2023

Gender Disparities: 5 Year Survival Rates of Elderly Colorectal Cancer Patients with Depression

Riya Joshi
Walden University

Walden University

College of Health Sciences and Public Policy

This is to certify that the doctoral dissertation by

Riya Joshi

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. Hadi Danawi, Committee Chairperson, Public Health Faculty
Dr. Hebatullah Tawfik, Committee Member, Public Health Faculty
Dr. Chinaro Kennedy, University Reviewer, Public Health Faculty

Chief Academic Officer and Provost
Sue Subocz, Ph.D.

Walden University
2023

Abstract

Gender Disparities: 5 Year Survival Rates of Elderly Colorectal Cancer Patients with Depression

by

Riya Joshi

MPH, University of North Texas Health Science Center, 2013

MBBS, Kathmandu University, 2008

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Public Health

Walden University

May 2023

Abstract

Colorectal cancer (CRC) is a life-threatening health condition known to greatly affect the elderly than younger populations. CRC, if comorbid with mental condition such as depression, can seriously affect patients' quality of life with a direct influence on the survival rate. There is a connection between depression and worsened cancer progression and survival rate of CRC, little is known on gender differences in survival rates on elderly CRC patients with pre-existing depression (PD). A socioeconomic model study was used to evaluate gender, initial diagnostic stage, and marital status (MS) at diagnosis as major independent variables and 5-years survival rate (5-YS) as the dependent variable for this retrospective study. Upon evaluating secondary data from SEER Medicare with total sample size of 28,278 patients, the 5-YS rate for women (35.8 %) and men (30.8 %) was significantly different as per Log Rank test (p <0.001). Further, Cox proportional hazard model (p<0.001) showed association between gender and 5-YS rates for elderly CRC patients with PD with Hazard ratio (HR) and 95% CI as 0.891(0.865, 0.919). There is also a significant difference in 5-YS between men and women for initial stage of diagnosis (I, II and III) with HR and 95% CI of 0.754 (0.688, 0.826), 0.823 (0.758, 0.894), 0.827 (0.759, 0.902) respectively but not at Stage IV. Recommendations include establishing gender specific treatment approaches for populations dealing with mental disease and terminal illness, creating a strong foundation for gender-based intervention in psycho-oncology. The positive social change of this study is to have a strong advocacy on development of gender specific treatment strategies and protocols if life ending condition such as cancer is comorbid with mental condition like depression.

Gender Disparities: 5 Year Survival Rates of Elderly Colorectal Cancer Patients with Depression

by

Riya Joshi

MPH, University of North Texas Health Science Center, 2011

MBBS, Kathmandu University, 2008

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Public Health

Walden University

May 2023

Dedication

I am dedicating this dissertation first to my parents, Dr. Shanker Lal Joshi and Kapil Joshi for giving me a dream and inspiring me to achieve it. Second to my husband, Dipendra Gyawali for all the support and encouragement throughout this long dissertation journey. I would also like to dedicate to my children Aakar and Aadir, for all the hugs and kisses. You guys are the inspiration for me to reach my goal. Last but not the least, I would like to dedicate this dissertation to my family and friends for all the support and love.

Acknowledgments

First, I would like to thank my Committee Chair, Dr. Hadi Danawi for the guidance and support throughout my dissertation journey. I made it through this journey because of the continuous encouragement. It was pleasure working with you and I am thankful for your time and positive words. I am equally thankful for my Committee Member, Dr. Hebatullah Tawfik and University Research Reviewer, Dr. Chinaro Kennedy for guidance and comprehensive feedback to me throughout this process.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction to the Study

Colorectal cancer (CRC) is defined as an abnormal cell growth of colon and rectum (Centers of Disease Control and Prevention [CDC], 2022). It is considered as a leading cancer next to lung and breast cancer globally (World Cancer Research Fund, 2018). As per National Cancer Institute (NCI, 2020-a, just second to lung/bronchus cancer, CRC has claimed more lives than any other cancer with roughly 150,000 more cases. The American Cancer Society (ACS, 2020 ac) estimated that more than 50,000 would die of CRC in 2020. . Based on information from Centers for Disease Control and Prevention (CDC, 2020) and a study by Murphy et al. (2011), CRC occurs in both men and women, but higher prevalence of CRC is reported in men than women. Similarly, there are mixed conclusions regarding gender dependency in the overall survival (OS) of CRC (Paulson et al., 2009; Koo & Leong, 2010).

In addition to gender, age is also determined as the prognostic factors in CRC patients (Itatani et al., 2018). Although numerous cases have been reported in younger people, CRC is more prevalent in people above the age of 50. Thus, the U.S. Preventive Services Task Force (2020) encourages initiating colonoscopy screening at the age 50 with follow up every 10 years interval.

Obesity, inactive lifestyle, processed meat and red meat, and smoking and alcohol use are considered as primary risk factors for CRC (Cancer Care, 2021). Chronic diseases like cardiovascular disease, Type II diabetes, lung disease, and depression are common in older people, but they are considered as independent risk factors for CRC incidence (Luque-Fernandez et al., 2020). However, such comorbidities have a substantial effect on

the OS of CRC (Luque-Fernandez et al., 2020). Depression, a mental disorder that is prevalent in older people, is reported to have higher impact in women than men of similar age (Girgus et al., 2017). According to the literature, cancer mortality can be influenced by mental illness (see Das Manushi et al., 2017; Cunninghal et al., 2015; Kim et al., 2019; Musuuza et al., 2013).

With the goal of evaluating the influence of pre-existing depression (PD) on the 5-YS of elderly CRC patients and if there is any gender disparity, I conducted an extensive literature review. I found minimal to no research on specific groups of patients who are elderly, depressed, and diagnosed with CRC. As stated above, there is no concrete conclusion on gender disparity in 5-YS of CRC, but depression shows greater impact on elderly women (see Girgus et al., 2017). So, it is important to understand if there is any gender bias on the survival of CRC patients with PD.

In this section I discuss CRC and depression related topics. The problem statement includes statistical information and addresses the void found in the understanding on this issue and explain the purpose of such a study. The research question identifies the variables of interest. Theoretical framework section outlines the approach used in the study. The nature section of this chapter provides information on study design and lists all the variables of interest, the methods and procedure to collect data, and the type of data I used. The main variables are described in the definition section. The assumption section list the assumptions to be used for the study. The scope and delimitation section point out the boundaries of the study. The shortcomings of the

study are discussed in the limitation section. The significance section explains the implication and the positive social change with respect to study.

Background

CRC is considered a serious health issue affecting both genders. The complex nature of colon biology and the cancer diagnosed stage impact the OS of an individual (Jochim et al., 2019). Specific to CRC, prevalence of acquiring cancer is significantly higher in older age (above 50). Six out of 10 patients are diagnosed with CRC at age above 70 years (elderly; Millan et al., 2015). Mistry et al. (2014) performed multivariate Cox regression analysis with an aim to predict the risk of dying across patients of various ages. People below 50 are considered young patients and are categorized as lower risk population for dying with CRC, diagnosed at any stage of diagnosis as compared to older patients (above 50 years old).

More recent studies tend to lean towards females being affected less with CRC (White et al., 2018). Although controversial, Yang et al. (2017) stated that OS and cancer-specific survival (CSS) was also comparatively greater in women. Further, the lifetime risk of CRC is also relatively higher in men (ACS, 2021a). Yang et al. (2017) studied on the gender differences on CRC patients on the OS concluded that women have better OS and CSS.

It is also worth mentioning that most female patients have a greater incidence of right sided tumors than their male counterparts (Quirt et al., 2017). Previous studies suggested that men and women with no pregnancy showed a similar survival rate but significantly differ from women who had been pregnant. (Koch et al., 1982). Further, the

estrogen receptor expression in colonic mucosa was also linked to the protective effect against CRC (Campbell-Thompson et al., 2001). These findings tend to suggest that overall endocrine functioning may influence CRC patients' survival (Campbell-Thompson et al., 2001). In addition to this, the cancer experience can be devastating, especially to the patients already dealing with mental issues such as depression (Campbell-Thompson et al., 2001). For the past 40 years, numerous studies have concluded that depression has a direct impact on immune function, endocrine function, and tolerance to the treatment of various cancer metastasis (Jia et al., 2017). For example, there is a significant increase in plasma cortisol levels in patients with depression (Jia et al., 2017). Further, cortisol is well known to be linked to the cell cycle regulation and modulation of cell growth (Jia et al., 2017).

Counter intuitive to the reasoning that cancer patients may go through mental health challenges, less than 20% of cancer patients are diagnosed with mental diseases (NCI, 2018b). Often the symptoms of depression may be overlooked in cancer patients due to similarity in symptoms like fatigue, loss of appetite, physical pain (Peng et al., 2019). A recent extensive literature review was done by Yu-Ning Peng et al. (2019) to understand the occurrence of depression particularly focusing on CRC. A total of only 15 studies published during 1967 to 2018 across 10 countries were found to be closely relevant to understanding the involvement of depression in CRC patients. Among these CRC patients, overall prevalence of depression ranged from 1.6% to 57%. Some of these studies tried to evaluate the age-associated prevalence of depression and reported a mixed conclusion. Only one study by Lavdaniti et al. (2012) reported on the gender bias on

prevalence of depression among CRC patients. Lavdaniti et al. studied the occurrence of depression in cohort of 79 CRC and pointed out a greater prevalence rate of depression in all CRC patients but indifference to gender.

