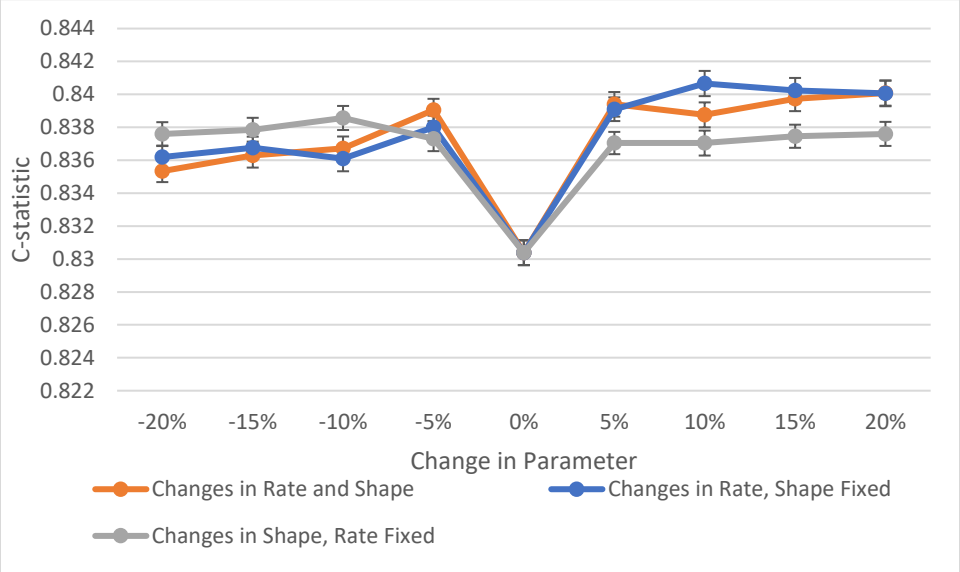


Figure 9: Log-Logistic Distribution-Effect of Baseline Survival Function Mis-calibration on Calibration Slopes at 5-years

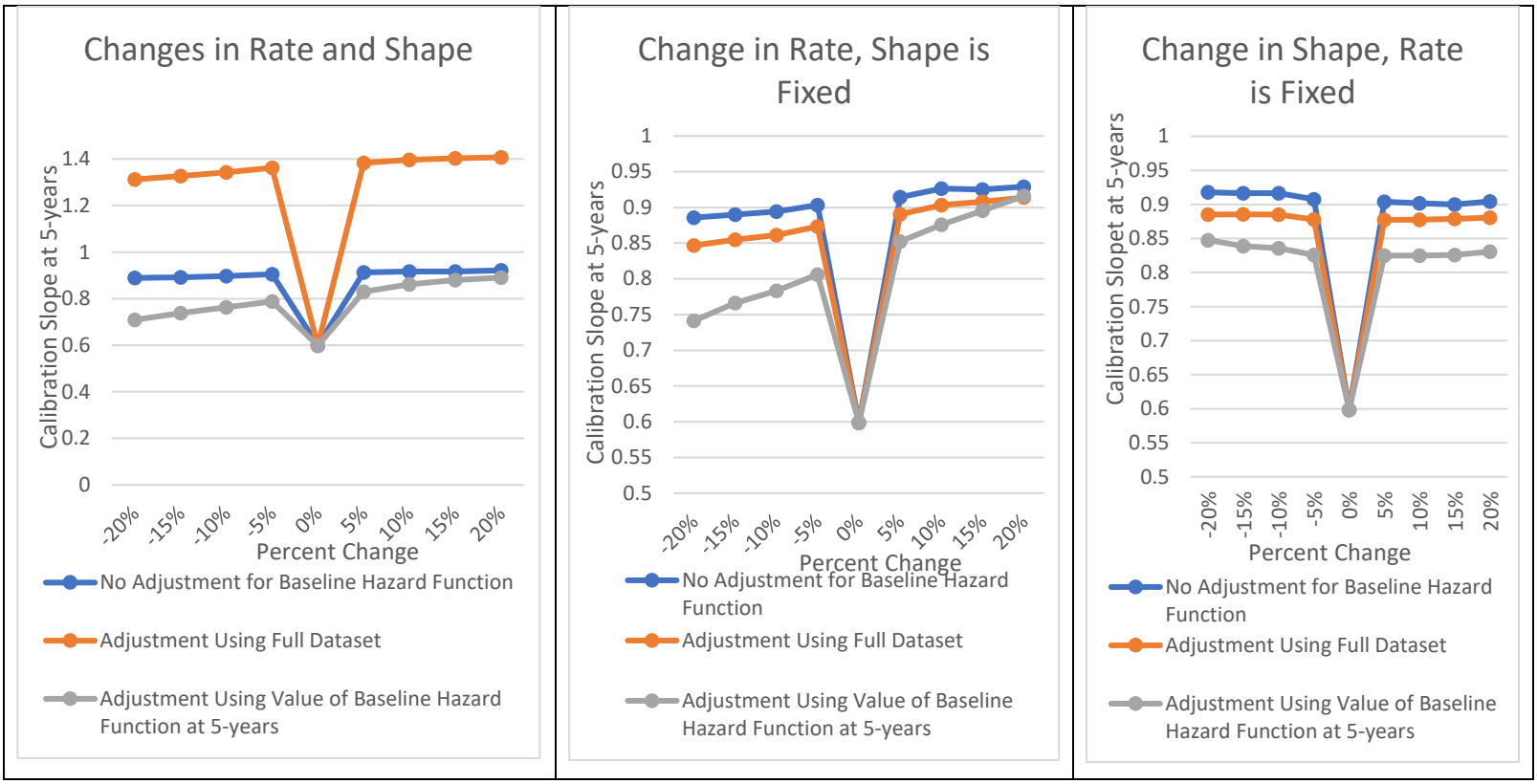


Figure 10: Log-Normal Distribution Baseline Survival Functions Simulated

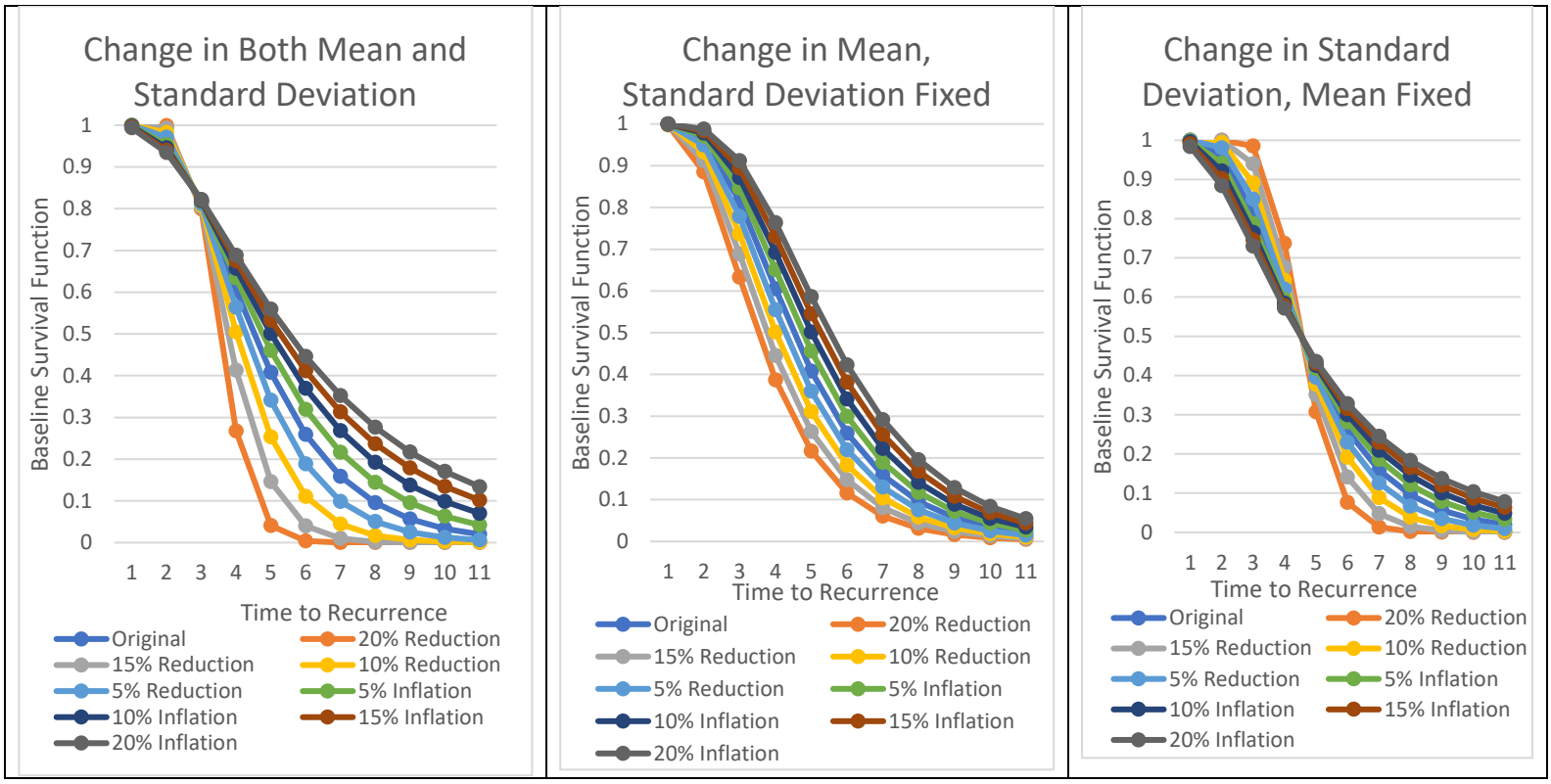


Figure 11: Log-Normal Distribution- Effect of Baseline Survival Function Mis-calibration on Overall C-Statistics

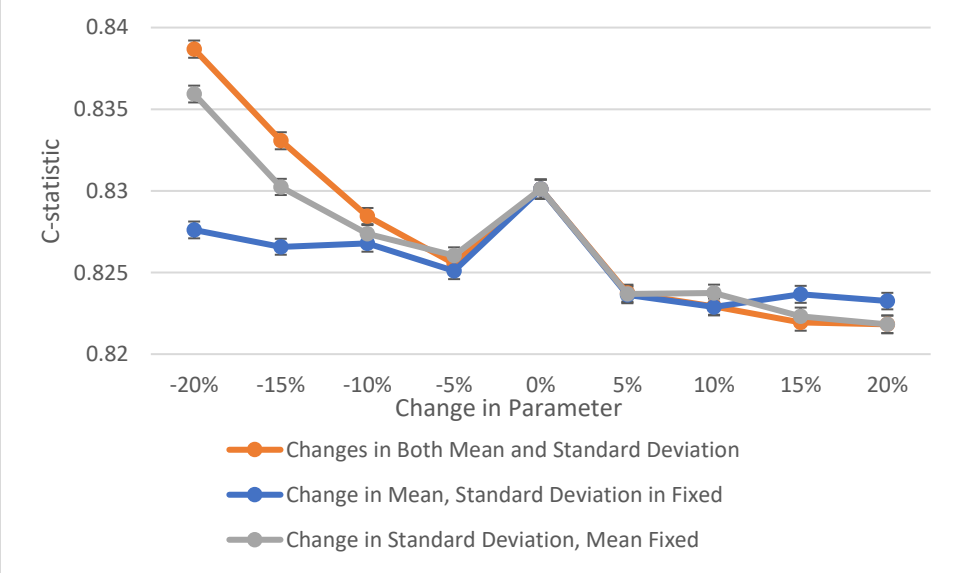


Figure 12: Log-Normal Distribution-Effect of Baseline Survival Function Mis-calibration on Calibration Slopes at 5-years

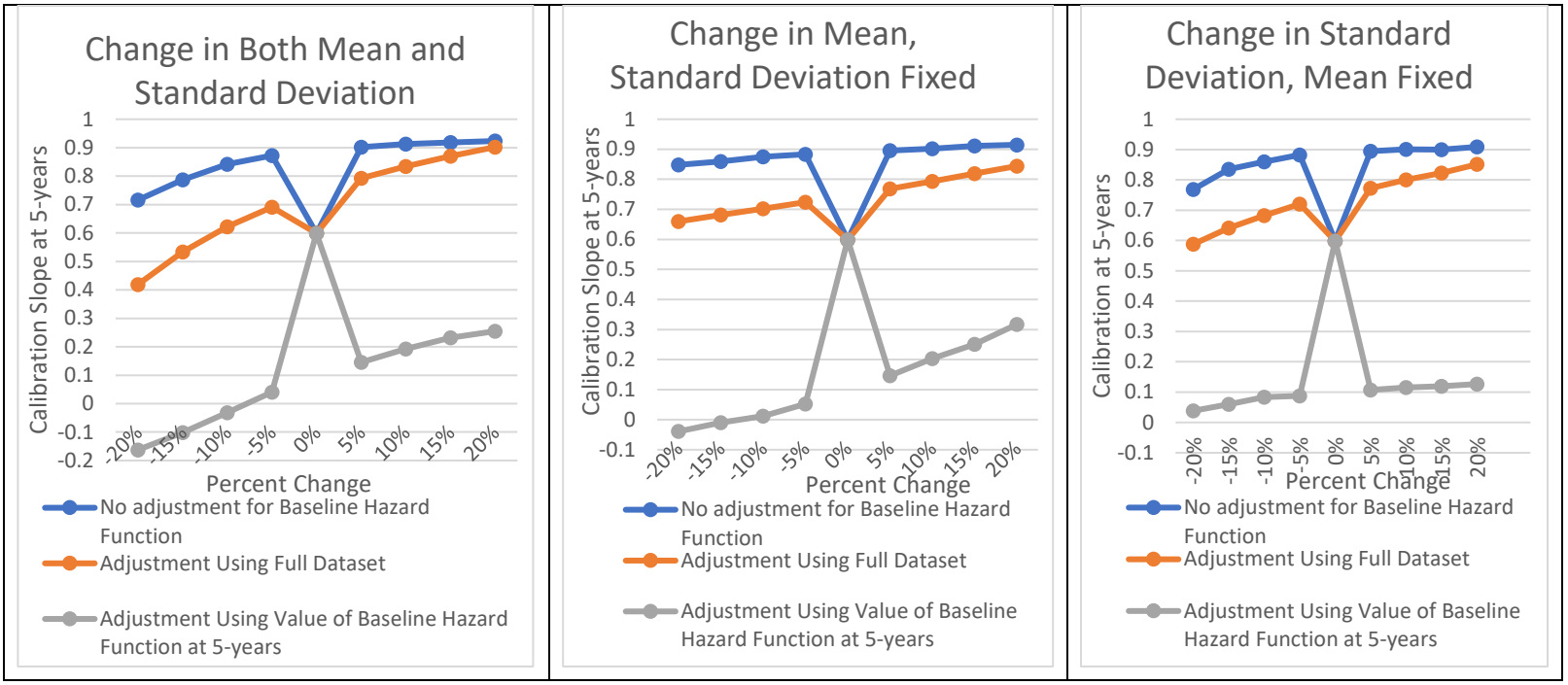


Figure 13: Effect of Differences in Development and External Validation Distribution Types

a) C-statistic

		External Validation Dataset Distributions			
		Weibull	Exponential	Log-Logistic	Log-Normal
Development Dataset Distributions	Weibull	0.8366219 (95% CI: 0.8352965-0.8379474)	0.8367112 (95%CI: 0.8353407-0.8380817)	0.8399242 (95%CI: 0.8384101-0.8414383)	0.8246607 (95%CI: 0.8236325-0.8256889)
	Exponential	0.8355315 (95%CI: 0.8340785-0.8369845)	0.8346368 (95%CI: 0.833159-0.8361147)	0.8383645 (95%CI: 0.8368202-0.8399088)	0.8246120 (95%CI: 0.8236087-0.8256153)
	Log-Logistic	0.8355114 (95%CI: 0.8340431-0.8369797)	0.8359539 (95%CI: 0.8345125-0.8373953)	0.8386561 (95%CI: 0.8372321-0.8400802)	0.8255020 (95%CI: 0.8244950-0.8265090)
	Log-Normal	0.8350423 (95%CI: 0.8335981-0.8364866)	0.8358257 (95%CI: 0.8343493-0.8373022)	0.839214 (95%CI: 0.8377163-0.8407117)	0.8242428 (95%CI: 0.8232628-0.8252228)

b) Calibration Slope-No Adjustment for Baseline Hazard Function

	External Validation Dataset Distributions			
	Weibull	Exponential	Log-Logistic	Log-Normal

Development Dataset Distributions	Weibull	0.8974972 (95%CI: 0.8912849-0.9037096)	0.8953464 (95%CI: 0.8891743-0.9015185)	0.9142518 (95%CI: 0.9079561-0.9205474)	0.8926545 (95%CI: 0.8878884-0.8974207)
	Exponential	0.8943893 (95%CI: 0.8878173-0.9009612)	0.8925387 (95%CI: 0.885869-0.8992083)	0.9067699 (95%CI: 0.9001691-0.9133707)	0.8909848 (95%CI: 0.8863231-0.8956465)
	Log-Logistic	0.8879654 (95%CI: 0.8814591-0.8944718)	0.8923066 (95%CI: 0.8859198-0.8986934)	0.9076416 (95%CI: 0.9007281-0.9145551)	0.8919431 (95%CI: 0.8875239-0.8963623)
	Log-Normal	0.8913411 (95%CI: 0.8849376-0.8977447)	0.8948727 (95%CI: 0.8885210-0.9012244)	0.9075134 (95%CI: 0.9007023-0.9143245)	0.8919861 (95%CI: 0.8872721-0.8967001)

c) Calibration Slope-Adjustment Using Full Dataset

		External Validation Dataset Distributions			
		Weibull	Exponential	Log-Logistic	Log-Normal
Development Dataset Distributions	Weibull	0.8974972 (95%CI: 0.8912849-0.9037096)	0.8716971 (95%CI: 0.8658569-0.8775374)	0.9045238 (95%CI: 0.8979923-0.9110552)	0.4854874 (95%CI: 0.4818572-0.4891175)
	Exponential	0.8707862 (95%CI: 0.8648595-0.8767128)	0.8925387 (95%CI: 0.885869-0.8992083)	0.902426 (95%CI: 0.8959831-0.9088690)	0.4852297 (95%CI: 0.4815944-0.4888649)
	Log-Logistic	0.8458899 (95%CI: 0.8397953-0.8519844)	0.8490318 (95%CI: 0.8428966-0.8551671)	0.9076416 (95%CI: 0.9007281-0.9145551)	0.4746717 (95%CI: 0.4710191-0.4783242)
	Log-Normal	1.6269679 (95%CI: 1.6070278-1.6469081)	1.6383971 (95%CI: 1.6183713-1.6584230)	1.5832442 (95%CI: 1.5642418-1.6022466)	0.8919861 (95%CI: 0.8872721-0.8967001)

d) Calibration Slope- Adjustment Using Value of Baseline Hazard Function at 5-years

		External Validation Dataset Distributions			
		Weibull	Exponential	Log-Logistic	Log-Normal
Development Dataset Distributions	Weibull	0.8974972 (95%CI: 0.8912849-0.9037096)	0.7557693 (95%CI: 0.7491592-0.7623794)	0.8281120 (95%CI: 0.8212282-0.8349957)	0.2396345 (95%CI: 0.2327671-0.2465018)
	Exponential	0.7478681 (95%CI: 0.7410791-0.7546571)	0.8925387 (95%CI: 0.885869-0.8992083)	0.8234851 (95%CI: 0.8165245-0.8304457)	0.2308400 (95%CI: 0.2240087-0.2376712)

	Log-Logistic	0.7530837 (95%CI: 0.7466801-0.7594873)	0.7562195 (95%CI: 0.7500628-0.7623762)	0.9076416 (95%CI: 0.9007281-0.9145551)	0.2418188 (95%CI: 0.2346658-0.2489717)
	Log-Normal	0.8173281 (95%CI: 0.8123343-0.8223220)	0.8158687 (95%CI: 0.8110724-0.8206649)	0.8684009 (95%CI: 0.8633253 0.8734766)	0.8919861 (95%CI: 0.8872721-0.8967001)

Abbreviation: CI=Confidence Interval

Table 8: Calibration Slope at 5-years For Pre-Existing Prognostic Models for All Patients

Distribution	Pre-Existing Prognostic Model	Calibration Slope at 5-years	Rank
Weibull	Kattan et al, 2001	0.025	5
	Cindolo et al, 2003	0.85	2
	Yaycioglu et al, 2013	0.032	4
	van der Mijn et al, 2019	1.071	1
	Yaycioglu et al, 2001	0.656	3
Exponential	Kattan et al, 2001	0.025	5
	Cindolo et al, 2003	0.868	2
	Yaycioglu et al, 2013	0.033	4
	van der Mijn et al, 2019	1.101	1
	Yaycioglu et al, 2001	0.67	3
Log-Logistic	Kattan et al, 2001	0.025	5
	Cindolo et al, 2003	0.848	2
	Yaycioglu et al, 2013	0.032	4
	van der Mijn et al, 2019	1.061	1
	Yaycioglu et al, 2001	0.656	3
Log-Normal	Kattan et al, 2001	0.025	5
	Cindolo et al, 2003	0.809	2
	Yaycioglu et al, 2013	0.032	4
	van der Mijn et al, 2019	1.057	1
	Yaycioglu et al, 2001	0.656	3

Chapter 5: Predication of Clear Cell Renal Cell Carcinoma Recurrence using the Canadian Kidney Cancer Information System

5.1 Abstract

Introduction The risk of clear-cell renal cell carcinoma cancer recurrence following surgery has been shown to vary considerably based on prognostic factors. Previous external validation of models for kidney cancer recurrence have shown a wide range in model performance, potentially due to single institution data, analysis of non-contemporary patients, or absence of known prognostic factors. The purpose of this analysis is to create, and internally validate, a new prognostic model for recurrence free survival using a large contemporary cohort of Canadian Clear Cell Renal Cell Carcinoma patients.

Methods Data from the Canadian Kidney Cancer Information System (CKCis) was used. CKCis is a nationwide prospective cohort of patients treated since Jan 1, 2011. Prognostic models for recurrence free survival using p-value selection, backwards-bootstrap selection, and elastic net penalization selection were used. The minimum required sample size for model creation was calculated to assess statistical power. Prognostic models were validated using optimism corrected bootstrapping and assessed for overall c-statistic, c-index over time, calibration slopes at 2-, 3- and 5-years and calibration curves at 2-,3- and 5-years post-surgery. Decision curve analysis was used to assess clinical utility. Risk groups were developed using several techniques and compared.

Results 3371 patients met inclusion criteria and 2555 were used to develop and internally assess the CKCis prognostic model. The sample size was adequate for model creation (minimum 261). The best performing prognostic model for recurrence free survival included lymph node stage, tumour stage, tumour size, tumour necrosis, sarcomatoid differentiation, surgical margin status, serum neutrophils, and performance status. This model had an overall c-statistic of 0.85, calibration slopes at 2-, 3- and 5-years of close to 1, and good calibration performance based on calibration curves at 2-, 3-, and 5- years. Based on 5-year recurrence decision curve analysis, this model is useful for clinically important recurrence thresholds (up to 35%). Minimal differences in performance were identified in sensitivity analyses using stage components (instead of stage groupings), multiple imputation for missing data, and k-fold validation. Sensitivity analyses when p-values for entry/exit in models of 0.2 and 0.1 yielded less parsimonious models with only slight differences in prognostic performance. The risk group classification system had a c-statistic of 0.83.

Conclusions Using a large cohort of contemporary Canadian patients with clear-cell carcinoma, a model to predict recurrence was developed that incorporates prognostic variables universally available in clinical practice. This model should be externally validated.

5.2 Introduction

The purpose of prognostic modelling is to predict the risk of clinical outcomes (1). Prognostic modelling can be used for patient counselling and delivery of personalized targeted interventions, such as testing or treatment (1). Prognostic models are also useful for clinical trial efficiency and patient stratification (1). Incorporating prognostic models in clinical decisions should allow high-risk patients to receive timely treatments and low-risk patients to avoid unnecessary investigations or treatments (1). Prognostic models should be evaluated over time and updated, as needed, due to changes in contemporary detection methods and treatments (2).

Various anatomical, histological, clinical, laboratory, and genetic factors have been associated with recurrence or progression of disease in patients with kidney cancer. Anatomical factors include local tumour invasion and nodal metastases (3). Histological features include nuclear grade, sarcomatoid features, histological subtype, and tumour necrosis (3). Clinical features include performance/functional status and symptoms (3). Laboratory factors include serum hemoglobin, neutrophils, platelets, and calcium(4). There may also be genetic and protein-based markers that are associated with kidney cancer prognosis (3). Prognostic factors have different strengths of association; for instance, pathological factors such as tumour grade have been demonstrated to be stronger predictors than clinical factors, such as hematuria (5).

Numerous prognostic models for renal cell carcinoma recurrence have been created. However, assessments have revealed varied performance and a narrow range of decision thresholds with clinical utility. The purpose of this study was to create and internally validate a prognostic model based on a large cohort of contemporary Canadian kidney cancer patients with Clear Cell Renal Cell Carcinoma (ccRCC).

5.3 Methods

5.3.1 Study Population

CKCis is a prospective cohort of kidney cancer patients enrolled at 14 academic tertiary care institutions across Canada. Patients have been enrolled and followed prospectively since January 1, 2011 and data cutoff for this study was June 30, 2021. We included patients with incident diagnosis of a clinically localized ccRCC treated with surgery (radical or partial nephrectomy). Patients were excluded if they were aged <18 years at surgery, those with clinical metastases, those treated with methods other than partial or radical

nephrectomy, those with multiple tumours or congenital renal cell carcinoma syndromes, and those with non-clear cell RCC histology. Non-clear cell histology was excluded because prognostic factors do not seem to be consistent across RCC subtypes and models that include all RCC subtypes have worse performance. ccRCC accounts for greater than 70% of incident RCC histology. Figure 1 outlines the development of the patient cohort used in prognostic model development.

5.3.2 Outcome

Time to recurrence was defined as time from surgery date to date of kidney cancer recurrence (usually detected via surveillance imaging).

5.3.3 Prognostic Variables

Potential model parameters were assessed from the CKCis cohort. Parameters were assessed through imaging, pathological observation of tumour specimens and laboratory assessment of blood work. These variables were selected based on availability in clinical practice, availability in CKCis, inclusion in previous models, and on clinical intuition. Variables with a high proportion of missing data or obvious collinearity issues (e.g clinical T stage and pathological T stage) were removed with the most reliable variables (e.g. pathologic stage) retained. Pathologic stage for kidney cancer incorporates composite features (eg pT3a is assigned for either sinus fat invasion, renal vein thrombus, or invasion of peri-nephric fat). While pathological T stages T1a and T1b were combined as T1 and T2a and T2b were combined as T2, all other prognostic variables were retained as is without any further reclassification.As a sensitivity analysis, stage components were included in models instead of tumour stage groupings to determine if stage deconstruction improved model performance. Given that surgery could transiently impact laboratory parameters, pre-operative results were used.

5.3.4 Prognostic Model Development

Prognostic models were constructed using variables with greater than 70% complete data. Methods of parameter selection were p-value (forwards, backwards, and stepwise), backwards bootstrap, and elastic net penalization. The benefits and limitations of different methods are outlined in chapter 2. Prognostic models were created using Cox Proportional Hazards Models. Sensitivity analyses for model development included varying entry and exit p-value thresholds (for p-value parameter selection methods). Model development and internal validation for the selection of the best performing prognostic model was completed using complete case analysis.

5.3.5 Internal Validation

Internal validation was performed to assess the performance of the model on indicators of accuracy: overall c-statistic, c-index over time, calibration slope (at 2-, 3- and 5-years), calibration curve (at 2-, 3- and 5-years) and decision curve analysis (5-years). The Cox Proportional Hazards Model was used to generate predicted recurrence free survival probabilities and Kaplan-Meier survival analysis was used to generate observed probabilities of recurrence free survival. The models were validated on the same cohort as model development. Optimism-corrected bootstrapping was used for internal validation using 500 samples with replacement. Overall c-statistic, c-index over time, and calibration slope were calculated by using the mean of the optimism values from each of these parameters in the 500 samples. An alternate method of internal validation using k-fold groupings (the dataset is divided into k partitions and the mean of each partition is used) was also used(6). K-fold analysis using k=10 was used to construct calibration plots at 2-, 3- and 5-years. Calibration plots from both bootstrap analysis and k-fold analysis were compared.

The overall c-statistic is a measure of the discriminative ability of the model as measured by the proportion of pairs in which the higher risk observation is identified correctly (7). C-statistics range from 0 to 1 with values closer to 1 signifying more optimal discrimination and those ≥0.7 signaling clinical usefulness (7). C-index measures discriminative ability over time by measuring the discriminative ability of all observations from baseline (time=0) to the time point of interest (i.e. for c-index at 5-years post-surgery, observations with time to follow-up between 0 and 5 years would be included) (8). Calibration slope at 2-,3-, and 5-years is assumed to have a slope of 1. The optimism correction is estimated using bootstrapping, and the optimism corrected calibration slope is obtained by subtracting it from 1. Where predicted survival probabilities were obtained from the prognostic model, observed survival probabilities were obtained using Kaplan-Meier survival analysis (9). Calibration curves for 2-, 3- and 5-years were created using LOESS curves. To generate calibration curves 100 groups were used. Calibration was also assessed by determining the proportion of patient groups within 5% and within 10% of predicted risk (i.e. $observed\ event\ free\ survival - predicted\ event\ free\ survival$). For instance, if a patient group had a predicted recurrence free survival probability of 0.75 at 5-years post-surgery and the observed recurrence free survival probability was 0.79, it satisfied the 'within 5%' agreement. If the observed recurrence free survival probability was 0.81, the patient group would not meet such agreement. The same principle was applied for the 'within 10%' agreement outcome. Lastly, decision curve analysis at 5-years was used as an evaluation of clinical utility by obtaining the range of patient risk at which the prognostic model yielded a clinical net benefit (10). The net benefit in a decision curve is the weighted

difference between true positive and false positives at a given threshold probability (10). Given that this model may be used to select patients for adjuvant therapy (which would likely have toxicity), we believe that a clinically useful threshold range would be between 20% and 30%. In all assessments, a comparison referent model was created using tumour stage as the sole prognostic variable.

