University of Calgary

PRISM Repository **https://prism.ucalgary.ca**

The Vault Open Theses and Dissertations

2020-09-23

The association between 'compliance with colonoscopy surveillance' after primary treatment and healthcare utilization

Qaedi, Atena

Qaedi, A. (2020). The association between 'compliance with colonoscopy surveillance' after primary treatment and healthcare utilization (Master's thesis, University of Calgary, Calgary, Canada). Retrieved from https://prism.ucalgary.ca.

http://hdl.handle.net/1880/112585

Downloaded from PRISM Repository, University of Calgary

UNIVERSITY OF CALGARY

The association between 'compliance with colonoscopy surveillance' after primary treatment and healthcare utilization

by

Atena Qaedi

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF ARTS

GRADUATE PROGRAM IN ECONOMICS

CALGARY, ALBERTA

SEPTEMBER, 2020

© Atena Qaedi 2020

Abstract

Choosing Wisely Canada recommends surveillance with colonoscopy for colorectal cancer patients undergoing curative-intent treatment. Although surveillance with colonoscopy after surgery is beneficial in terms of early detection of recurrence and survival, there is limited real-world evidence on the compliance of recommended colonoscopy surveillance, and the health utilization and costs associated with it. This retrospective study uses existing administrative data sets from Alberta Health Services and Alberta Health, which includes 7120 observations for the 2004-2015 period. The study sample consisted of colorectal cancer patients at stages I and II, who underwent curative-intent surgery. This project compared healthcare utilization (measured by cost) and health outcomes (measured by survival) for patients who complied with colonoscopy surveillance, versus those who did not comply. Cost and survival analysis were conducted, employing multivariate analyses via COX and logistic regressions. For the purposes of this study, cost data was calculated using the physician claims or the physician's payment. In total, 6,962 patients were eligible for analysis. The median age was 67 (range: 18-104) years old. The proportion of patients with stage I and II colorectal cancer was 42.46% and 57.54%, respectively. A total of 2,812 (40.39%) patients had a one-year compliance, and 275 patients (3.95%) had two-to-five-year compliance. The average healthcare utilization of one-year and two-to-five-year compliance per person was 3,762 and 4,758 in CAD dollars, respectively. Compliance with colonoscopy surveillance after a primary treatment was associated with lower age, earlier cancer stage (stage I), lower cancer grade (grade 1), lower CCI, and higher income. In addition, the overall death ratio and cancer-related death ratio was lower for those patients with compliance in each category (one-year and two-five-year follow-up), compared to those with no compliance. The results of this study suggest that colonoscopy surveillance compliance following primary treatment for early-stage colorectal cancer is associated with lower healthcare utilization and better cancer-specific survival.

Keywords: Healthcare Utilization, Economics, Cost, Logistic Regression, Cox Regression.

Preface

This thesis is an original work by the author Atena Qaedi. No part of this thesis has been previously published.

Acknowledgments

There are many individuals I would like to thank who supported me while I completed this thesis. I would like to express my sincerest gratitude towards my supervisor, Prof. Mingshan Lu, for her continuous guidance, encouragement, and support during my M.Sc. program. Her ideas, feedback, and vision helped me shape my research career. Without her wisdom, this thesis would not have been possible. Working with Prof. Mingshan Lu has provided me with a great opportunity to collaborate with the ACRON(Alberta cancer outcome and research network) team during my project, as well as write a peer-reviewed paper.

I would like to sincerely thank Dr. Yuan Xu from the Departments of Oncology, Community Health Sciences and Surgery. He supported the paper writing process for my thesis, and provided me with material, without which this thesis would never have come to fruition.

I also would like to acknowledge Dr. Winson Cheung from the Oncology department who leads the ACORN for providing guidance and data access.

I also would like to thank Shiying Kong from the Centre for Health Informatics (CHI), who kindly helped me with data preparation and analysis, which has enriched the content of this thesis.

I would like to especially extend my gratitude to my committee members, Dr. Lasheng Yuan and Dr. Yuan Xu again. I am greatly appreciative of your support throughout my journey. I would like to thank them for inspiring me and for making my defense day a very memorable one.

A special thank you goes out to my wonderful family: my parents and brothers, my lovely aunt, and my brother-in-law. I would like to thank them for being the best family they could ever be, for being an inspiration to how I understand caregiving, and for always having high hopes for me. Without their loves, support, and prayers, I would not have been able to complete my M.Sc.

Finally, I would like to express my deepest gratitude to my husband, Meysam Moradizaniani, who

has always encouraged and supported me to follow my ambitions and goals, whatever they may be.

Tables of Contents

List of Tables

List of Figures and Illustrations

List of Abbreviations

CRC: Colorectal cancer

FOBT: Faecal Occult Blood Test

FIT: Fecal Immunochemical Test

FSIG: Flexible Sigmoidoscopy

AJCC: American Joint Committee on Cancer

GP: General Practitioner

ACR: Alberta Cancer Registry

DAD: Discharge Abstract Database

NACRS: National Ambulatory Care Reporting System

CCI: Charlson Comorbidity Index

OR: Odds Ratio

HR: Hazard ratio

1 Introduction

The cancer burden continues to grow globally, and the disease is becoming a major economic expenditure for all developed countries. The global cancer burden is anticipated to reach 27.5 million new cancer cases by 2040, and 16.3 million cancer deaths solely due to the population's growth and its aging [1]. In addition, cancer treatments exerting tremendous physical, emotional, and, more importantly, financial strain on individuals, families, communities, and health systems [2]. Among all types of cancer, the most common cancer that stands in 3rd place is colorectal cancer globally. Subsequently, Colorectal cancer (CRC) has become a worldwide problem due to its high incidence rate and increasing burden [3],[4].

Cancer that originates from the colon and rectum (parts of the large intestine) is termed colorectal cancer (CRC), also known as bowel cancer or colon cancer. Based on Canadian Cancer Statistics 2019 reports [5], colorectal cancer is the third most common cause of cancer in both men and women in Canada. More specifically, according to the Alberta Health Services' report [6], about 27,000 new diagnoses were reported in 2017 in Canada. Of those, an incidence of 79.6 per 100,000 was identified for men and 54.9 per 100,000 for women. Colorectal cancer also remains the third leading cause of death from cancer in women, and the second leading cause of death in men in Canada [7]. In 2017, there were 9,400 CRC-related deaths in Canada [6]. Advances in cancer prevention and treatments have contributed to the steady decline in mortality rates in almost all types of cancer over the past decade, including CRC [8]. For instance, based on evidence from observational studies and randomized controlled trials, CRC screening with Faecal Occult Blood Test (FOBT) or endoscopy effectively reduced colorectal cancer incidence and mortality [9], [10].

After a diagnosis of cancer, universally, the primary treatment for about 80% of CRC patients with a non-metastatic diagnosis is surgical resection. Subsequent to having potentially curative surgery and using radiotherapy and/or chemotherapy, follow-up treatment is in the interest of all types of cancer patients to increase the chance of survival and remission. Follow-up programs are important for four main reasons

improved survival, psychological support, quality control, and discovering new cytostatic drug treatments [11]. In fact, a follow-up program helps to detect cancer recurrences at an early stage, which in turn may improve the effectiveness of treatments and improve survival overall. Studies have reported that over 40% of patients experience disease recurrence within the first five years following primary therapy, mainly in the liver patients [6]. In contrast, patients who have customary visits with specialists after cancer surgery report that they experience improved mental prosperity and psychological well-being [11]. In addition, patients who complied with a regular visit after surgery allowed their surgeons to track their improvement quality through recording medical complications, the frequency of recurrence, and survival.