Disregarding cancer, women at all ages are reported to have higher chances of diagnosing with depression than men of similar ages (Girgus et al., 2017). Girgus et al. (2017) provided a holistic review especially on yje old age (above 60 years) cohort and concluded that there was a clear difference in unipolar depression or depression-like symptoms in elderly women compared to male of similar age group. Although there is a clear difference between prevalence of depression between the genders, there is no study to the author's knowledge comparing the survival of CRC patients with depression among both genders. The outcome of this study will add to literature and has the potential of illustrating the importance of gender appropriate mental health intervention in addition to cancer management.

I this study, I focused on the gender biased in 5-YS rate for elderly CRC patients with initial depression diagnostic stage, and Marital status at diagnosis (MS) at diagnosis. Future studies can be focused on etiological factors or other risk factors for gender disparities in survival. This information on gender disparities in depressed elderly CRC patients may advocate the addition of psychological care for cancer patients. Gender specific strategies can also be incorporated during screening, treatment methods, as well as in the prevention protocol which might help in reduction of mortality and increase in quality of life (see Kim et al., 2015).

Problem Statement

Depression as a comorbid condition affects the cancer patients' quality of life and there is an association between longer hospital stay, progression of disease, and mortality rate, and with worst cancer progression (Saracino et al., 2015; Yang et al., 2018). Symptoms of depression can be easily ignored in elderly cancer patients due to its similar nature to other old age symptoms (Saracino et al., 2015). There are some studies that suggested that the cancer related mortality is higher in individuals diagnosed with mental health illness like schizophrenia, schizoaffective disorder, and bipolar disorders versus than the individual without mental illness at diagnosis (Cunningham et al., 2015; Das-Manushi et al., 2017). More studies suggested there maybe be gender disparity in CRC survival (Yang et al., 2017). But there are not many studies done on gender differences in survival on CRC patients with depression in elderly. Understanding this gap in literature, the objective of the study was to see if there is gender bias in 5-YS rate between elderly CRC patients with PD. It is important to understand the gender disparities to come up with gender specific strategies for treatment if there is disparity.

Purpose of the Study

In this study, I aimed to investigate the difference in 5-YS rates for elderly pre-existing depressed CRC by gender. The major independent variables were gender, initial diagnostic stage, and MS at diagnosis. 5-YS rate was the dependent variable. The covariates were race, ethnicity, socioeconomic status, and primary site of tumor.

Research Questions and Hypothesis

RQ1: What is the association between gender and 5-YS rates of elderly CRC patients with PD?

$H_0$1: There is no association between gender and 5-YS rates of elderly CRC patients with PD.

H_a1: There is an association between gender and 5-YS rates of elderly CRC patients with PD.

RQ2: What is the association between initial diagnostic stage (Stage I, Stage II, etc.) and 5-YS rates of elderly CRC patients with PD?

$H_0$2: There is no association between initial diagnostic stage (Stage I, Stage II, etc.) and 5-YS rates of elderly CRC patients with PD?

H_a2: There is an association between initial diagnostic stage (Stage I, Stage II, etc.) and 5-YS rates of elderly CRC patients with PD?

RQ3: What is the association between MS at diagnosis and 5-YS rates of elderly CRC patients with PD?

$H_0$3: There is no association between MS at diagnosis and 5-YS rates of elderly CRC patients with PD?

H_a3: There is an association between MS at diagnosis and 5-YS rates of elderly CRC patients with PD?

RQ4: What is the association between gender, initial diagnosis stage and MS at diagnosis and 5-YS rates of elderly CRC patients with PD after adjusting for race, ethnicity, socioeconomic status, primary site of tumor?

$H_0$4: There is no association between gender, initial diagnosis stage and 5-YS rates of elderly CRC patients with PD after adjusting for race, ethnicity, socioeconomic status, primary site of tumor.

H_a4: There is an association between gender, initial diagnosis stage and 5-YS rates of elderly CRC patients with PD after adjusting for race, ethnicity, socioeconomic status, primary site of tumor.

Theoretical Framework

The socioecological model (SEM) was first introduced in 1970 by Bronfrenbrener and McLeroy which later was formalized into theory in the 1980s by Belsky and Steuart (Kilanoswki, 2017). Theory suggests that an individual is in the middle and is surrounded by various factors (Kilanoswki, 2017). SEM is a theory-based framework that helps researchers to understand that there are various levels that affect health. The hierarchical levels of included in SEM are individual, interpersonal, community, organizational, and policy level. This theory has various levels and offers guidance to evaluate gender inequalities at various levels. I aimed to evaluate the gender inequalities at various levels like initial stage at diagnosis at individual level, treatment type at organizational level.

Table 1

Table 1 Construct of SEM

Construct of SEM	Independent variable/covariates
Individual	Independent variables: Gender, initial stage of diagnosis of cancer Covariates: Ethnicity, socioeconomic status, race, and primary site of tumor
Interpersonal	Independent variable: Marital status

Nature of the Study

In this retrospective study, I collected data from records and the outcomes have already occurred in the past (see Setia, 2016). In retrospective study, a resseacher start with exposure, and follows the cohort during the follow-up period (Setia, 2016). Retrospective studies offer quick turnaround time and relatively inexpensive than other studies like prospective cohort (Setia, 2016). In addition, the objective will be to look for survival time and there will be no need to manipulate variables in the study. So, the study design is chosen as appropriate study design.

The Surveillance, Epidemiology, and End Results Medicare Linked (SEER-ML) database was used for secondary data analysis. Data from various cancer registries are collaborated and various information like clinicals, demographics, cause of death, cancer diagnosis are collected in SEER-ML database (NCI, 2018-c). New diagnosis of cancer patients is available in the SEER Program. Medicare data has

information on claims since an individual's eligibility of Medicare till their death . The data is linked between two major databases . The linked SEER-Medicare Database is a very large and complex file. The cancer data file has information about the cancer diagnosis (NCI, 2019- d). Other files include chronic conditions flags and census tracts file. These files were requested from SEER- Medicare. Gender, PD status and initial stage of diagnosis are independent variables whereas race, marital status, and primary site of tumor are the covariates.

CRC patients who had claims for depression with Medicare Part A and Part B available in the reference timeframe of 1 year were flagged as depression in SEER-ML database (Center for Medicare and Medicaid [CMS], 2020). During data analysis missing, incomplete, and duplicate data was removed from analysis.

I used quantitative methods as the outcome of interest was 5-YS rate of elderly CRC patients with PD. The 5-YS rate of CRC was termed as the percentile of individual who live for 5 years after cancer diagnosis. The main research question was to determine if there was gender disparity in 5-YS rates of patients with PD and how much was the difference between men and women. Statistical methods like KM measure the 5-YS rates and LR test measures the difference in 5-YS rates (Bland & Altman, 2004). LR test considers the follow up period and no information is required regarding the curve appearance or distribution of the survival time, so it is advantageous (Bland & Altman, 2004). The null hypothesis for LR test is that probability of event difference at any point of time between the population is null (Bland & Altman, 2004). The LR test does not provide the estimate of difference on survival time or confidence interval as it is just a

test of significance (Bland & Altman, 2004). In this study I used KM estimates and a LR test.

CP hazard model and hazard ratio should be used to estimate the size difference and evaluate if there is any association between covariates and 5-YS in men and female separately (Bland & Altman, 2004). In this study, I used bivariate models at first the used multivariate models with gender, initial stage of diagnosis, and MS at diagnosis as independent variable and race, ethnicity, socioeconomic status, primary site of tumor, and primary site of tumor as covariates. In this study, the subjects who were alive at 5 year-follow up period were censored. The quantitative findings are generalizable to the population included in study (see Rahman, 2017). In this study, I attempted to quantify if there was difference in 5-YS rates, so the quantitative method was the appropriate method of choice. The data was evaluated with IBM Statistical Packaging for Social Science (SPSS) 25 version.

Definitions

CP hazard model: The hazard rate is calculated with this model which is the number of new cases of disease at risk per unit time (The Pennsylvania State University, 2021). It also shows the hazard function which is the probability of person to survive at time (The Pennsylvania State University, 2021)

Colorectal cancer (CRC): Tumor or abnormal growth in cells of colon and rectum (American Cancer Society, 2020-b).

Depression: A common and serious mood disorder. If individuals have the signs and symptoms like persistent sad and anxious mood, hopelessness, irritability, feeling of

guilt, worthlessness, loss of interest, decrease in energy, restlessness, difficulty in concentration, loss of appetite, or suicidal thoughts for more than 2 weeks then diagnosed with depression (National Institute of Mental Health [NIMH], 2018-b).

Elderly: In this study, CRC patient greater than or equal to 67 years are considered as elderly

5-years survival rate: The percentage of individuals alive after 5 years after they were diagnosed with disease or cancer (NCI, 2019- e).

Gender: the state of being female or male.

Hazard ratio: Relative risk of the event occurring at time *t* (The Pennsylvania State University, 2021).

Incidence: New cases during the specified time interval (NCI, 2019- a).*Initial Diagnosis Stage*: Stage of CRC disease at diagnosis

Kaplan Meier (KM): It is the statistical method to see the survival curve (Bland, & Altman, 2004).

LR test: Statistical method to evaluate difference in the survival between 2 groups (Bland, & Altman, 2004).

Mortality: Refers to the total deaths in a certain group during a period (NCI, 2019- f).

Prevalence: the proportion of individuals with disease (NCI, 2019- d).

Race: Race if social constructs that are used to categorize and characterize distinct populations based on biology and physical characteristics (Blakemore, 2019).