The best prognostic model (CKCis model) was subjectively determined by selecting the model with an adequate c-statistic (≥0.7), acceptable calibration, reasonable parsimony, and the largest range of patient risk thresholds for which the model yielded net clinical benefit.

After the best model (CKCis model) was selected, the potential impact of missing data was assessed using multiple imputation, assuming data were missing completely at random with the Substantive Model Compatible Fully Conditional Specification. Lastly, we assessed if our cohort had adequate sample size to assess performance. The minimum sample size required for prognostic model construction given the mean follow-up time, event rate, and other relevant parameters was calculated (11).

5.3.6 Risk Group Development Based on the CKCis Model

There were several methods used for the creation of risk groups. For consistency with previous RCC prognostic models, we decided to create 4 risk groups (Low, Intermediate, High, and Very High). The first method to generate risk groups was the 'quartile method' based on dividing patients into quartiles of risk scores. For instance, the Low-risk group in this method would contain all patients with risk scores from zero until the first quartile. The second method was the 'mean risk score method' where the predicted recurrence free survival probability was divided into three ranked groups from highest to lowest and the mean risk score for each group calculated. The risk group thresholds were the mean risk scores for each ranked group (eg low risk was all scores from 0 to the mean of the first ranked group). The third method was the 'percentile method' using the 16^{th}, 50^{th} and 84^{th} percentiles of risk scores as thresholds (12). For example, patients with risk scores above the risk score at the 84^{th} percentile were considered to be 'very high risk.' The fourth method was the 'face validity' method based on risk-labels that would be likely to correspond to a patient and clinicians' perception of risk: Model scores corresponding with <10% 5-year recurrence risk was considered low risk, 10%-<30% for Intermediate risk, 30%-70% for High risk and ≥70% for Very High risk. All recurrence-free probabilities were converted into risk scores by determining the risk score that corresponded to the survival probability from the 5-year nomogram using polynomial regression. Risk score cutoffs for each risk group are presented in Table 10.

5.3.7 Risk Group Assessment

To compare risk group classification systems, we determined the proportion of patients in each risk group (to assess distribution), the Kaplan Meier curves (to assess stratification and label-to-risk concordance), overall c-statistic (to assess discrimination) and predicted vs. observed number of events (to assess calibration) (13). Label-to-risk concordance means that the label (low, intermediate, high, or very high) should correspond to what most patients and clinicians would consider the numerical recurrence risk. For example, most patients would likely consider 5% risk of recurrence to be low risk and 75% risk of recurrence to be very high risk.

Statistical analyses were performed using SAS Enterprise Guide 8.2, Microsoft Excel 2016 and R 4.0.2.

5.4 Results

After exclusions 3,731 CKCis cohort patients had a sporadic, solitary, clear cell RCC treated with surgery. Of these, 1,933 (51.8%) underwent radical nephrectomy, 2,448 (65.6%) were male, mean age at surgery was 60.8 years (95%CI: 60.4 years-61.2 years). The mean Charlson co-morbidity score was 4.4 (95%CI:4.4-4.5) with 1,900 (52.7%) patients having hypertension and 824 (22.8%) patients having diabetes. (Table 1). At clinical presentation 1,029 (27.6%) patients had local symptoms (eg hematuria) and 226 (6.1%) had systemic symptoms (eg weight loss). There was a high distribution of lower risk tumors, with 1,559 (41.8%) pathological T1a stage and 1,642 (44.6%) grade 2 (out of 4). Coagulative tumour necrosis was present in 849 (22.8%), 129 (3.5%) had sarcomatoid features, and 248 (6.9%) had positive surgical margins.

5.4.1 Recurrence Free Survival Prognostic Model

Overall, 761 patients experienced recurrence at last follow-up with estimated recurrence risk of 16.4%, 20.2%, and 26.9% at 2-, 3- and 5-years post-surgery. Mean follow-up time was approximately two-years. Figure 2 shows the recurrence free survival of patients over time post-surgery. Assuming complete follow-up truncated at 5-years, median follow-up in those without recurrence was 2 years. Specific recurrence risks may vary slightly across prognostic models due to exclusion of patients with incomplete data for each individual prognostic model. Loss-to-follow up was 11.2% at 2-years post-surgery, 17.4% at 3-years post-surgery and 17.7% at 5-years post-surgery. Low pre-operative hemoglobin was found to be associated with increased risk of recurrence on univariable analysis. All other prognostic variables were found to have positive associations with increased recurrence risk. Smoking, pre-operative serum creatinine concentration, and pre-operative platelet concentrations were not significantly associated with risk of recurrence. (Table 2).

The construction of prognostic models for recurrence free survival is as outlined in Table 3a and Table 4b. Depending on the parameter selection method, the parameters included corresponding magnitude of hazard ratios varied. In some scenarios the p-value section methods (forward, backward, and selection) produced the same parameters.

5.4.2 Internal Validation Using Optimism Corrected Bootstrap

Prognostic models were created including pathological T stage and excluding pathological T stage as a sensitivity analysis. Overall, prognostic models including pathological T stage had better discrimination, calibration, and clinical utility. The p-value selection model including pathological T stage (CKCis model) was the best performing model. The overall c-statistic, calibration slopes at 2-, 3- and 5-years post-surgery, sample size available for corresponding model development are presented in Table 6a for prognostic models including pathological T stage and Table 6b for prognostic models using stage components rather than T stage. The overall c-statistic was 0.851 (Tables 5a&5b; Figure 3))

LOESS smoothed calibration curves for recurrence free survival 2-, 3- and 5-years post-surgery are shown in Figures 4-6 and "within 5%" and "within 10%" agreement proportions in Tables 7a&7b. Calibration performance was reasonable despite some under- and over-estimation of recurrence free survival. In general, calibration performance over time decreased from 2- to 5-years post-surgery. Calibration performance was also better at higher recurrence free survival probabilities, indicating better performance in lower-risk patients.

From decision curve analysis, p-value selection model had a wide range of clinical utility. All models had net-benefit in the range of threshold probabilities which we believed would have clinical utility for selection of adjuvant therapy (20% to 30% treatment thresholds). All prognostic models had better performance than pathological T stage alone (Tables 6a &6b, Figure 7)

5.4.3 Sensitivity Analysis Based on p-values for Automated Parameter Selection

A sensitivity analysis was performed for automated parameter selection models (forward, backward and stepwise selection) with more liberal p-values for model entry and exit (p=0.2, p=0.1). The models are shown in Tables 7 and resulted in more parameters with minimal change in model performance compared to the automated parameter selection models generated using p-values of 0.05, these models have marginally improved discrimination and calibration. (Tables 8-9, Figure 8-12).

5.4.4 Internal Validation Using k-fold Validation Compared to Optimism Corrected Bootstrap

K-fold validation was compared to optimism corrected bootstrap sampling. Slight differences were observed for overall c-statistic, calibration slope, calibration plots, and proportion agreement. (Figures 4-6 for internal validation using bootstrap, Figure 13 for k-fold validation).

5.4.5 Sensitivity Analysis for Missing Data Using Multiple Imputation

To examine the potential effect of missing data, multiple imputation using 14 imputed datasets was used. Internal validation using optimism corrected bootstrapping on the imputed datasets resulted in an overall c-statistic of 0.845 (95%CI:0.845-0.845). Calibration slopes were similar at 1.000 (95%CI:0.999-1.000), 0.999(95%CI:0.999-1.000) and 0.999 (95%CI:0.999-1.000) at 2-, 3- and 5- years post-surgery (Table 5a). C-index over time was also similar at all time points compared to those observed when complete data were used (Figure 14). While 100% agreement within 10% was observed at all timepoints in both methods, the proportion of patient groups with agreement within 5% was lower when multiple imputation was used (57% vs. 100% at 2-years post-surgery,42% vs. 100% at 3-years post-surgery and 51% vs. 56% at 5-years post-surgery). This is corroborated in calibration plots at 2-, 3- and 5-years post-surgery (Figure 15).

5.4.6 Minimum Required Sample Size for New Model Creation

All prognostic models had sample sizes larger than the minimum required sample size. This indicates adequate statistical power for model creation in all models (Tables 5a&5b).

5.4.7 Final Model Selection

Overall, the p-value selection model including pathological T stage is recommended for the prediction of recurrence free survival prognosis based on good discrimination, calibration, and parsimony. The nomogram for this model is shown (Figure 16). Risk scores would be determined for each prognostic variable, individual scores for each variable would be added and the total risk score would be mapped to its' corresponding recurrence free survival probability at 2-, 3- and 5-years post-surgery.

The cumulative baseline hazard function was determined to follow a Weibull distribution and have values of 0.096 at 2-years post-surgery, 0.129 at 3-years post-surgery, and 0.198 at 5-years post-surgery.

5.4.8 Risk Groups

Using the CKCis model, the risk score cutoffs for each risk-group classification method is given in Table 10. Risk group assignments for individual patients and the distribution of risk groups varied based on the method used for risk group classification. There was good distribution of patients into risk groups for all methods, with most patients classified as low-risk for recurrence using the mean-risk-score and face-validity methods. (Figure 17).

On Kaplan-Meier survival curves, considerable overlap is observed between the low and intermediate risk groups in all methods except the face-validity method. (Figure 18). Recurrence rates were observed to have varied based on the method used to classify patients into risk groups (Figures 19a-c). The distributions of prognostic variables by risk groups are presented in Tables 11 to 17 and Figures 20 to 28 below.

All risk group methods provided good discrimination, with the face validity method having the highest c-statistic (0.83). Similarly, all risk group methods provided good calibration with only minor under- and over-estimation of recurrences (Figure 29). The face-validity risk group method was felt to be optimal based on the combination of patient distribution, separation of survival-curves, label-to-risk concordance, discrimination, and calibration.

5.5 Discussion

In this study, the best performing prognostic model for recurrence free survival was the p-value selection model that incorporates pathological T stage (CKCis model). This model had an overall c-statistic of 0.85 and calibration slopes of 1.0. This model also had the highest proportion of patients within 5% and 10% predictive error and had net-benefit within clinically useful recurrence-probability thresholds.

In general, prognostic models for recurrence free survival including pathological T stage had better performance than those without pathological T stage. When results from internal validation using k-fold validation were compared to those from optimism corrected bootstrapping, the performance of the model was similar. Multiple imputation yielded slightly lower, but similar, c-statistics, c-index over time and calibration.

We previously published the performance of prognostic models for recurrence in our CKCis cohort. Not surprisingly, compared to previous models for recurrence, the present model had better discrimination and calibration. The better performance is expected given the model was generated and assessed from the same cohort. In general, comparisons between studies should be discouraged and model performance and comparisons are best assessed within an independent cohort where all predictor variables are available.

Based on external validation studies, the best performing models for ccRCC recurrence were from Mayo Clinic and Memorial Sloan Kettering Cancer Centre (14). A comparison of parameters from these models to the CKCis model is shown in Figure 30. Consistent prognostic factors across all models include tumour size, grade, and necrosis (15-17). However, there are a number of parameters unique to the CKCis model,

including surgical margin status, Charlson Co-morbidity Score, and pre-operative serum neutrophil concentration. The presence of a positive surgical margin (presence of aggregated tumour cells at the inked margin) is intuitively associated with recurrence risk, but has not been included in previous models (18). It is possible that more aggressive use of partial nephrectomy in contemporary patients may explain why this has become a stronger predictor. The CKCis model also includes neutrophils and sarcomatoid differentiation. Both findings have biologic plausibility (19-22). Charlson comorbidity score has not been included in previous models. The reasons that patient co-morbidity is associated with recurrence is unclear, but perhaps patients with greater comorbidity have lower host immune function or less tolerance to treatments (ie. comorbidity bias) (23).

5.5.1 Limitations

There were several potential limitations in this study. Some patients were excluded due to missing data and some potential prognostic markers (e.g. pre-operative serum calcium) could not be evaluated because too few patients had these tested prior to surgery. While the inclusion of contemporary patients was a strength, follow-up was relatively short, highlighting the need to revise prognostic models over time as more years of follow-up and data are accumulated.

5.5.2 Strengths

This study also has relative strengths. The CKCis cohort is representative of Canadian kidney cancer patients, comprising of roughly one-fifth of all kidney cancer patients in Canada (24). The sample sizes used in the creation of prognostic models in this study were greater than most sample sizes reported in pre-existing prognostic models. We used multiple approaches to prognostic model building which allowed for the selection of the best performing model and examine how model construction and performance varied across different methods. Sensitivity analyses using k-fold cross-validation and multiple imputation demonstrate how techniques used in data validation and missing data affect assessment of model performance. This study also compared the effect of including pathological T stage as a prognostic factor compared to solely including the components of pathological T stage without pathological T stage itself. This study was the first study to report values of the cumulative baseline hazard function for prognostic models constructed using the Cox Proportional Hazards model, allowing for more accurate assessment in external validation studies. The inclusion of multiple methods to create risk group classification systems to ensure that any variation due to differences in the definition of risk groups across different methods (i.e. Low-risk definition varying across different methods) is captured. The use of multiple methods in evaluating risk group classification systems allowed for clinically useful selection based on distribution, discrimination, calibration, and label concordance.

5.6 Conclusion

The p-value selection model including pathological T stage was the best performing model to predict recurrence free survival in Canadian ccRCC patients (CKCis model). The CKCis model had good discrimination and calibration. The results were robust as multiple sensitivity analyses revealed similar performance. The CKCIs models must now be externally validated in other populations to ensure clinical utility outside of Canadian patients. If externally valid, the CKCis risk groups could be applied broadly to clinical practice and research.

5.7 Other Information

There is no supplementary information available and this study did not receive any external funding.

5.8 References

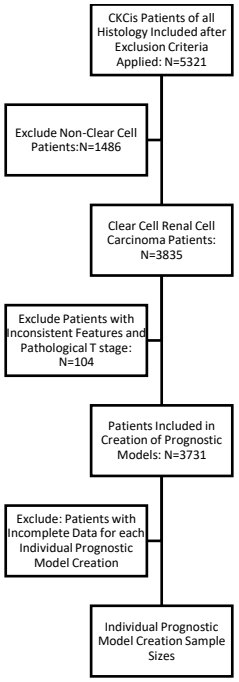
1. Vogenberg FR. Predictive and prognostic models: implications for healthcare decision-making in a modern recession. Am Health Drug Benefits. 2009 Sep;2(6):218-222.
2. Lenert MC, Matheny ME, Walsh CG. Prognostic models will be victims of their own success, unless…. J Am Med Inform Assoc. 2019 Dec 1;26(12):1645-1650. doi: 10.1093/jamia/ocz145.
3. Volpe A, Patard JJ. Prognostic factors in renal cell carcinoma. World J Urol. 2010 Jun;28(3):319-27. doi: 10.1007/s00345-010-0540-8.
4. Heng DY, Xie W, Regan MM, Warren MA, Golshayan AR, Sahi C, Eigl BJ, Ruether JD, Cheng T, North S, Venner P, Knox JJ, Chi KN, Kollmannsberger C, McDermott DF, Oh WK, Atkins MB, Bukowski RM, Rini BI, Choueiri TK. Prognostic factors for overall survival in patients with metastatic renal cell carcinoma treated with vascular endothelial growth factor-targeted agents: results from a large, multicenter study. J Clin Oncol. 2009 Dec 1;27(34):5794-9. doi: 10.1200/JCO.2008.21.4809.
5. Lam JS, Klatte T, Kim HL, Patard JJ, Breda A, Zisman A, Pantuck AJ, Figlin RA. Prognostic factors and selection for clinical studies of patients with kidney cancer. Crit Rev Oncol Hematol. 2008 Mar;65(3):235-62. doi: 10.1016/j.critrevonc.2007.08.003.
6. Kattan MW, Reuter V, Motzer RJ, Katz J, Russo P. A postoperative prognostic nomogram for renal cell carcinoma. J Urol. 2001 Jul;166(1):63-67.
7. Caetano SJ, Sonpavde G, Pond GR. C-statistic: A brief explanation of its construction, interpretation and limitations. Eur J Cancer. 2018 Feb; 90:130-132. doi: 10.1016/j.ejca.2017.10.027.
8. Correa AF, Jegede O, Haas NB, Flaherty KT, Pins MR, Messing EM, Manola J, Wood CG, Kane CJ, Jewett MAS, Dutcher JP, DiPaola RS, Carducci MA, Uzzo RG. Predicting Renal Cancer Recurrence: Defining Limitations of Existing Prognostic Models With Prospective Trial-Based Validation. J Clin Oncol. 2019 Aug 10;37(23):2062-2071. doi: 10.1200/JCO.19.00107.
9. Steyerberg EW. Clinical Prediction Models A Practical Approach to Development, Validation, and Updating. 2nd ed. Cham (Switzerland):Springer Nature Switzerland;c2019.558p.
10. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. Diagn Progn Res. 2019 Oct 4;3:18. doi: 10.1186/s41512-019-0064-7.

11. Riley RD, Snell KI, Ensor J, Burke DL, Harrell FE Jr, Moons KG, Collins GS. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. Stat Med. 2019 Mar 30;38(7):1276-1296. doi: 10.1002/sim.7992.
12. Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. BMC Med Res Methodol. 2013 Mar 6;13:33. doi: 10.1186/1471-2288-13-33.
13. Heng DY, Xie W, Regan MM, Harshman LC, Bjarnason GA, Vaishampayan UN, Mackenzie M, Wood L, Donskov F, Tan MH, Rha SY, Agarwal N, Kollmannsberger C, Rini BI, Choueiri TK. External validation and comparison with other models of the International Metastatic Renal-Cell Carcinoma Database Consortium prognostic model: a population-based study. Lancet Oncol. 2013 Feb;14(2):141-8. doi: 10.1016/S1470-2045(12)70559-4.
14. Rosiello G, Larcher A, Fallara G, Giancristofaro C, Martini A, Re C, Cei F, Musso G, Tian Z, Karakiewicz PI, Mottrie A, Bertini R, Salonia A, Necchi A, Raggi D, Briganti A, Montorsi F, Capitanio U. Head-to-head comparison of all the prognostic models recommended by the European Association of Urology Guidelines to predict oncologic outcomes in patients with renal cell carcinoma. Urol Oncol. 2022 Jun;40(6):271.e19-271.e27. doi: 10.1016/j.urolonc.2021.12.010.
15. Leibovich BC, Blute ML, Cheville JC, Lohse CM, Frank I, Kwon ED, Weaver AL, Parker AS, Zincke H. Prediction of progression after radical nephrectomy for patients with clear cell renal cell carcinoma: a stratification tool for prospective clinical trials. Cancer. 2003 Apr 1;97(7):1663-1671. doi: 10.1002/cncr.11234.
16. Sorbellini M, Kattan MW, Snyder ME, Reuter V, Motzer R, Goetzl M, McKiernan J, Russo P. A postoperative prognostic nomogram predicting recurrence for patients with conventional clear cell renal cell carcinoma. J Urol. 2005 Jan;173(1):48-51. doi: 10.1097/01.ju.0000148261.19532.2c.
17. Leibovich BC, Lohse CM, Cheville JC, Zaid HB, Boorjian SA, Frank I, Thompson RH, Parker WP. Predicting Oncologic Outcomes in Renal Cell Carcinoma After Surgery. Eur Urol. 2018 May;73(5):772-780. doi: 10.1016/j.eururo.2018.01.005.
18. Bensalah K, Pantuck AJ, Rioux-Leclercq N, Thuret R, Montorsi F, Karakiewicz PI, Mottet N, Zini L, Bertini R, Salomon L, Villers A, Soulie M, Bellec L, Rischmann P, De la Taille A, Avakian R, Crepel M, Ferriere JM, Bernhard JC, Dujardin T, Pouliot F, Rigaud J, Pfister C, Albouy B, Guy L, Joniau S, van Poppel H, Lebret T, Culty T, Saint F, Zisman A, Raz O, Lang H, Spie R, Wille A, Roigas J, Aguilera A, Rambeaud B, Martinez Piñeiro L, Nativ O, Farfara R, Richard F, Roupret M, Doehn C, Bastian PJ, Muller SC, Tostain J, Belldegrun AS, Patard JJ. Positive surgical margin appears to have negligible impact on survival of renal cell carcinomas treated by nephron-sparing surgery. Eur Urol. 2010 Mar;57(3):466-471. doi: 10.1016/j.eururo.2009.03.048.
19. Shao Y, Wu B, Jia W, Zhang Z, Chen Q, Wang D. Prognostic value of pretreatment neutrophil-to-lymphocyte ratio in renal cell carcinoma: a systematic review and meta-analysis. BMC Urol. 2020 Jul 6;20(1):90. doi: 10.1186/s12894-020-00665-8.
20. Shen M, Hu P, Donskov F, Wang G, Liu Q, Du J. Tumor-associated neutrophils as a new prognostic factor in cancer: a systematic review and meta-analysis. PLoS One. 2014 Jun 6;9(6):e98259. doi: 10.1371/journal.pone.0098259.
21. Shuch B, Bratslavsky G, Linehan WM, Srinivasan R. Sarcomatoid renal cell carcinoma: a comprehensive review of the biology and current treatment strategies. Oncologist. 2012;17(1):46-54. doi: 10.1634/theoncologist.2011-0227.

22. Xiong S, Dong L, Cheng L. Neutrophils in cancer carcinogenesis and metastasis. J Hematol Oncol. 2021 Oct 21;14(1):173. doi: 10.1186/s13045-021-01187-y.
23. Kang HW, Kim SM, Kim WT, Yun SJ, Lee SC, Kim WJ, Hwang EC, Kang SH, Hong SH, Chung J, Kwon TG, Kim HH, Kwak C, Byun SS, Kim YJ; KORCC (KOrean Renal Cell Carcinoma) Group. The age-adjusted Charlson comorbidity index as a predictor of overall survival of surgically treated non-metastatic clear cell renal cell carcinoma. J Cancer Res Clin Oncol. 2020 Jan;146(1):187-196. doi: 10.1007/s00432-019-03042-7.
24. Tajzler, C., Tanguay, S., Mallick, R., Ahrens, B., Ly, T. L., Breau, R. H., Basappa, N. S., Kapoor, A., Heng, D. Y., Pouliot, F., Finelli, A., Lavallee, L. T., So, A. I., Drachenberg, D. E., Soulieres, D., Bjarnason, G., Richard, P. O., Maloni, R., Power, N. E., Haan, M., & Wood, L. A. Determining generalizability of the Canadian Kidney Cancer information system (CKCis) to the entire Canadian kidney cancer population. Can Urol Assoc J. 2020 Jun 16; 14(10). doi: 10.5489/cuaj.6716.

5.9 Figures and Tables

Figure 1: Summary of Patients Included for Analysis



Abbreviations: N=number of patients; CKCis=Canadian Kidney Cancer Information System

Table 1: Patient Characteristics of Patients Included in the CKCis Cohort for Analysis[1]

Features	**N (%)**	**Mean (95%CI); Median (IQR)**
Age at Surgery (years)		60.8 (60.4-61.2); 61.3 (53.2-69.1)
Sex		
Female	1283 (34.4)	
Male	2448 (65.6)	
Local Symptom (Yes)	1029 (27.6)	
Systemic Symptom (Yes)	226 (6.1)	

Charlson Co-morbidity Score		4.4 (4.4-4.5); 4.0 (3.0-5.0)
Ever Smoker (Yes)	1943 (58.6)	
Diabetic (Yes)	824 (22.8)	
Hypertension (Yes)	1900 (52.7)	
Hemoglobin (g/L)		134.3 (133.5-134.9); 137.0 (122.0-148.0)
Neutrophil (x 10^9 cells/L)		6.0 (5.8-6.2); 4.9 (3.5-7.2)
Platelet (x 10^9 cells/L)		247.6 (243.9-251.4); 234.0 (193.0-281.0)
Creatinine (µmol/L)		91.2 (89.2-93.2); 81.0 (69.0-98.0)
Extent Partial Radical	 1798 (48.2) 1933 (51.8)	
Positive Margin (Yes)	248 (6.9)	
Tumour Grade G1 G2 G3 G4	 241 (6.5) 1642 (44.6) 1353 (36.7) 449 (12.2)	
Necrosis (Yes)	849(22.8)	
Sarcomatoid (Yes)	129(3.5)	
Pathological T stage T1a T1b T2a T2b T3a T3b T3c T4	 1559(41.8) 757(20.3) 179(4.8) 60(1.6) 1059(28.4) 79(2.1) 20(0.5) 18(0.5)	
Pathological N stage NX N0 N1	 3008(80.6) 661(17.7) 62(1.7)	
Tumour Size (cm)		5.2(5.1-5.4);4.5(3.0-7.0)
Invasion Adrenal Gland (Yes)	15(0.4)	
Invasion Perinephric Fat (Yes)	417(11.8)	
Invasion Sinus Fat (Yes)	546(15.5)	
Tumour Thrombus No Level 0 Level 1-4	 3046(82.9) 469(12.8) 160(4.4)	
Extension Outside Gerota's fascia (Yes)	8(0.2)	

Abbreviations: N=number of patients; CI=confidence interval; IQR=interquartile range; g=gram; L=litre; µmol=micromole; cm=centimeter; T1a=pathological T1a stage; T1b=pathological T1b stage; T2a=pathological T2a stage; T2b=pathological T2b stage; T3a=pathological T3a stage; T3b=pathological T3b stage; T3c=pathological T3c stage; T4=pathological T4 stage; G1= tumour grade 1; G2=tumour grade 2;G3=tumour grade 3; G4=tumour grade 4; N0=pathological N0 stage; NX=nodal resection not assessed; N1=pathological N1 stage

[1]Data provided as count (%) or mean (95%CI);median (IQR) as needed

Figure 2: Recurrence Free Survival in Patients Included

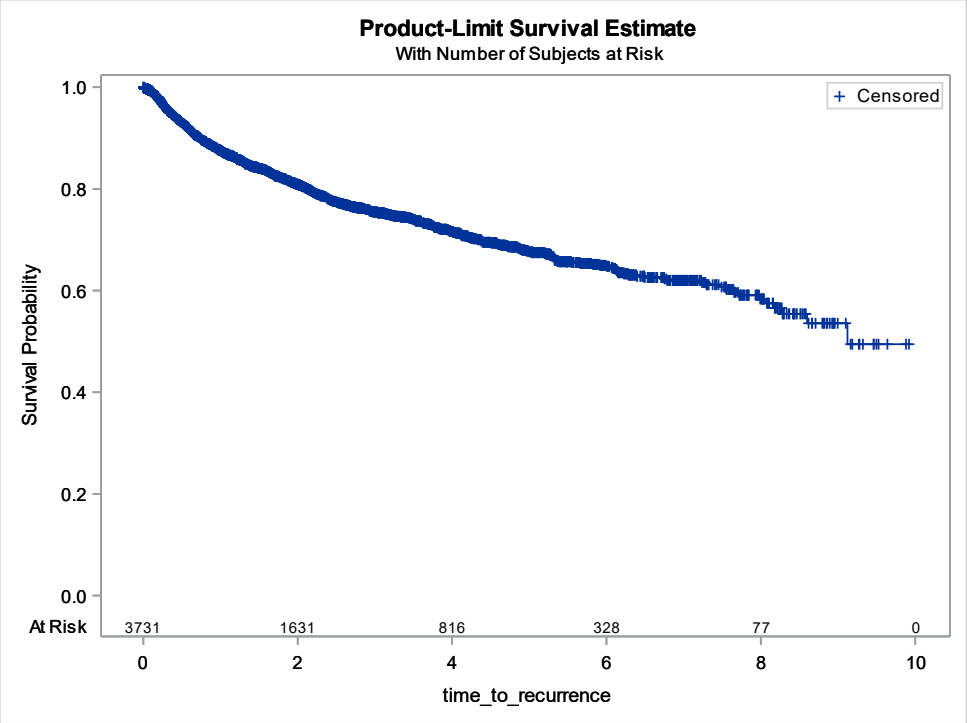


Table 2: Univariate Recurrence Free Survival Models

Features	**Hazard Ratio (95%CI)**
Age at Surgery (years)	1.02(1.02-1.03)
Sex Female Male	 Ref 1.67(1.41-1.96)
Local Symptom (Yes)	1.67(1.44-1.93)
Systemic Symptom (Yes)	2.55(2.07-3.16)
Charlson Co-morbidity Score	1.19(1.15-1.23)
Ever Smoker (Yes)	0.88(0.76-1.03)
Diabetic (Yes)	1.32(1.12-1.55)
Hypertension (Yes)	1.37(1.18-1.59)
Hemoglobin (g/L)	0.98(0.98-0.98)
Neutrophil (x 10^9 cells/L)	1.02(1.00-1.04)
Platelet (x 10^9 cells/L)	1.00(1.00-1.00)
Creatinine (µmol/L)	1.00(1.00-1.00)
Positive Margin (Yes)	2.09(1.67-2.62)
Tumour Grade G1 G2	 Ref 1.27(0.78-2.06)