There is a wide consensus on screening programs and their role in managing colorectal cancer [12], [13]. In this regard, various methods are available for colorectal cancer screening including stool-based tests (highly sensitive fecal immunochemical test (FIT), highly sensitive guaiac-based fecal occult blood test (gFOBT), multi-targeted stool DNA test (mt-sDNA)), and visual exams of the colon and rectum(colonoscopy, CT colonography (virtual colonoscopy), flexible sigmoidoscopy (FSIG) [14],[10]. Above all, colonoscopy is considered a gold standard screening tool; it is the primary recommendation for CRC screening according to its predifined surveillance guideline [12],[13]. Yet, colonoscopy is considered as one of the more expensive tests per life-year gained in the recommended cost-efficient screening strategies [15],[16].

In practice, treatment decisions are likely made by both the patient and health care provider. Although valuable, some oncologists prefer not to recommend the follow-up program for patients with CRC at the healthy stages (stage I & II based on AJCC Stage [17]) for cost-effectiveness concerns. While it is widely accepted that surveillance (specifically with colonoscopy/coloscopy) after surgery is beneficial for survival, little is known about the resulting healthcare utilization and its associated costs. There is a

growing body of literature that addresses the importance of the "Choosing Wisely" approach for healthcare decision-makers, including clinicians, patients, hospitals, health care system designers, and providers. This approach aims to mitigate the challenge of having limited resources and high costs that are typically associated with healthcare utilization [18].

2 Literature review

Population-based studies have been conducted that provide evidence to guide the *Choosing Wisely* approach, including the cost-effectiveness of different follow-up strategies after primary cancer curative-intent treatment. One such study evaluated the cost-effectiveness of multimodal ovarian cancer screening with a serum cancer antigen in the United States, and ultimately determined that multimodal screening reduced the mortality rate among women with ovarian cancer [18]. M. Barbier et al. [19] conducted a systematic literature review to investigate the cost-effectiveness of intensive follow-up strategies of patients previously treated for cancer. They suggested that intensive follow-up of patients with colorectal disease is likely to be cost-effective, however the opposite holds true for breast cancer. In a study conducted by Renehan AG et al. [20], the incremental cost-effectiveness ratios is estimated for each life-year gained for various trials designed for early detection of extramural recurrences (targeted surveillance) among patients with colorectal cancer. The investigators found that targeted surveillance is more cost-effective, and that a higher number of life years (0.83) was achieved with intensive follow-up. As a broad generalization, having regular screening and visits with specialists is beneficial to cancer patients. However, there is no consensus about the necessity of follow-up from an effectiveness point of view [21]. There is a gap in the literature regarding the costs associated with healthcare utilization of colorectal cancer patients, specifically of those who are discharged after primary treatment in stages I and II, and those who are not.

The aim of this study is to provide real-world evidence on the compliance with colonoscopy surveillance after primary curative-intent treatment, and to understand the associated health outcomes and healthcare utilization. To do this, patients in the early stages of their colorectal cancer disease trajectory (stage I & II) were identified to investigate the differences in healthcare utilization between patients who follow-up after curative surgery and those who do not.

Patients with stages I and II colorectal cancer were targeted for this study. Those patients at higher stages

were excluded, because they have treatment plans overseen by specialists who have substantial grounds to keep patients, irrespective of cost incurred. Nonetheless, this approach does not necessarily hold for patients at stages I and II. Based on the clinical practice guidelines provided by Alberta Health Services for Colorectal Cancer Surveillance [6], all non-metastatic patients should be considered a potential candidate for additional treatment, but it is not a necessity. Current follow-up treatment protocols for CRC patients at stages I, II and III recommend colonoscopy (a test with high diagnostic accuracy) at one-year post-surgery, and every three to five years after that. Nevertheless, for stages I and II patients, the guidelines are still unclear. Patients in stages I and II are considered relatively healthy, and specialists make the decision to keep them or not based on personal preferences and practice style. Some oncologists prefer to keep the patients in their own practice, while some might discharge patients of similar disease stages to general practitioners (GPs). As a comparison, patients in stage III will be monitored for five years, but for stage II, some patients are discharged immediately following surgery or chemotherapy.

The contribution of the current project is to expand our knowledge from an economic perspective. There are costs associated with healthcare utilization under the follow-up cancer treatment, which are not always considered an essential treatment. This can also have substantial economic implications in order to provide potentially unnecessary healthcare services. The main objective for this study is to determine how compliance with *Choosing Wisely* impacts cost/healthcare utilization. It is also of importance to understand the relationship between compliance with *Choosing Wisely* and survival.

3 Methodology

3.1. Data and study setting

Administrative data from Alberta Health Services and Alberta Health was used for this study, which included 7120 observations. To the best of our knowledge, this was the first study that compares healthcare utilization in terms of cost. The study sample included CRC patients who complied with colonoscopy surveillance, versus those who did not, after curative-intent surgery. The study sample was identified by linking the Alberta Cancer Registry (ACR) records to the Discharge Abstract Database (DAD) and the National Ambulatory Care Reporting System (NACRS). All records of visits were retrieved for CRC surgical and curative procedures following a CRC diagnosis. The DAD is a national database for information on all separations from acute care institutions, including discharges, deaths, sign-outs, and transfers, within a fiscal year (April 1 to March 31). Over time, the DAD has also been used to capture data on day surgery procedures, long-term care, and rehabilitation, among others. The NACRS database contains information for all hospital-based and community-based ambulatory care visits, including day surgery, outpatient and community-based clinics, and emergency departments. The identified data are used to perform a cost comparison of the *Choosing Wisely* approach. After cleaning the data, the total number of patients in the study was 6962, which consists of both male and female patients who were diagnosed with colorectal cancer. The understudy data were on Alberta residents in different cities and hospitals from 2004-2015.

3.1.1. Data variables

Our study sample consisted of individuals who have colorectal cancer at an early stage (stages I & II). Costs were divided into three categories: direct, indirect, and out-of-pocket. However,

given the constraints of the data available, to calculate the costs associated with healthcare utilization under each definition of compliance we focused on physician claims' data as a proxy for a partial or subset of the total healthcare utilization. Two definitions were used that were both considered patient compliance for colonoscopy: 1. A patient who had a colonoscopy within one year after surgery; and 2. A patient who had a colonoscopy within one year after surgery, and had a second colonoscopy two to five years after the first colonoscopy.

3.2. Analysis

Descriptive analyses were carried out for demographic characteristics, which included age, sex, income, and education (see Table 1). Descriptive cancer-related statistics were also carried out, which included tumor stage, CCI, death by cancer, death, and grade (see Table 2). Frequency of each variable was calculated based on Compliance-1y and Compliance-5y variables separately. The primary statistics are provided below.

Table 1. Demographic characteristics of patients with CRC

Cancer Statistic		**No Compliance**		**One- year Compliance**		**Two-to-Five-year Compliance**	
category	level	Patients	Percentage	Patients	Percentage	Patients	Percentage
Sex	Female	1791	59%	1144	37%	115	4%
	Male	2359	60%	1393	36%	160	4%
Average age		69.03	60%	65.34	36%	68.77	4%
Income	≥ 49k	1553	58%	1011	38%	122	4%
	≤ 49k	30	81%	7	19%	0	0%
Education	≥ 80%	2318	58%	1506	38%	158	4%
	≤ 80%	1832	61%	1031	34%	117	5%

i)Total 6962 colorectal cancer patients who were diagnosed from 2004 through 2015 were identified from the Alberta Cancer Registry. ii) Average income of cohort is equal to 49614 CAD. iii) Average educational attainment of the cohort is 80% (Percentage of high school graduates or higher).