Primary site of tumor: Original or first part of the body where cancer arises from (NCI, 2019- i)

Assumptions

Secondary analysis was done based on SEER- ML database. For this study, I assumed that the secondary data from SEER -ML database was accurate with minimal error. Additional assumptions included that the required variables were available in the secondary data. Missing data was handled by removing the observations from analysis. The data was de-identified for analysis.

Scope and Delimitations

Elderly CRC patients diagnosed with depression before CRC cancer diagnosis were the scope of the study. I decided to limit the scope of the study on elderly patients because it is challenging to distinguish the symptoms of depression and signs of old age due to these signs of depression being often neglected in elderly. Since the study is focused on elderly patients, the results from this study may not be generalized to the young group of individuals. Since the goal of the study is to see if there is difference in 5-YS rates by gender in CRC patients with depression, additional research will be needed to determine the cause if there are gender disparities. The drawback of a retrospective study design is that the data may not be very accurate as the exposure and outcome are already collected before the study is started (Setia, 2016).

Limitations

One of the limitations of the data included limiting it to patients aged 67 and older. One year of Medicare data before diagnosis of cancer was needed to categorize the

patients with pre-existing diagnosis of depression. Depression diagnosis was based on chronic condition flag file from SEER-Medicare database which includes, yearly, midyear and ever flag (CMS, 2021-a). Depression diagnosis is based on presence of treatment using claim-based algorithms created by CMS (CMS, 2021). During the 1-year reference period, if a patient has valid ICD-9, ICD-10, CPT4 or HCPCS Codes on at least one inpatient, SNF (Skilled Nursing Facility), as well as on HHA (Home Health Agency), HOP or Carrier claim then the patients were flagged to have depression (CMS, 2021). Data was limited to patients with Medicare Part A which includes inpatient/hospital coverage and Part B coverage includes outpatient/medical coverage (see CMS, 2021). The claim data from managed care and HMOs are not included for the depression flag (CMS, 2021). The diagnosis of depression in the patient was based on diagnosis mentioned on physician's professional charges which may have a chance of underreporting. There were chances of undetected selection bias as well. In addition, the newly eligible Medicare beneficiaries may have only a partial year of fee-for-service (FFS) coverage and may not have the depression flag (NCI, 2019- d). To overcome this limitation, only CRC patients of age 67 and older were included There were charges associated with requesting the database. It was essential to understand the approval process to request the data from SEER-ML database. The timeline for research also incorporated the time from understanding the data structure, request data, and time to obtain data from approval.

Significance

It is evident that depression affects the quality of life of cancer patients as it can predict cancer prognosis and is negatively associated with cancer (Ko et al., 2019). There is clear association between PD and a worse cancer prognosis (Yang et al., 2018). The cancer incidence is comparable in patients with or without diagnosis of mental illness, but the cancer-related mortality is seen greater in patients with diagnosis of mental illness than with those without (Cunningham et al., 2015). There is an association between depression and an increase in all cancer mortality rate (Ko et al., 2019). Therefore, PD is one of the factors that needs to be evaluated in newly diagnosed cancer cases. The assessment of depression in the elderly is difficult compared to other age groups due to age related physical changes, signs of normal aging, and presence of comorbid conditions in the elderly age group and the symptoms of depression can easily be overlooked (Saracino et al., 2015). In terms of public health and clinical importance, it is important that depression and anxiety are detected early, and early start of intervention is initiated (Parpa et al., 2015). Emotional support is essential in addition to regular cancer treatment (Parpa et al., 2015). Depression is common in women, but the survival of CRC is better than men (Maajani et al., 2019; NCI, n.d.-a). There is an inadequate understanding about gender difference in CRC cancer survival in d elderly CRC patients with PD. Knowledge of difference in 5-YS rates in CRC patients with PD can increase awareness and help in coming up with strategies and resources. This study can be useful for understanding the gender differences and integration of psychological care along with cancer treatment and consider gender differences and address the specific needs of all patients. If there is

difference in 5-YS, then future researchers should focus on cause of inequalities and cause of inequalities to improve outcomes in men and women.

Collaboration between oncologists, psychiatrists, primary care practitioners, psychologists, social workers, and nurses is necessary to deliver comprehensive and holistic care (University of Washington, 2021). Such collaboration in cancer management had been effective in cancer care and in many research trials (University of Washington, 2021). If health care professionals are aware of gender differences, the care team can evaluate the specific needs as per gender in case of depressed elderly patients. The improvement of care quality of cancer patients, medical support, and social support should be provided based on taking into consideration gender-based medicine and psycho-oncology (Koyama et al., 2016). In addition, by understanding the gender related differences, gender specific treatment strategies and protocols can be developed with the aim to reduce mortality and make the life of patient better (Kim et al., 2015).

Summary

The prognosis of cancer and life quality of such patients depends on depression diagnosis. Depression symptoms are often ignored in elderly cancer patients due to similar nature of symptoms of depression and old age. Studies have shown gender disparities in survival of CRC and there are no studies on gender difference on survival of CRC patients with PD. The study investigated the difference in survival time by gender in elderly patients with PD diagnosed before cancer diagnosis. Additionally, the difference in 5-YS was evaluated in terms of initial diagnosis stage and marital stage at diagnosis. SEM was the basis for this study. Retrospective study was planned based on

secondary data from SEER-ML database. The statistical test was based on the belief that the data was precise and accurate. The limitation of the study was that only patients age 67 or older were included in the study. Part A and Part B Medicare claim data was used to identify the chronic condition flag depression status of the elderly CRC patients. There was limited knowledge in 5-YS rate by gender in patients with PD and the result of this study was useful integration of psychological care in cancer patients. The treatment of cancer patients can be focused on collaborative care between oncologist, psychiatrist, primary care physician, psychologist, social workers, and other care team members. The next chapter will elaborate on the literature review on CRC. There will be additional focus on various study variables and the existing gap in the literature

Chapter 2: Literature Review

In this chapter I reviewed information on CRC and its association with PD, epidemiology and survival, and gender disparity based on peer-reviewed literature. I also aimed to establish the unmet need of understanding on gender disparity in survival of CRC in PD elderly patients and provide the justification for the need of further research on gender disparities in 5-YS in CRC patients with PD.

Literature Search Strategy

Through literature review search, I gathered information that enabled me to review CRC and depression, the survival of the disease, and the gender disparity. I searched peer-reviewed articles through Medline, PubMed, PsycINFO, CINAHL, Medscape, Science Direct, Google, and Google Scholar. Most of the peer-reviewed articles were published between 2015 and 2020. Some of the papers that were important but older than 2015 were also selectively included in the literature review. The keyword used were *CRC, CRC survival, overall survival, cancer specific survival, gender differences, improvement in survival in CRC, sex effect on cancer, cancer mortality, depression, mental health disorder,* and *preexisting mental health and cancer*. I included peer-reviewed journals published in English language primarily focused on cancer survival and depression. This literature review aims to gain in-depth understanding on CRC and depression and to design a study to identify and fill the gap in gender disparity in 5-YS in depressed CRC patients.

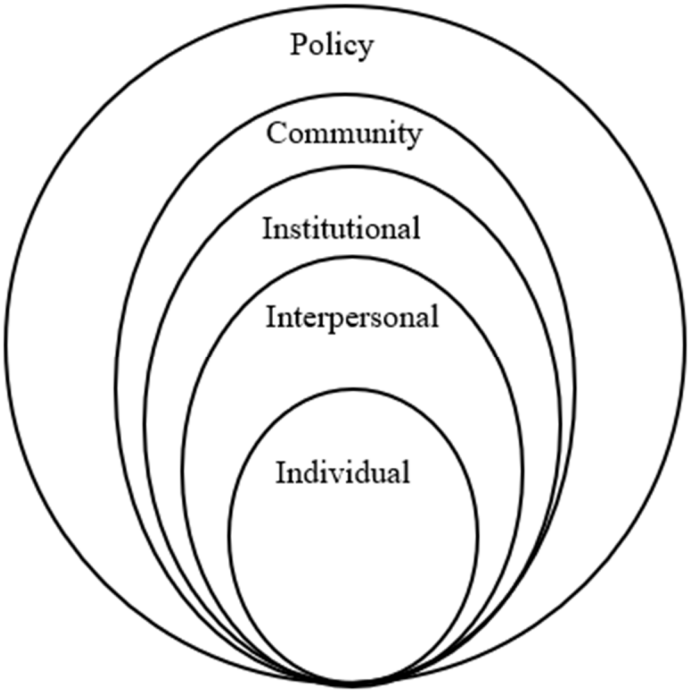
Theoretical foundation

SEM-based theoretical framework research was originally conceptualized by Bronfrenbrener and McLeroy (Moore et al., 2015). SEM focuses on understanding the relationship between various determinants. SEM helps researchers to understand the complex and multilevel system of influences on survivorship (Moore, Buchanan, Fairley, & Lee, 2015). The initial theory of Bronfrenbrener suggested that a behavior is influenced by various factors and the level of influence is measured as micro-, meso-, exo-, and macrosystem (McLeroy et al., 1988). Microsystem refers to face- to- face influence, interaction with family members, social networks, and groups. Mesosystem refers to the influence of various settings like school, church, peer group on individuals whereas exo system refers to a larger social system, and finally macro system is the cultural belief. Using Bronfrenbrener's theory as framework a lot of work are done in the areas like child abuse, assessing healthy lifestyle, studies related to problems confronting community psychology. Mostly for health promotion the model is used. The importance of SEM is that it focuses the problem from different levels, and it can help to come up with intervention according to the levels (McLeroy et al., 1988). A slight modification was done by McLeroy et al. (1988) on Bronfrenbrener's model by Belsky and Steuart (1970) in which the outcome is influenced by various behaviors at five different levels which includes intrapersonal, interpersonal, institutional, community, and public policy. SEM emphasizes that there are various factors that affect the outcome (Moore et al., 2015). This model helps to understand the relationship between individuals and other

determinants of health (Moore et al., 2015). The figure below shows the five concepts of SEM.