G3	3.37(2.10-5.42)
G4	11.35(7.03-18.32)
Necrosis (Yes)	3.69(3.20-4.26)
Sarcomatoid (Yes)	6.00(4.75-7.57)
Pathological T stage	
T1a	Ref
T1b	3.64(2.67-4.97)
T2a	8.84(6.23-12.57)
T2b	16.43(10.71-25.20)
T3a	13.52(10.33-17.69)
T3b	23.86(16.22-35.10)
T3c	25.41(13.95-46.28)
T4	35.70(18.76-67.95)
Pathological N stage	
NX	Ref
N0	2.47(2.10-2.89)
N1	12.66(9.32-17.22)
Tumour size (cm)	1.27(1.25-1.29)
Invasion Adrenal Gland (Yes)	4.79(2.39-9.63)
Invasion Perinephric Fat (Yes)	4.04(3.45-4.75)
Invasion Sinus Fat (Yes)	3.74(3.20-4.37)
Tumour Thrombus	
No	Ref
Level 0	3.47(2.92-4.12)
Level 1-4	5.90(4.70-7.41)
Extension outside Gerota's fascia (Yes)	15.16(7.18-32.02)

Abbreviations: N=number of patients; CI=confidence interval; g=gram; L=litre; µmol=micromole; cm=centimeter; T1a=pathological T1a stage; T1b=pathological T1b stage; T2a=pathological T2a stage; T2b=pathological T2b stage; T3a=pathological T3a stage; T3b=pathological T3b stage; T3c=pathological T3c stage; T4=pathological T4 stage; G1= tumour grade 1; G2=tumour grade 2;G3=tumour grade 3; G4=tumour grade 4; N0=pathological N0 stage; NX=nodal resection not assessed; N1=pathological N1 stage

Table 3a: Parameters and Hazard Ratios (95%CI) for Recurrence Free Survival Including Pathological T stage [1]

Risk Factor	**Forward Selection (p-Value Selection)**	**Backwards Selection**	**Stepwise Selection**	**Backwards Bootstrap Selection[1]**	**Elastic Net Selection[1]**
Tumour size (cm)	1.13(1.09-1.17)	1.13(1.09-1.17)	1.13(1.09-1.17)	1.14	17.42
Neutrophil (x 10^9 cells/L)	1.03(1.01-1.05)	1.03(1.01-1.05)	1.03(1.01-1.05)	Not Applicable	Not Applicable
Charlson Co-morbidity Score	1.19(1.13-1.25)	1.19(1.13-1.25)	1.19(1.13-1.25)	Not Applicable	Not Applicable
Pathological N Stage					
NX	Ref	Ref	Ref	Ref	Ref
N0	1.12(0.90-1.40)	1.12(0.90-1.40)	1.12(0.90-1.40)	1.10	1.26
N1	2.71(1.72-4.27)	2.71(1.72-4.27)	2.71(1.72-4.27)	3.89	1.58
Pathological T Stage					
T1	Ref	Ref	Ref	Ref	Ref

T2	2.13(1.40-3.23)	2.13(1.40-3.23)	2.13(1.40-3.23)	2.39	1.34
T3a	2.75(2.03-3.73)	2.75(2.03-3.73)	2.75(2.03-3.73)	2.53	1.80
T3b	3.38(2.08-5.50)	3.38(2.08-5.50)	3.38(2.08-5.50)	3.49	2.41
T3c	4.44(2.18-9.02)	4.44(2.18-9.02)	4.44(2.18-9.02)	3.50	3.23
T4	2.41(0.99-5.87)	2.41(0.99-5.87)	2.41(0.99-5.87)	2.01	4.32
Tumour Grade					
G1	Ref	Ref	Ref	Ref	Ref
G2	1.40(0.64-3.07)	1.40(0.64-3.07)	1.40(0.64-3.07)	1.11	1.61
G3	2.24(1.04-4.84)	2.24(1.04-4.84)	2.24(1.04-4.84)	1.66	2.61
G4	3.96(1.80-8.71)	3.96(1.80-8.71)	3.96(1.80-8.71)	2.93	4.21
Necrosis					
No	Ref	Ref	Ref	Ref	Ref
Yes	1.29(1.04-1.61)	1.29(1.04-1.61)	1.29(1.04-1.61)	1.38	1.38
Positive Margin				Not Applicable	Not Applicable
No	Ref	Ref	Ref		
Yes	1.56(1.16-2.10)	1.56(1.16-2.10)	1.56(1.16-2.10)		
Sarcomatoid Features					
No	Ref	Ref	1.56(1.16-2.02)	Ref	Ref
Yes	1.56(1.16-2.02)	1.56(1.16-2.02)		1.41	1.40
Systemic Symptom	Not Applicable	Not Applicable	Not Applicable	Not Applicable	
No					Ref
Yes					1.05

Abbreviations: N=number of patients; CI=confidence interval; L=litre; cm=centimeter; T1=pathological T1 stage; T2=pathological T2 stage; T3a=pathological T3a stage; T3b=pathological T3b stage; T3c=pathological T3c stage; T4=pathological T4 stage; G1= tumour grade 1; G2=tumour grade 2;G3=tumour grade 3; G4=tumour grade 4; N0=pathological N0 stage; NX=nodal resection not assessed; N1=pathological N1 stage

[1]95% Confidence Interval not available for this parameter selection methods.

Table 3b: Parameters and Risk Score for Recurrence Free Survival Including Pathological T stage

Risk Factor	**p-Value Selection**	**Backwards Bootstrap Selection**	**Elastic Net Selection**
Tumour size (cm)	4.5pts/cm for each cm above 0.7cm between 0.7cm-23cm	4.5 pts/cm for each cm above 0.7cm between 0.7cm-23cm	4.5pts/cm for each cm above 0.7cm between 0.7cm-23cm
Neutrophil (x 10^9 cells/L)	1.1pts/(x 10^9cells/L) for each unit above 0.367	Not Applicable	Not Applicable
Charlson Co-morbidity Score	6.5pts/Charlson Co-morbidity Score above 2 pts	Not Applicable	Not Applicable

Pathological N Stage			
NX	0pts	0pts	0pts
N0	4pts	3pts	8pts
N1	37pts	48pts	16pts
Pathological T Stage			
T1	0pts	0pts	0pts
T2	28pts	31pts	10pts
T3a	38pts	33pts	20pts
T3b	46pts	44pts	31pts
T3c	56pts	44pts	41pts
T4	33pts	25pts	51pts
Tumour Grade			
G1	0pts	0pts	0pts
G2	13pts	4pts	17pts
G3	30pts	18pts	34pts
G4	52pts	38pts	50pts
Necrosis			
No	0pts	0pts	0pts
Yes	10pts	11pts	11pts
Positive Margin		Not Applicable	Not Applicable
No	0pts		
Yes	17pts		
Sarcomatoid Features			
No	0pts	0pts	0pts
Yes	14pts	12pts	12pts
Systemic Symptom	Not Applicable	Not Applicable	
No			0pts
Yes			2pts

Abbreviations: N=number of patients; L=litre; cm=centimeter; T1=pathological T1 stage; T2=pathological T2 stage; T3a=pathological T3a stage; T3b=pathological T3b stage; T3c=pathological T3c stage; T4=pathological T4 stage; G1= tumour grade 1; G2=tumour grade 2;G3=tumour grade 3; G4=tumour grade 4; N0=pathological N0 stage; NX=nodal resection not assessed; N1=pathological N1 stage;pts=Points

Table 4a: Parameters and Hazard Ratios (95%CI) for Recurrence Free Survival Excluding Pathological T stage[1]

Risk Factor	**Forward Selection (p-Value Selection)**	**Backwards Selection**	**Stepwise Selection**	**Backwards Bootstrap Selection[1]**	**Elastic Net Selection[1]**
Tumour size (cm)	1.18(1.15-1.22)	1.18(1.15-1.22)	1.18(1.15-1.22)	1.18	1.17
Neutrophil (x 10^9 cells/L)	1.03(1.02-1.05)	1.03(1.02-1.05)	1.03(1.02-1.05)	Not Applicable	Not Applicable
Charlson Co-morbidity Score	1.21(1.14-1.27)	1.21(1.14-1.27)	1.21(1.14-1.27)	Not Applicable	Not Applicable
Pathological N Stage					

NX	Ref	Ref	Ref	Ref	Ref
N0	1.23(0.98-1.54)	1.23(0.98-1.54)	1.23(0.98-1.54)	1.05	1.28
N1	2.96(1.85-4.75)	2.96(1.85-4.75)	2.96(1.85-4.75)	3.19	1.65
Tumour Grade					
G1	Ref	Ref	Ref	Ref	Ref
G2	1.29(0.59-2.81)	1.29(0.59-2.81)	1.29(0.59-2.81)	1.14	1.67
G3	2.43(1.13-5.23)	2.43(1.13-5.23)	2.43(1.13-5.23)	1.94	2.79
G4	4.56(2.08-10.00)	4.56(2.08-10.00)	4.56(2.08-10.00)	3.65	4.67
Necrosis					
No	Ref	Ref	Ref	Ref	Ref
Yes	1.40(1.11-1.76)	1.40(1.11-1.76)	1.40(1.11-1.76)	1.40	1.42
Positive Margin				Not Applicable	Not Applicable
No	Ref	Ref	Ref		
Yes	1.87(1.40-2.49)	1.87(1.40-2.49)	1.87(1.40-2.49)		
Sarcomatoid					
No	Ref	Ref	Ref	Ref	Ref
Yes	1.47(1.07-2.03)	1.47(1.07-2.03)	1.47(1.07-2.03)	1.38	1.35
Invasion Adrenal Gland				Not Applicable	Not Applicable
No	Ref	Ref	Ref		
Yes	3.21(1.06-9.68)	3.21(1.06-9.68)	3.21(1.06-9.68)		
Invasion Perinephric Fat					
No	Ref	Ref	Ref	Ref	Ref
Yes	1.32(1.04-1.68)	1.32(1.04-1.68)	1.32(1.04-1.68)	1.38	1.33
Extension Outside Gerota's fascia				Not Applicable	
No	Ref	Ref	Ref		Ref
Yes	4.19(1.60-10.98)	4.19(1.60-10.98)	4.19(1.60-10.98)		1.54
Tumour Thrombus	Not Applicable	Not Applicable	Not Applicable		
No				Ref	Ref
Level 0				1.18	1.03
Level 1-4				1.53	1.03
Local Symptom	Not Applicable	Not Applicable	Not Applicable	Not Applicable	
No					Ref
Yes					1.05
Systemic Symptom	Not Applicable	Not Applicable	Not Applicable	Not Applicable	
No					Ref
Yes					1.10

Invasion Sinus Fat No Yes	Not Applicable	Not Applicable	Not Applicable	Not Applicable	 Ref 1.18

Abbreviations: N=number of patients; CI=confidence interval; L=litre; cm=centimeter; T1=pathological T1 stage; T2=pathological T2 stage; T3a=pathological T3a stage; T3b=pathological T3b stage; T3c=pathological T3c stage; T4=pathological T4 stage; G1= tumour grade 1; G2=tumour grade 2;G3=tumour grade 3; G4=tumour grade 4; N0=pathological N0 stage; NX=nodal resection not assessed; N1=pathological N1 stage

[1]95% Confidence Interval not available for this parameter selection methods.

Table 4b: Parameters and Risk Scores for Recurrence Free Survival Excluding Pathological T stage

Risk Factor	**p-Value Selection**	**Backwards Bootstrap Selection**	**Elastic Net Selection**
Tumour size (cm)	4.5pts/cm for each cm above 0.7cm between 0.7cm-23cm	4.5pts/cm for each cm above 0.7cm between 0.7cm-23cm	0.5pts/cm for each cm above 0.7cm between 0.7cm-23cm
Neutrophil (x 10^9 cells/L)	0.9pts/(x 10^9cells/L) for each unit above 0.367	Not Applicable	Not Applicable
Charlson Co-morbidity Score	5pts/Charlson Co-morbidity Score above 2 pts	Not Applicable	Not Applicable
Pathological N Stage NX N0 N1	 0pts 5pts 29pts	 0pts 1pts 32pts	 0pts 16pts 32pts
Tumour Grade G1 G2 G3 G4	 0pts 7pts 24pts 41pts	 0pts 4pts 18pts 35pts	 0pts 33pts 67pts 100pts
Necrosis No Yes	 0pts 9pts	 0pts 9pts	 0pts 23pts
Positive Margin No Yes	 0pts 17pts	Not Applicable	Not Applicable
Sarcomatoid No Yes	 0pts 10pts	 0pts 9pts	 0pts 19pts
Invasion Adrenal Gland No Yes	 31pts 0pts	Not Applicable	Not Applicable

Invasion Perinephric Fat			
No	0pts	0pts	0pts
Yes	8pts	6pts	18pts
Extension Outside Gerota's fascia		Not Applicable	
No	0pts		0pts
Yes	39pts		28pts
Tumour Thrombus	Not Applicable		
No		0pts	0pts
Level 0		5pts	1pts
Level 1-4		12pts	2pts
Local Symptom	Not Applicable	Not Applicable	
No			0pts
Yes			3pts
Systemic Symptom	Not Applicable	Not Applicable	
No			0pts
Yes			6pts
Invasion Sinus Fat	Not Applicable	Not Applicable	
No			0pts
Yes			11pts

Abbreviations: N=number of patients; L=litre; cm=centimeter; T1=pathological T1 stage; T2=pathological T2 stage; T3a=pathological T3a stage; T3b=pathological T3b stage; T3c=pathological T3c stage; T4=pathological T4 stage; G1= tumour grade 1; G2=tumour grade 2;G3=tumour grade 3; G4=tumour grade 4; N0=pathological N0 stage; NX=nodal resection not assessed; N1=pathological N1 stage;pts=Points

Table 5a: Overall C-statistic, Calibration Slopes and Sample Sizes in Prognostic Models Built Including Pathological T Stage

	Overall C-statistic	**Calibration Slope 2-years post-surgery**	**Calibration Slope 3-years post-surgery**	**Calibration Slope 5-years post-surgery**	**Sample Size in Model Development**	**Minimum Required Sample Size in Model Development at 5-years**
p-Value Selection	0.8510	1.0030	1.0003	0.9997	2555	261
Backward Bootstrap Selection	0.8368	0.9986	0.9995	0.9995	3666	183
Elastic Net Penalization	0.8336	1.0012	0.9995	0.9961	3666	228

Table 5b: Overall C-statistic, Calibration Slopes and Sample Sizes in Prognostic Models Built Excluding Pathological T Stage

	Overall C-statistic	**Calibration Slope 2-years post-surgery**	**Calibration Slope 3-years post-surgery**	**Calibration Slope 5-years post-surgery**	**Sample Size in Model Development**	**Minimum Required Sample Size in Model Development at 5-years**
p-Value Selection	0.8432	0.9942	0.9954	0.9978	2441	328

Backward Bootstrap Selection	0.8305	0.9990	1.0012	0.9951	3452	231
Elastic Net Penalization	0.7930	0.9995	1.0024	0.9998	3452	459

Figure 3: C-Index over Time in Prognostic Models Including and Excluding Pathological T Stage

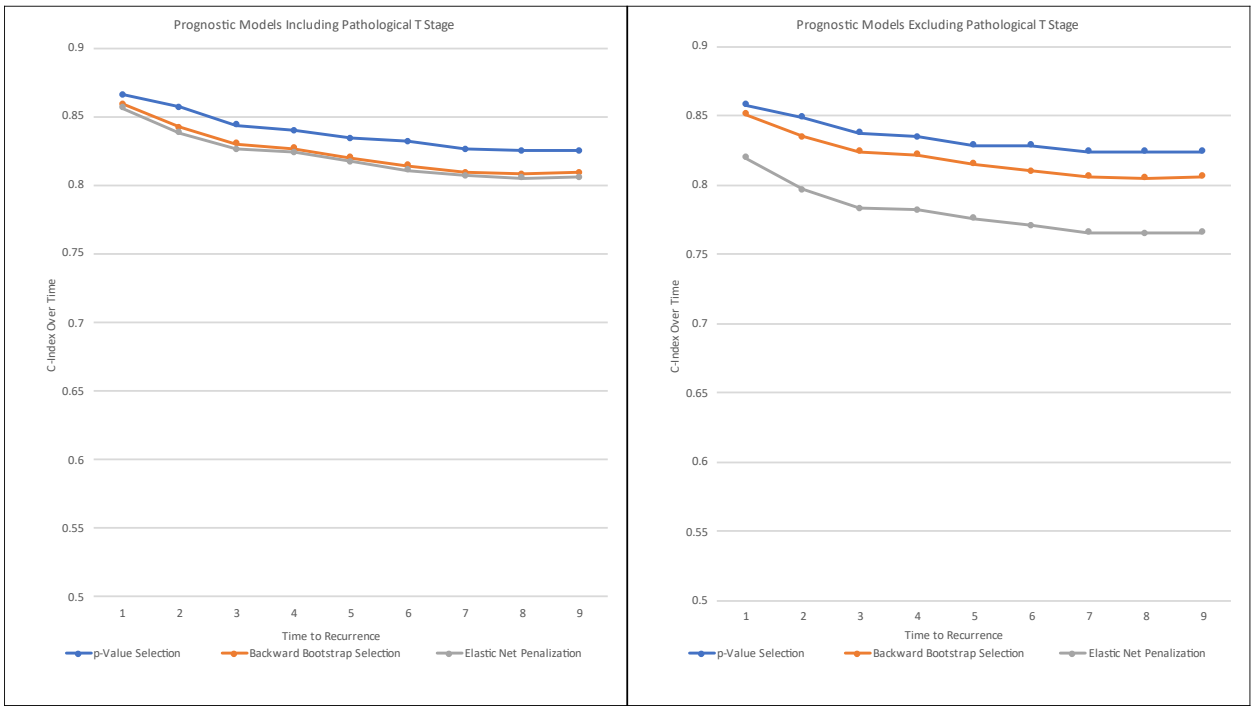


Figure 4: Local Regression (LOESS) Smoothed Calibration Curve for Recurrence Free Survival at 2-years Post-Surgery

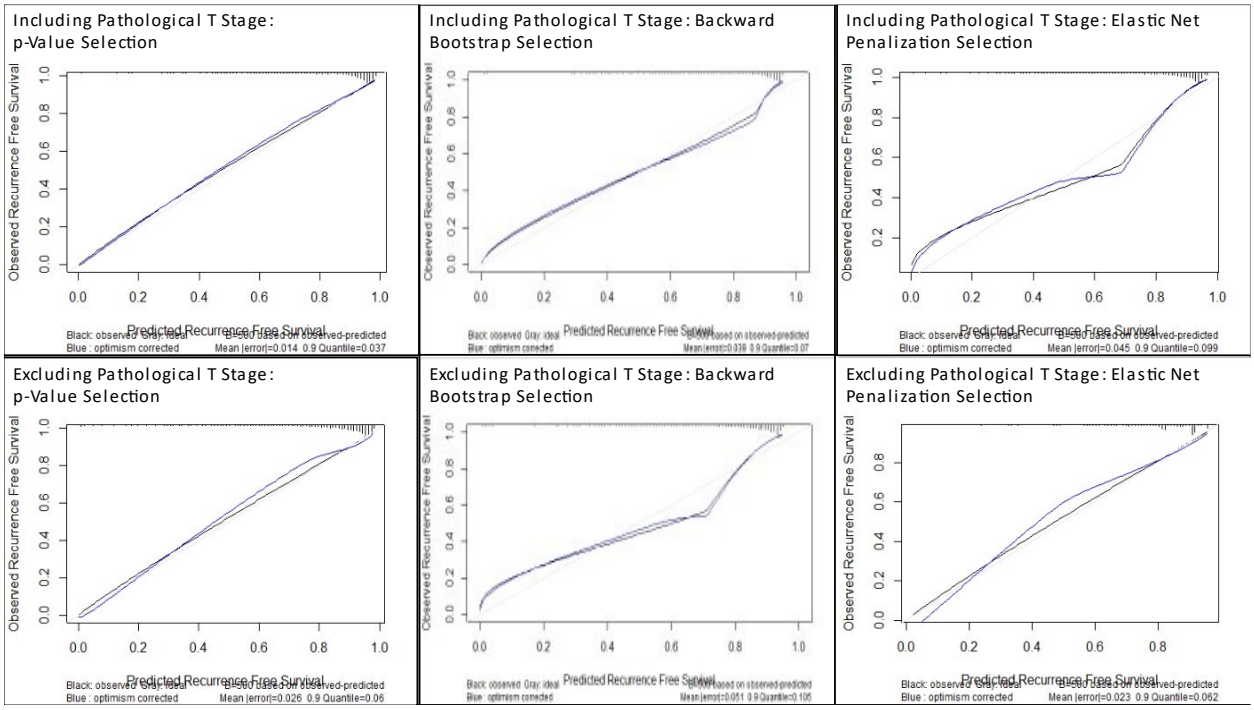


Figure 5: Local Regression (LOESS) Smoothed Calibration Curve for Recurrence Free Survival at 3-years Post-Surgery

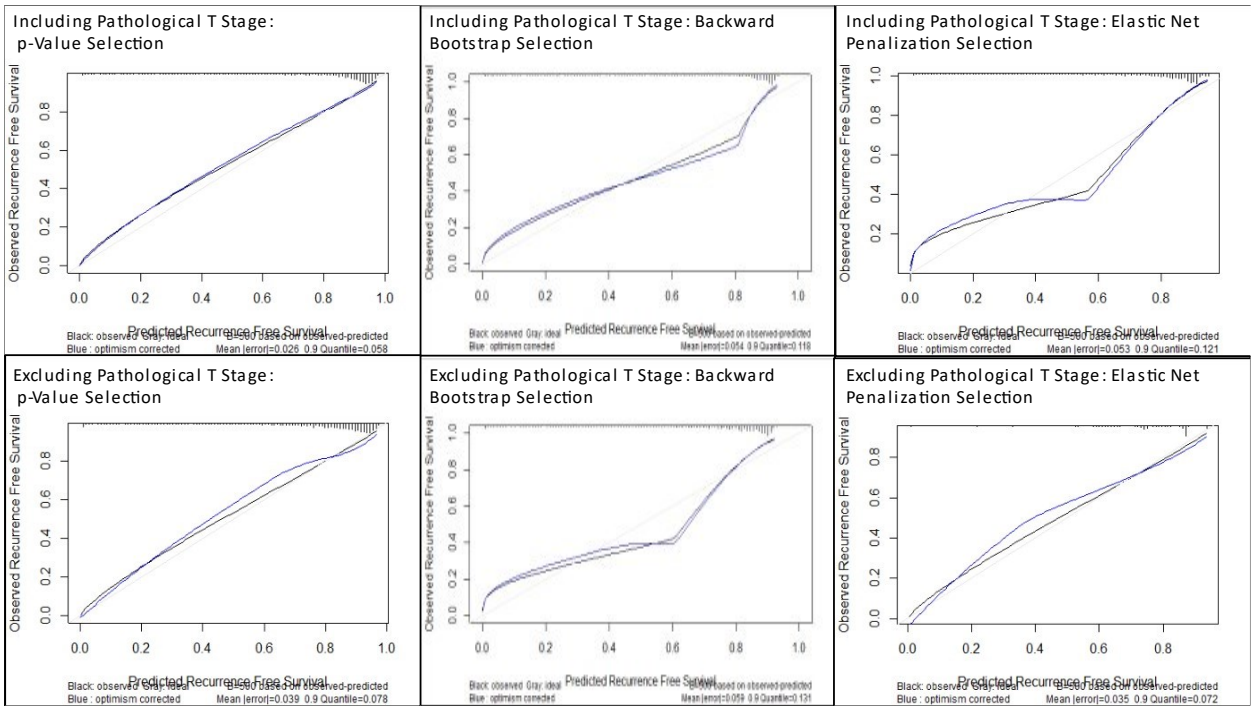


Figure 6: Local Regression (LOESS) Smoothed Calibration Curve for Recurrence Free Survival at 5-years Post-Surgery

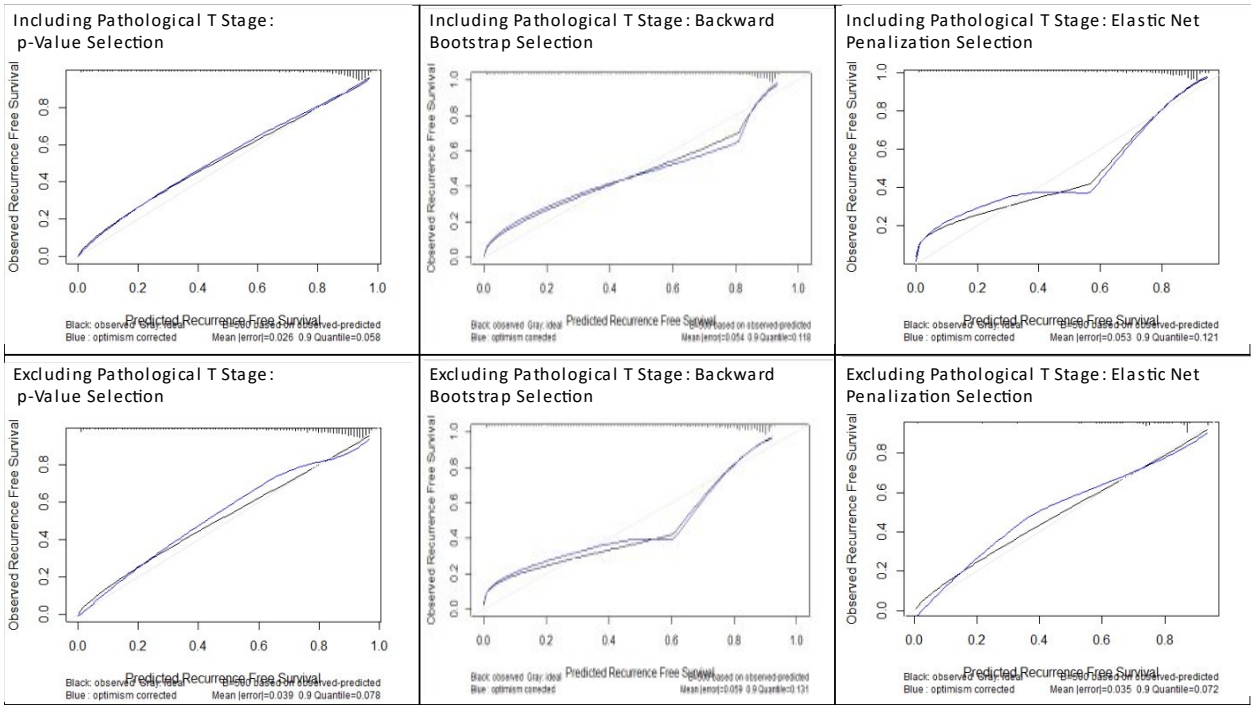


Table 6a: Decision Curve Analysis and Agreement within 5% and 10% for Prognostic Models Including Pathological T Stage

	Proportion of Patients with Agreement Within Absolute Range						
	2-years post-surgery		**3-years post-surgery**		**5-years post-surgery**		
Prognostic Model	**Within 5%**	**Within 10%**	**Within 5%**	**Within 10%**	**Within 5%**	**Within 10%**	**Recurrence Risk Threshold for Clinical Utility**
p-Value Selection	100%	100%	100%	100%	56%	100%	0%-35.5%
Backward Bootstrap Selection	100%	100%	52%	100%	35%	82%	0%-37.5%
Elastic Net Selection	65%	91%	43%	87%	39%	64%	0%-37.5%

Table 6b: Decision Curve Analysis and Agreement within 5% and 10% for Prognostic Models Excluding Pathological T Stage

	Proportion of Patients with Agreement Within Absolute Range						
	2-years post-surgery		**3-years post-surgery**		**5-years post-surgery**		
Prognostic Model	**Within**	**Within**	**Within**	**Within**	**Within**	**Within**	**Recurrence Risk Threshold for Clinical Utility**

	5%	10%	5%	10%	5%	10%	
p-Value Selection	100%	100%	69%	100%	44%	100%	0%-35.5%
Backward Bootstrap Selection	53%	92%	46%	87%	41%	68%	0%-37.5%
Elastic Net Selection	59%	85%	58%	100%	56%	88%	0-31.25%

Figure 7: Decision Curve Analysis for Prognostic Models Including and Excluding Pathological T Stage