Table 2. Main cancer characteristics of patients with CRC

Cancer Statistic		**No Compliance**		**One-year Compliance**		**Two-to-Five-year Compliance**	
category	level	Patients	Percentage	Patients	Percentage	Patients	Percentage
stage	Ⅰ	1729	58%	1118	38%	109	4%
	Ⅱ	2421	60%	1419	35%	166	5%
grade	1	255	58%	167	38%	18	4%
	2	3106	58%	1981	37%	224	5%
	3	215	66%	96	29%	15	5%
	UNK	574	65%	293	33%	18	2%
*CCI	0	1491	53%	1165	41%	147	6%
	1	604	55%	445	41%	47	4%
	+2	2055	67%	927	30%	81	3%
death by cancer		673	79%	166	19%	8	2%
death		1337	78%	351	20%	23	2%

CCI: Charlson Comorbidity Index

3.3. Multivariate analysis

3.3.1. Logistic regression

Logistic regression or logit regression was first introduced by Pearl and Reed (1920) and was further developed and expanded by Reed and Berkson's (1929) (for further details on origin of logistic regression and its history, see Cramer (2002)). Logistic regression is mainly used to investigate the relationship between predictors or explanatory variables, and the specific categorical outcomes. In fact, this statistical model uses a logistic function in its basic form to model a binary dependent variable[22],[23]. The logistic model or logit model can then be used to derive the probability of a specific class or binary event, such as success or failure, presence or absence of disease, alive or dead, and healthy or sick. It should be mentioned that the dependent variable in logistic regression model is categorical, however the predictors can be either continuous or categorical. The broad definition of logistic regression considers two main categories based on the types of dependent variables, whether it is binary, multinomial, or ordinal logistic regression; it also considers the numbers of predictors, which can be simple or multivariable (or "multiple") logistic regression. Hence, if the categorical dependent variable has two levels or two possible outcomes such as "0" and "1", this model is referred to as a binary logistic regression. Subsequently, if the dependent variable has more than two levels like treatment "A", treatment "B", and treatment "C", it is considered a multinomial regression. If the categories of the dependent variable in multinomial regression are ordered, it is called ordinal logistic regression [24]. In the second category of logistic regression, if there is only one predictor, the model is called a simple logistic regression; if there is more than one predictor variable (including categorical and continuous variables) the model is called multivariable or multiple logistic regression [25].

In order to estimate the parameters of a logistic regression or assess the relationship between predictors (n) and the dependent variable, logistic regression uses an odds ratio (OR) [25]. The OR is the ratio of two odds, which is generally expressed in terms of the probability of an event occurring [26], [25]. So, if we consider "p" as the probability of an event, then the odds can be defined as:

$$Odds = \frac{p}{(1-p)}$$

We can also derive probabilities from odds by using the following equation:

$$p = Odds/(1 + Odds)$$

This transformation is also referred to as the logit transformation of the probability "p." This ratio $p/(1-p)$ in the transformation is called the odds (further details is provided later in this chapter). Based on the above formula, the odds can vary from "0" to "∞" as probability goes from "0" to "1". There are a number of different terms used to describe ORs, depending on the number of predictors. If there is only one predictor, the OR is called an unadjusted or crude OR. Otherwise, when there is more than one predictor, the OR is called the adjusted OR, which can quantify the effect of a predictor, while the effect of the rest of the predictors holds constant.

To obtain the odds, logistic regression first estimates the parameters of the corresponding logistic model through the maximum-likelihood method (further details on this topic are provided in Douglas et. Al. (2010)) [23]. To achieve this, the logit or log-odds needs to be defined, which is the logarithm of odds.

$$Logit\ p = ln\frac{p}{1-p} \qquad for \quad 0 < p < 1$$

The measurement unit for the scale of log-odds is called a logit (from the logistic unit), and the subsequent probability of the value labeled "1" can range from 0 and 1. If the value of the event

or the dependent variable is labeled “1”, we can assume that the log-odds or logarithm of odds is a linear combination of predictors or independent variables.

The standard logistic function is expressed by “$\sigma: \mathrm{R} \rightarrow (0,1)$” and is given by the following formula [25], [27]:

$$\sigma(t) = \frac{e^{\sigma}}{e^{\sigma} + 1} = \frac{1}{1 + e^{-1}}$$

Logistic function can take any real input “t” ($t \in R$) and output with a value between “0” and “1”. So, if we assume “t” is a linear combination of multiple explanatory variables like "x”, we can define “t” as:

$$t = \beta_0 + \beta_1 x$$

Subsequent logistic function can be defined as:

$$p(x) = \sigma(t) = \frac{1}{1+e^{-(\beta_0+\beta_1 x)}}$$

Where $p(x)$ is the probability of a binary dependent variable like "Y,” which in turn has two levels of value like dead or alive [23]:

$$\begin{cases} Pr\ (Y = 1|x) = P(x) \\ Pr\ (Y = 0|x) = 1 - P(x) \end{cases}$$

Moreover, the defined logit function can be written as the inverse of the standard logistic function [27], [26]:

$$g(p(x)) = \sigma^{-1}\big(p(x)\big) = logit\ p(x) = ln\left(\frac{p(x)}{1 - p(x)}\right) = \beta_0 + \beta_1 x$$

In the above equation, the variables are defined as:

- g: the logit function, and equally $g(p(x))$ is the transformer to the linear regression;
- $p(x)$: the probability or a quantitative expression of the chance, given any linear combination of predictors, that the dependent variable is equal to an event. In fact, $p(x)$

acts as a transformation tool that can express the value of the linear regression based on a probability range from 0 to 1;

- β_0: the intercept or the constant term of the linear regression equation;
- β_1: is the regression coefficient that captures the changes of dependent values by one-unit change in the corresponding explanatory variable.

Afterwards, by taking the exponentiate of both sides of the logit function, we derive the odds equation [27], [25]:

$$Odds = \frac{p(x)}{1 - p(x)} = e^{\beta_0 + \beta_1 x}$$

As mentioned above, the OR is the ratio of the two odds for the continuous independent variables, which can be written as follows:

$$OR = \frac{Odds(x+1)}{Odds(x)} = \frac{\left(\frac{F(x+1)}{1-F(x+1)}\right)}{\left(\frac{F(x)}{1-F(x)}\right)} = \frac{e^{\beta_0 + \beta_1 (x+1)}}{e^{\beta_0 + \beta_1 x}} = e^{\beta_1}$$

In the formula above, "β_1" provides an interpretation for the exponential relationship between predictors and odds. For each one-unit increase in the corresponding predictor "x", the odds will change to "e^{β_1}".

Also, ORs for categorical or binary independent variables can be defined using a contingency table. A contingency table, also known as a cross tabulation or crosstab, is a type of table where a matrix represents the frequency or multivariate distribution of each variable. For example, if we consider the table below, the overall probability of patients with CRC diagnosis is equal to $(a + b)/n = 12/35 = 0.342$ or 0.34.

	Female = 1	**Female = 0**	**Total**
CRC = 1	a = 5	b = 7	(a+b) = 12
CRC = 0	c =10	d =13	(c+d) = 23
Total	(a+c) = 15	(b+d) = 20	n = 35

Subsequently, the odds of a patient with a CRC diagnosis is equal to $p/(1-p)$=0.34/0.66=0.51.