Figure 1

Five Concepts of SEMl



There are five concepts in the SEM as listed below.

1. Individual level: It includes individual patient's characteristics like demographics, gender, race, ethnicity, socioeconomic status, an individual's attitudes, and behaviors and so on.
2. Interpersonal level: It includes social networks of the patient including family members, coworkers, friends, and other direct social interactions.

3. Organizational/ Institutional level: It includes characteristics of social organizations.
4. Community level: it includes the relationship between various organizations and networks.
5. Policy level: It includes local, state, national, global laws, and policies.

This model was widely used by various reputed organizations of public health. More specifically, the CRC Control Program (CRCCP) used the SEM model and addressed multiple factors that can influence CRC prevention (CDC, 2021). In Project HEART, change agents were identified at five levels which included HEART participants at individual level, HEART promoters, participant's family, friends, social networks in interpersonal level; parks and recreational facilities in organizational level; and the Community Health Advisory Council (CHAC) at policy level (Balcazar et al., n.d.). SEM assumes that there is connection between individuals and other factors that influence the individual.

I used SEM for this study to understand factors that might affect the gender difference in 5-YS of CRC in individuals with PD. Out of five levels of SEM, the factors that influence 5-YS gender difference are as follows.

- Individual level: Demographic characteristics like gender, race, socioeconomic status, primary site of tumor and initial stage of diagnosis of cancer are included.
- Interpersonal level: It includes marital status.

CRC Overview

CRC is a terminology collectively referred to the cancer associated with colon or rectum (CDC, 2022). It is also known as bowel cancer. The colon constitutes the large intestine whereas rectum refers to the area connecting colon to anus. In most cases, cancer is initiated as a benign growth of tissue within these areas and later progresses as a cancer. As per NC (n.d.-b), the most widespread type of CRC is adenomatous CRC which constitutes -95 % of all large intestinal tumors. The cancer later spreads through the lymphatic system and commonly metastasizes to the liver, lungs, and brain. CRC has also been reported spreading to adjacent pelvic organs like prostate in men and ovaries in women and to bones (NCI, n.d.-b).

The most used strategy for pointing out the stage of CRC is known as TNM classification system , which refers to tumor (T), node (N), and metastasis (M) (ACS, 2023a). At stage T, size and growth of the tumor within the transverse section of intestinal tissue governs the staging. For example, if the tumor is restricted within the inner layer of the bowel, it is considered stage T1; if the tumor is grown into the underlying muscle layer, it is considered stage T2; if it reaches the outer lining of the bowel, it is stage T3, and finally if it outgrows the bowel wall and progress to peritoneum or nearby organ, it is considered stage T4. At stage N, confirmation of cancerous cells on the total number of lymph nodes directs the staging. For example, stage N1 suggests cancer cells found in one to three nearby nodes whereas stage N2 suggests cancer cells are found in four or more lymph nodes. Finally, stage M is characterized as the progress of cancer to different parts or organs of the body.

To properly stage upcoming CRC cases via the TNM staging system, various medical tools are used for proper diagnosis (NCI, 2021-i). Although a colonoscopy remains the gold standard for diagnosis, intestinal obstruction and perforation may signify a poor prognosis. A positive colonoscopy is usually followed by more sophisticated techniques such as CT scan, surgical exploration, pathologic examination of resected tissue, evaluation of the tumor penetration into the wall, and evaluation and involved lymph nodes site. Although most of the CRC is localized during diagnosis, elevated alkaline and positive magnetic resonance imaging may link metastases to CRC (Thanikachalam & Khan, 2019).

Some of the most common symptoms that led to the diagnosis of CRC were change in bowel movement, bleeding from rectum, iron deficiency anemia, abdominal pain, weight loss, and loss of appetite (ACS, 2023-b). But if the CRC progressed to stage M, site dependent symptoms may appear as well (Thanikachalam, & Khan, 2019). The most common risk factors of CRC are believed to be gender (male), age (older population), hereditary, environmental whereas positive family history counted approximately 10% to 20% cases (Dekker et al., 2019). Some of the environmental risk factors included smoking, alcohol, obesity, and Type 2 diabetes and, in some cases, bacterial species such as *Fusobacterium nucleatum* and *Bacteroides fragilis* were also linked to the risk (Dekker et al., 2019).

Treatment options for CRC are chosen according to the spread of the cancer cells. If it is loco-regional, surgical resection is the preferred treatment whereas for locally developed CRC chemo-radiation is further recommended upon surgery. CRC with

involvement of lymph nodes or metastases to distant sites, chemotherapy is the treatment of choice (Thanikachalam & Khan, 2019). In recent years, more advancement in the field of immunotherapy has enabled cancer patients including CRC with antibody treatment such as Pembrolizumab. Even though carcinoembryonic antigen (CEA) is a known tumor indicator, it is not used as a diagnosis tool but for posttreatment surveillance.

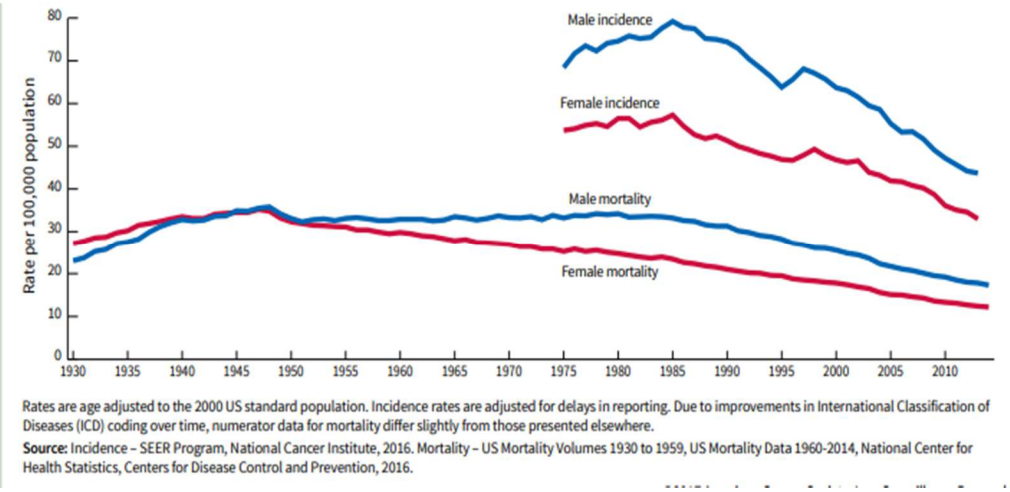
As stated by U.S. Preventive Services Task Force (2016), CRC was found to be more frequently diagnosed and the median age of death was 68 years and range 65 to 74 years. But according to ACS (2017), CRC was diagnosed mainly at 68 (men)and 72 (women) years. The risk of CRC increases with age, so screening benefits mostly for asymptomatic age groups of 50 to 75. It is advocated by the U.S. Preventive Services Task Force that colonoscopy screening should be initiated at age 50 with follow up every 10 years interval.

Epidemiology of CRC

CRC constitutes nearly one in 10 of all diagnosed cancer and cancer associated deaths annually (Itatani et al., 2018). As per ACS (2020a), over 100,000 cases of colon and nearly 50,000 rectal cancers were estimated alone in the United States in 2020. CRC claimed about 53,200 deaths in the year 2020 in the United States alone (ACS, 2020 - a). However, looking at historical data, there was a decrease in the CRC incidence after the year 1985, although the previous decade (1975 to 1985) saw a gradual increase in the incidence of CRC. Similarly, overall mortality had a downward trend in the United States as illustrated in Figure 2 below (ACS, 2017-c).

Figure 2

CRC Incidence (2009-2013) and Mortality (2010-2014) in United States by Sex



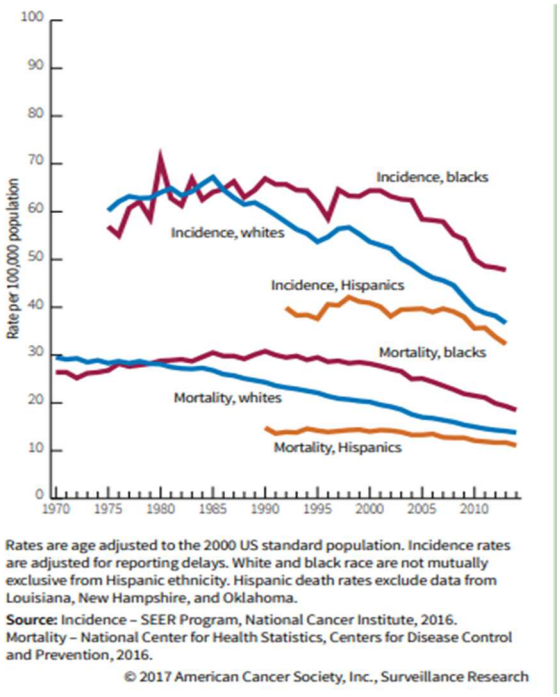
The NCI (n.d.-b) stated that "Americans have about a one in 20 lifetime risk of developing CRC" (p.3). One in 23 men have the chance of developing CRC while one in 25 in women have the risk (ACS, 2020-a). There was an increasing risk of developing CRC after 40 years but seemed to affect people over 65 years old; however, around 3 % of malignancies occurred below the age of 40. The incidence of CRC looked unbiased to gender until age of 50, but after the age of 50 the incidence was slightly higher in men (NCI, n.d-a.).