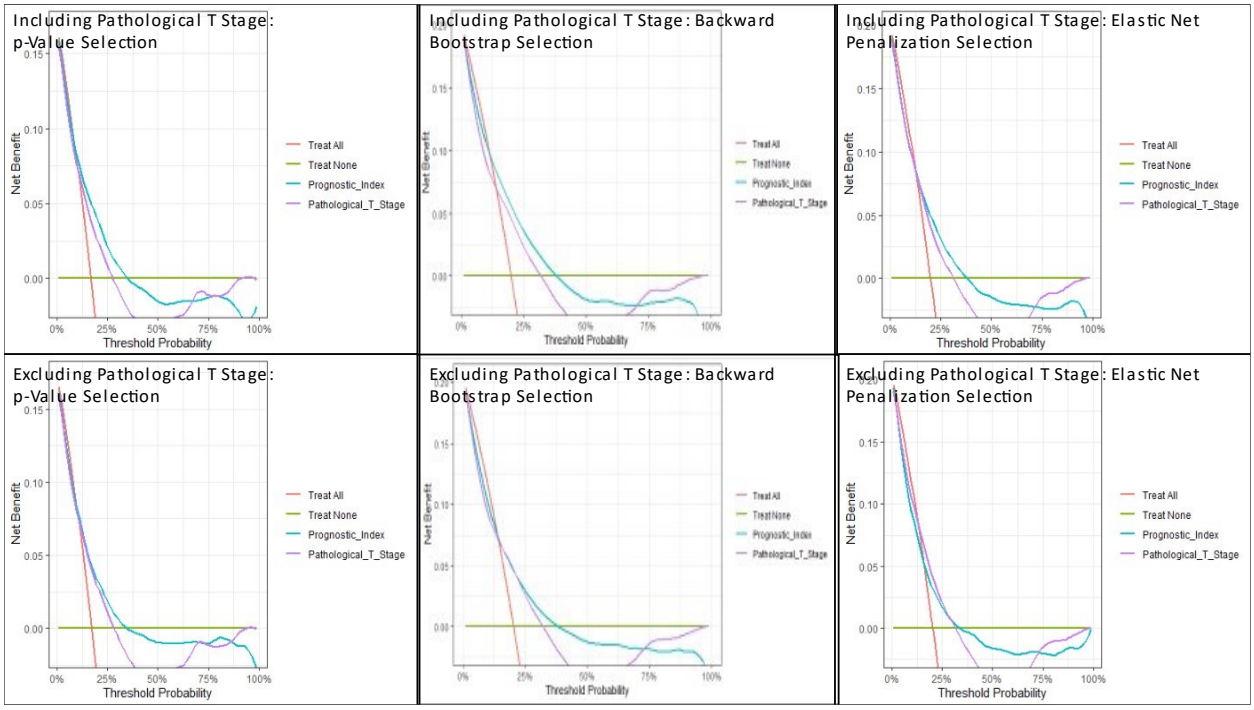


Table 7: Prognostic Models Developed When Entry p-value=0.2 and Exit p=value=0.1 Including Pathological T Stage

Parameters	Forward Selection		Backward Selection		Stepwise Selection	
	Hazard Ratio (95%CI)	Risk Score Points	Hazard Ratio (95%CI)	Risk Score Points	Hazard Ratio (95%CI)	Risk Score Points
Tumour size (cm)	1.12(1.08-1.16)	3.7pts/cm for each cm above 0.7cm between 0.7cm-23cm	1.13(1.09-1.16)	4.5pts/cm for each cm above 0.7cm between 0.7cm-23cm	No new model as same parameters selected as Backward Selection model including pathological T stage	
Neutrophil (x 10^9 cells/L)	1.03(1.01-1.05)	1.0pts/(x 10^9cells/L) for each unit above 0.367	1.03(1.01-1.05)	1.1pts/(x 10^9cells/L) for each unit above 0.367		
Platelet (x 10^9 cells/L)	1.00(1.00-1.00)	0.03pts/(x 10^9cells/L)	Not Applicable	Not Applicable		

		for each unit above 19			
Charlson Co-morbidity Score	1.21(1.14-1.28)	6.1pts/ Charlson Co-morbidity Score above 2 pts	1.19(1.13-1.26)	6.7pts/ Charlson Co-morbidity Score above 2 pts	
Pathological N Stage					
NX	Ref	0pts	Ref	0pts	
N0	1.12(0.90-1.41)	4pts	1.13(0.90-1.41)	5pts	
N1	2.83(1.80-4.48)	34pts	2.74(1.74-4.31)	38pts	
Pathological T Stage					
T1	Ref	0pts	Ref	0pts	
T2	2.20(1.44-3.35)	26pts	2.18(1.43-3.32)	30pts	
T3a	2.72(2.01-3.69)	32pts	2.74(2.03-3.71)	38pts	
T3b	3.28(2.02-5.33)	38pts	3.34(2.06-5.43)	46pts	
T3c	4.63(2.28-9.42)	50pts	4.46(2.19-9.07)	57pts	
T4	2.34(0.94-5.79)	28pts	3.78(1.71-8.33)	36pts	
Gender					
Female	Ref	0pts	Ref	0pts	
Male	1.30(1.04-1.62)	8pts	1.23(0.99-1.53)	8pts	
Tumour Grade					
G1	Ref	0pts	Ref	0pts	
G2	1.37(0.63-3.00)	10pts	1.38(0.63-3.02)	12pts	
G3	2.15(0.99-4.64)	25pts	2.17(1.00-4.70)	30pts	
G4	3.56(1.61-7.87)	41pts	3.78(1.71-8.33)	51pts	
Necrosis					
No	Ref	0pts	Ref	0pts	
Yes	1.31(1.05-1.64)	9pts	1.30(1.04-1.62)	10pts	
Positive Margin					
No	Ref	0pts	Ref	0pts	
Yes	1.55(1.15-2.08)	14pts	1.53(1.14-2.06)	16pts	
Sarcomatoid					

No Yes	Ref 1.44(1.04-1.99)	0pts 12pts	Ref 1.51(1.09-2.07)	0pts 16pts	
Diabetic No Yes	 1.22(0.96-1.55) Ref	 6pts 0pts	Not Applicable	Not Applicable	
Hypertension No Yes	 Ref 1.15(0.93-1.41)	 0pts 4pts	Not Applicable	Not Applicable	

Abbreviations: N=number of patients; CI=confidence interval; L=litre; cm=centimeter; T1=pathological T1 stage; T2=pathological T2 stage; T3a=pathological T3a stage; T3b=pathological T3b stage; T3c=pathological T3c stage; T4=pathological T4 stage; G1= tumour grade 1; G2=tumour grade 2;G3=tumour grade 3; G4=tumour grade 4; N0=pathological N0 stage; NX=nodal resection not assessed; N1=pathological N1 stage;pts=Points

Table 8: Overall C-statistic, Calibration Slopes and Sample Sizes in Prognostic Models Built When Entry p-value=0.2 and Exit p=value=0.1

	Overall C-statistic	**Calibration Slope 2-years post-surgery**	**Calibration Slope 3-years post-surgery**	**Calibration Slope 5-years post-surgery**	**Sample Size in Model Development**	**Minimum Required Sample Size in Model Development at 5-years**
Prognostic Models Including Pathological T stage						
Forward Selection	0.8529	1.0028	1.0000	0.9994	2555	371
Backward Selection	0.8525	0.9996	1.0016	0.9975	2555	288
Stepwise Selection	No new model as same parameters selected as Backward Selection model including pathological T stage					

Figure 8: C-Index over Time in Prognostic Models Built When Entry p-value=0.2 and Exit p=value=0.1

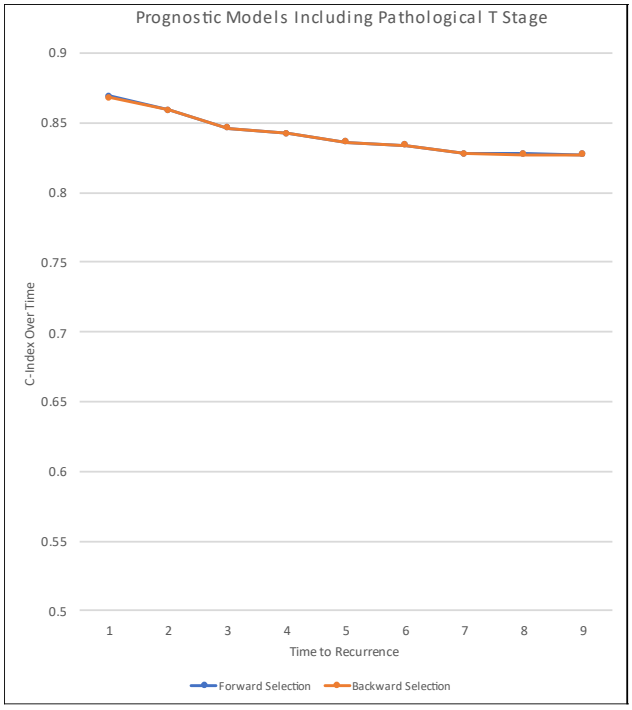


Figure 9: Calibration Curve 2-years Post-Surgery Recurrence Free Survival Prognostic Models Built When Entry p-value=0.2 and Exit p=value=0.1

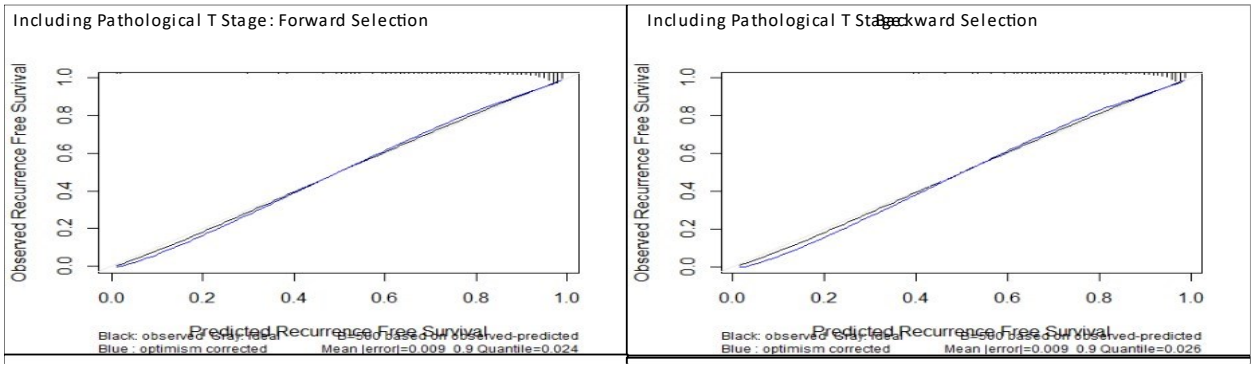


Figure 10: Calibration Curve 3-years Post-Surgery Recurrence Free Survival Prognostic Models Built When Entry p-value=0.2 and Exit p=value=0.1

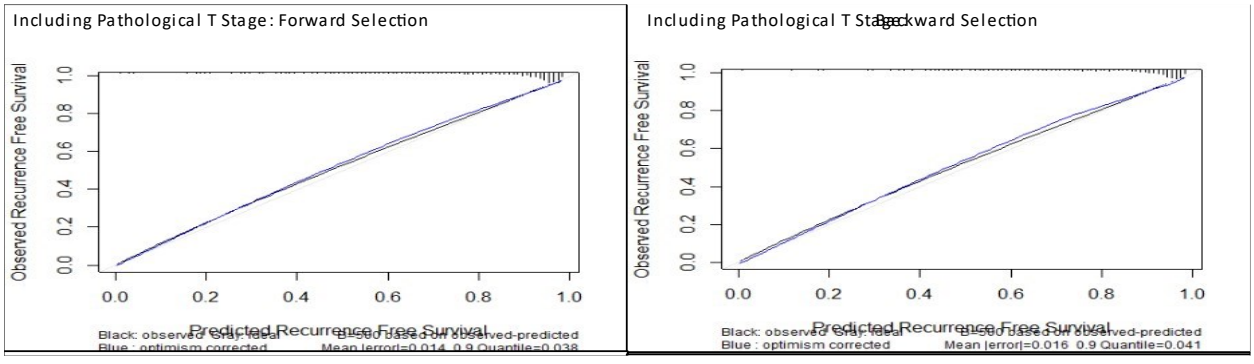


Figure 11: Calibration Curve 5-years Post-Surgery Recurrence Free Survival Prognostic Models Built When Entry p-value=0.2 and Exit p=value=0.1

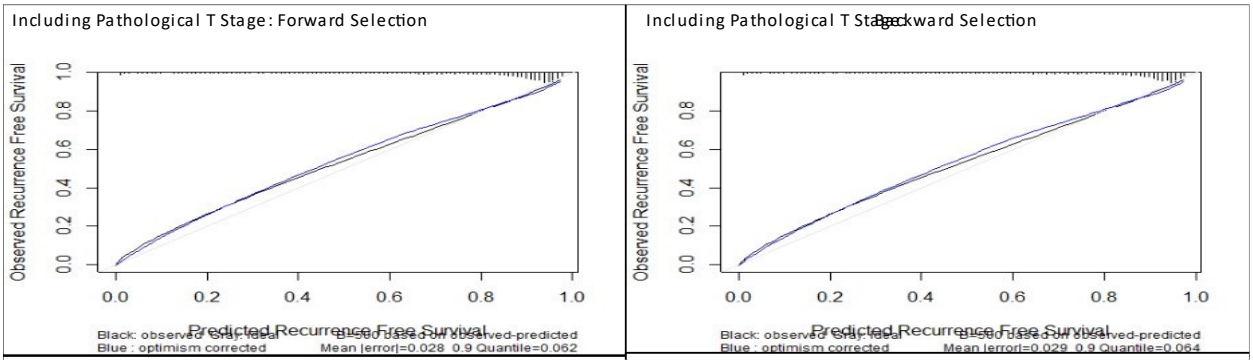


Table 9: Decision Curve Analysis and Agreement within 5% and 10% for Prognostic Models Built When Entry p-value=0.2 and Exit p=value=0.1

	Proportion of Patients with Agreement Within Absolute Range						
	2-years post-surgery		**3-years post-surgery**		**5-years post-surgery**		
Prognostic Model	**Within 5%**	**Within 10%**	**Within 5%**	**Within 10%**	**Within 5%**	**Within 10%**	**Recurrence Risk Threshold for Clinical Utility**
Prognostic Models Including Pathological T Stage							
Forward Selection	100%	100%	100%	100%	50%	100%	0%-36.5%
Backward Selection	100%	100%	100%	100%	49%	100%	0%-36.5%
Stepwise Selection	No new model as same parameters selected as Forward Selection model including pathological T stage						

Figure 12: Decision Curve Analysis at 5-years Post-Surgery for Survival Prognostic Models Built When Entry p-value=0.2 and Exit p=value=0.1

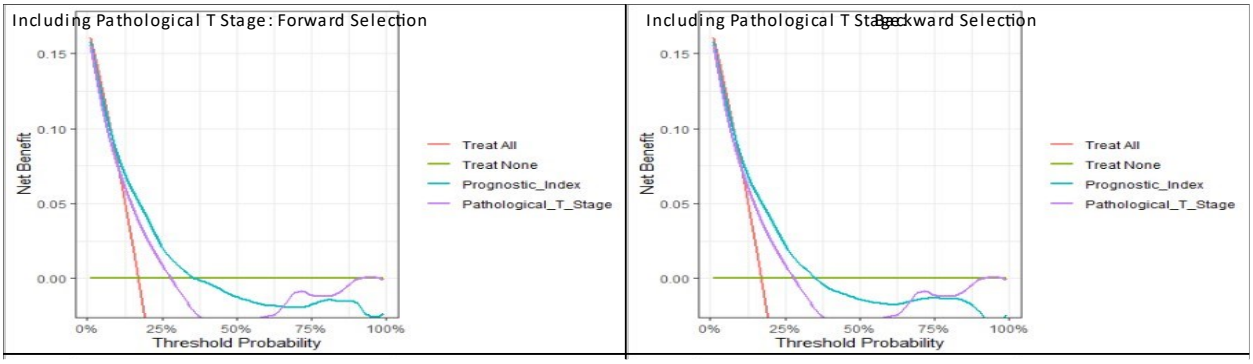


Figure 13: Calibration Plots Generated Using k-fold Validation where k=10 for p-Value Selection Model Including Pathological T Stage

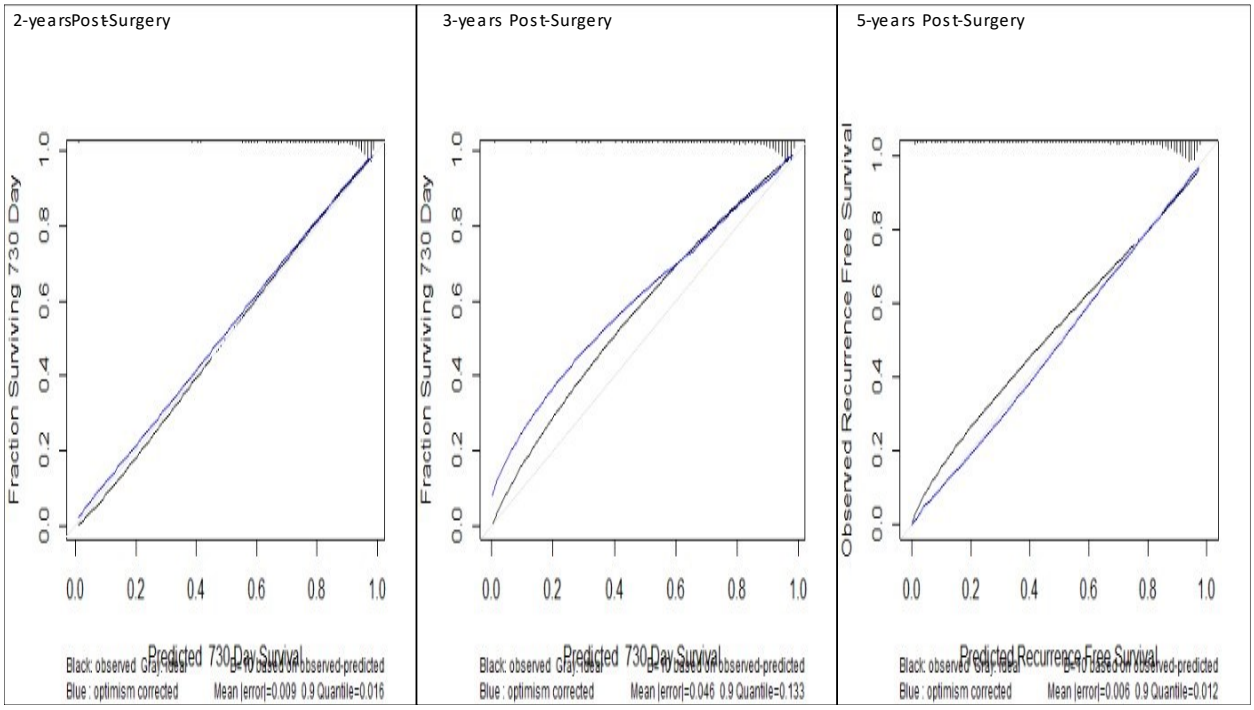
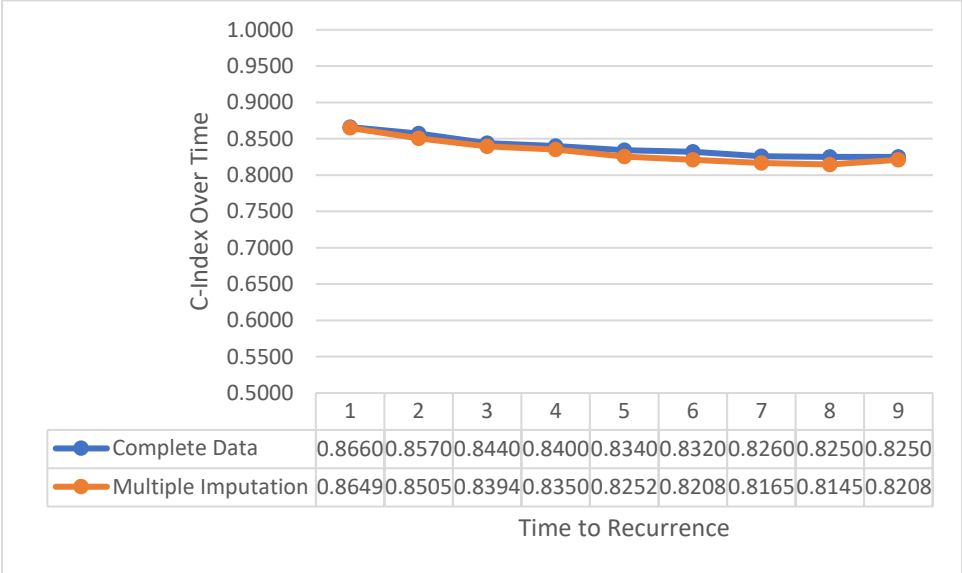


Figure 14: Effect of Missing Data-C-Index Over Time in p-Value Selection Model Including Pathological T Stage



	1	2	3	4	5	6	7	8	9
Complete Data	0.8660	0.8570	0.8440	0.8400	0.8340	0.8320	0.8260	0.8250	0.8250
Multiple Imputation	0.8649	0.8505	0.8394	0.8350	0.8252	0.8208	0.8165	0.8145	0.8208

Figure 15: Recurrence Free Survival Calibration Curves Generated Using Multiple Imputation

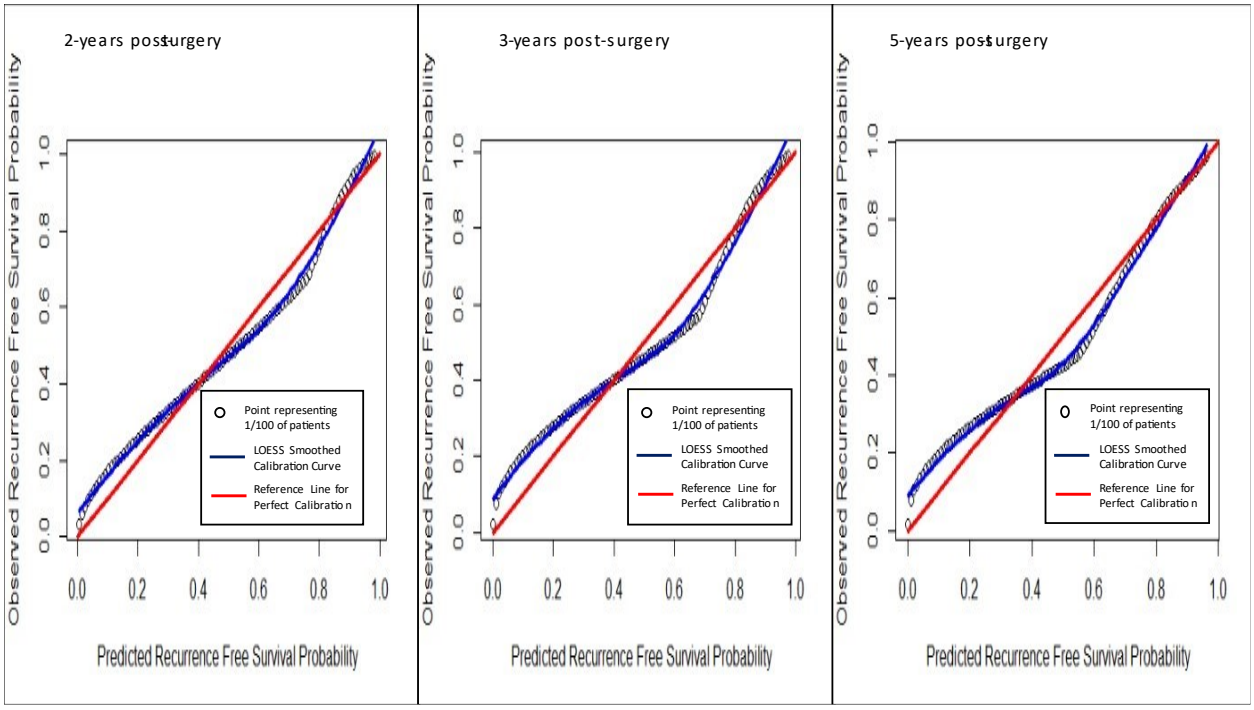
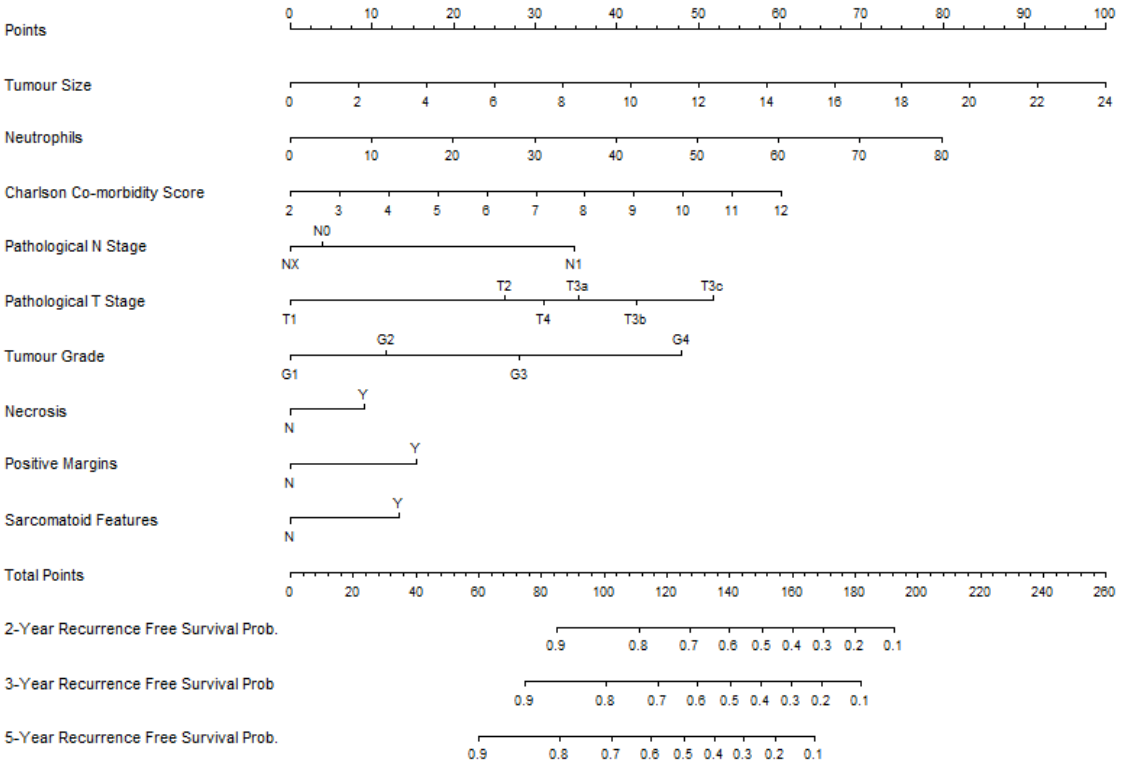


Figure 16: p-Value Selection Model Including Pathological T Stage for Recurrence Free Survival Nomogram



Abbreviations: N0=pathological N0 stage; NX=nodal resection not assessed; N1=pathological N1 stage; T1=pathological T1 stage; T2=pathological T2 stage; T3a=pathological T3a stage; T3b=pathological T3b stage ; T3c=pathological T3c stage;T4=pathological T4 stage;G1= tumour grade 1; G2=tumour grade 2;G3=tumour grade 3; G4=tumour grade 4;N=no; Y=yes; Recurrence Free Survival Prob.=Recurrence Free Survival Probability
Tumour size is measured in cm and neutrophil concentration is measured in $\times 10^9$ cells/µL.