The odds of a CRC diagnosis in female patients is then equal to:

$$Odds(CRC|Female) = \frac{P(CRC|Female)}{1-P(CRC|Female)} = \frac{a/(a+c)}{c/(a+c)} = \frac{5/15}{10/15} = \frac{1}{2} = 0.5$$

The odds of a CRC diagnosis in male (female=0) patients is equal to:

$$Odds(CRC|Male) = \frac{P(CRC|Male)}{1-P(CRC|Male)} = \frac{b/(b+d)}{d/(b+d)} = \frac{7/20}{13/20} = \frac{7}{13} = 0.538$$

Now, the OR of having a CRC diagnosis is equal to:

$$OR = \frac{Odds(CRC|Male)}{Odds(CRC|Female)} = \frac{0.538}{0.5} = 1.076$$

An OR with a value of 1.076 indicates that the odds of having a CRC diagnosis is 7.6% more likely for males than females.

As all of the essential equations have been defined, the multivariate logistic regression will now be explained. Multivariate logistic regression is the same as the standard logistic regression, but considers more than one predictor. Mathematically, the logit of the multiple logistic regression model is written as [27]:

$$g(X) = ln\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$$

In the above equation, "X" is a collection of "n" independent variables denoted by the vector $\acute{X} = (x_1, x_2, \ldots, x_n)$, considering the conditional probability that the outcome is present given by

$Pr(Y = 1|X) = p(X)$. Moreover, through the logistic function, we can convert the log odds to a probability. Hence, by exponentiating both sides of above equation and rearranging it, we have [27], [23]:

$$p(X) = \frac{e^{g(X)}}{1 + e^{g(X)}}$$

If any of the independent variables are nominal or discrete such as sex, cancer stage, and so forth, we cannot treat them like interval scale variables. In this case, they would be assigned a dummy variable or design variable. For instance, consider the variable "sex" as a one of the independent variables, which can be coded as "female" and "male". In this situation, we can assign "1" to female and "0" to male. As a general rule, in the case of an existing nominal scaled variable with "k" possible values, we can design a "$k - 1$" variable (including an intercept) with a related assigned numerical value. By taking nominal scaled variables into consideration and assuming that there are "x_j" nominal independent variables with "k_j" levels, the logit of multivariate logistic regression is given by the following equation [27]:

$$g(X) = \beta_0 + \beta_1 x_1 + \sum_{l=1}^{k_j - 1} \beta_{jl} D_{jl} + \cdots + \beta_n x_n$$

Where;

- D_{jl}: represents the "$k_j - 1$" design variables for "k_j" levels of "x_j," and ($l = 1,2, \dots, k_j - 1$);
- β_{jl}: the coefficients for the design variables.

So as to identify factors independently associated with colonoscopy compliance programs after receiving a CRC diagnosis for patients who are at stages I and II, we performed multivariable logistic regression on the observed characteristics of the understudy patients. Our dependent categorical variable (having compliance) has two levels: if a patient complies with colonoscopy

within the specified period (one year after surgery or within two-to-five-years after surgery), this variable takes on the value of "1," or "0" if they did not comply (for details on the variables used for the study, refer to Table 3).

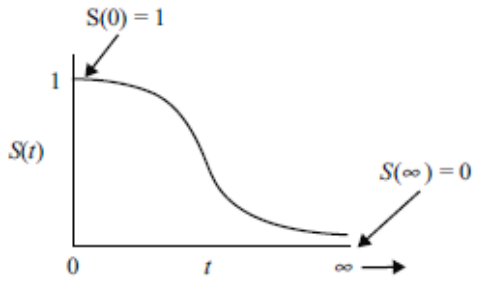
3.4. Survival analysis

Studying how patients respond over time to specific follow-up treatments is important in identifying how therapies impact the patient's quality of life and their disease progression. Time-to-event studies, such as survival analysis, utilize the Kaplan-Meier model (a univariate model) and the Cox proportional hazards model (a multivariate model). These are closely related statistical approaches [28]. There are some specific variables in survival analysis or time to event analysis that need to be defined. Survival function is denoted by S(t), and gives the probability of survival at a given time:

$$S(t) = Pr(T > t) = 1 - F(t) = \int_t^{\infty} f(u)\, du$$

In the above equation, "T" denotes the random variable for time of death, "t" is time, "Pr" is the probability, and F(T) is the probability distribution function of a random variable "T". The survival function is the probability that the time of death is later than some specified time t.

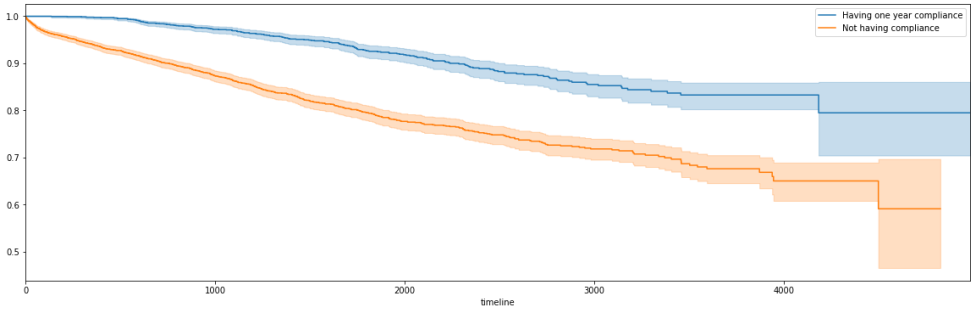
Figure 1. the lifetime distribution function and event density (F and f below).



As depicted in the figure above, the survival function is a smooth curve and does not increase with "t". In the survival function, "t" can take margin values from 0 to infinity. When t = 0, then S(0)=1(i.e., S(t) →1 as t → 0), meaning that the event has not happened yet. Conversely, when t=∞ then S(∞) = 0 (i.e., S(t) → 0 as t → ∞), indicating that no one will survive and the survival curve will fall to 0. So, falling between these two extreme values of S(t) (i.e., S(0)>S(t)>S(∞)) indicates that there is the possibility of immediate death or failure for any value less than 1 (S(t) < 1). For any value greater than 0 (S(t)>0), there is the possibility of survival.

The survival function is estimated with the Kaplan-Meier estimator. The Kaplan-Meier (K-M) curve is a visual representation of survival function, which shows the probability of an event at a specified time interval. The event can be survival, or the occurrence of disease, death, sensor failure, etc. In this study, the event of interest is death, and the time is measured in days lapsed between the date of the CRC diagnosis and the date of the follow-up. Figure 2 shows the comparison between two groups of patients and their death risk exposure. The two groups compared were those who had a one-year colonoscopy compliance after curative surgery (Group 1), and those who did not have compliance (Group 2). As a result, because the survival curve of Group 2 decreases faster than the curve for Group 1, we can infer that Group 2 has a higher risk of mortality compared to Group 1.

Figure 2. Kaplan-Meier curve for two groups of patients based on whether having one-year compliance with colonoscopy after resection surgery



One of the limitations of the Kaplan-Miere estimator is that it is a univariate model, and therefore cannot estimate survival while considering multiple covariates. To overcome this limitation, the current study employed an alternative method called the Cox proportional hazards regression. This method is able to consider the influence of covariates, whether they are categorical, binary, or continuous variables, on hazards or survival rates for survival data [29]. Such covariates may include demographic factors, the severity of diseases and treatments received.