Analyzing the long-term cancer incidence data available for White and non-White population in the United States, the incidence was similar for both populations until the 1980s then it started declining in Whites but remained the same for Blacks creating a racial gap (ACS, 2017). Similarly, Hispanic communities reported lower incidence and

mortality (Figure 3). Data also suggested that since the 1990s, the incidence of CRC has declined in all three communities (ACS, 2017).

Figure 3

CRC Incidence and Mortality Rates in United States



Globally, CRC mortality and incidence rates have been reported greater in developed countries (Arnold et al., 2017). This trend of CRC was stable or downward in developed countries but increasing in developing countries (Arnold et al., 2017). This decreasing trend in both incidence and mortality rate may have been an impact of proactive cancer screening.

Epidemiology of Depression

Depression is a common mental health issue worldwide in recent decades (Lim et al., 2018). According to the National Institute of Mental Health (NIMH; n.d.), the

prevalence of CRC alone was approximately 11.3 million in the United States. Lim et al. (2018) performed meta-analysis on prevalence of depression in 30 countries over the period of 1994 to 2014. Data showed point, lifetime, and 1 year prevalence was 12.9%, 10.8%, 7.2 % respectively (Lim et al., 2018). The prevalence of depression tends to be biased towards gender (Brody et al., 2018). Unlike CRC, women are twice as vulnerable for depression when compared to male counterparts and in all age groups (Brody et al., 2018). Some of the major risk factors include the influence of sex hormones, stress, extreme violence, childhood sexual abuse, lack of gender equality at the societal level, and discrimination (Riecher-Rössler, 2016). The prevalence of depression in the United States in adult women was 8.7% compared to men 5.3% as per data from year 2017 (NIMH, n.d.). The National Survey on Drug Use and Health from 2005 to 2015 showed that the prevalence of depression has been significantly increased in the United States mainly in younger and older age groups (Weinberger et al. 2018). Survey data from 2005-2016 showed an upward trend of depression for age groups greater than 65 years old (Yu et al., 2019).

CRC and Depression

Even though the occurrence of cancer is similar in patients with and without mental illness, death from cancer appears accelerated in patients with mental illness in developed countries (Cunningham et al., 2015). The quality of life of a patient diagnosed with cancer is influenced by depression (Saracino et al., 2015). Depression also decreases the ability of individuals to deal with problems (Saracino et al., 2015). Depression, longer hospital stays, and rapid disease progression are linked to greater mortality rate in such

patients (Saracino et al., 2015). Depression, especially at an older age, may likely prevent people from seeking preventive screening and demotivate them from regular activity (Pinquart & Duberstein, 2010). Most vegetative symptoms such as sleep, appetite, energy level, and concentration can be further affected by the treatment side effects of cancer. In addition to this, physical effects, especially in immunological and endocrine function, of depression have a direct effect on Cancer survival (Pinquart & Duberstein, 2010). Assessment of depression is harder in patients diagnosed with cancer as it is difficult to distinguish whether the symptoms are due to mood disorder or due to illness and medical treatment (Saracino et al., 2015). PD in CRC is also associated with longer diagnostic intervals as well (Walter et al., 2016).

Das-Manushi et al. (2017) did a longitudinal study from the UK and National Statistics in England and Wales from Jan 1, 2007, to Dec 31, 2014. The aim was to estimate the mortality in individuals with mental conditions for various racial subgroups. Various covariates related to demographics were included in the study. Shenfield residual plots were used to assess the interaction with survival time and likelihood ratio test was done to assess the statistical interaction. As per Das-Manushi et al. (2017), results showed that age standardized cancer mortality is higher in groups with mental illness than the age standardized cancer related mortality in the general population. In this study, severe mental illness in minority group has lower mortality than white British people.

Musuuza et al. (2013) did another study on Ohio residents who died of cancer between 2004 to 2007 and with the aim to compare the death rates of the group with psychological conditions to the population of Ohio. This study used data from a public

database in the Ohio mental health system. The authors linked this data to death certificate records in Ohio. For the authors, the presence and absence of mental illness was the main variable of interest, and the other variables of interest were year of birth, race, sex. They concluded that the cancer mortality is higher in individuals with comorbid mental conditions across all races and gender.

Similarly study by Cunningham et al. (2015) also showed cancer related mortality rate is relatively higher in cancer patients with mental illness. The data from the cancer registry in New Zealand were linked to records from psychiatric hospitals. The researchers compared the 5-YS of breast cancer and CRC patients and compared them to individuals with cancer diagnosis but without history of mental illness. Cancer specific survival was compared using Cox regression . Patients with such psychological conditions were individuals who had assessment and treatment for mental illness 5 years before cancer diagnosis . Cancer specific survival was the outcome of the study . Age at diagnosis, ethnic group, comorbidity index, stage at diagnosis, and deprivation status were considered as covariates. The study concluded that the survival is worst in mentally ill patients and both breasts as well as CRC cohort. In the CRC cohort, after adjusting for demographics the patients with mental illness have approximately three-fold increase in death due to cancer. The strong point of this paper was the completeness of data as all deaths were captured and minimal loss to follow-up). The weakness of this paper was that it overlooked factors like treatment, smoking status, as they were not available.

Study done on Finnish population by Yang et al. (2018) showed worse prognosis in cardiovascular disease and cancer when the patients have PD. Patients above the age of

24 who had conditions like of cardiovascular disease, stroke, or cancer were under evaluation from the diagnosed date to year 2012 for cause specific mortality. Logistic and Cox regression models evaluated short or long-term mortality graded by depression status. The study concluded that patients with PD do not have a good prognosis if patients have conditions like coronary heart disease, cancer, or stroke.

Ko et al. (2019) did another research in 5-year cancer survivors from 2004 to 2009. This study aimed to evaluate PD and its association with mortality after cancer. PD were identified as any psychiatric admission or outpatient visit within 2 years before depression. Very interestingly depression was linked with higher cancer mortality and is a predictor of prognosis of cancer. Lack of screening, delay in diagnosis of cancer, poor access to treatment is some of the contributing factors for high mortality for pre-existing mental illness cancer patients. This study concluded that there was a pronounced mortality in men with PD. Some of the drawbacks of this study were that the stage, type of cancer, medication used for psychiatric illness, marital status, level of education, status of employment, stress level was unavailable in the data so, the effect of stage and cancer treatment on mortality could not be evaluated.

A study by Lin et al. (2016) aimed to see if pre-existing mental illness was linked with the diagnosis stage, type of treatment. Data analysis was done on data from the U.S. Military Health System (MHS). The U.S. MHS data was linked with cancer registry data. Patients included were individuals who were histologically confirmed with non– small cell lung cancer (NSCLC). Two NSCLC groups categorized with or without pre-existing mental illness were similar in terms of cancer treatment, stage of diagnosis. Even though

there was no difference in cancer treatment, not much difference in advanced stage of disease between 2 groups, results showed increased mortality in lung cancer patients with PD. In this study there were higher percentages of smokers and survival analysis was also done adjusting for tobacco use but still, the survival was low in patients with pre-existing mental health disorder. Lin et al. (2016) pointed out one of the disadvantages that secondary data sources were used so there was a lack of detailed information and may have a chance of inaccurate records suggesting further research to be done on bigger sample.

A cross-sectional study was done on survey data from metastatic breast cancer, lung, and CRC patients regarding quality of life, comorbidity medication adherence, symptoms related to cancer then further linked it to claims data (Drzayich, et al., 2018). The study concluded low comorbidity medication adherence is one of the factors linked to increase in cancer symptoms and unhealthy days for cancer patients. The limitation mentioned by the author included generalizability of the study to younger and uninsured populations as Medicare claim data was used.

Kaerlev et al. (2018) did a study on Danish CRC group databases and did investigation on CRC patients with pre-existing mental conditions. The study aimed to see if having preexisting psychiatric illness affected the cancer stage at the time of surgery. The study concluded that patients with pre-existing mental conditions had a higher cancer stage at the time of operation even though no difference in incidence of CRC in the general population and patients with pre-existing psychiatric illness. So, it

was important that we give attention to CRC patients with pre-existing serious psychiatric illness.

Survival for CRC

Survival rate defines the proportion of patients living in a specified period following the diagnosis of a disease (Maajani et al., 2019). Such survival rate studies of a disease help to evaluate the effect of a particular treatment and other healthcare controls. To compare cancer survival across the globe, the first study of its kind known as CONCORD was conducted by participating 31 countries in 2008 (Coleman et al., 2008). This study was contributed by 42% of American patients diagnosed from 1990 to 1994 and concluded that 5-YS particularly to CRC was 60.1% . This study was further carried out under the program CONCORD-2 for the population from 67 countries during the period of 1995 to 2009 in 2015 (Allemani et al., 2015). Based on this study, there was an appreciable improvement in 5-YS for CRC for American patients when compared between the calendar intervals of 1995-1999 (61%) to 2005-2009 (65%). Such improvement in survival has been attributed to the advancement of early detection and treatment technology.