Table 10: Risk Group Creation Based on Risk Scores

	Risk Group			
Risk Group Classification System	**Low**	**Intermediate**	**High**	**Very High**
Quartile Method	Risk Score <47.68	47.68≤Risk Score<70.84	70.84≤Risk Score<116.76	Risk Score≥116.76
Mean Risk Score Method	Risk Score <61.15	61.15≤Risk Score<68.35	68.35≤Risk Score<125.04	Risk Score≥125.04
Percentile Method	Risk Score <40.03	40.03≤Risk Score<70.84	70.84≤Risk Score<138.34	Risk Score≥138.34
Face Validity Method	Risk Score <64.75	64.75≤Risk Score<98.42	98.42≤Risk Score<147.82	Risk Score≥147.82

Figure 17: Risk Group Categorizations Based on Risk Group Classification Method

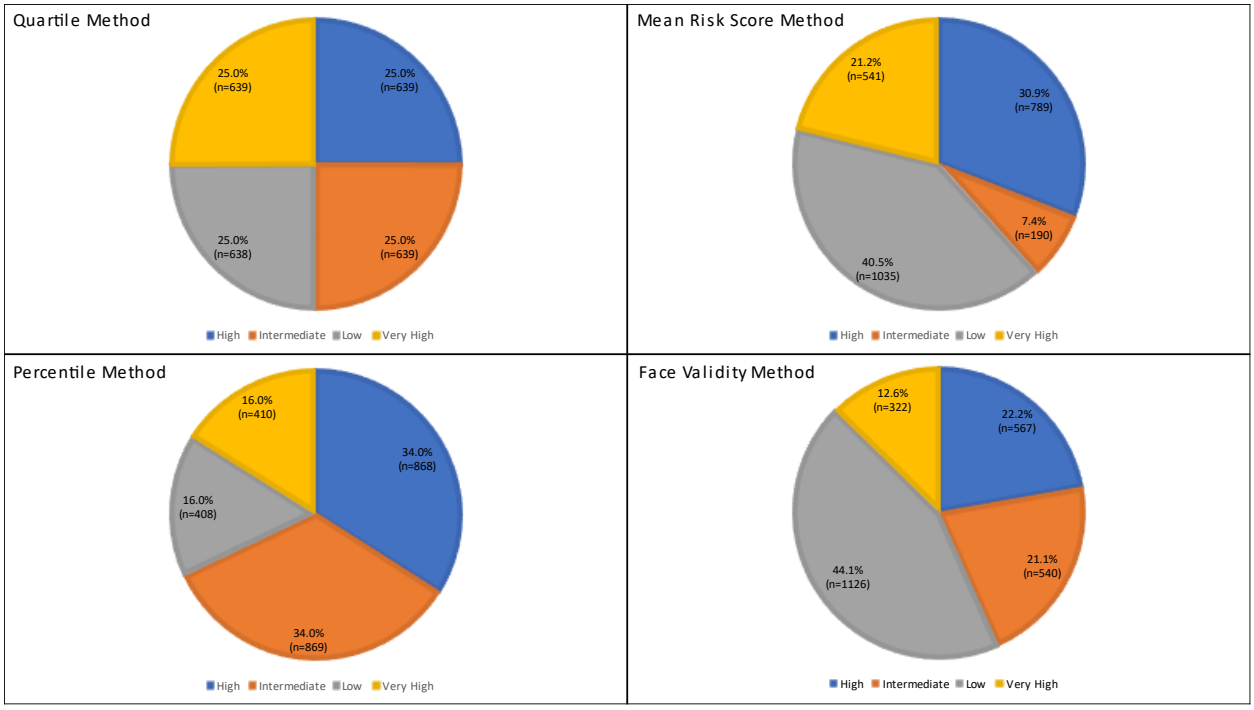


Figure 18: Kaplan-Meier Survival Curves for Risk Group Stratification Systems

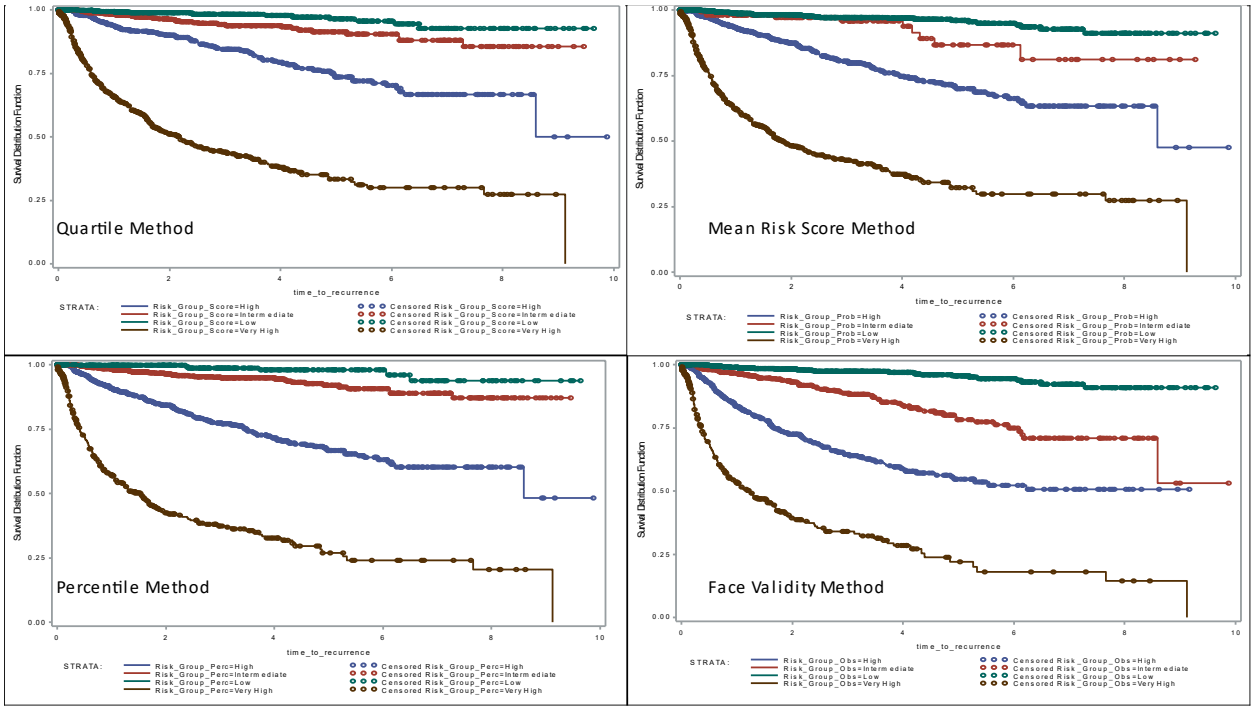


Figure 19a: Kaplan-Meier Survival Curves for Risk Group Stratification Systems at 2-years Post-Surgery

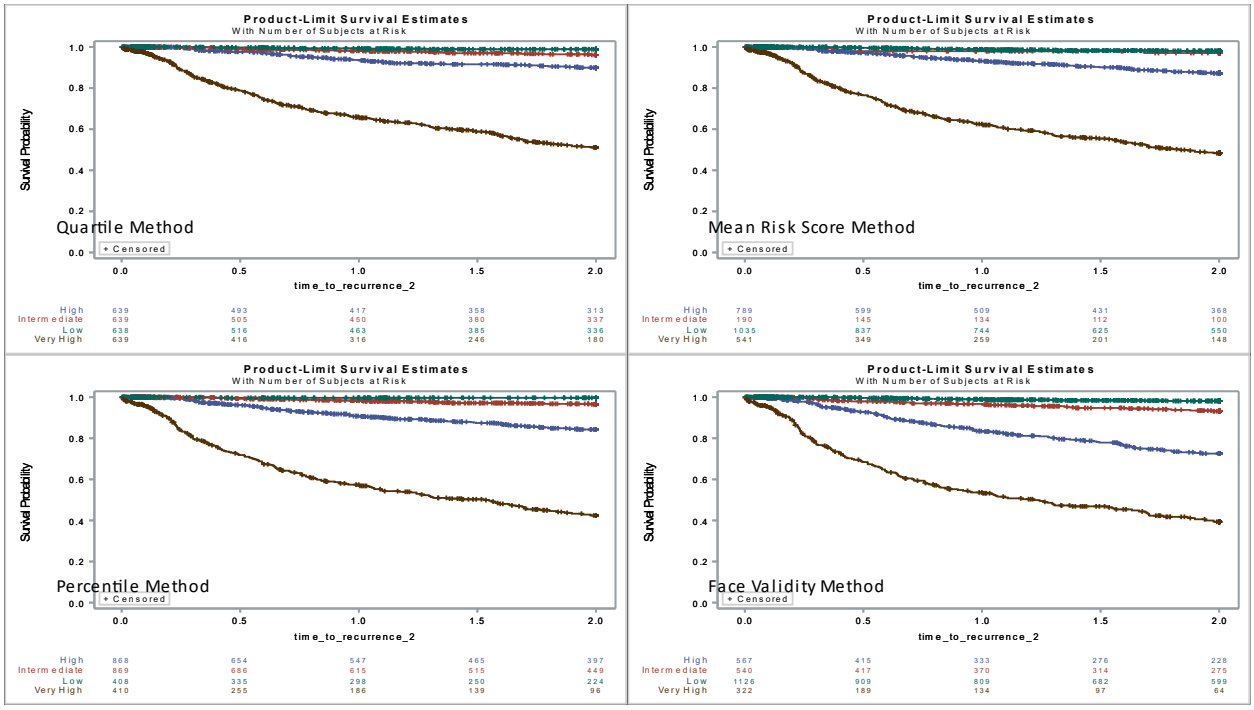


Figure 19b: Kaplan-Meier Survival Curves for Risk Group Stratification Systems at 3-years Post-Surgery

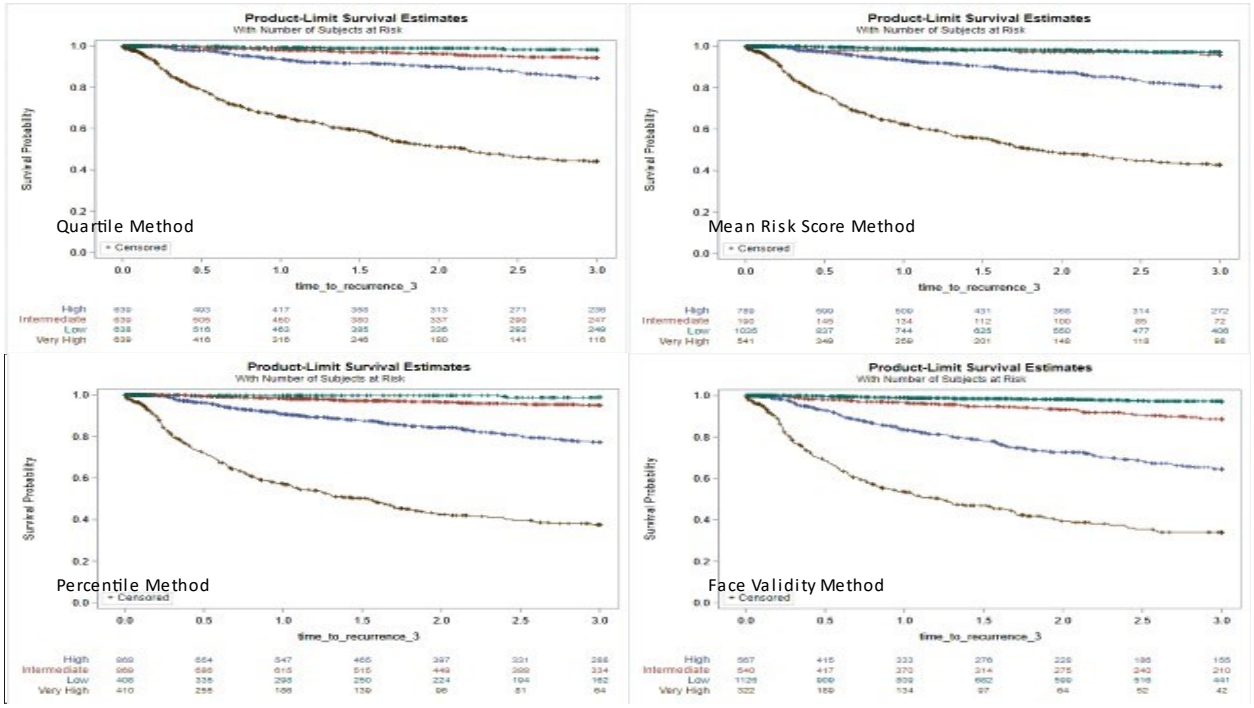


Figure 19c: Kaplan-Meier Survival Curves for Risk Group Stratification Systems at 5-years Post-Surgery

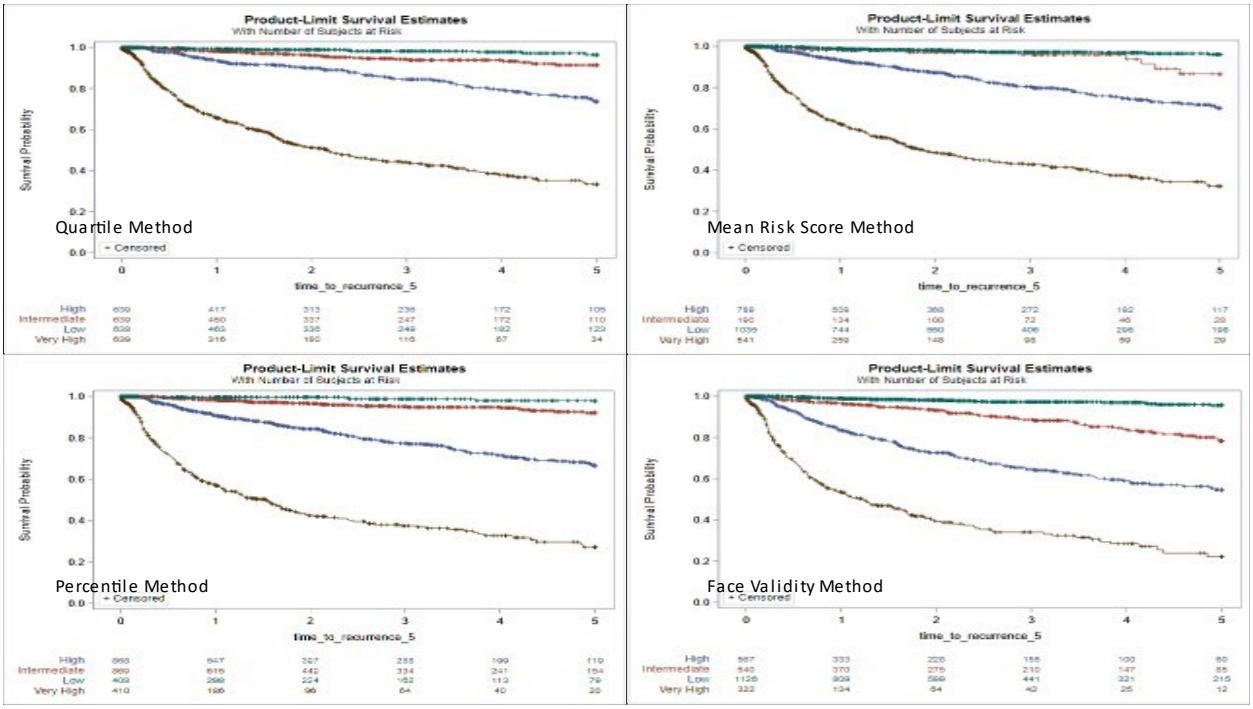


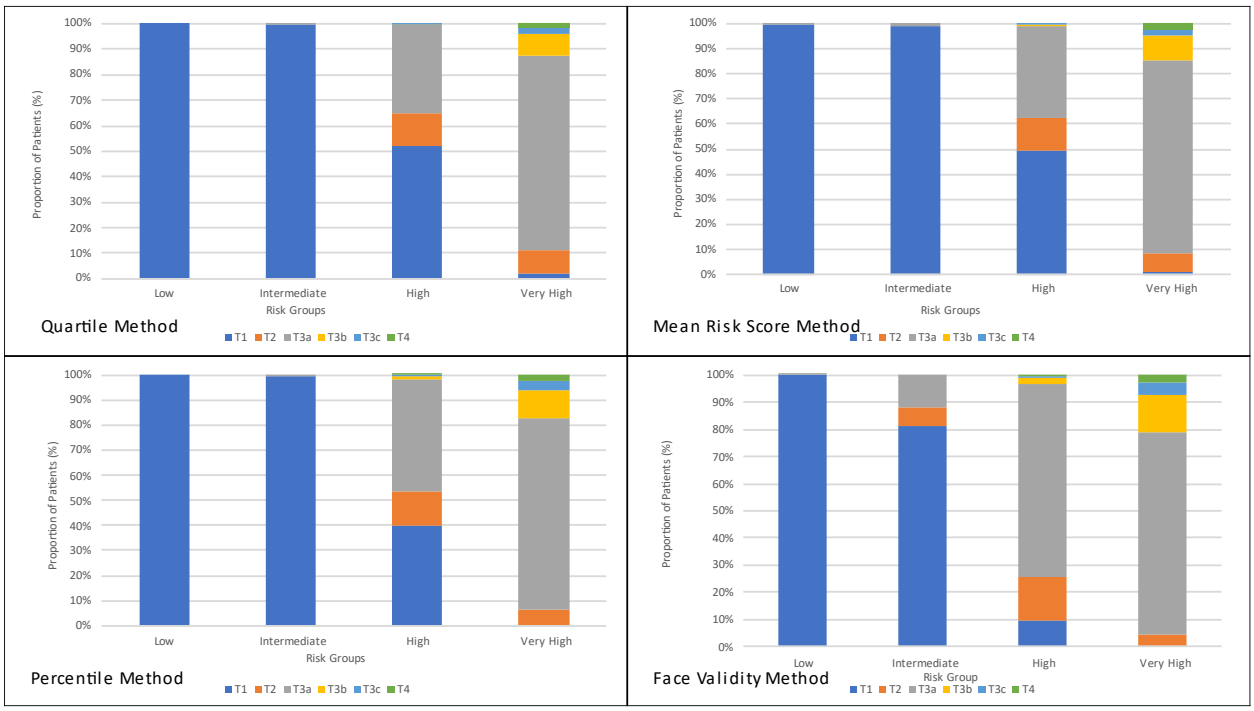
Table 11: Distribution of Pathological T Stage across Risk Groups

Pathological T Stage	Low		Intermediate		High		Very High		p-value
	N	%	N	%	N	%	N	%	
Quartile Method									
T1	638	100.0%	635	99.4%	332	52.0%	14	2.2%	<0.0001
T2	0	0.0%	0	0.0%	84	13.1%	58	9.1%	
T3a	0	0.0%	4	0.6%	219	34.3%	485	75.9%	
T3b	0	0.0%	0	0.0%	3	0.5%	55	8.6%	
T3c	0	0.0%	0	0.0%	1	0.2%	16	2.5%	
T4	0	0.0%	0	0.0%	0	0.0%	11	1.7%	
Mean Risk Score Method									
T1	1034	99.9%	188	98.9%	390	49.4%	7	1.3%	<0.0001
T2	0	0.0%	0	0.0%	102	12.9%	40	7.4%	
T3a	1	0.1%	2	1.1%	290	36.8%	415	76.7%	
T3b	0	0.0%	0	0.0%	4	0.5%	54	10.0%	
T3c	0	0.0%	0	0.0%	3	0.4%	14	2.6%	
T4	0	0.0%	0	0.0%	0	0.0%	11	2.0%	
Percentile Method									
T1	408	100.0%	865	99.5%	346	39.9%	0	0.0%	<0.0001
T2	0	0.0%	0	0.0%	117	13.5%	25	6.1%	

T3a	0	0.0%	4	0.5%	391	45.0%	313	76.3%	
T3b	0	0.0%	0	0.0%	10	1.2%	48	11.7%	
T3c	0	0.0%	0	0.0%	3	0.3%	14	3.4%	
T4	0	0.0%	0	0.0%	1	0.1%	10	2.4%	
Face Validity Method									
T1	1124	99.8%	439	81.3%	56	9.9%	0	0.0%	<0.0001
T2	0	0.0%	37	6.9%	91	16.0%	14	4.3%	
T3a	2	0.2%	64	11.9%	402	70.9%	240	74.5%	
T3b	0	0.0%	0	0.0%	13	2.3%	45	14.0%	
T3c	0	0.0%	0	0.0%	3	0.5%	14	4.3%	
T4	0	0.0%	0	0.0%	2	0.4%	9	2.8%	

Abbreviations: N=number of patients;T1=pathological T1 stage; T2=pathological T2 stage; T3a=pathological T3a stage; T3b=pathological T3b stage ; T3c=pathological T3c stage;T4=pathological T4 stage

Figure 20: Pathological T Stage Distribution by Risk Groups



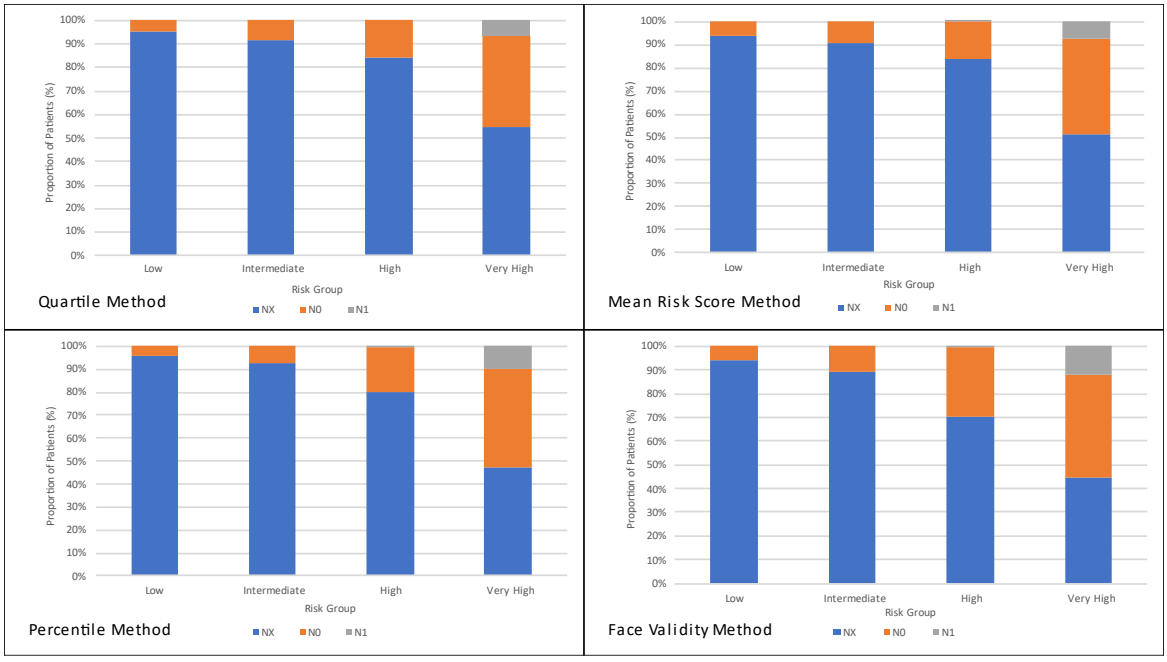
Abbreviations: T1=pathological T1 stage; T2=pathological T2 stage; T3a=pathological T3a stage; T3b=pathological T3b stage ; T3c=pathological T3c stage;T4=pathological T4 stage

Table 12: Distribution of Pathological N Stage across Risk Groups

Pathological N Stage	Low		Intermediate		High		Very High		p-value
	N	%	N	%	N	%	N	%	
Quartile Method									
NX	610	95.6%	585	91.5%	540	84.5%	348	54.5%	<0.0001
N0	28	4.4%	54	8.5%	99	15.5%	250	39.1%	
N1	0	0.0%	0	0.0%	0	0.0%	41	6.4%	
Mean Risk Score Method									
NX	974	94.1%	172	90.5%	661	83.8%	276	51.0%	<0.0001
N0	61	5.9%	18	9.5%	127	16.1%	225	41.6%	
N1	0	0.0%	0	0.0%	1	0.1%	40	7.4%	
Percentile Method									
NX	391	95.8%	804	92.5%	694	80.0%	194	47.3%	<0.0001
N0	17	4.2%	65	7.5%	173	19.9%	176	42.9%	
N1	0	0.0%	0	0.0%	1	0.1%	40	9.8%	
Face Validity Method									
NX	1060	94.1%	482	89.3%	398	70.2%	143	44.4%	<0.0001
N0	66	5.9%	58	10.7%	167	29.5%	140	43.5%	
N1	0	0.0%	0	0.0%	2	0.4%	39	12.1%	

Abbreviations: N=number of patients; N0=pathological N0 stage; NX=nodal resection not assessed; N1=pathological N1 stage

Figure 21: Pathological N Stage Distribution by Risk Groups



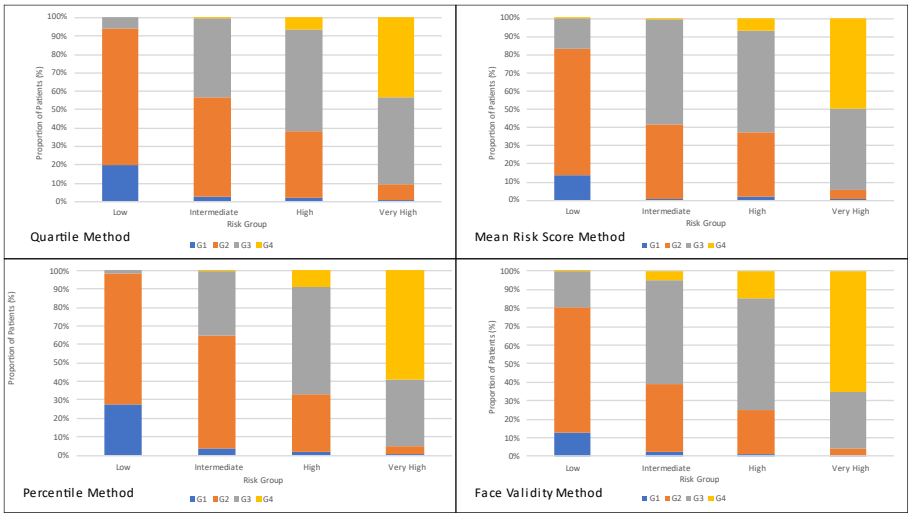
Abbreviations: N0=pathological N0 stage; NX=nodal resection not assessed; N1=pathological N1 stage

Table 13: Distribution of Tumour Grade across Risk Groups

Tumour Grade	Low		Intermediate		High		Very High		p-value
	N	%	N	%	N	%	N	%	
Quartile Method									
G1	128	20.1%	17	2.7%	15	2.3%	4	0.6%	<0.0001
G2	474	74.3%	346	54.1%	230	36.0%	58	9.1%	
G3	36	5.6%	273	42.7%	352	55.1%	302	47.3%	
G4	0	0.0%	3	0.5%	42	6.6%	275	43.0%	
Mean Risk Score Method									
G1	143	13.8%	1	0.5%	17	2.2%	3	0.6%	<0.0001
G2	722	69.8%	78	41.1%	279	35.4%	29	5.4%	
G3	169	16.3%	110	57.9%	442	56.0%	242	44.7%	
G4	1	0.1%	1	0.5%	51	6.5%	267	49.4%	
Percentile Method									
G1	112	27.5%	33	3.8%	17	2.0%	2	0.5%	<0.0001
G2	289	70.8%	531	61.1%	270	31.1%	18	4.4%	
G3	7	1.7%	302	34.8%	505	58.2%	149	36.3%	
G4	0	0.0%	3	0.3%	76	8.8%	241	58.8%	
Face Validity Method									
G1	143	12.7%	12	2.2%	7	1.2%	2	0.6%	<0.0001
G2	761	67.6%	200	37.0%	135	23.8%	12	3.7%	
G3	221	19.6%	302	55.9%	342	60.3%	98	30.4%	
G4	1	0.1%	26	4.8%	83	14.6%	210	65.2%	

Abbreviations: N=number of patients;G1= tumour grade 1; G2=tumour grade 2;G3=tumour grade 3; G4=tumour grade 4

Figure 22: Tumour Grade Distribution by Risk Groups



Abbreviations: G1= tumour grade 1; G2=tumour grade 2;G3=tumour grade 3; G4=tumour grade 4

Table 14: Distribution of Necrosis across Risk Groups

Necrosis	Low		Intermediate		High		Very High		p-value
	N	%	N	%	N	%	N	%	
Quartile Method									
No	623	97.6%	570	89.2%	507	79.3%	278	43.5%	<0.0001
Yes	15	2.4%	69	10.8%	132	20.7%	361	56.5%	
Mean Risk Score Method									
No	988	95.5%	162	85.3%	615	77.9%	213	39.4%	<0.0001
Yes	47	4.5%	28	14.7%	174	22.1%	328	60.6%	
Percentile Method									
No	404	99.0%	789	90.8%	642	74.0%	143	34.9%	<0.0001
Yes	4	1.0%	80	9.2%	226	26.0%	267	65.1%	
Face Validity Method									
No	1068	94.8%	438	81.1%	377	66.5%	95	29.5%	<0.0001
Yes	58	5.2%	102	18.9%	190	33.5%	227	70.5%	

Abbreviation: N=number of patients

Figure 23: Necrosis Distribution by Risk Groups

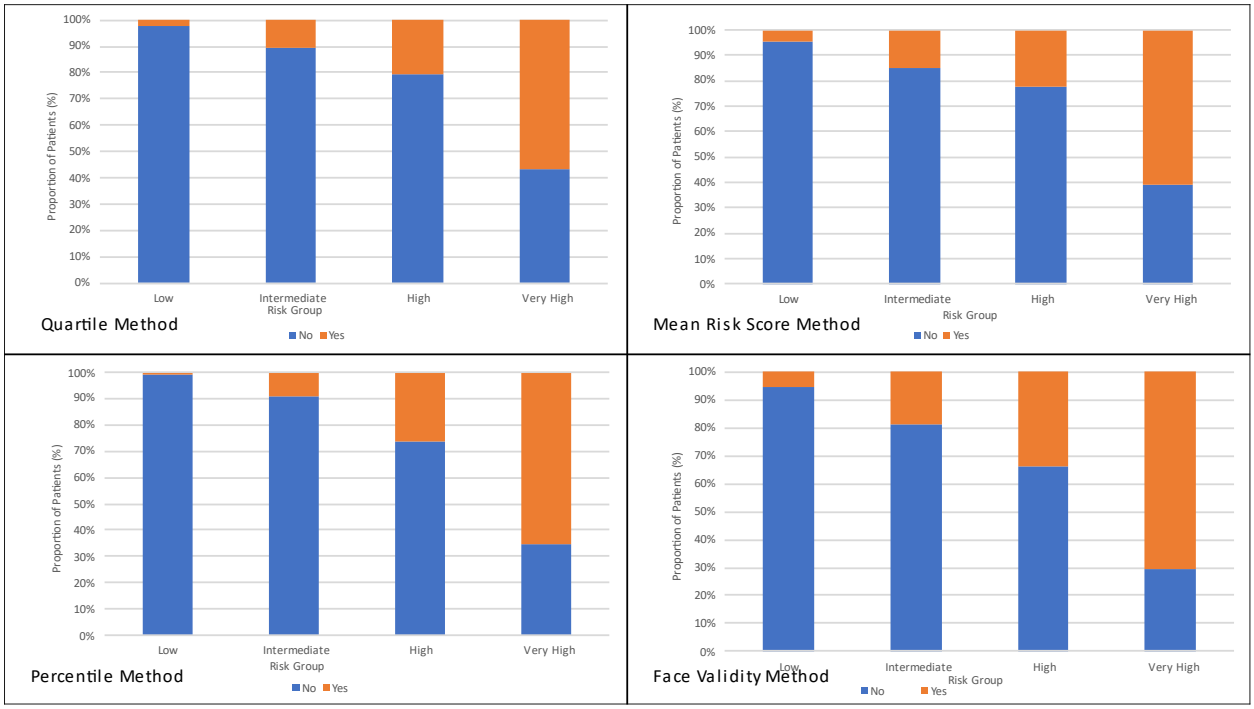


Table 15: Distribution of Positive Margins across Risk Groups

Positive Margins	Low		Intermediate		High		Very High		p-value
	N	%	N	%	N	%	N	%	
Quartile Method									
No	625	98.0%	601	94.1%	592	92.6%	552	86.4%	<0.0001
Yes	13	2.0%	38	5.9%	47	7.4%	87	13.6%	
Mean Risk Score Method									
No	1000	96.6%	178	93.7%	734	93.0%	458	84.7%	<0.0001
Yes	35	3.4%	12	6.3%	55	7.0%	83	15.3%	
Percentile Method									
No	399	97.8%	827	95.2%	808	93.1%	336	82.0%	<0.0001
Yes	9	2.2%	42	4.8%	60	6.9%	74	18.0%	
Face Validity Method									
No	1084	96.3%	496	91.9%	538	94.9%	252	78.3%	<0.0001
Yes	42	3.7%	44	8.1%	29	5.1%	70	21.7%	