3.4.1. Cox regression

Cox proportional hazards regression or Cox regression was developed by Sir David Cox [30]. It is a technique used for exploring the impact that determined factors have upon the time a predetermined event takes place.

Generally, Cox regression is used for survival analysis in the context of an outcome such as death This regression model simultaneously assesses the effects that several risk factors have on survival time through a hazard function, which measures the association between the survival time of patients and predictor variables. Survival analysis regression therefore measures the relationship between the event occurrence and a set of covariates through the hazard function [31]. The method assumes that the effects that the predictor variables have upon survival are constant over time and are additive on one scale. This semi-parametric method has been used to deal with time-to-event data in the presence of censoring, and assumes the ratio of hazard functions for any two individuals is determined by covariate values, which is constant over time [31]. A main feature that differentiates survival from other research areas is that survival data is typically censored. In survival analysis censoring occurs when information is not available for all research participants on time to outcome events, such as death, a disease recurrence or cancer metastasis [32]. A patient

is said to be censored when there is no information on time to event as a result of loss to follow-up or non-occurrence of the outcome event before the study ends. Conversely, a subject is called uncensored when the event occurs within the time period of observation. Censoring is usually classified as either point censoring or interval-censoring [32].

Point censoring refers to a situation when, regardless of constant monitoring of an outcome event, the patient is lost to follow-up, or the event does not occur within the study duration. Point censoring also falls under two categories: right-censoring and left-censoring. Generally, there are three reasons that can give rise to right-censoring: 1) if a patient does not undergo the event before the end of study; 2) if a patient fails or is lost to follow-up during the study period; and 3) if a patient dies (in the case of not considering death as the event of interest).

Some notations have been used commonly for defining censoring. In right-censoring, we assume that there is a lifetime "X" for a specific individual understudy, and a fixed censoring time, "C_r" (right-censoring). The lifetime of individuals (X's) is assumed to be independent and distributed identically with a probability density function $f(X)$, and a survival function S(X). When "X" is less than or equal to "C_r", this means that there is information on the exact lifetime "X" of an individual or patient. If "X" is greater than "C_r", the event time of the patient is censored at "C_r". In other words, we can express it differently by using these random variables "T" and "δ," where δ =1 indicates that the lifetime X corresponds to an event, and δ=0 indicates that it is censored. If the lifetime observed "T" is equal to "X" and δ=1, otherwise, "X" is censored and "T" is equal to "C_r", i.e. $T = \min(X, C_r)$ [27][33][34].

In left-censoring, if the event of interest occurs for an individual or a patient prior to study entry at a specific time, then we can say the patient is left-censored. Similar to right-censoring, we assume that there is a lifetime "X" for a specific understudy individual, and a fixed censoring time,

"C_L" (left-censoring). The lifetime of individuals (X's) is assumed to be independent and distributed identically with a probability density function $f(X)$, and a survival function S(X). So, if the "X" of a patient is less than the censoring time "C_L", as contrast to the right-censoring, the patient is left-censored. For these patients, we know that they have experienced the event sometime ahead of time "C_L", but there is no information on their exact event time or it is unknown. So, if the "X" is greater than or equal to "C_L", the exact lifetime "X" will be known. Similar to right-censoring, we can express it differently by using these random variables "T" and "ε," where ε =1 indicates that the lifetime X is observed, and ε =0 indicates that it is censored. Hence, if the lifetime is observed, "T" is equal to "X" and $\varepsilon = 1$, otherwise, "X" is censored and "T" is equal to "C_L", i.e. $T = \max(X, C_r)$ [34].

Interval-censoring also deals with an incomplete data structure. In this case, a random variable of interest is known only to lie within an interval (between a left endpoint and right endpoint in the study) rather than being observed precisely. To put it differently, when patients' event time is only known to fall within an interval in a case of having a periodic follow-up in a clinical trial or longitudinal study, interval-censoring may occur [34]. In this case, certain observations can be missed because of pre-scheduled findings for a clinically measurable change in disease or health condition and a return with changed health status. Therefore, it is only known that the actual event time lies somewhere between the last observation when the change did not occur, and the first observation time when the change was detected. This results in an interval that includes the actual time of the change incidence, which is not observed [35].

However, left-censoring and interval-censoring are beyond the scope of this study (see [35], [34] for an overview). This project focused solely on right-censoring. In this study, the event of interest is death, and the time is measured in days lapsed from the date of the CRC diagnosis to the date of

first follow-up. In addition, patients with CRC at stages I and II who were still alive during the follow-up period were considered as the right-censored variables.

Mathematically, the COX regression model is written as:

$$h(t|x) = h_0(t)e^{\beta_1 x_1 + \cdots + \beta_n x_n}$$

The variables are explained below:

- $h(t|x)$= the hazard at time t for a subject with a set of predictors $x_1, \ldots, x_n$;
- $h_0(t)$= the baseline hazard function; and
- $\beta_1, \ldots, \beta_n$= the model parameters describing the effect of the predictors on the overall hazard.

Hazard function is one of the indices used for comparison purposes in Cox regression, which conveys the relative risk in terms of the HR. So, hazards are considered proportional under the Cox model assumption. According to Heman, the Hazard ratio (HR) is "defined as the hazard in the exposed (treated) groups divided by the hazard in the unexposed (control) groups." For practical purposes, hazards can be thought of as incidence rates, and thus the HR can be roughly interpreted as the incidence ratio [36]. The hazard is "the instantaneous event probability at a given time, or the probability that a patient or an individual under observation experiences the event in a period centered around that point in time" [31]. The Cox model's HR is defined as the ratio of the predicted hazard function upon two different predictor variable values. An HR with a value greater than 1 means that an event is more likely to occur, and a ratio less than 1 means that an event is less likely to occur. A HR of 1 means that the indicator has no effect on the hazard of the occurrence [37].

In summary:

- HR = 1: No effect
- HR < 1: Reduction in the hazard
- HR > 1: Increase in hazard

In most contexts, Hazard function, h(t) is expressed as:

$$h(t) = lim_{\Delta t \rightarrow} \frac{Pr\ (t < T \leq t + \Delta t | T > t)}{\Delta t}$$

or

$$h(t) = \frac{f(t)}{(1 - F(t))} = \frac{f(t)}{S(t)}$$

In health literature, an HR is treated as a probability of death. An HR is the probability that a patient dies somewhere between t and $t + \Delta t$, divided by the probability that the patient survived beyond time t.

For instance, an HR of 2 means that a group has twice the chance of dying than a comparison group [38].

Another advantage of using Cox proportional hazard regression is that it allows us to impose specific parametric assumptions for hazard functions when one is unsure of the functional form of hazard function. As an example, we can estimate the survival proportion function in an exponential functional form as follows:

$$S(t) = S_0(t)^{exp\ (\gamma)}$$

where $S_0(t)$ is the baseline survival, representing the survival proportion when all covariates are equal to zero, and γ is equal to $\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$. After deriving the value of the baseline survival at a particular time, we can obtain the predicted survival probabilities for patients with any specified covariate of x_n values [31]. For details on employed variables in each regression

model, please refer to Table 3.

Table 3. Outcomes, the corresponding explanatory variables and the regression models.