As an extension of CONCORD-2 study, White et al. (2017) led another independent study solely focusing on CRC patients of 37 states of America diagnosed between 2001 to 2009. This United State government led work concluded that when data year 2001 to 2003 was compared with year 2004 to 2009, there was a 0.9% increase in 5-YS rate in 2004 to 2009. The 5-YS rate was 63.7 % during 2001 to 2003 whereas the 5-YS rate was 64.6% during the year 2004-2009 . Pohar Perme estimator after controlling

mortality by other causes method was used to calculate the 1-, 3-, and 5-years net survival along with 95% confidence interval and compared for race, stage, and stage at diagnosis. In this state wise comparison, authors acknowledge a limitation that it should not be over interpreted due to some state estimates being outside the limits. Authors did a rigorous comparison between patients at various age groups, it would be interesting if they have extended their study further to understand gender dispersity and comorbidity especially to the older cohort.

Although cutting edge technology such as immuno-oncology, personalized cancer vaccines are emerging as an intervention to increase the survival rate of various cancer patients including CRC, surgical resection is the foremost choice of intervention of cancerous tumors. To understand the 5-YS of Radical (R0)-resected colon cancer patients from periods 1991–1995 and 1995–2000, Štor et al. (2019) another comparative study. They concluded that 5-YS for CRC was increased from 1995 to 2000 compared to the period from 1991 to 1995 in these patients. CRC patients who went through R0-resection in the period 1991 to 2000 were selected for analysis. Based on the Cancer Registry, the researchers evaluated the survival rate for the population. Difference in survival curves was evaluated and adjusted for potential covariates. The Likelihood test was done to check for the goodness of fit for multivariate models. Multivariate analysis results showed that early diagnosis and adjuvant aided chemotherapy were linked with higher survival upon surgery.

Another study focused on Iran led by Maajani et al. (2019) evaluated the survival rate of CRC. They performed a systematic review that led to a meta-analysis from

various international databases up to December 1st, 2017. All studies that had survival rate for patients with CRC in Iranian population were included in this study. At 5% significance level, Chi-square test was used. Results showed 1, 3, and 5-YS were 84%, 64% and 54% respectively for patients with CRC. When evaluated separately, 1-, 3-, and 5-YS rate for colon cancer was 90%, 69%, and 60% respectively whereas for rectum was 88%, 73%, and 54%. The study suggested the survival rate in Iran is between developing and developed countries. They also concluded that the survival of CRC is better in women compared to men. Not taking consideration of different stages of disease was the limitation of the study.

Furthermore, Yu et al. (2020) concluded that majority of recurrence, metastasis, death in CRC occur mostly before 6 years of diagnosis and is considered as a critical window time for CRC patients. This study pointed out that if the patient survived the initial 6 years period after diagnosis, the risk of disease outcome is less compared to the window critical period and the long-term consequence is minimal.

Factors Affecting CRC Survival Rates

Maajani et al. (2019) mentioned age, gender, race, histology type, tumor grade, tumor size, tumor stage, regional lymph nodes metastasis, pathologic stage, and location of tumor are some of the factors that influence the survival rate of CRC. Gender was considered as a most important factor and there were better chances of survival in rectal cancer compared to colon cancer. Further details on these factors are included in sections below.

Age and CRC Survival

The process of ageing is unique to individuals and the elderly age group patients have highly heterogeneous physical and mental attributes (Myint, & Gérard, 2020). It can be challenging to treat the elderly group because of comorbidities (Bojer, & Roikjær, 2015). It is difficult to distinguish depression in elderly cancer patients. Saracino et al. (2016) did a literature review with the purpose of identifying depression in elderly patients with cancer and searched for studies that addressed depression in cancer patients. The study concluded that it is hard to distinguish between depression ages related changes in elderly population due to its similar nature and the current DSM criteria may not be sufficient to evaluate depression for elderly cancer patients so recommended other additional criteria to evaluate depression. By now, it has been established that the incidence of CRC rises with age. About 50% of the patients are greater than age 80 (Itatani et al., 2018). It has been noted that CRC is more common on the right sided colon compared to left side in elderly patients. The incidence of right sided colon cancer was also interestingly higher in women in the age greater than 80. They further confirmed that age was one of the prognostic factors and should be considered when determining the strategy for CRC treatment.

The trend of CRC mortality also varies with age group (Liu et al., 2019). Liu et al. (2019) illustrated the trend is declining in younger subjects but increasing in aged patients. The researchers did the analysis on WHO mortality data from year 1987 to 2000 and data from China Statistical Yearbook from 2002 to 2016. The average yearly change in mortality rates were estimated by joinpoint regression analysis method. Results

showed that mortality rate declined in both men and women in the younger age group (below 45) but in the older age group (above 80) the mortality rate was increased in both men and women for the last 30 years. Mortality risk increased with age in all the four-age groups in the study). The exponential growth trend of advancing age might be influenced by biological factors. This study was more descriptive rather than causal inference and screening factor was incorporated in the analysis model, this was noted as a limitation of the study.

Feng et al. (2020) did a study based on SEER database with one cohort from the SEER 9 registries database from 1975 to 2009 and another cohort was based on SEER 18 registries database including CRC patients from 1973 - 2014 (Feng et al., 2020). The SEER 9 database showed no survival gain since 1990. SEER 18 database showed steady increase in survival but for the age group >85 the survival was not better.

Gender Difference in Survival in CRC

Gender was also a significant factor that affected the survival in CRC patients (Maajani et al., 2019, Yang et al., 2017). Many studies showed gender differences in CRC survival and women have higher CRC survival compared to men (Maajani et al., 2019). Afshar et al. (2018) did a population-based study to evaluate the survival difference by gender based on Victorian Cancer Registry. They included all new cases of 25 different invasive cancers from January1982 to December 2015 of age 15- 99 years, diagnosed in Victoria, Australia. Information extracted from the registry were demographics, tumor characteristics. Patients with multiple cancer were excluded from analysis. Results showed that most of the cancers along with CRC were more common in

men, but the survival disadvantage was for men. For CRC, the gender difference had been decreased since 1982. Afshar et al. (2018) also mentioned that the conclusions from their study are like other studies done in Canada, the United States, Europe, and Korea. Limitations mentioned was the inability to investigate the potential mediators for gender difference.

Similarly, another population-based study was done in Canada based on Ontario Cancer Registry (ocr), specific to colon cancer from year 2000–2008, showed the prevalence of disease was more on the right side for women, but there was no difference in disease stage, grade, treatment but still showed gender difference in long term survival (Quirt et al., 2017). The incidence of disease was greater in men and had a worse prognosis than women. They did a random sampling of all the patients from 2000-2008 with resection for primary CRC. Although the study showed gender difference in survival, the study was not sufficient to evaluate the causal mechanism for gender difference. Since random sampling was done in the study and because of large sample size selection biases were minimized in this study.

A meta-analysis was done on studies conducted from 1960 to 2017 to compare the difference in OS between male and female (Yang et al., 2017). Adjustment for survival analysis for CRC patients was done for covariates related to patients' demographics, progression of cancer, prognosis of cancer and the treatment received by patients. Literature search was done on databases including studies related to gender differences in OS along with cancer specific survival. Information was extracted including basic characteristics of the study, patient characteristics, age, TNM staging,

diagnosis, comparative outcomes. The meta-analysis result showed the OS and CSS were better in women than men. Gender was concluded as an independent prognostic predictor for CRC patients used. Only the studies available in Chinese and English were included in this meta-analysis so they may be eligible papers that may not have been included in this meta-analysis. The advantage of this meta-analysis is that the quality of papers included were satisfactory and this study overcome the limitation of small sample size.

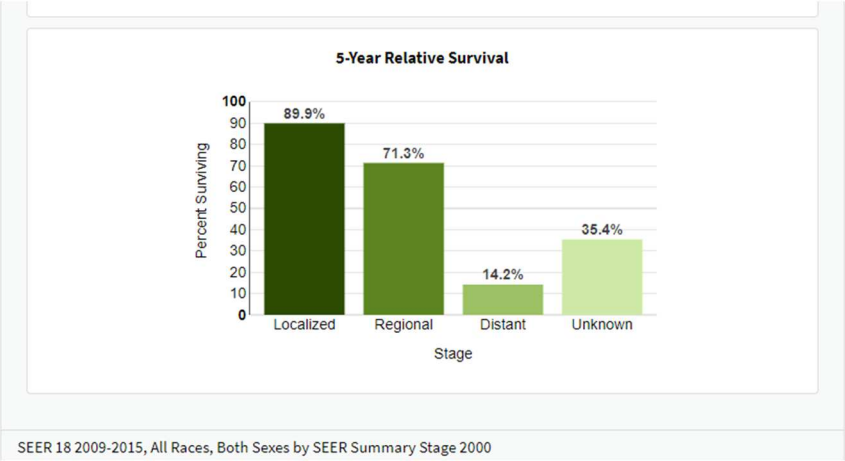
Kotake et al. (2016) analyzed data using the database from the hospital-based CRC registration system and included CRC patients who underwent surgery. The data was collected for the period of 1985 to 2004 for CRC patients with surgery. Age, gender, tumor characteristics were collected from the database. Gender was the main primary independent variable in the study. This Japan based study concluded that there was only a slight significant difference in survival when men and women compared. They also stated gender as an independent prognostic factor for CRC. Adjuvant therapy, and other potential confounders like socioeconomic status, life status in gender differences were not considered in this study. If any subject who did not have follow- up information were excluded which might have overestimated or underestimated the survival probabilities.