Abbreviation: N=number of patients

Figure 24: Positive Margins Distribution by Risk Groups

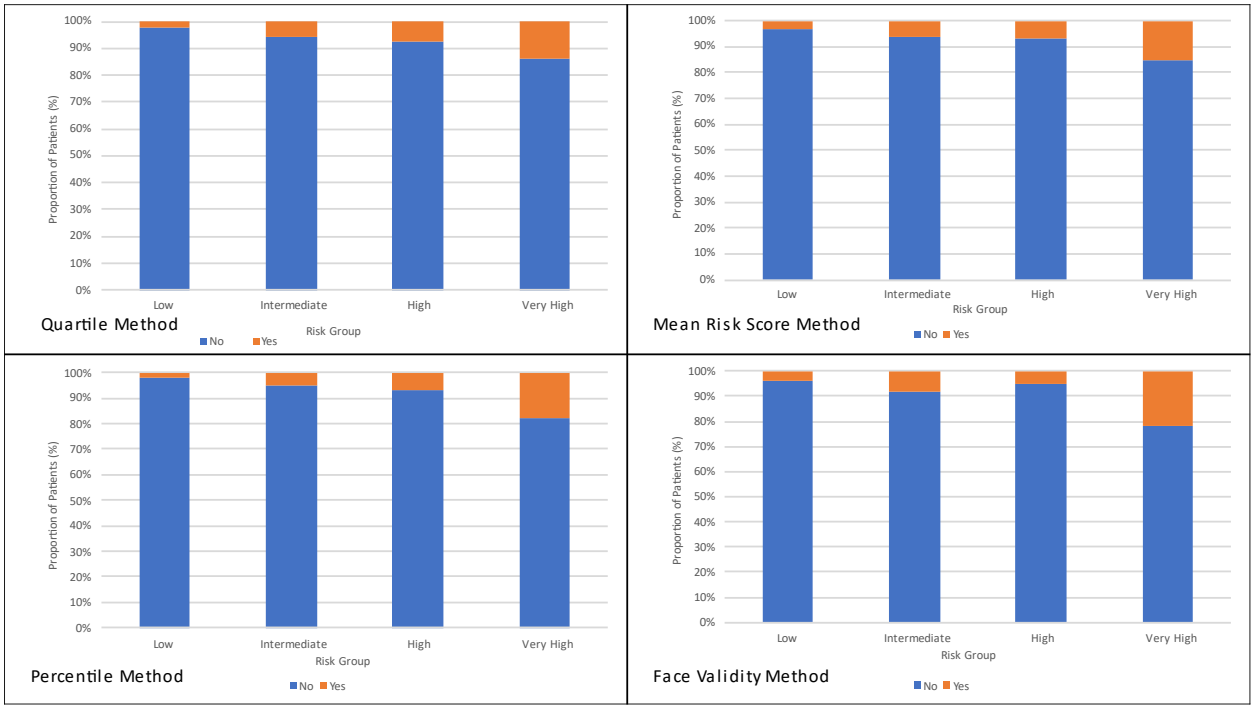


Table 16: Distribution of Sarcomatoid Features across Risk Groups

Sarcomatoid Features	Low		Intermediate		High		Very High		p-value
	N	%	N	%	N	%	N	%	
Quartile Method									
No	638	100.0%	639	100.0%	635	99.4%	550	86.1%	<0.0001
Yes	0	0.0%	0	0.0%	4	0.6%	89	13.9%	
Mean Risk Score Method									
No	1035	100.0%	190	100.0%	781	99.0%	456	84.3%	<0.0001
Yes	0	0.0%	0	0.0%	8	1.0%	85	15.7%	
Percentile Method									
No	408	100.0%	869	100.0%	853	98.3%	332	81.0%	<0.0001
Yes	0	0.0%	0	0.0%	15	1.7%	78	19.0%	
Face Validity Method									
No	1126	100.0%	539	99.8%	549	96.8%	248	77.0%	<0.0001
Yes	0	0.0%	1	0.2%	18	3.2%	74	23.0%	

Abbreviation:N=number of patients

Figure 25: Sarcomatoid Distribution by Risk Groups

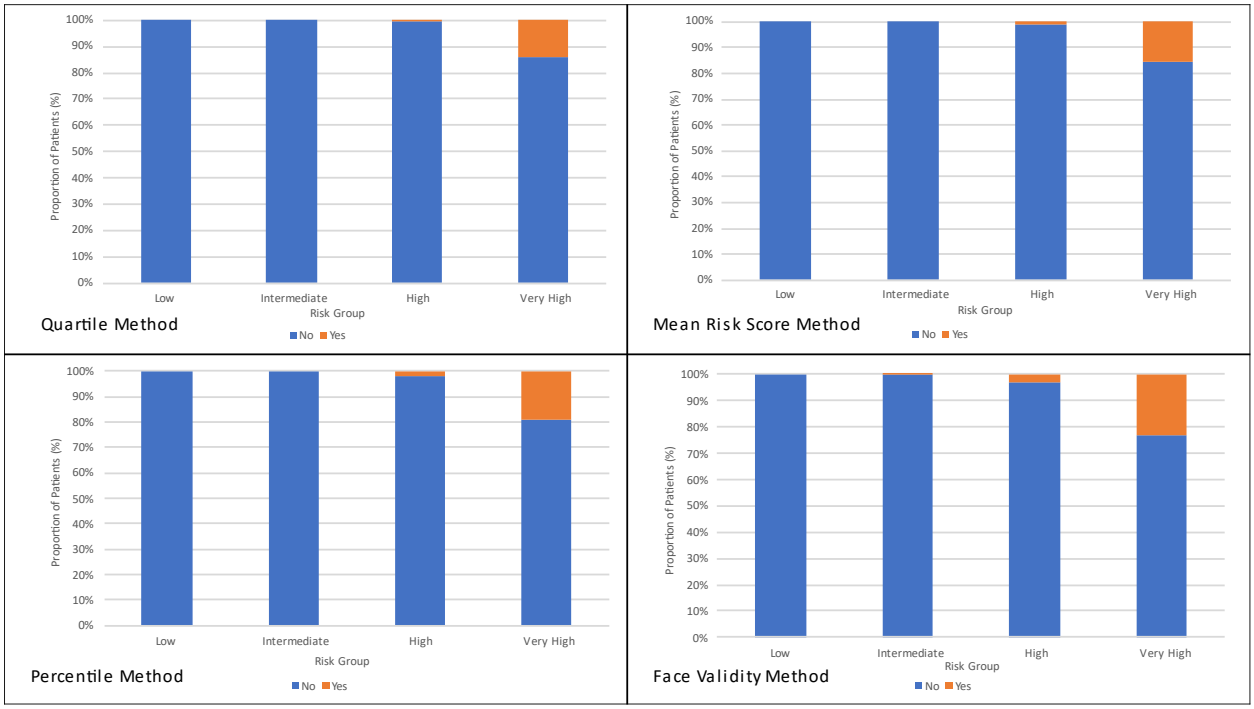
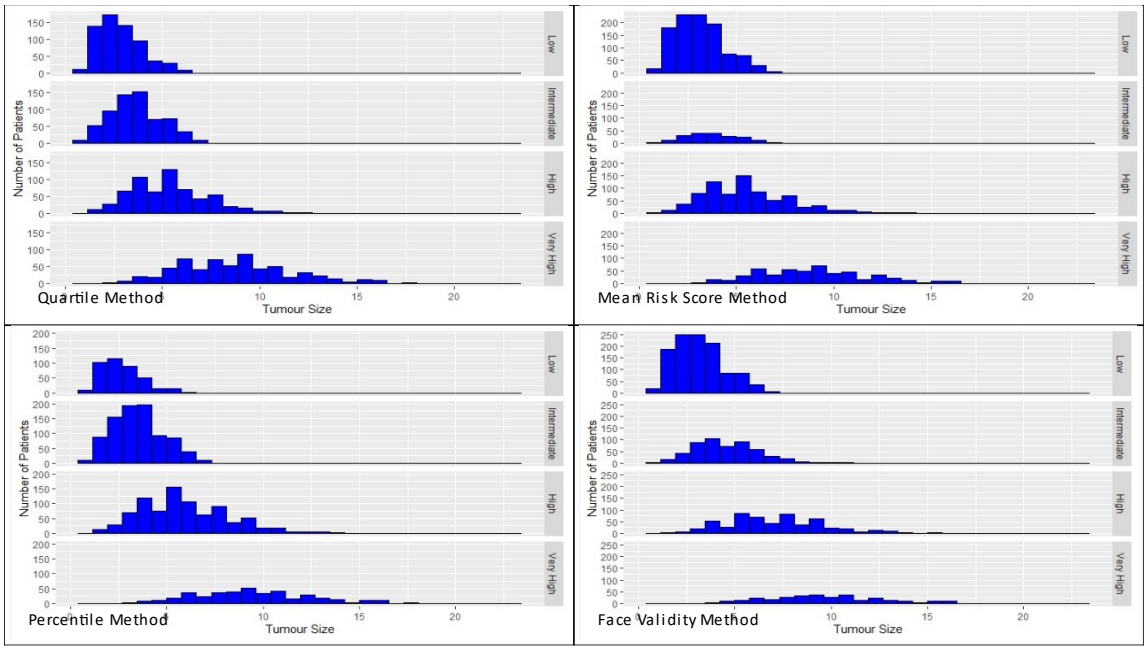


Table 17: Distribution of Tumour Size, Neutrophils and Charlson Co-morbidity Scores across Risk Groups[1]

Risk Group Stratification System	Low	Intermediate	High	Very High	p-value
Tumour size (cm)					
Quartile Method	2.799 (95%CI:2.713-2.886)	3.612 (95%CI:3.509-3.714)	5.362 (95%CI:5.200-5.245)	8.801 (95%CI:8.565-9.037)	<0.0001
Mean Risk Score Method	3.086 (95%CI:3.009-3.162)	3.741 (95%CI:3.548-3.933)	5.493 (95%CI:5.337-5.649)	9.068 (95%CI:8.808-9.327)	<0.0001
Percentile Method	2.632 (95%CI:2.528-2.736)	3.475 (95%CI:3.389-3.561)	5.912 (95%CI:5.757-6.067)	9.558 (95%CI:9.252-9.864)	<0.0001
Face Validity Method	3.130 (95%CI:3.056-3.204)	4.545 (95%CI:4.399-4.692)	7.096 (95%CI:6.883-7.307)	9.760 (95%CI:9.413-10.107)	<0.0001
Neutrophil ($x10^9$ cells/μL)					
Quartile Method	5.082 (95%CI:4.801-5.363)	6.091 (95%CI:5.777-6.406)	6.401 (95%CI:5.995-6.808)	6.388 (95%CI:6.014-6.762)	<0.0001
Mean Risk Score Method	5.411 (95%CI:5.177-5.645)	6.357 (95%CI:5.796-6.918)	6.354 (95%CI:6.007-6.702)	6.441 (95%CI:6.017-6.865)	<0.0001
Percentile Method	4.707 (95%CI:4.396-5.018)	6.000 (95%CI:5.728-6.272)	6.350 (95%CI:6.009-6.691)	6.488 (95%CI:6.020-6.956)	<0.0001
Face Validity Method	5.452 (95%CI:5.230-5.675)	6.771 (95%CI:6.307-7.236)	5.953 (95%CI:5.602-6.306)	6.631 (95%CI:6.079-7.182)	<0.0001
Charlson Co-morbidity Score					
Quartile Method	3.281 (95%CI:3.193-3.368)	4.549 (95%CI:4.433-4.665)	4.941 (95%CI:4.798-5.083)	5.241 (95%CI:5.092-5.390)	<0.0001
Mean Risk Score Method	3.672 (95%CI:3.592-3.753)	4.816 (95%CI:4.608-5.024)	4.967 (95%CI:4.841-5.093)	5.307 (95%CI:5.140-5.474)	<0.0001
Percentile Method	2.946 (95%CI:2.853-3.039)	4.371 (95%CI:4.274-4.467)	4.906 (95%CI:4.787-5.024)	5.483 (95%CI:5.286-5.694)	<0.0001
Face Validity Method	3.762 (95%CI:3.682-3.842)	5.026 (95%CI:4.878-5.174)	4.810 (95%CI:4.665-4.954)	5.680 (95%CI:5.452-5.910)	<0.0001

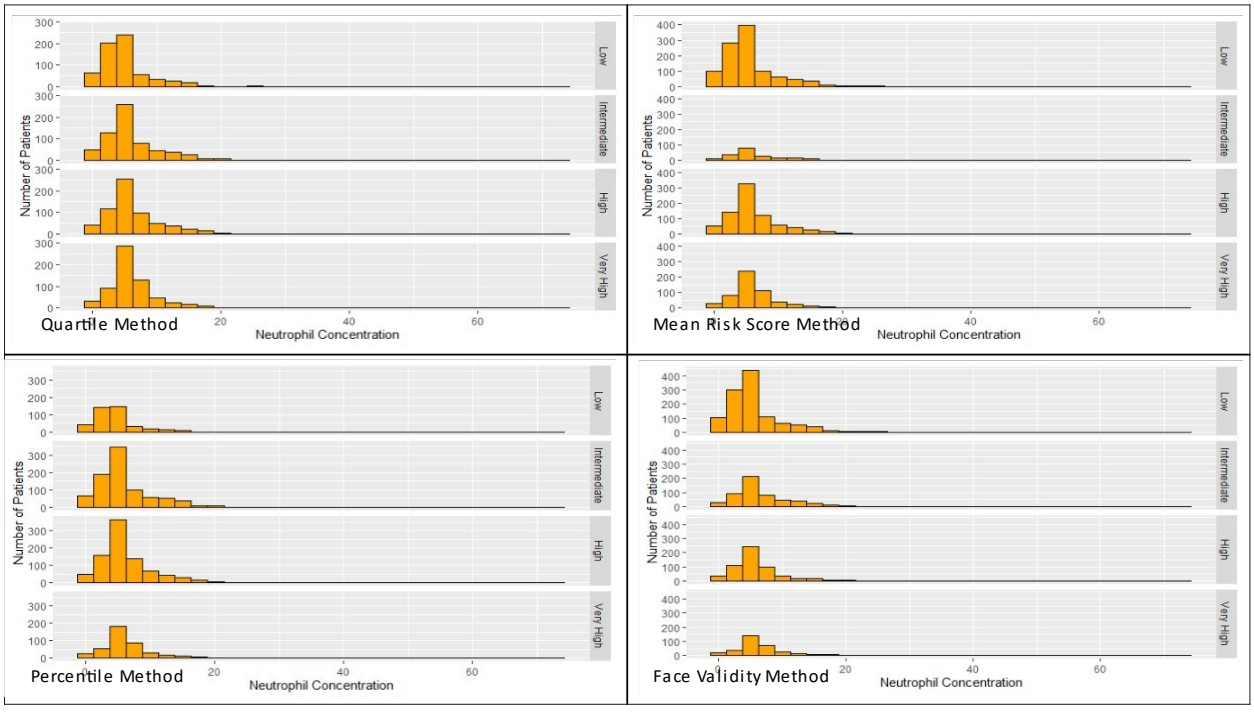
Abbreviations: CI=confidence interval;µL=microlitre;cm=centimetre
The mean (95%CI) is given in this table.

Figure 26: Distribution of Tumour Size by Risk Group[1]



[1]Tumour size is measured in cm.

Figure 27: Distribution of Neutrophil Concentration by Risk Group[1]



[1]Neutrophil concentration is measured in $x10^9$ cells/µL.

Figure 28: Distribution of Charlson Co-morbidity Score by Risk Group

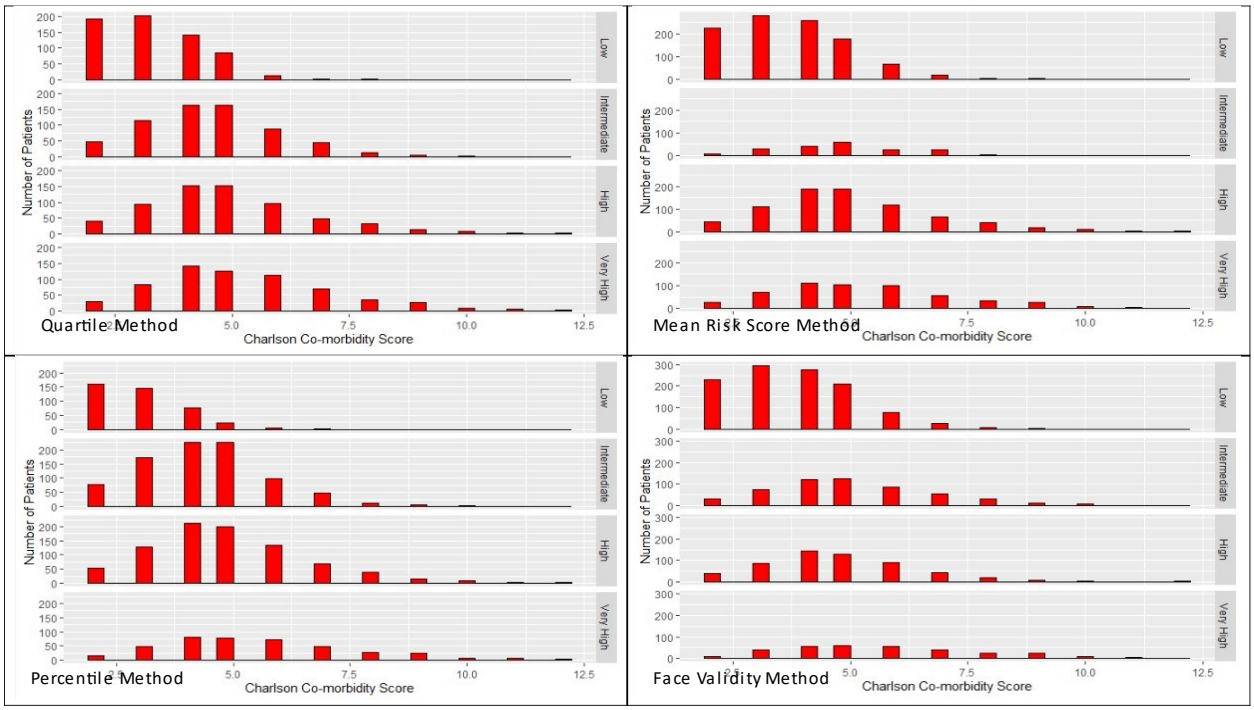


Table 18: Risk Group Classification System Comparison of Performance Characteristics

Risk Group Classification Systems	C-Statistic Overall
Quartile Method	0.8162
Mean Risk Score Method	0.8158
Percentile Method	0.8173
Face Validity Method	0.8295

Figure 29: Risk Group Recurrence Prediction Performance by Risk Group Classification Systems

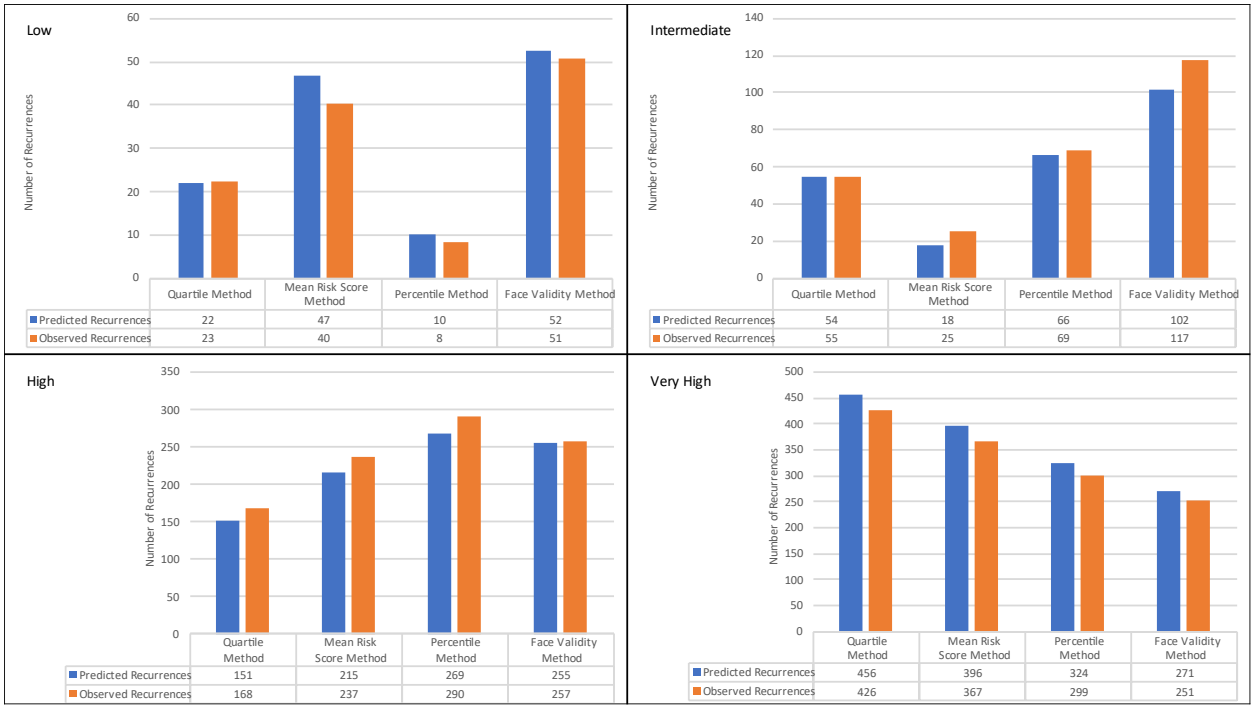


Figure 30: Prognostic Factors Included in Prognostic Models for Recurrence Free Survival in Clear Cell Carcinoma Patients

		Clear Cell Carcinoma Prognostic Models			
		New Proposed Model	Leibovich et al, 2003 (15)	Sorbellini et al, 2005 (16)	Leibovich et al, 2018- Clear Cell RCC (17)
Prognostic Factors	Tumour Size	Included	Included	Included	Included
	Neutrophil	Included	Not Included	Not Included	Not Included
	Charlson Co-morbidity Score	Included	Not Included	Not Included	Not Included
	Pathological N Stage	Included	Included	Not Included	Included
	Pathological T Stage	Included	Included	Included	Not Included
	Tumour Grade	Included	Included	Included	Included
	Necrosis	Included	Included	Included	Included
	Positive Margin	Included	Not Included	Not Included	Not Included
	Sarcomatoid Features	Included	Not Included	Not Included	Included
	Lymphovascular Invasion	Not Included	Not Included	Included	Not Included
	Clinical Presentation	Not Included	Not Included	Included	Included
	Perinephric or Renal Sinus Fat Invasion	Not Included	Not Included	Not Included	Included
	Tumour Thrombus	Not Included	Not Included	Not Included	Included
	Extension Outside Gerota's fascia	Not Included	Not Included	Not Included	Included

Included
Not Included

Chapter 6: Post-Surgery Surveillance Imaging Recommendations for Clear-Cell Carcinoma Patients Based on CKCis Risk Groups compared to the Canadian Urologic Association Guideline

6.1 Abstract

Introduction The current Canadian Urological Association (CUA) guideline provides follow-up schedules for patients based on risk-groups according to pathological tumour stage and pathological lymph node stage. A new prognostic model has been developed based on data from Canadian kidney cancer patients (CKCis model). In addition to pathologic tumour and lymph node stages, the CKCis model incorporates tumour grade, tumour size, tumour necrosis, sarcomatoid differentiation, surgical margin status, serum neutrophil concentration, and Charlson co-morbidity score. The purpose of this analysis were to compare CKCis risk groups to CUA risk groups and generate a follow up surveillance strategy based on the CKCis model.

Methods Clear cell renal cell carcinoma patients enrolled in the Canadian Kidney Cancer Information System (CKCis) between January 2011 and 2021 were used for this analysis. The primary outcome was post-surgery RCC recurrence. The CKCis and CUA risk groups were compared by assessing patient distribution, label concordance (5-year recurrence-free survival), discrimination (c-statistic), calibration (predicted vs observed events), risk-group agreement (Cohen's kappa), and the net reclassification improvement. New follow-up schedules based on the CKCis risk groups were created for ≥1% and ≥2% interval recurrence incidence, and modified for clinical use.

Results Both the CKCis and CUA risk groups stratified patients with different prognosis. However, the distribution of patients within each risk group were substantially different s (Low 44.1% vs 63.3%; Intermediate 21.1% vs 5.6%; High 22.5% vs. 29.5%; Very-high 12.6% vs. 1.6%) with weak agreement (Kappa Statistic: 0.43 (95%CI:0.41-0.45). The label-concordance was also different, with 5-year risk of recurrence between CKCis and CUA risk groups (Low risk 2.2% vs 5.1%; Intermediate 10.2% vs 23.2%; High 26.8% vs 35.9%; Very-high 55.9% vs 63.4%). The CKCis risk group classification system had better discriminative ability (c-statistic =0.83 versus 0.78). Calibration was similar between risk group systems. The net reclassification improvement using the CKCis risk groups was 0.47.

Conclusions CKCis risk groups include more parameters compared to the CUA risk groups. However, the CKCis risk groups had improved distribution, label-concordance, and discrimination. Electronic tools based on CKCis risk groups should be developed to facilitate use of CKCis risk-groups in clinical care.

6.2 Introduction

In 2020, there were an estimated 7500 patients with a new diagnosis of Renal Cell Carcinoma (RCC) and the majority of patients will be treated with surgery (1). The recurrence free survival by 5 years post-operative is approximately 95% for patients with pathological T1 stage tumours, 79% for patients with pathological T2 stage tumours, 70% for patients with pathological T3 stage tumours, and 50% for patients with pathological T4 stage tumours (2). While the current Canadian Urological Association (CUA) guideline for the management of non-metastatic kidney cancer acknowledges the existence of prognostic models and several prognostic factors for kidney cancer, it stratifies patient risk based on pathological tumour and pathological lymph node stage (3). The Canadian Urological Association guidelines uses the following risk group classification system:low-risk (pT1+pN0/X stage), intermediate-risk (pT2+pN0/X stage), high-risk (pT3-4+pN0/X stage) and very high-risk (any pT stage+pN1 stage)(3).To detect recurrence, within 5 years of surgery, the CUA guideline recommends abdominal imaging twice for patients in the low-risk group and 8 times for patients in the high or very-high risk (3).The full imaging schedule for patients based on the current CUA guidelines is provided in Table 1.

The current CUA guidelines are easy to use in clinical practice because risk groups are based only on pathological T and N stages. However, CUA risk groups do not consider other factors independently associated with recurrence, such as tumour grade. The CKCis recurrence risk model was developed based on a large cohort of Canadian patients and incorporates many more prognostic variables. This model had an overall c-statistic of 0.85 and was well calibrated. The purpose of this study was to compare CKCis risk groups to the CUA guideline risk groups and develop new follow-up imaging recommendations.

6.3 Patients and Methods

6.3.1 Study Population

This study used data from the Canadian Kidney Cancer Information System (CKCis), a prospective cohort of patients treated at one of 14 academic centres in Canada. CKCis has been shown to be representative of the Canadian kidney cancer population (4). This study included surgically treated patients with clear cell renal cell carcinoma (ccRCC) between January 1, 2011 and June 30, 2021. The following patients were excluded: non-Clear Cell RCC patients, patients <18 years of age, patients treated with surveillance or non-surgical methods, patients with multiple tumours, or patients with germ-line kidney cancer syndromes. Patients with incomplete information to determine risk-group classification were also excluded.

6.3.2 Comparing CKCis and CUA Risk Groups

Patients were categorized into risk groups based on CKCis or CUA methods. Kaplan Meier survival curves were generated to assess recurrence risk for each risk group. Cohen's kappa statistic was used to assess

agreement between the CUA and CKCis risk categories (5). Kappa statistic values range from 0 to 1 and can be interpreted as:0.0-0.20 no agreement, 0.21-0.39 minimal, 0.40-0.59 weak, 0.60-0.70 moderate, 0.8-0.9 strong and >0.9 almost perfect (5).

To compare risk group classification systems we determined the proportion of patients in each risk group (to assess distribution), the Kaplan Meier curves (to assess stratification and label-to-risk concordance), overall c-statistic (to assess discrimination) and predicted vs. observed number of events (to assess calibration) (6). Net reclassification improvement (NRI) was used to assess how well the CKCis model correctly reclassified patient risk. Net reclassification improvement is the sum of the proportion of patients who were correctly reclassified (moved to a higher risk group and experienced a recurrence or moved to a lower risk group and did not experience a recurrence) (7). An NRI ranges from -1 to +1 with -1 indicating completely incorrect reclassification and +1 indicating completely correct reclassification. The advantages of the net reclassification improvement measure are that it is a simple and easily estimated measure (8). The disadvantages of the net reclassification improvement measure are that positive values do not always signify improved classification of intermediate risk groups when good and poor risk groups are well classified; overfitting or model misspecification can also lead to overestimation of the net reclassification improvement statistic (8).

Label-to-risk concordance means that the label (low, intermediate, high, or very-high) corresponds to what most patients and clinicians would consider the numerical risk. For example, most patients would likely consider 5% risk of recurrence to be low and 75% risk of recurrence to be very-high.