Variable	Definition	Regression
Age67	Dummy variable = 1 if patient had 67-year old or higher, 0 otherwise.	Logistic & COX
Sex	Dummy variable = 1 if patient was male, 0 female	Logistic & COX
Stage	Categorical variable including stage I and stage II. It refers to the cancer stage of patients with CRC.	Logistic & COX
Grade	Categorical variable ranging from 1, 2, 3, and unknown grade. It refers to the cancer grade of patients with CRC.	Logistic & COX
CCI	Categorical variable ranging from 0, 1, and +2. The unit was Charlson Comorbidity Score.	Logistic & COX
income49	It refers to neighborhood income. Dummy variable = 1 if patient had on average 49k $ yearly income, 0 otherwise.	Logistic
Education 80%	It refers to neighborhood education (% with high school or above). Dummy variable = 1 if patient holds 80% education or higher, 0 otherwise.	Logistic
compliance_1y	Dummy variable = 1 if patient had colonoscopy within 1 year after surgery, 0 otherwise	Logistic & COX
compliance_5y	Dummy variable = 1 if patient had first colonoscopy within 1 year after surgery, and had the second colonoscopy 3 years after the first colonoscopy (i.e. 4 years from surgery), 0 otherwise	Logistic & COX
Death	censoring indicator = if patient is alive(censored); 1otherwise (died =event)	COX
Cancer death	Dummy variable = 1 if patient was dead because of cancer, 0 otherwise	COX
Lenfol	Numerical variable, it is length of follow-up from diagnosis date of CRC, the unit is day.	COX

https://doi.org/

In the next step, a cost comparison was done comparing patients based on the two compliance definitions. Analyses are performed with SAS statistical software version “9.4” SAS Institute, Inc., Cary, NC) and RStudio version “3.5.3”.

4 Results

4.1. Patient characteristics

In total, 6,962 patients were eligible to participate in the study. The median age was 67 (range: 18-104). Of 6,962 patients, 42.46% had stage I CRC, and 57.54% had stage II. A total of 2,812 (40.39%) patients had a one-year compliance, and 275 patients (3.95%) had a five-year compliance. There were 2,803 (40.26%) patients who had a Charlson Comorbidity score of 0, 1,096 (15.74%) patients who had a score of 1, and 3,063 (44%) who had a score of 2 or more (Tables 1 & 2).

4.2. Multivariate analyses of the factors associated with compliance for one year and five-years after surgery.

Patient compliance one year after surgery was found to be significantly associated with extremes of age ($P<0.0001$), lower tumor grade ($P<0.0041$), higher Charlson Comorbidity score ($P<0.0001$), costs ($P<0.03$), and hospitalization within the first year after surgery ($P<0.007$). The odds of one-year compliance in patients who were 67 years of age or above was 0.7 (95% CI: 0.633-0.775), which indicates that patients of a lower age (age<67) were more likely to comply. Patients with the following features had a higher odds of having one-year compliance after surgery (Table 4): Females, with an OR of 1.089 (95% CI: 0.986-1.201, p-value ($p=0.0913$)) compared to males; patients at stage II versus stage I, (1/0.958=1.044) OR (95% CI: 0.865-1.06); patients at grade 1 versus grade 2 or more with (1/0.774=1.34) OR (95% CI: 0.571-1.05); patients with lower CCI versus higher CCI with 1.65 (1/0.607) OR (95% CI: 0.543-0.679)**;** high income patients versus low income with 1.038 OR (95% CI: 0.929- 1.160); and possessing higher education with 1.069 OR (95% CI: 0.958-1.193). It should be noted that based on the cost OR (1), all levels of cost have the same effect on compliance.

Table 4. Multivariable logistic regression odds ratios (95% CI) of factors associated with one-year compliance after surgery, 2004-2015, Alberta.

Parameter	Odds Ratio (95% CI)
Age (years)	
≥ 67	0.7 (0.63-0.77)***
< 67	1 (reference)
Sex	
Male	1 (reference)
Female	1.09 (0.99-1.2)*
Tumor grade	
Grade 1	1 (reference)
Grade 2	0.996 (0.81-1.22)***
Grade 3	0.77 (0.57-1.05)
Unknown	0.74 (0.58-0.94)**
Charlson comorbidity score	
0	1 (reference)
1	0.99 (0.86-1.14)***
≥ 2	0.61 (0.54-0.68)***
Educational attainment of neighborhood (% high school or greater)	
< 80%	1 (reference)
≥ 80%	1.07 (0.96-1.19)
Median neighborhood income	
< $49,000	1 (reference)
≥ $49,000	1.04 (0.93-1.16)
AJCC Stage	
Ⅱ	1 (reference)
Ⅰ	0.96 (0.86-1.06)

CI: confidence interval; AJCC: American Joint Committee on Cancer

*** indicates that coefficient is significant at 1% level;

** indicates that coefficient is significant at 5% level; and

* indicates that coefficient is significant at 10% level.

Having compliance within two-to-five-years after surgery is significantly associated with stage of

cancer (P<0.0304), tumor grade (P<0.0047), a higher Charlson Comorbidity score (P<0.0001), costs (P<0.0003), and hospitalization (P<0.001). The odds of two-to-five-year compliance in patients who are 67 years or older is 0.811 (95% CI: 0.631-1.043) which indicates that patients younger than 67 are more likely to comply. The results of our regression suggest that patients with the following features have a higher probability of having two-to-five-year compliance after surgery (Table 5): Males with 1.12 (1/0.886) OR (95% CI: 0.962-1.135) versus females; Patients at stage II versus patients at stage I with 1.29 (1/0.755) OR (95% CI: 0.586-0.974); Patients at grade "2" or higher versus grade "1" or more with 1.015 and 1.159 OR respectively; Patients with lower CCI versus higher CCI (e.g. having CCI "0" versus CCI "1") with 1.16 (1/0.855) OR (95% CI: 0.608-1.202); High income patients versus low income ones with 1.288 OR (95% CI: 0.976-1.1699); and having lower levels of education versus higher with 1.21(1/0.822) OR (95% CI: 0.622-1.086). It should be noted that based on the cost OR "1," all levels of cost have the same effect of having compliance.

Table 5. Multivariable logistic regression odds ratios (95% CI) of factors associated with two-to-five-year compliance after surgery, 2004-2015, Alberta.

Parameter	**Odds Ratio (95% CI)**
Age (years)	
≥ 67	1 (reference)
< 67	0.811 (0.631-1.043)
Sex	
Male	1 (reference)
Female	0.886 (0.692-1.135)
Tumor grade	
Grade 1	1 (reference)
Grade 2	1.015 (0.618-1.668)
Grade 3	1.159 (0.568-2.366)
Unknown	0.493

	(0.253-0.964)***
Charlson comorbidity score	
0	1 (reference)
1	0.588 (0.608-1.202)
≥ 2	0.521 (0.391-0.695)***
Educational attainment of neighborhood (% high school or greater)	
< 80%	1 (reference)
≥ 80%	0.822 (0.622-1.086)
Median neighborhood income	
< $49,000	1 (reference)
≥ $49,000	1.288 (0.976-1.699)*
AJCC Stage	
Ⅱ	1 (reference)
Ⅰ	0.755 (0.586-974)**

CI: confidence interval; AJCC: American Joint Committee on Cancer.

*** indicates that coefficient is significant at 1% level;

** indicates that coefficient is significant at 5% level; and

* indicates that coefficient is significant at 10% level.