Stage of Initial Diagnosis of CRC and Survival

OS of CRC predominantly relies on the stage of the initial diagnosis (ACS, 2021-b). The relative 5- YS rate for colon or rectal cancer is as high as 91% when diagnosed at localized stage but if the cancer is spread to distal organs, the survival rate is drastically reduced to 14% (ACS, 2021-b). Figure 4 below shows the five year survival for various stages of CRC.

Figure 4

CRC 5-Year Relative Survival



A study from cancer registries in France (Martinique) between 1993 to 2012 evaluated the survival data over time in CRC patients (Joachim et al., 2019). In this study, the information extracted from the registry were gender, zone of residence, histology, subsite of cancer, stage at diagnosis. 62.4% of CRC patients included in the data had stage III -IV group. This study evaluated the difference in survival using the LR test. Results showed difference in OS by sex, age at diagnosis and stage. Stage III-IV at the stage of diagnosis had a hazard ratio of 3.7 and was considered one of the prognostic factors for OS. The primary tumor site or zone of residence were not prognostic factors. Limitations of this study was that it did not consider various factors like screening process, pattern of care in CRC, treatment method. Similarly, White et al. (2017) showed that the 5-YS was highest in patients with localized disease compared to distant stages of CRC. The survival rates for the periods 2001-2003 and 2004-2009 were not different for

the localized stage of CRC. There was an increase in survival probability when data from 2004-2009 compared to period 2001-2003 in distant and regional stages.

Marital Status and Survival in CRC

MS was identified as one of the factors associated with better outcome on CRC patients (Wang et al., 2011). Married individuals had increased social support compared to unmarried cancer patients (Gomez et al., 2016). Moreover, married individuals had higher chance to be diagnosed in an earlier stage of disease and received early treatment . In addition, married individuals were involved in healthy lifestyles including healthy diets, exercise, participation in preventive measures like screening, better access to health insurance etc. Unmarried individuals had higher chances of non-adherence to prescribed medication as well (Zhang et al., 2017).

In a retrospective study by Wang et al. (2011) based on a SEER database which included 127,753 CRC patients diagnosed between 1992 and 2006 and examined the relationship between MS and 5-YS rate. In this study MS was categorized into various categories. The covariates used for this study included demographics, stage of cancer and surgery treatment. The researchers found out that early diagnosis of CRC was seen in married group compared to other groups (Wang et al., 2011). In addition, the married individuals received surgical treatment compared to individuals of other groups. The Hazard ratio with confidence intervals for men and women were 0.86 (0.82–0.90) and 0.87 (0.83–0.91) respectively when compared to single after adjusting for the covariates mentioned above. Therefore, it was concluded that married patients showed higher survival for both genders.

Similar result was concluded by Ching-Chieh et al. (2019) on the MS and 5-YS in individuals treated surgically for CRC. This study evaluates the influence of MS on survival of patients who were treated surgically for colon cancer. Data analysis was performed on 6 years of data from 2010 to 2015 from a cancer registry of the medical center. This data included 80 percent of married patients. The researchers found better outcomes in married individuals compared to single individuals. The analysis after further stratification for age greater than 65 years, female, advanced stage of tumor showed negative effect of unmarried status. Some of the limitations of this study were not adjusting for comorbidity status of the CRC patients, insurance status, education level, and change in MS during cancer.

Similarly, Gomez et al. (2016) studied a cancer cohort from California (not just s CRC specific) and established better survival outcome in married cancer patients compared to single. Interestingly, the study also showed that there is a greater protective effect of marriage on survival in men compared to women and protective effect of MS decreases with age. So, it would be interesting to see if MS is linked with survival in elderly CRC patients with PD.

Race and CRC Survival

Many studies had shown racial disparities in CRC survival (Coleman et al., 2008; White et al., 2017; Al-Husseini et al. 2018). Al-Husseini et al. (2018) mentioned that the difference in access to medical care may be one of the factors in CRC survival in race. CONCORD study by Coleman et al. (2018) also evaluated the racial difference in cancer survival on a total of 31 countries of five continents including the US. But this study

included various cancers like breast, colon, rectum, and prostate. Specific to the US, the survival was less in black compared to white population. When we compare the survival of black women was less than compared to white women. Survival in black was less when compared to the general United States population. Furthermore, survival for black women was less than the mean survival for the United States population. Similar patterns of less survival in black population were seen when evaluated specific to colon cancer and rectal in the United States. White et al. (2017) performed one of the largest studies showing the trend in 5-YS by stage and race in the United States for Colon cancer using CONCORD-2 data. In this study the researchers examined the survival trend in the United States based on 37 different states' cancer registries funded by CDC and SEER programs (White et al., 2017). The CRC patients included were invasive colon cancer diagnosed patients between the age range of 15 to 99 years diagnosed between 2001-2009 and followed up was done until December 31, 2009. The researchers estimated the survival of colon cancer on 2 calendar periods 2001-2003 and 2004-2009. The authors categorized the patients into five age groups and estimated the age standardized net survival. The study concluded that 5 years of survival has increased over the years for both black and white races when 2 calendar periods were compared. Black population has lower survival compared to whites in both calendar periods. The net survival in blacks was less compared to whites (White et al., 2017) and the result was similar in all states. The difference in survival rates between these 2 races was lower in 2004-2009 compared to the year 2001-2003. Also, the survival was poor in black race in all stages of disease compared.

Al-Husseini et al. (2018) studied the racial disparities in CRC survival among races in the United States using the SEER database for patients diagnosed between 1973 and 2014.The result showed that there was significant difference in CRC OS and cancer specific survival between races. Black had the worst survival compared to other races (Al-Husseini et al., 2018). Asian and Pacific Islanders have the best cancer specific survival but there was not much difference in survival when whites and American Indians/Alaska Natives were compared. When the adjustment was done for covariates like age, marital status, stage, grade, surgery black population had the worst OS and cancer specific survival, and Asian or Pacific Islanders had better survival than whites and blacks. Therefore, race is one of the factors that needs to be considered in survival of CRC.

Socioeconomic Status and CRC

In general, socioeconomic status is considered as one of the factors that has an association with mortality and incidence of cancer. The socioeconomic factor was also one of prognostic factors for CRC (Zhang et al., 2017). There were various determinants of behavioral health of cancer like dietary factors, smoking, alcohol consumption, occupational status, environmental exposure that are linked to socioeconomic status (Singh, & Jemal, 2017). Singh and Jemal (2017) conducted extensive study on US mortality, incidence, survival of all cancers from 1950 to 2014 and socioeconomic status and racial/ethnic disparities. The researchers found that cancer patient survival was less in the most disadvantaged group. The study reported that during the period of 1988 to 1999, the 10-year survival rate for patients with all cancers combined in the most

disadvantaged group was 41% whereas 60.4% reported for the least-disadvantaged group. To CRC, the result showed the 10-year survival rates as 49.2% for most deprived groups versus 61.5% for the least deprived group. One of the limitations mentioned in this study was that socioeconomic status indices were used at country level rather than neighborhood level . The influence of socioeconomic status on cancer mortality might be underestimated in this study. In addition, there might be correlation between the indices for socioeconomic status. Statistical tests used in this study were least square, risk ratios, rates, log -linear and Cox regression. Since the large number of tests and multiple comparisons have been done in the study and the level of significance has not been corrected has been mentioned as a limitation of this study.

Zhang et al. (2017) did study using data from SEER diagnosed with CRC between 2007 and 2013 and assessed the effect of socioeconomic status on OS of CRC. Various lifestyle information, comorbidities, receipt of chemotherapy as treatment, change in insurance status over the period factors were not considered for analysis. Multivariate analysis showed that the lowest education group had poor prognosis. One of the limitations was the generalizability. Socioeconomic study in this study was census county level measures instead of individual level socioeconomic levelSimilar study showing the impact of socioeconomic status on survival of CRC based on data from Netherland specifically on patients who did curative surgery for CRC also showed an improved OS with increase in socioeconomic status (Van den Berg et al., 2019).

Tumor Side and CRC Survival

Colon cancer is developed from the epithelial tissue of the colon either on the right side or left side of colon (Baran et al., 2018). The right vs. left side of the colon has different embryological origins, so the tumors exhibit different histology. Right sided and left sided colon do have different immunology and gut flora (Petrelli et al., 2017). Even the morphology of right vs. left sided tumor is different in that right sided CRC is difficult to be detected early in colonoscopy due to its flat morphology (Baran et al., 2018). Moreover, right-sided CRC have advanced and bigger tumors. Even the metastases differ depending on the side of the tumor where the left sided CRC tends to metastasis in liver and lung whereas right sided CRC metastasis on the peritoneal cavity. Moreover, there are differences in immunologic response to tumor depending on right and left side of tumor origin (Petrelli et al., 2017). There has been a debate going on whether right sided or left sided cancer has better prognosis and whether it is a prognostic factor (Warschkow et al., 2016). Warschkow et al. (2016) mentioned that some of the past studies suggested right side tumors have poor prognosis, but other studies showed no contradictory finding that there is no difference between locations of tumor. Warschkow et al. (2016) evaluated data collected from the SEER database between 2004 to 2012. Results in univariate analysis demonstrated that right -sided cancer had worse OS than left sided but with propensity score matching, specifically for stage I and stage II cancer the OS was better for right sided tumor. For stage III cancer, the survival for right and left sided colon cancer were similar. Petrelli et al. (2017) did a meta-analysis on 66 published papers and evaluated 1437846 colon cancer patients who had survival data available.