6.3.3 Creation of Post-Surgical Surveillance Protocol

Patients were divided into risk groups based on the CKCis risk group classification system derived from the new CKCis prognostic model for Clear Cell Renal Cell Carcinoma patients as outlined in chapter 5. Timepoints for imaging were based on Kaplan-Meier survival curves with imaging recommend when the interval recurrence free survival difference exceeded 1% or 2% (9). To modify for clinical practice, these timepoints were rounded to the nearest 3-month interval (ie. Imaging recommendations were a minimum of 3 months apart and were at 3-month intervals).

6.4 Results

6.4.1 Comparing Risk Group Stratification Systems: CKCis vs. CUA

We observed good stratification between risk groups in both the CKCis and CUA systems (Figure 2). However, the proportions of patients within each risk group were significantly different between CKCis and CUA categories as very few patients are included in the CUA intermediate and very-high-risk groups

(Low 44.1% vs 63.3%; Intermediate 21.1% vs 5.6%; High 22.5% vs. 29.5%; Very-high 12.6% vs. 1.6%) (Figure 3). The label-to-risk concordance was also different between CKCis and CUA risk groups (5-year recurrence: Low risk 2.2% vs 5.1%; Intermediate 10.2% vs 23.2%; High 26.8% vs 35.9%; Very-high 55.9% vs 63.4%) (Figure 2). Weak agreement (Kappa Statistic: 0.43 (95%CI:0.41-0.45) was observed between the CUA risk groups and the CKCis risk groups. Many (494;30.5%) of the patients considered low risk in CUA classification were intermediate (n=439;81.3%) or high (n=55;9.7%) risk based on CKCis system (Figure 4).

Overall, the CKCis risk groups had better discriminative accuracy compared to the CUA risk groups (c=0.83 vs. c=0.78). Net Reclassification Improvement analysis showed a 0.47 (95%CI:-0.04-0.55) improvement in prediction accuracy using the CKCis risk groups compared to the CUA risk groups.

Figure 5 showed the predicted vs observed number of recurrences in Low, Intermediate, High and Very-high-risk groups. After considering relative error, Low (3.4% vs. 9.4%), High (0.8% vs. 2.1%) and Very-high-risk (7.7% vs.10.3%) groups were found to have improved calibration using the CKCis risk groups compared to the CUA risk groups, while the CKCis intermediate-risk group had worse calibration (13.2% vs. 3.6%).

6.4.2 Creation of New Recommendations for Post-Surgical Imaging

The timepoints at which recurrence incidence changed by ≥1% and ≥2% are shown in Figures 6 and 7. As expected, there are more timepoints when a 1% threshold is used compared to 2% threshold. There are also more timepoints for lower risk patients compared to higher risk patients. The 2% timepoint model was modified into clinically applicable (easy to use) intervals at least 3 months apart and at 3-month intervals (Figure 8).

6.5 Discussion

In this study, we observed that the CKCis risk groups outperformed CUA risk groups for assessing prognosis and we recommend the CKCis model be preferred for patient counseling, stratification in clinical trials, and surveillance imaging protocols. Our findings were anticipated given that the CUA risk group stratification only incorporates tumour and lymph node stage.

One of the most important findings is the significant proportion of CUA low-risk patients reclassified as CKCis intermediate- or high-risk. With any risk-group label system, the risk of recurrence should coincide with what the patient perceives is “low”, “intermediate”, etc (label-to-risk concordance). In the CUA guideline, a patient with “intermediate risk” has a 25% risk of experiencing recurrence within 5 years (compared to a 10% risk in the CKCis intermediate group). We believe the risk of recurrence in the CKCis risk groups are more concordant with clinician and patient expectations. Compared to the CUA guideline,

the CKCis recommendation is fewer scans for low-risk patients and more frequent scans for higher risk patients. Alternative schedules using higher or lower interval detection risk could also be considered based on the patients risk tolerance.

Interestingly, clinicians in Canada may already be deviating from CUA guidelines because they are incorporating other prognostic factors, such as grade. In a recent analysis, only about 35% of patients had surveillance imaging compliant with CUA recommendations. Patients with low-risk (pT1) were the most common group "over-imaged" and patients at high-risk (pT3) were the most common group "under-imaged"(10). It is possible that the CKCis based surveillance model may provide a more valid assessment of risk, and hence we may observe more compliance with follow up imaging recommendations if the CKCis risk groups are adopted.

The American Urological Association's (AUA) and European Association of Urology (EAU) Guidelines provide different recommendations from the CUA, and from each other. The AUA divides patient risk groups based on pathological T stage, grade, and other tumour features (11). The EAU separates risk groups based on the Leibovich et al 2003 prognostic model (12). Compared to the CKCis surveillance schedule presented here, the AUA and EAU guidelines suggest more imaging for low risk patients (AUA 12mths, 24 mths, 48mths, and 60mths; EAU 6mths, 18mths, 30mths and once every two years beyond three years post-surgery) and fewer imaging tests for high risk patients (AUA 6mths,12mths,18mths,24mths,30mths,36mths,48mths and 60mths; EAU 3mths,6mths, 12mths,18mths,24mths,36mths and annually until 5 yrs) (11,12).

In addition to local recurrence, the most common sites for ccRCC recurrence are the lungs (70%), lymph nodes (45%), bones (32%), liver (18%) and adrenal gland (10%) (13). The CUA guidelines differentiates between thoracic and abdominal imaging (3) whereas the new proposed guideline recommends imaging of the chest and abdomen at each interval. Our recommendation to combine chest and abdominal imaging are consistent with AUA and EAU guidelines. (11,12).

6.5.1 Limitations

The main limitation of these new surveillance guidelines is that we do not know if any surveillance schedule will ultimately improve patient-important outcomes. We assume that early detection and treatment of recurrence at the lowest burden of disease will ultimately optimize outcomes such as treatment burden, disease symptoms, or survival. However, this assumption may be wrong. Furthermore, approximately 30% of RCC patients have non-clear cell histology. The CKCis model likely does not perform

well in non-clear cell patients given the differences in prognostic factors for patients with non-clear cell tumour types.

6.6 Conclusions

The CKCis-based risk group classification system incorporates more prognostic information compared to the CUA risk-groups and this results in improved discrimination, improved label-to-risk concordance, positive net reclassification improvement value. The current proposal recommends fewer surveillance scans for low risk and more scans for patients at higher risk. Future directions for research include a clinical trial evaluating the comparative effectiveness of the new proposed guidelines to the current guidelines in ccRCC patients and the subsequent development of clinical guidelines in non-Clear Cell RCC patients in addition to the creation of electronic medical record-based tools for clinical use.

6.7 References

1. Brenner DR, Weir HK, Demers AA, Ellison LF, Louzado C, Shaw A, Turner D, Woods RR, Smith LM; Canadian Cancer Statistics Advisory Committee. Projected estimates of cancer in Canada in 2020. CMAJ. 2020 Mar 2;192(9):E199-E205. doi: 10.1503/cmaj.191292.
2. Kim SH, Park B, Hwang EC, Hong SH, Jeong CW, Kwak C, Byun SS, Chung J. Retrospective Multicenter Long-Term Follow-up Analysis of Prognostic Risk Factors for Recurrence-Free, Metastasis-Free, Cancer-Specific, and Overall Survival After Curative Nephrectomy in Non-metastatic Renal Cell Carcinoma. Front Oncol. 2019 Sep 4;9:859. doi: 10.3389/fonc.2019.00859.
3. Kassouf W, Monteiro LL, Drachenberg DE, Fairey AS, Finelli A, Kapoor A, Lattouf JB, Leveridge MJ, Power NE, Pouliot F, Rendon RA, Sabbagh R, So AI, Tanguay S, Breau RH. Canadian Urological Association guideline for followup of patients after treatment of non-metastatic renal cell carcinoma. Can Urol Assoc J. 2018 Aug;12(8):231-238. doi: 10.5489/cuaj.5462.
4. Tajzler, C., Tanguay, S., Mallick, R., Ahrens, B., Ly, T. L., Breau, R. H., Basappa, N. S., Kapoor, A., Heng, D. Y., Pouliot, F., Finelli, A., Lavallee, L. T., So, A. I., Drachenberg, D. E., Soulieres, D., Bjarnason, G., Richard, P. O., Maloni, R., Power, N. E., Haan, M., & Wood, L. A. Determining generalizability of the Canadian Kidney Cancer information system (CKCis) to the entire Canadian kidney cancer population. Can Urol Assoc J. 2020 Jun 16; 14(10). doi: 10.5489/cuaj.6716.
5. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb). 2012;22(3):276-82.
6. Heng DY, Xie W, Regan MM, Harshman LC, Bjarnason GA, Vaishampayan UN, Mackenzie M, Wood L, Donskov F, Tan MH, Rha SY, Agarwal N, Kollmannsberger C, Rini BI, Choueiri TK. External validation and comparison with other models of the International Metastatic Renal-Cell Carcinoma Database Consortium prognostic model: a population-based study. Lancet Oncol. 2013 Feb;14(2):141-8. doi: 10.1016/S1470-2045(12)70559-4.
7. Leening MJ, Vedder MM, Witteman JC, Pencina MJ, Steyerberg EW. Net reclassification improvement: computation, interpretation, and controversies: a literature review and clinician's guide. Ann Intern Med. 2014 Jan 21;160(2):122-31. doi: 10.7326/M13-1522.
8. Jewell ES, Maile MD, Engoren M, Elliott M. Net Reclassification Improvement. Anesth Analg. 2016 Mar;122(3):818-824. doi: 10.1213/ANE.0000000000001141.

9. Siddiqui SA, Frank I, Cheville JC, Lohse CM, Leibovich BC, Blute ML. Postoperative surveillance for renal cell carcinoma: a multifactorial histological subtype specific protocol. BJU Int. 2009 Sep;104(6):778-785. doi: 10.1111/j.1464-410X.2009.08499.x.
10. Dragomir A, Aprikian A, Kapoor A, Finelli A, Pouliot F, Rendon R, Black PC, Moore R, Breau RH, Kawakami J, Drachenberg D, Lattouf JB, Tanguay S. Follow-up imaging after nephrectomy for cancer in Canada: urologists' compliance with guidelines. An observational study. CMAJ Open. 2017 Dec 11;5(4):E834-E841. doi: 10.9778/cmajo.20170005.
11. Campbell SC, Uzzo RG, Karam JA, et al: Renal Mass and Localized Renal Cancer: Evaluation, Management, and Follow-up: AUA Guideline: Part II. J Urol 2021; 206: 209.
12. EAU Guidelines. Edn. presented at the EAU Annual Congress Milan 2021. ISBN 978-94-92671-13-4.
13. Dudani S, de Velasco G, Wells JC, Gan CL, Donskov F, Porta C, Fraccon A, Pasini F, Lee JL, Hansen A, Bjarnason GA, Beuselinck B, Pal SK, Yuasa T, Kroeger N, Kanesvaran R, Reaume MN, Canil C, Choueiri TK, Heng DYC. Evaluation of Clear Cell, Papillary, and Chromophobe Renal Cell Carcinoma Metastasis Sites and Association With Survival. JAMA Netw Open. 2021 Jan 4;4(1):e2021869. doi: 10.1001/jamanetworkopen.2020.21869.

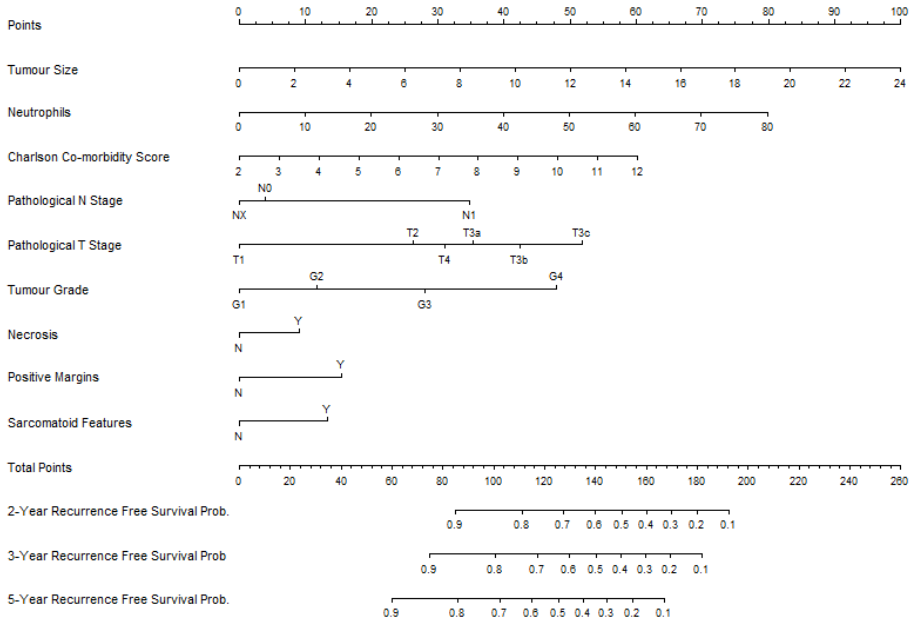
6.8 Figures and Tables

Table 1: Patient Follow-up Schedule Based on Patient Risk Group (3)

Months Post-Operation	**3**	**6**	**12**	**18**	**24**	**30**	**36**	**48**	**60**
Low-Risk (pT1+pN0/X stage)									
Chest X-Ray			X		X		X	X	X
Abdominal CT/MRI/Ultrasound					X				X
Intermediate-Risk (pT2+pN0/X stage)									
Chest X-Ray		X	X	X	X	X	X	X	X
Abdominal CT/MRI/Ultrasound			X		X		X		X
High-Risk (pT3-4+pN0/X stage)									
Chest X-Ray		X	X	X	X	X	X	X	X
Abdominal CT/MRI/Ultrasound		X	X	X	X		X		X
Very High-Risk (any pT stage+pN1 stage)									
Chest X-Ray	X	X	X	X	X	X	X	X	X
Abdominal CT/MRI/Ultrasound	X	X	X	X	X	X	X	X	X

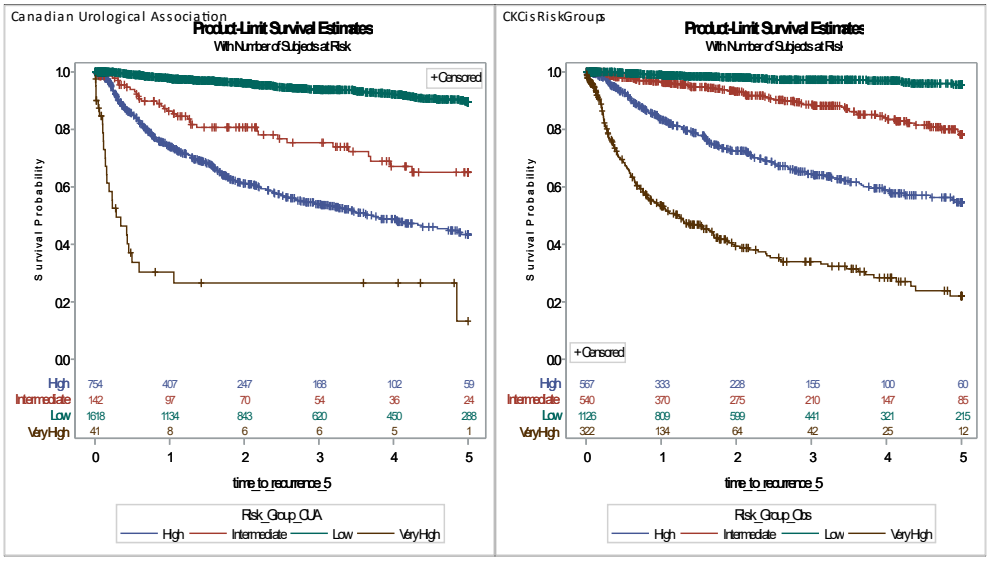
Abbreviation: X=Imaging required at timepoint**;** pN0/X=pathological N0 or NX stage; pN1=pathological N1 stage; pT1=pathological T1 stage; pT2=pathological T2 stage; pT3-4=pathological T3 or T4 stage; CT=Computed Tomography, MRI=Magnetic Resonance Imaging

Figure 1: CKCis Recurrence Free Survival Nomogram



Abbreviations: N0=pathological N0 stage; NX=nodal resection not assessed; N1=pathological N1 stage; T1=pathological T1 stage; T2=pathological T2 stage; T3a=pathological T3a stage; T3b=pathological T3b stage ; T3c=pathological T3c stage;T4=pathological T4 stage;G1= tumour grade 1; G2=tumour grade 2;G3=tumour grade 3; G4=tumour grade 4;N=no; Y=yes; Recurrence Free Survival Prob.=Recurrence Free Survival Probability
Tumour size is measured in cm and neutrophil concentration is measured in x 10^9 cells/µL.

Figure 2: Kaplan-Meier Survival Curves for CUA and CKCis risk groups



Abbreviation: CKCis=Canadian Kidney Cancer Information System

Table 2: Recurrence Rates Overall, 2- and 5-years Post-Surgery by Risk Group Stratification System

Risk Groups	Recurrence at 2-years		Recurrence at 5-years	
	N	%	N	%
Canadian Urological Association Guideline Risk Groups				
Low	45	3.9%	82	10.5%
Intermediate	23	19.3%	33	34.9%
High	223	38.8%	271	56.6%
Very High	25	73.4%	26	86.7%
CKCis Risk Groups				
Low	15	1.8%	25	4.5%
Intermediate	26	6.8%	55	21.8%
High	113	27.5%	152	45.4%
Very High	162	60.6%	180	78.0%

Abbreviations: N=number of patients; CKCis=Canadian Kidney Cancer Information System

Figure 3: Proportion of CKCis patients in CUA and CKCis risk groups

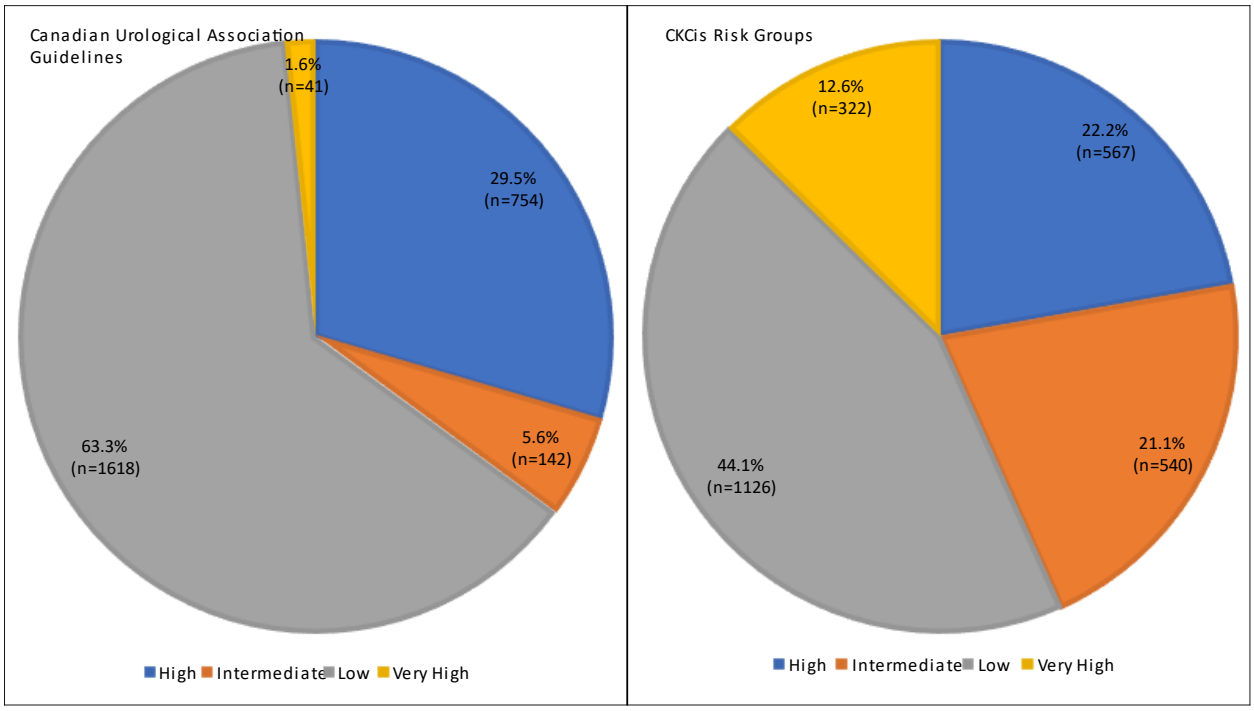


Figure 4: Reclassification of patients in Canadian Urological Association Risk Groups into CKCis Risk Groups

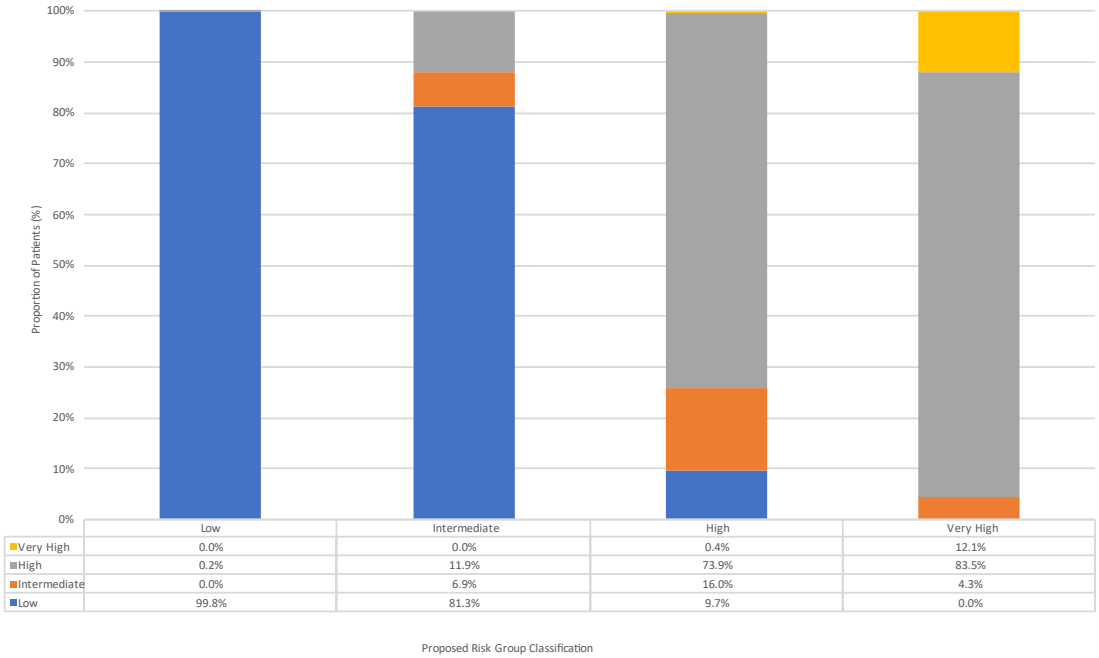
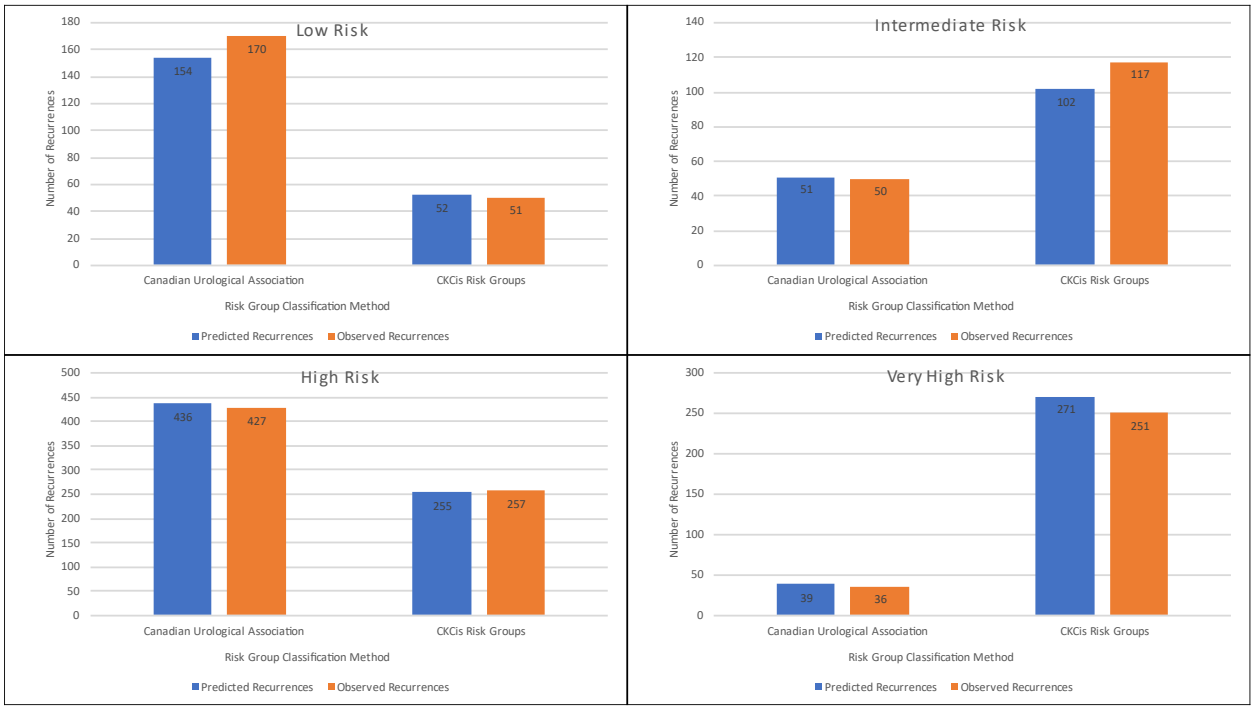


Figure 5: Predicted and Observed Number of Recurrences 5-years Post-Surgery



Abbreviation: CKCis=Canadian Kidney Cancer Information System

Figure 6: Timepoints where Recurrence Incidence Increased by at Least 1%[1]

Risk Group	Time Post-Surgery (Months)																																	
	0	1	3	5	6	7	8	9	10	11	12	15	16	18	19	20	22	25	26	29	33	35	36	38	41	43	46	47	50	51	54	55	58	59	60
Low									X										X										X				X		
Intermediate			X			X				X			X			X		X		X	X		X		X		X			X	X			X	
High		X		X			X				X	X		X			X		X	X		X		X		X		X				X	X		
Very High	X		X		X			X			X	X			X		X		X	X				X		X				X			X		

Abbreviations: X=Imaging required at timepoint

[1]For Very High-risk, the first timepoint is 4 days post-surgery.

Figure 7: Timepoints where Recurrence Incidence Increased by at Least 2%[1]

Risk Group	Time Post-Surgery (Months)																											
	0	2	3	5	6	10	13	14	18	19	21	23	25	26	28	30	33	38	41	43	44	47	49	51	52	55	56	60
Low													X											X				
Intermediate				X			X				X		X				X		X				X				X	
High		X			X	X		X		X			X			X		X		X		X				X		
Very High	X		X		X	X		X	X			X			X			X			X		X		X			

Abbreviations: X=Imaging required at timepoint

[1]For Very High-risk, the first timepoint is 7 days post-surgery.

Figure 8: Recommended CKCis imaging schedule for non-metastatic patients with ccRCC

	Follow-up Time Post-Surgery (months)													
Risk Groups	**3**	**6**	**9**	**12**	**15**	**18**	**21**	**24**	**30**	**36**	**42**	**48**	**54**	**60**
Low								X				X		
Intermediate		X		X				X		X		X		X
High	X	X	X	X		X		X	X	X	X	X	X	X
Very High	X	X	X	X	X	X	X	X	X	X	X	X	X	X

Abbreviations: ccRCC=Clear Cell Renal Cell Carcinoma;X=Imaging required at timepoint

Chapter 7: Discussion

7.1 Discussion Introduction

Prognostic models are developed to predict future patient outcomes. If we can accurately predict outcomes, we can appropriately counsel patients, offer preventative therapeutics (if they exist), and recommend risk concordant imaging surveillance. Accurate prognostic models also allow for more efficient observational studies and clinical trials. Specifically, prognostic models provide estimates to calculate endpoint distributions across clinical trial arms for sample size calculations, determining generalizability, and comparing prognosis in treated and untreated arms (1). In addition to this, prognostic model-based risk groups are useful to stratify patients prior to randomization in clinical trials. This is certainly relevant in kidney cancer research as previous studies examining the effect of post-surgery adjuvant therapy have shown inconsistent findings – potentially due to an absence of standardized risk categorization, or due to use of risk groups that are not accurate (2).

The goal of this thesis was to understand existing prognostic models and potentially improve on existing models for patients with kidney cancer. We accomplished these goals by validation of published models, examination of the effect of baseline hazard function on validation, creation of a new prognostic model for recurrence, and applying the new model to establish risk groups, and corresponding imaging surveillance recommendations. This chapter will summarize our work, review clinical applications, and suggest directions for future research.

7.2 Summary of Findings

The first objective of the thesis was the examine the accuracy of published prognostic models for kidney cancer recurrence, cancer related death, and death from any cause in Canadian kidney cancer patients. 15 models for recurrence, 6 models for cancer specific death, and 4 models for overall survival were identified from literature searches. Some models were specific for kidney cancer sub-groups based on histology or tumour stage. In general, we observed a wide range in accuracy. For prognostic models examining recurrence in all patients, Yaycioglu et al 2013 had the best discriminative ability, Kattan et al 2001 had the lowest Brier Scores over time and van der Mijn et al 2019 had the best calibration. Among models specific to clear cell renal cell carcinoma (ccRCC) recurrence, Leibovich et al 2003 had the best discrimination and lowest Brier Scores and Leibovich et al 2018 had the best calibration. For papillary renal cell carcinoma (pRCC) recurrence, Klatte et al 2019 was the best performing model.