4.3. Multivariate survival analyses for CRC patients

The results of the COX regression analysis suggest that although females had an approximate 7% decrease in the HR (0.923) compared to males, this decrease was not significant. However, with each year of age, the HR increased by 68% (HR of patients younger than 67 years of age = 0.317), which was a significant change. For grades “2”, “3,” and “unknown,” the HRs were 1.35, 1.771, and 2.22, respectively, and all with p-values of less than 0.01. This indicates a strong relationship between a higher tumor stage and an increased risk of death. Similarly, the significant HRs of CCI (p-value <0.01) suggests that there is a positive relationship between a higher CCI and increased risk of death. Those patients with stage II cancer were also found to have an increased HR by a

factor of 1.18, or 18%. More importantly, given the HRs of complying a year after surgery and/or two-to-five-years after surgery, complying with colonoscopy will reduce the risk of death by 61% and 56% respectively. See Table 6 for further details.

Table 6. Multivariate COX regression Hazard ratios with 95% CI of CRC patients, 2004-2015, Alberta.

Parameter	**Hazard Ratio (95% CI)**
Age (years)	
≥ 67	1 (reference)
< 67	0.317 (<.0001)***
Male	1 (reference)
Female	0.923 (0.1058)
Tumor grade	
Grade 1	1 (reference)
Grade 2	1.350 (0.0052)***
Grade 3	1.771 (<.0001)***
Unknown	2.220 (<.0001)***
Charlson comorbidity score	
0	1 (reference)
1	1.255 (0.0100)***
≥ 2	2.082 (<.0001)***
Cancer Stage	
Ⅱ	1.186 (0.0016)***
Ⅰ	1 (reference)
compliance_1y	0.389 (<.0001)***
compliance_5y	0.444 (<.0001)***

CI: confidence interval. Numbers in parenthesis indicate P-value.
*** indicates that coefficient is significant at 1% level;
** indicates that coefficient is significant at 5% level; and
* indicates that coefficient is significant at 10% level.

5 Main findings

In summary, the results of our study suggest that having a compliance program increases the odds of a patient going into remission and improves survival. In addition, the death ratio (both cancer-related death and overall death) among patients who were considered compliant with colonoscopy in each category (one-year and two-to-five-year follow-up) was lower compared to those who were not considered compliant. Finally, patients who were compliant with colonoscopy surveillance after primary treatment were found to have lower healthcare utilization (cost) and a better health outcome (higher survival rate).

6 Discussion

This retrospective cohort study used administrative data sets derived from Alberta Health Services and Alberta Health, which included 7,120 observations from 2004-2015 in Alberta, Canada. The post-operative colonoscopy dates were extracted for CRC patients by looking at the length of follow-up time after being diagnosed with CRC and undergoing colorectal resection. As mentioned earlier, this study only captured data from CRC patients who were at stages I or II. The cohort dataset encompasses patient factors such as age, sex, education level, neighborhood income level, CCI, and also after-surgery data, including hospitalizations, emergency department visits, and general physician visits.

We observed that the cost (measured by taking the average of physician claims over the specified period) was lower for those who comply, versus those who do not comply. One possible interpretation is that compliance to a cancer patients' follow-up care could lead to better disease management and prevention, therefore reducing complications and relapses and ultimately lowering costs. In order to explore the factors related to having compliance, we use multivariate logistic regression models. A population-based study in China used a similar estimation method, and found that patients who 1) had a positive FOBT result, 2) had a family history of CRC, 3) had a history of intestinal polyps, 4) had higher education, and 5) were aged 60-69 were more likely to undergo colonoscopy [39]. The results of our analyses suggest that patients who were at stage II, who were in the younger age group, and who had a lower CCI had a higher odds of having one-year and also two-to-five-year compliance after surgery. We further observed that there was a positive link between high income and having compliance. Based on the derived p-value and HRs, patients with lower HRs were more likely to live longer, but we do not know how much longer. As we expected, we observed that younger patients who

had a lower CCI, lower cancer grade, and were at stage I had higher chances of survival compared to their baseline category. However, the representativeness of the data is unclear as there is a lack of information regarding the costs of each type of healthcare utilization and total cost variable, which would reflect the actual overall treatment cost. We estimated costs associated with healthcare utilization using only physician claims, which is a subset of total healthcare utilization. This limitation can result in potential biases of the presented cost comparison. In fact, the total cost variable information on healthcare utilization would be a more precise and better reflection of the tradeoff between complying with the recommended follow-up program and its associated healthcare utilization, which in turn may change the direction or the sign of the association. Because if we want patients to comply more and better with the recommended follow-up program, the direct cost or the actual healthcare cost would be higher. For the reason that being more adherent to the recommended follow-up treatment program means more treatments would be needed or recommended that leads to higher healthcare utilization and then higher costs. It would be of value to mention that the argument always is that the tradeoff will lead to saving more down the road by preventing some patients from progressing into more severe cases, which would cost the health system more money.

7 Conclusion and policy implication

The empirical evidence of this study shows that there are potential cost-saving and positive externalities from encouraging people to comply with the follow-up program recommendation if the clinical practice being reinforced. Also, the success of screening depends on high participation rates in the recommended groups. However, it is worthwhile noticing that in reality and based on our data the compliance rate is so low. So in order to increase screening compliance, more interventions are needed to boost both patient eagerness to get screened as well as organizational structures within the health system to expedite identifying patients who require initial screening and to provide effective information on screening tests and positive result of follow-up. In this regard, one of the most powerful incentives for patients to get screened is physician recommendation. Hence, more investments would be of value to put in increasing the physician incentives to guide and recommend more patients towards getting screened within follow-up program. Then, as a future study, we can investigate the reason why the compliance rate is low, what are the barriers from both patients' and physicians' sides, which need to be enhanced.

Bibliography

[1] American Cancer Society, “Global Cancer Facts & Figures 4th Edition,” Atlanta, 2018. doi: 10.1787/health_glance_eur-2018-graph47-en.

[2] World Health Organization, “Cancer, key facts,” 2018. https://www.who.int/health-topics/cancer#tab=tab_1.

[3] P. Favoriti, G. Carbone, M. Greco, F. Pirozzi, R. E. M. Pirozzi, and F. Corcione, “Worldwide burden of colorectal cancer: a review,” *Updates Surg.*, vol. 68, no. 1, pp. 7–11, 2016, doi: 10.1007/s13304-016-0359-y.

[4] P. Rawla, T. Sunkara, and A. Barsouk, “Epidemiology of colorectal cancer: Incidence, mortality, survival, and risk factors,” *Prz. Gastroenterol.*, vol. 14, no. 2, pp. 89–103, 2019, doi: 10.5114/pg.2018.81072.

[5] L. A. Gaudette, E. M. Illing, and G. B. Hill, “Canadian Cancer Statistics 1991.,” *Health Rep.*, vol. 3, no. 2, pp. 107–135, 1991.

[6] Alberta Health Services, “Stages Ii and Iii Colorectal Cancer Surveillance,” pp. 1–5, 2014.

[7] Canadian Cancer Society, “Colorectal Cancer Statistics,” 2019. https://www.cancer.ca/en/cancer-information/cancer-type/colorectal/statistics/?region=on.

[8] American Cancer Society, “Facts and figures of Cancer,” 2019. https://www.cancer.org/latest-news/facts-and-figures-2019.html.

[9] H. Brenner, C. Stock, and M. Hoffmeister, “Effect of screening sigmoidoscopy and screening colonoscopy on colorectal cancer incidence and mortality: Systematic review and meta-analysis of randomised controlled trials and observational studies,” *BMJ*, vol. 348, no. April, pp. 1–12, 2014, doi: 10.1136/bmj.g2467.

[10] P. Hewitson, P. Glasziou, E. Watson, B. Towler, and L. Irwig, “Cochrane systematic review of colorectal cancer screening using the fecal occult blood test (Hemoccult): An update,” *Am. J. Gastroenterol.*, vol. 103, no. 6, pp. 1541–1549, 2008, doi: 10.1111/j.1572-0241.2008.01875.x.

[11] K. Bibbins-Domingo *et al.*, “Screening for colorectal cancer: US preventive services task force recommendation statement,” *JAMA - J. Am. Med. Assoc.*, vol. 315, no. 23, pp. 2564–2575, 2016, doi: 10.1001/jama.2016.5989.

[12] R. Adult, I. Schedule, and U. States, “Annals of Internal Medicine Clinical Guidelines C HANGES IN THE S CHEDULE,” *Ann. Intern. Med.*, vol. 147, no. 10, pp. 1–5, 2008.

[13] P. Armaroli *et al.*, “European Code against Cancer, 4th Edition: Cancer screening,” *Cancer Epidemiol.*, vol. 39, pp. S139–S152, 2015, doi: 10.1016/j.canep.2015.10.021.

[14] H. Brenner, M. Kloor, and C. P. Pox, “Colorectal cancer,” *Lancet*, vol. 383, no. 9927, pp. 1490–1502, 2014, doi: 10.1016/S0140-6736(13)61649-9.

[15] A. G. Zauber, “Cost-effectiveness of Colonoscopy,” *Gastrointest. Endosc. Clin. N. Am.*, vol. 20, no. 4, pp. 751–770, 2010, doi: 10.1016/j.giec.2010.07.008.

[16] K. K. F. TSOI, S. S. M. NG, M. C. M. LEUNG, and J. J. Y. SUNG, “Cost-effectiveness analysis on screening for colorectal neoplasm and management of colorectal cancer in Asia,” *Aliment. Pharmacol. Ther.*, vol. 28, no. 3, pp. 353–363, 2008, doi: 10.1111/j.1365-2036.2008.03726.x.

[17] “Cancer Staging System,” *American Joint Committee on Cancer*. https://www.cancerstaging.org/references-tools/Pages/What-is-Cancer-Staging.aspx.

[18] A. B. Gillian D. Sanders, Matthew L. Maciejewski, “Overview ofCost-effectiveness Analysis,” *JAMA - J. Am. Med. Assoc.*, vol. 321, no. 24, pp. 1400–1401, 2019, doi: 10.1001/jama.2018.6072.

[19] M. Barbieri, G. Richardson, and S. Paisley, “The cost-effectiveness of follow-up strategies after cancer treatment: A systematic literature review,” *Br. Med. Bull.*, vol. 126, no. 1, pp. 85–100, 2018, doi: 10.1093/bmb/ldy011.

[20] A. G. Renehan, S. T. O’Dwyer, and D. K. Whynes, “Cost effectiveness analysis of intensive versus conventional follow up after curative resection for colorectal cancer,” *Br. Med. J.*, vol. 328, no. 7431, pp. 81–84, 2004, doi: 10.1136/bmj.328.7431.81.

[21] B. Ohlsson, U. Breland, H. Ekberg, H. Graffner, and K. G. Tranberg, “Follow-up after curative surgery for colorectal carcinoma - Randomized comparison with no follow-up,” *Dis. Colon Rectum*, vol. 38, no. 6, pp. 619–626, 1995, doi: 10.1007/BF02054122.

[22] J. Tolles and W. J. Meurer, “Logistic Regression: Relating Patient Characteristics to Outcomes,” *Jama*, vol. 316, no. 5, pp. 533–534, 2016, doi: 10.1001/jama.2016.7653.

[23] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis*. Hoboken, New Jersey: John Wiley & Sons, Inc., 2012.

[24] S. H. Walker and D. B. Duncan, “Estimation of the probability of an event as a function of several independent variables.,” *Biometrika*, vol. 54, no. 1, pp. 167–179, 1967, doi: 10.1093/biomet/54.1-2.167.

[25] T. G. Nick M and K. M. Campbell, “Logistic Regression,” in *Topics In Biostatistics*, vol. 404, no. 4, W. T. Ambrosius, Ed. Humana Press Inc., p. 113.

[26] C. E. M. Eric Vittinghoff, Stephen C. Shiboski, David V. Glidden, *Statistics for Biology and Health: Regression Methods in Biostatistics*, vol. 53, no. 9. San Francisco: Springer, 2005.

[27] D. W. HOSMER, S. LEMESHOW, and R. X. STURDIVANT, *Applied Logistic Regression Analysis.*, Third., vol. 45, no. 4. Hoboken, New Jersey: John Wiley & Sons, Inc., 2013.

[28] N. W. Dudley, R. Wickham, and N. Coombs, "An Introduction to Survival Statistics: Kaplan-Meier Analysis," *J. Adv. Pract. Oncol.*, vol. 7, no. 1, pp. 91–100, 2016, doi: 10.6004/jadpro.2016.7.1.8.

[29] D. Machin, C. Yin Bun, and P. Mahesh, *Survival Analysis: A Practical Approach*. 2006.

[30] D. R. Cox, "Regression Models and Life-Tables," *J. R. Stat. Soc. Ser. B*, vol. 34, no. 2, pp. 187–202, 1972, doi: 10.1111/j.2517-6161.1972.tb00899.x.

[31] M. J. Bradburn, T. G. Clark, S. B. Love, and D. G. Altman, "Survival Analysis Part II: Multivariate data analysis- An introduction to concepts and methods," *Br. J. Cancer*, vol. 89, no. 3, pp. 431–436, 2003, doi: 10.1038/sj.bjc.6601119.

[32] K. M. Leung, R. M. Elashoff, and A. A. Afifi, "Censoring issues in survival analysis," *Annu. Rev. Public Health*, vol. 18, pp. 83–104, 1997, doi: 10.1146/annurev.publhealth.18.1.83.

[33] S. E. Dietz, *SURVIVAL ANALYSIS Techniques for Censored and Truncated Data*, no. 10. 1985.

[34] J. P. Klein and M. L. Moeschberger, "Survival analysis: techniques for censored and truncated data." Springer, New York, pp. xiv, 502 p., 1997, [Online]. Available: file://catalog.hathitrust.org/Record/003165121.

[35] Z. Zhang and S. Jianguo, "Interval Censoring," *Stat. Methods Med. Res.*, vol. 19, no. 1, pp. 53–70, 2010, doi: 10.1177/0962280209105023.

[36] M. A. Hernan, "The Hazards of Hazard Ratios Miguel," *Epidemiology*, vol. 21, no. 1, pp. 13–15, 2010, doi: 10.1097/EDE.0b013e3181c1ea43.

[37] B. George, S. Seals, and I. Aban, "Survival analysis and regression models," *Am. Soc. Nucl. Cardiol.*, vol. 21, no. 4, pp. 686–694, 2014, doi: 10.1007/s12350-014-9908-2.

[38] L. D. Case, G. Kimmick, E. D. Paskett, K. Lohman, and R. Tucker, "Interpreting Measures of Treatment Effect in Cancer Clinical Trials," *Oncologist*, vol. 7, no. 3, pp. 181–187, 2002, doi: 10.1634/theoncologist.7-3-181.

[39] C. Hongda *et al.*, "Colonoscopy Screening Compliance and Related Factors in High-risk Population of Urban Colorectal Cancer in China," pp. 1–7, 2020.