Among prognostic models examining the outcomes of cancer specific survival, Kutikov et al 2010 had the best discrimination, calibration, and Brier Scores; in ccRCC patients, Leibovich et al 2018 had the best

discrimination and Brier Scores at 3-and 5-years post-surgery, Frank et al 2002 had the best calibration at 2-years post-surgery.

Lastly, when considering prognostic models examining overall survival, Zisman et al 2001 had the best discrimination and Brier Scores but Zisman et al 2002 had the best calibration. Interestingly, many prognostic models had lower discriminative ability than pathological T stage alone and the presence of over- and under- estimation of recurrence, cancer specific mortality and all-cause mortality risks suggested that room for improvement was present. Further, decision curve analysis demonstrated that prognostic models examining specific patient groups had a larger range of patient risk thresholds for which prognostic models were clinically useful compared to prognostic models examining all patients.

When validating previously published models, it became apparent that baseline hazard was frequently omitted from publications, and we wondered how this might impact predictive performance. The second objective of the thesis was to examine the effect of baseline hazard function variation on assessment of model discrimination and calibration. The Weibull, Exponential, Log-Logistic and Log-Normal distributions were examined with changes in the rate parameter, shape parameter, or both (the mean and standard deviation were used for the Log-Normal distribution) along with over- and under-estimates of 5%-20% and three methods for calculating the calibration slope was assessed: Cox Proportional Hazards model fitted as is, Poisson regression adjusting for model development dataset, and Poisson regression adjusting for the value of the baseline hazard function in the model development dataset at a single timepoint. We found that discrimination is robust to baseline hazard function miscalibration as the estimated c-statistic did not vary considerably across different deviations in the baseline hazard function. However, the calibration slope varied considerably, especially with Log-Logistic and Log-Normal distributions. It seems the ideal method to assess the calibration slope varies based on the shape of the baseline hazard function and whether rate and shape parameters are over- or under-estimated. When the calibration slopes of five models were compared across different assumed distribution shapes, the calibration slopes varied slightly and the relative ranking from best to worst calibration remained constant. In summary, assessing discrimination and choosing the best performing model is likely reasonable without baseline hazard information, but estimates of calibration have the potential to be inaccurate (overestimated) if the baseline hazard function is not available.

The third objective of this thesis was to create a new prognostic model for recurrence free survival in ccRCC patients. The p-value selection model including pathological T stage performed best. As expected,

the accuracy of the CKCis model was better on internal validation compared to external validation of pre-existing models.

Based on the CKCis model, risk groups were established and compared to the CUA guideline groups. The CKCis risk-groups had improved distribution, label-to-risk concordance, discrimination, and calibration. We applied the CKCis risk groups to provide a surveillance schedule for patients following surgery. Low risk patients are recommended to have imaging at 24- and 48-months post-surgery. Intermediate risk patients are recommended to have imaging performed at 6, 12,24,36,48 and 60 months post-surgery. High-risk patients are recommended to have imaging every 3 months until 1-year post-surgery and subsequent scans every 6 months until 5-years post-surgery; Very High-risk patients are recommended to have imaging at every 3 months until 2-years post-surgery and subsequent scans every 6 months until 5-years post-surgery. Compared to the CUA recommendations, our imaging schedule requires fewer scans for low-risk patients and more scans for High and Very-High risk patients.

7.3 Practical Considerations in Prognostic Modelling

7.3.1 Issues Related to Prognostic Factor Identification

Prognostic factors may be identified from previous research combined with biological reasoning and causal pathways. Prognostic factors should provide information that is independent from each other (3). A complete understanding of prognostic factors is usually impossible as we have an incomplete understanding of biology, biased publications, small sample sizes, inaccurate data capture, and inadequate statistical analysis (3). Ideally, prognostic factors are assessed in large cohort studies that are representative of the population, with sufficient statistical power, and appropriate statistical analysis (3).

7.3.2 Issues in Development and Assessment of Prognostic Models

There is a lifecycle of prognostic modes: creation, external validation, impact assessment, and update. Ideally, prognostic models are developed from large representative cohorts with analytic plans to account for common issues (missing data, clustering, etc) (4). The quality of the cohort data is obviously important and details about how the data was collected and codified should be clearly reported (4). The CKCis cohort is one of the largest and most diverse in the world, thus an ideal resource for model development and assessment.

Methods for internal validation may be important. Different methods of optimism correction in bootstrapping found that Harrell's optimism correction method was comparable to other methods except in small sample size settings, where the 0.632+ estimator was preferred (5). In chapter 7 we found that

optimism corrected bootstrapping yielded more optimistic measures of model performance than k-fold validation, a finding supported by previous research (6).

As outlined, external validation of prognostic models is required prior to applying the model in different populations (4) but prognostic models should also be updated over time when clinical practice changes or new information emerges (4). For instance, independent of tumour histology, there may be prognostic information from blood markers andthe Leibovich et al 2003 prognostic model has been expanded by adding mean platelet volume as a potentially new parameter (7). Another example in our work is the inclusion of neutrophil concentration. In all situations, the clinical impact of prognostic models should also be examined. In this thesis, the clinical utility was evaluated using decision curve analysis and various methods were used to examine the impact of CKCis risk groups on patient surveillance.

7.3.3 Issues in Applying Prognostic Models in Clinical Practice

The PROGRESS research group has provided recommendations for applying research in clinical practice. It is important to consider that prediction models are imprecise and there may be an imbalance in the consequences of a false negative compared to a false positive. For example, the CKCis model may be applied to the selection of adjuvant (preventative) therapy after surgery. An individual patient may place a higher value on prevention of cancer recurrence, while another may place a higher value on avoiding treatment-related toxicity. These values and preferences should be incorporated when discussing treatment thresholds to an individual patient considering their treatment options.

Given the significant considerations that must be taken into account when deciding whether a prognostic model should be used, a flowchart has been developed to guide the decision making process (Figure 1).

7.3.4 Limitations of the CKCis Prognostic Model

Despite the strengths of the CKCis cohort, there are several limitations. For instance, we could not assess some potentially prognostic markers, such as genetic profiles, as these data are not available for most patients. There may also be selection/referral bias in the CKCis sites, as they are all tertiary care institutions. Incomplete data, incomplete follow up, and short duration of follow-up are additional limitations with an unknown impact on model performance.

7.4 Future Directions for Research

The optimal method for determining the value of prognostic model-based care is through randomized clinical trials (RCTs). However, prediction model RCTs are rare because they are costly and difficult to implement (8). In the absence of RCTs, rigorous model development and external assessment is the most feasible method to determine if model incorporation improves patient counseling and delivery of care.

7.4.1 Need for External Validation of CKCis Model

The CKCis model must be validated externally prior to clinical use in other populations. Even if the CKCis model is found to be accurate in other populations, further validation in the future is also necessary. Prognostic models can often become "victims of their own success" as effective prognostic models used in treating patients may change outcome distributions over time, causing the prognostic model to lose effectiveness (9). Another reason for repeated model assessment is that diagnostic technology and practice patterns may change over time. Examples in kidney cancer are the increased use of 'active surveillance' (instead of surgery), increased use of partial nephrectomy (instead of radical nephrectomy), and imaging related 'stage migration' where tumours are detected at smaller sizes and lower stages (10).

7.4.2 Incorporation of Prognostic Model into Electronic Medical Record and Online Tools

Electronic medical records can potentially result in fewer medical errors, more standardized care (i.e. computerized reminders regarding guidelines), increased data availability, and improved organization (11). Online tools have been implemented for kidney prognostic models including the UISS model for non-metastatic patients or the IMDC criteria for metastatic patients (12). Incorporating CKCis risk models into electronic medical records and online nomograms should be possible.

A prognostic model is only useful if it is used. Despite the accuracy of nomograms, they historically have not been used in clinical practice because they are not understood, are time consuming, or are cumbersome. Data visualization using an interactive nomogram version of the International Metastatic Renal Cell Carcinoma Database Consortium (IMDC) criteria for metastatic kidney cancer patients has been shown to have ease of use and acceptability; increased accuracy in prognosis by physicians when presented with an updated version of the tool was also observed (13). It remains to be seen whether a data visualization tool integrating data from patient electronic medical records into the newly developed prognostic model for kidney cancer recurrence would have the same effectiveness and acceptability. Future incorporation of the CKCis model into online tools and electronic medical records may allow for higher use. Methods to encourage and facilitate the use of the CKCis models requires further research.

7.4.3 Incorporation of CKCis Model in Clinical Trials

In addition to cohort validation, prospective clinical trials can offer insight on how prognostic models predict clinical outcomes. Stratification of patients into risk groups prior to randomization in clinical trials facilitates balance between treatment arms while maintaining the benefits of randomization; this ensures that differences in outcome arising from differences in risk factors is virtually eliminated and observed differences between allocation arms is more likely due to the intervention being studied. Accurate

stratification also decreases variation between the clinical trial arms in all aspects aside from treatment provided, decreasing the sample size required for the clinical trial to provide a definitive answer (14,15).

7.4.4 Incorporation of Prognostic Models into Surveillance Guidelines

While the current Canadian Urological Association guidelines do not incorporate any prognostic model into its assessment of patient risk groups (16), the European Association of Urology recommends the division of patients into Low, Intermediate and High risk based on Leibovich et al 2003 risk scores for Clear Cell renal carcinoma patients and tumour grade, pathological T and N stage used for non-Clear Cell patients with imaging recommended every three months in high-risk patients and annually starting at six months post-surgery for Low-risk patients (17). Overall, guidelines for kidney cancer management vary in terms of risk group classification and imaging recommendations. Assuming that patient outcome is dependent on clinician's treatment decision, which may vary based on the guidelines used, it is important to evaluate guidelines prior to clinical use. Ideally, surveillance guidelines are based on evidence and then studied in a randomized trial. While surveillance trials are uncommon, they are possible. An example of a surveillance trial is the TRISST trial which randomized patients into four different trial arms representing different imaging types and schedules (18). The primary outcome was disease relapse and the authors also evaluated tumour size at relapse, overall survival, disease-free survival, prognostic factor assessment, and resource use (18). A similar approach may be possible for a kidney cancer surveillance trial.

7.4.5 New Predictive Biomarkers and the Improvement of the Prognostic Model

The addition of new predictive biomarkers may improve the performance of the CKCis model. Recent research has suggested blood, urine, or genetic factors may be useful. Nuclear renin, ubiquitin-specific protease 2, circulating tumour DNA, circulating tumour cells, and unique patterns of gene expression are some of the biomarkers under evaluation (19). A review of studies examining prognostic biomarkers for renal cell carcinoma identified the most commonly validated genetic markers are KI67, BIRC5, TP53, CXCR4 and CA9 (20). Therefore, the addition of new biomarkers to CKCis data collection and expansion of prognostic models has the potential to considerably improve prognostic performance.

7.4.6 Develop models for Papillary and Chromophobe Kidney Cancer Patients

Results from the external validation of prognostic models demonstrated that recurrence free survival prognosis varied by histological subtype. Non-clear cell RCC models do not perform well, perhaps due to small sample size. We are endeavouring to combine CKCis data with other databases worldwide to generate prognostic models for patients with rarer RCC histology.

7.4.7 Develop models for Cancer Specific Survival and Overall Survival

Results from the external validation of prognostic models demonstrated that the performance of cancer specific survival and overall survival models also varied considerably. Improvements in data quality for the collection of cancer specific survival data is needed. As room for improvement exists, new prognostic models to improve the prediction of cancer specific survival and overall survival may be needed for future trials in very high risk non-metastatic patients. As the CKCis cohort follow up matures, models predicting cancer related death and overall survival will be generated.

7.5 Conclusions

Many prognostic models for kidney cancer recurrence free survival, cancer specific survival and overall survival have been published. However, pre-existing prognostic model performance varied widely when applied to contemporary Canadian patients. We found that the reporting baseline hazard function and distribution should be encouraged to allow better assessment of calibration. Using a large cohort of Canadian patients, we developed a prognostic model with improved accuracy that can be applied to clinical care. External validation in other populations will be important prior to widespread adoption.

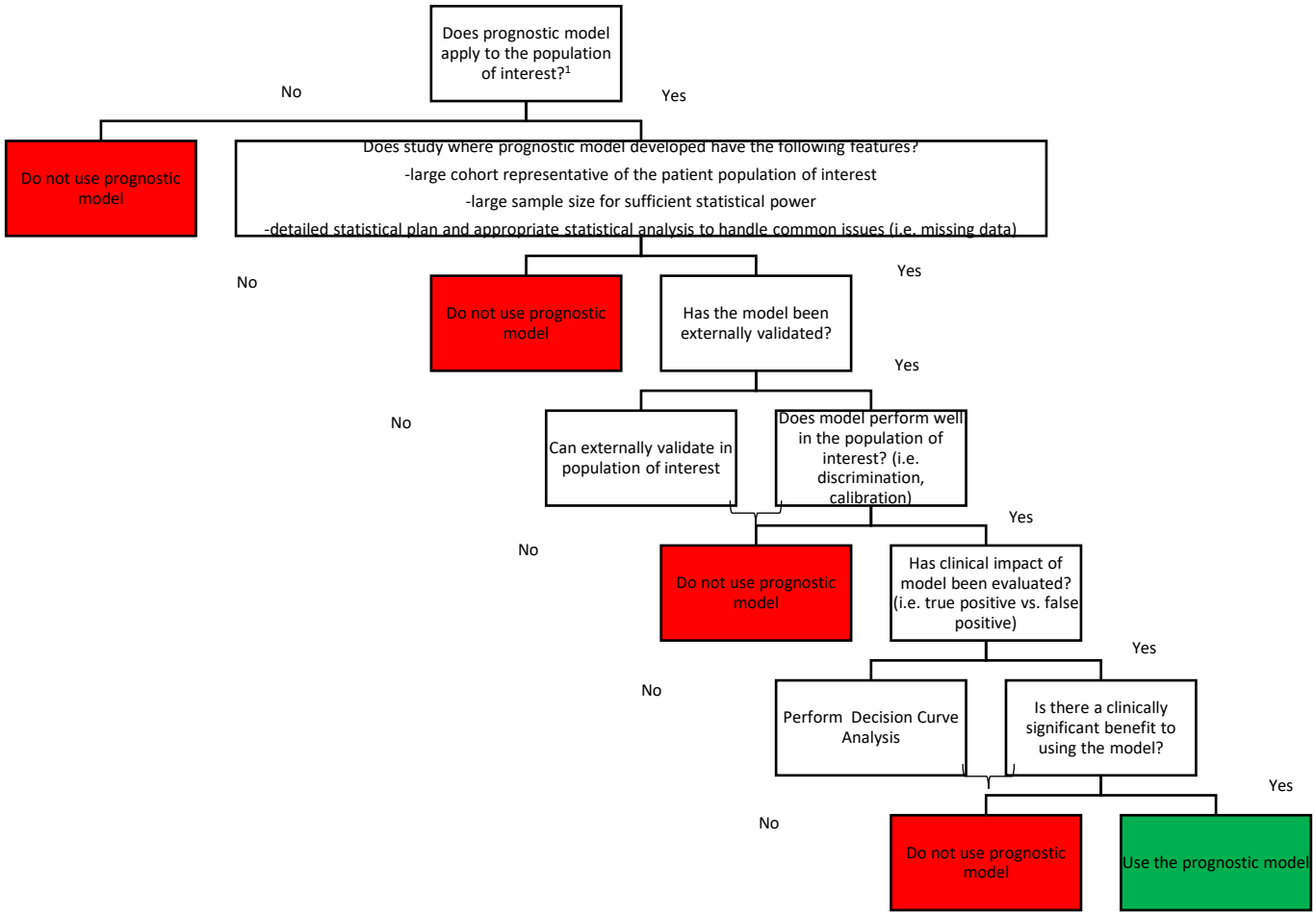
7.6 References

1. Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A, Briggs A, Udumyan R, Moons KG, Steyerberg EW, Roberts I, Schroter S, Altman DG, Riley RD; PROGRESS Group. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. BMJ. 2013 Feb 5;346:e5595. doi: 10.1136/bmj.e5595.
2. Sharma T, Tajzler C, Kapoor A. Is there a role for adjuvant therapy after surgery in "high risk for recurrence" kidney cancer? An update on current concepts. Curr Oncol. 2018 Oct;25(5):e444-e453. doi: 10.3747/co.25.3865.
3. Riley RD, Hayden JA, Steyerberg EW, Moons KG, Abrams K, Kyzas PA, Malats N, Briggs A, Schroter S, Altman DG, Hemingway H; PROGRESS Group. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. PLoS Med. 2013;10(2):e1001380. doi: 10.1371/journal.pmed.1001380.
4. Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, Riley RD, Hemingway H, Altman DG; PROGRESS Group. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. PLoS Med. 2013;10(2):e1001381. doi: 10.1371/journal.pmed.1001381.
5. Iba K, Shinozaki T, Maruo K, Noma H. Re-evaluation of the comparative effectiveness of bootstrap-based optimism correction methods in the development of multivariable clinical prediction models. BMC Med Res Methodol. 2021 Jan 7;21(1):9. doi: 10.1186/s12874-020-01201-w.
6. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. in: IJCAI'95: Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2, 1995. p. 1137-1143.

7. Seles M, Posch F, Pichler GP, Gary T, Pummer K, Zigeuner R, Hutterer GC, Pichler M. Blood Platelet Volume Represents a Novel Prognostic Factor in Patients with Nonmetastatic Renal Cell Carcinoma and Improves the Predictive Ability of Established Prognostic Scores. J Urol. 2017 Dec;198(6):1247-1252. doi: 10.1016/j.juro.2017.07.036.
8. Kappen TH, van Klei WA, van Wolfswinkel L, Kalkman CJ, Vergouwe Y, Moons KGM. Evaluating the impact of prediction models: lessons learned, challenges, and recommendations. Diagn Progn Res. 2018 Jun 12;2:11. doi: 10.1186/s41512-018-0033-6.
9. Lenert MC, Matheny ME, Walsh CG. Prognostic models will be victims of their own success, unless…. J Am Med Inform Assoc. 2019 Dec 1;26(12):1645-1650. doi: 10.1093/jamia/ocz145.
10. Patel HD, Gupta M, Joice GA, Srivastava A, Alam R, Allaf ME, Pierorazio PM. Clinical Stage Migration and Survival for Renal Cell Carcinoma in the United States. Eur Urol Oncol. 2019 Jul;2(4):343-348. doi: 10.1016/j.euo.2018.08.023.
11. Menachemi N, Collum TH. Benefits and drawbacks of electronic health record systems. Risk Manag Healthc Policy. 2011;4:47-55. doi: 10.2147/RMHP.S12985.
12. Zisman, A., 2022. UCLA Integrated Staging System (UISS) for Renal Cell Carcinoma (RCC) - MDCalc. [online] Mdcalc.com. Available at: <https://www.mdcalc.com/ucla-integrated-staging-system-uiss-renal-cell-carcinoma-rcc#evidence> [Accessed 27 August 2022].
13. Shee K, Pal SK, Wells JC, Ruiz-Morales JM, Russell K, Dudani S, Choueiri TK, Heng DY, Gore JL, Odisho AY. Interactive Data Visualization Tool for Patient-Centered Decision Making in Kidney Cancer. JCO Clin Cancer Inform. 2021 Aug;5:912-920. doi: 10.1200/CCI.21.00050.
14. Ganju J, Zhou K. The benefit of stratification in clinical trials revisited. Stat Med. 2011 Oct 30;30(24):2881-9. doi: 10.1002/sim.4351.
15. Kernan WN, Viscoli CM, Makuch RW, Brass LM, Horwitz RI. Stratified randomization for clinical trials. J Clin Epidemiol. 1999 Jan;52(1):19-26. doi: 10.1016/s0895-4356(98)00138-3.
16. Kassouf W, Monteiro LL, Drachenberg DE, Fairey AS, Finelli A, Kapoor A, Lattouf JB, Leveridge MJ, Power NE, Pouliot F, Rendon RA, Sabbagh R, So AI, Tanguay S, Breau RH. Canadian Urological Association guideline for followup of patients after treatment of non-metastatic renal cell carcinoma. Can Urol Assoc J. 2018 Aug;12(8):231-238. doi: 10.5489/cuaj.5462.
17. EAU Guidelines. Edn. presented at the EAU Annual Congress Milan 2021. ISBN 978-94-92671-13-4.
18. Joffe J, Cafferty F. TRISST Trial of Imaging and Schedule in Seminoma Testis [Internet]. London; 2015 May 19 [cited:2022 Mar 04]. Available from: https://www.mrcctu.ucl.ac.uk/media/1302/trisst-protocol.pdf
19. Marchioni M, Rivas JG, Autran A, Socarras M, Albisinni S, Ferro M, Schips L, Scarpa RM, Papalia R, Esperto F. Biomarkers for Renal Cell Carcinoma Recurrence: State of the Art. Curr Urol Rep. 2021 Apr 22;22(6):31. doi: 10.1007/s11934-021-01050-0.
20. Petitprez F, Ayadi M, de Reyniès A, Fridman WH, Sautès-Fridman C, Job S. Review of Prognostic Expression Markers for Clear Cell Renal Cell Carcinoma. Front Oncol. 2021 Apr 28;11:643065. doi: 10.3389/fonc.2021.643065.

7.7 Figures and Tables

Figure 1: Decision Making Process on Whether a Prognostic Model should be Used[1]



[1]While the prognostic model should be validated in the clinical population of interest, this does not have to be exactly the specific population being examined and can be used with caution in other contexts (i.e. the CKCis prognostic model was developed in Canadian Clear Cell Renal Cell Carcinoma (CCRC) patients and can be used in CCRC patients in a specific non-CKCis site hospital).

Appendix

Appendix 1: Ethics Approval

The Ottawa Hospital | L'Hôpital d'Ottawa
RESEARCH INSTITUTE | INSTITUT DE RECHERCHE

March 28, 2022

Dr. Rodney Breau

Ottawa Hospital - General Campus
Department of Surgery
Division of Urology
501 Smyth Road, Box #222
Ottawa, ON K1H 8L6

Re: OHRI Institutional Approval for Ottawa Health Science Network Research Ethics Board (OHSN-REB) Submission

Protocol ID#: 20220143-01H;

Estimating Prognosis of Patients with Kidney Cancer

Dear Dr. Rodney Breau,

This letter serves as **Ottawa Hospital Research Institute (OHRI)** Institutional Approval for the above-referenced study. Please maintain this documentation in your investigator study file.

Based on the information you provided about this study through the Clinical Research Registration Form, you have satisfied the requirements for institutional (OHRI) approval. This includes initial research ethics approval by OHSN-REB, appropriate departmental/service area notifications and execution (fully signed versions) of all agreement(s) required to begin the study locally. Please note there may be additional agreement(s) pending execution that are required to send funds, samples, or data to external sites, but are not required for you to begin your study locally.

Changes and/or additions to your study that may require additional agreement(s) or revisions to existing agreement(s) must be communicated to the OHRI Contracts Office. This should be undertaken simultaneously with any related OHSN-REB amendment submission.

Changes and/or additions to your study that affect various hospital/institution departments (e.g., pharmacy, Department of Medical Imaging, EORLA, EEG, etc.) must be communicated to the relevant departments.

As mentioned in the 'Response' tab of the Ethics application, you have 3 months from the date of initial OHSN-REB approval to submit French documents including the translation certificate to OHSN-REB through the Translated Documents section of the ethics application (if applicable).

Should you have any questions, please contact REBadministration@ohri.ca or 613-798-5555 extension 16719.

Signature Blocked Out

Nancy Carnack, BScN RN MBA CCRP
Director, Clinical Research Administration
Ottawa Hospital Research Institute | Institut de recherche de l'Hôpital d'Ottawa

Civic Campus, Box 675, 725 Parkdale Avenue, Ottawa, Ontario, K1Y 4E9
613-798-5555 extension 16719 Fax : 613-761-4311 http://www.ohri.ca/ohsn-reb

Appendix 2: Testing Assumptions of Cox Proportional Hazards Models

Figure 1: Schoenfeld Residuals for Proportional Hazards Assumption Test in New CKCis Prognostic Model for Clear Cell Renal Cell Carcinoma Patients

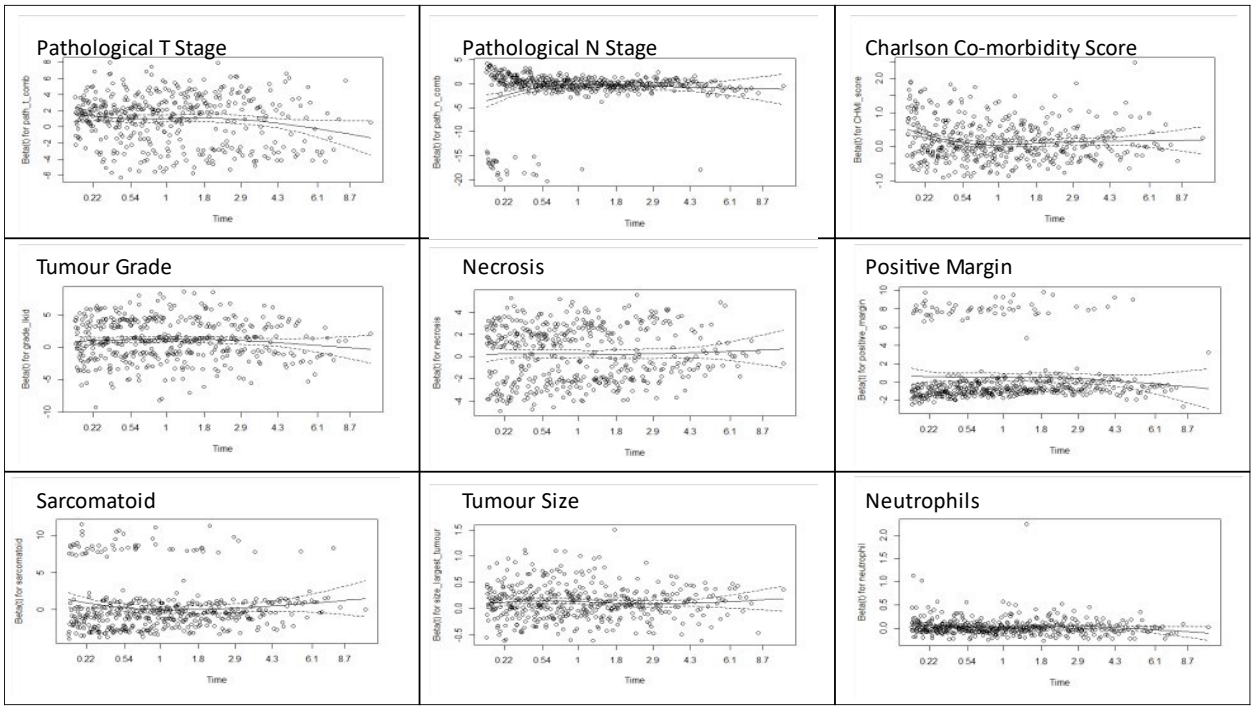
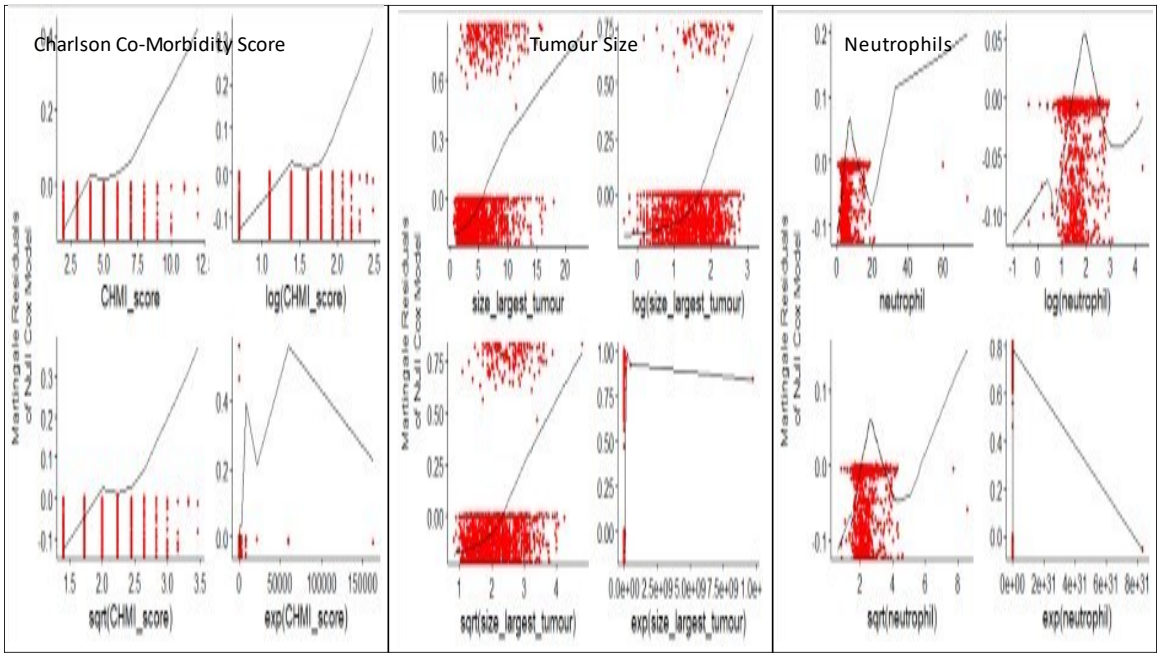


Figure 2: Martingale Residuals for Linearity Assumption Test in New CKCis Prognostic Model for Clear Cell Carcinoma Patients



Abbreviations: sqrt=Square Root; log=Natural Logarithm; exp=Exponential