This study was the first meta- analysis showing the site of colon was one of the prognostic factors in early and advanced colon cancer. The researchers showed that there was 19% reduced risk of death when the tumor arised from the left side. Another study was done based on data from Mayo clinic on adenocarcinoma patients from 1972 to 2017 showed higher survival in patients with tumors on the left side (Wang et al., 2019).

Overview of Methodology-Related Literature

There are various studies mentioned in this literature review section for this study that compared the mortality in cancer patients with and without mental illness. Study by Musuuza et al. (2013) did a cross-sectional study to evaluate mortality in cancer patients with and without mental illness. There are many papers with cohort studies used as methodology (Yang et al., 2018; Ko et al., 2019). Some papers did meta-analysis and literature search review (Saracino et al., 2015; Maajani et al., 2019; Yang et al., 2017; Joachim et al., 2019). Das-Manushi et al. (2017) did a longitudinal study and concluded that mortality is higher in patients with mental illness compared to the general population irrespective of ethnicity. In retrospective study we look for exposure and follow to look for outcome (Setia, 2016). So, to see if there is a gender biased in the 5-YS rate in elderly CRC patients who had PD the exposure is elderly CRC patients with PD and the outcome is 5-YS. This retrospective study will use de-identified individual level data. This method of retrospective study is relatively less time consuming, inexpensive compared to case control or prospective cohort study and easy to conduct. One of the limitations of this retrospective study is that the exposure and outcome has already been collected before the initiation of the study. There is a chance that the measurement may not be accurate.

Most of the papers mentioned in this literature review did secondary data analysis. Data is collected from various sources in these papers. Liu et al. (2019) did analysis using WHO mortality data. A lot of studies like studies by Lin et al. (2016), Quirt et al. (2017), Coleman et al., (2008), Afshar et al. (2018) used cancer registry data. Some of the papers used national cancer registry data whereas a study by Gomez et al. (2016) used state specific cancer registry like California Cancer Registry. Study by Cunningham et al. (2015) linked cancer registry data to hospital data. Similarly, Ching-Chieh et al. (2019) used registry data and linked it to Electronic Health Records (EHRs). Instead of registry, only hospital data has been used in studies like a study by Kotake et al. (2016). Yu et al. (2020) did secondary analysis on NHANES dataset. There are various studies in this chapter that did secondary data analysis from SEER database. Al-Husseini et al. (2018) used SEER database to review CRC patients between 1973 and 2014. Some studies have used more than one database like study by Singh, and Jemal (2017) used the national mortality database and SEER database. Similarly, a study by Zhang et al. (2017) which investigated the impact of socioeconomic status on OS (OS) of CRC patients also used the SEER database. Warschkow et al. (2016) also did a study to compare if prognosis is better in right sided versus left sided localized colon cancer using SEER database. Moreover, a study by Wang et al. (2011) examined MS and CRC using the SEER database. Many studies have used secondary analysis of SEER databases. I will use the SEER- ML database and will do secondary data analysis for this study. The advantage of secondary data analysis is that the data has already been collected so we do not need to spend time on collecting data in secondary data analysis and it is cost effective. In the

secondary data collection method, we utilize existing data to test new research questions and generate new knowledge. SEER database collects information about cancer since 1973 and is funded by NCI which has data from 17 cancer registries that covers 30% of the United States population (Daly, & Paquette, 2019). Collaboration of NCI, SEER registries, and the Centers for Medicare and Medicaid Services (CMS) collaborated and linked SEER data to Medicare data (Daly, & Paquette, 2019). The advantage of SEER-ML database is the high quality of data as NCI conducts the quality control program annually to ensure good quality of data. The limitation of the SEER- ML database is that cancer data cannot be obtained for individuals who are not enrolled in Medicare.

Overview of Statistical Analysis Plan-Related Literature

Most of the papers mentioned in this chapter used KM estimates to calculate the survival and LR test to evaluate if there were significant differences in survival between various groups. Multivariate Cox regression models were also used in many papers. In the study by Afshar et al. (2018) to see the difference in cancer survival by gender, Pohar-Perme method for survival probability was mentioned in the statistical section of the paper Age standardized survival was also calculated and compared between men and women. Year of diagnosis, age, gender, age of diagnosis was some of the covariates mentioned in the study. The goodness of the model was checked using Pearson Chi-square ($\chi 2$) statistics and the residual degrees of freedom. Pohar Perme estimator for survival was also used by White et al. (2017) in extension of CONCORD-2 study. White et al. (2017) also used life table methods using Pohar Perme estimator method. Coleman et al. (2008) used a life table method for estimation of survival using Pohar Perme

estimator. Comparison of proportions were done using chi-square test and survival analysis done using KM method, and association of disease and treatment related factors were evaluated using CP hazards regression models in study by Quirt et al. (2017) which aimed to see gender difference in CRC survival. As statistical methods, Hazard ratio (HR) and CI were analyzed for continuous variables, and chi square test was used for study by Yang et al. (2017) to evaluate gender difference in CRC. Sensitivity analysis was done to investigate any potential bias (Yang et al., 2017). Ching-Chieh et al. (2019) used KM method to calculate the 5-YS in married versus single surgically treated colon cancer patients. In addition, univariate and multivariate Cox regression models were used to find a significant independent variable to determine if MS has benefit in patients with colon cancer. Al-Husseini et al. (2018) used unadjusted KM test and multivariate covariate- adjusted Cox models as statistical tests to calculate the overall and cancer specific survival for different races. to compare the OS by gender, subsite, age group, geographical zone and stage of diagnosis LR test was done in study by Joachim et al. (2019).

To investigate the factors associated with survival outcomes Hazard ratio with confidence interval can be used as calculated. The researchers identified the prognostic factors for OS with the Multivariate Cox model in study by Joachim et al. (2019). The multivariate Cox model evaluated the variables sex, tumor location, age category, geographical location, and stage and the variable with p value less than 0.20 in the univariate model were used in the multivariate model. Wang et al. (2011) calculated the hazard ratio and confidence interval to see if MS is associated with survival outcome.

The researchers calculated the Hazard ratio confidence interval for male and female separately when compared to single after adjusting for covariates (Wang et al., 2011). Zhang et al. (2017) also did multivariate analysis to evaluate if socioeconomic factor is a prognostic factor in CRC. Van den Berg et al. (2019) also performed CP hazards regression to analyze the effect of socioeconomic status on OS. In addition to CP hazards regression, Van den Berg et al. (2019) also did a LR test to see if there is significant difference in survival between income quartiles. A similar approach of Hazard ratio and confidence interval was used by Gomez et al. (2016) when they compared the survival outcome in married versus unmarried cancer patients. In addition, to rule out the potential correlation between the categorical variables, pairwise Pearson's correlation method was used in study by Yu et al. (2019). Yu et al. (2019) did univariate Cox models and multivariate Cox models for the variables to calculate the hazard ratio. Petrelli et al. (2017) did meta-analysis to evaluate the survival in left sided and right sided colon cancer using pooled Hazard ratio with 95% CI from multivariate analysis. Univariate and multivariate Cox regression were also used as a statistical model in study by Warschkow et al. (2016) with the aim to see if there was difference in survival in right sided versus left sided tumor. To investigate whether right sided tumor or left sided tumor is an independent prognostic factor, the researchers did propensity score matching (Warschkow et al., 2016). Other study that used KM, LR test, Cox regression model is study by Wang et al. (2019).

Based on the review of statistical methods used in the reviewed papers, KM estimate will be used as a statistical method to calculate the 5-YS rate of elderly CRC men with PD and elderly CRC women with PD.

Most of the study used the 2-sided alpha, p value of <0.05 as level of significance and 95% confidence interval. Hazard ratio of <1 signifies better survival. KM is one of the best and simplest tests that would measure the survival of subjects (Goel, Khanna, & Kishor, 2010). In this method if any subject loss to follow up, withdrew from study or may not experience event or death are censored for analysis of survival (Goel et al., 2010). One of the good things about this method is that it takes into consideration some patients who are lost to follow up as they do provide some survival informationGoel, et al., 2010). To evaluate if there is a difference in survival between 2 groups, LR tests can be done. Edward L Kaplan and Paul Meier in 1958 innovated the KM method for calculation of survival time (Stalpers, & Kaplan, 2018). KM is used widely in epidemiology and medical studies, especially oncology. It is also incorporated in various statistical software packages. Before the KM method, the life table method was used. Though the KM method was published in 1956, it was not recognized until 1969. In 1969, Gehan published a paper and introduced a term KM product- limit estimator. Life table method was considered for large samples and clinical studies mostly do not have large sample size of studies started using KM method for survival analysis. The KM method was developed for non-biological purposes but later started to be used on oncological studies widely. LR test shows the difference in survival times but does not check for the effect of independent variables (Goel et al., 2010). To test the effect of

independent variables, the CP hazard model is used. CP hazard model is like multiple regression and calculates the Hazard ratio. One of the drawbacks of the KM method is that it is limited to one factor analysis rather than multivariate analysis (Jager et al., 2008).

Summary and Gap in Literature

In this chapter, I did a literature search in Colorectal cancer and pre-existing depression. CRC is a common cancer with staging of cancer based on TNM staging. The signs and symptoms of CRC depend on progression of the cancer. There were various demographics, socioeconomic, hereditary, environmental risk factors for CRC. Treatment relies on cancer stage at diagnosis. The literature search strategy which included databases, keywords used for the search were mentioned in this chapter. There had been a decreasing trend in incidence and mortality of CRC however developed countries have higher incidence and mortality. In addition, men have more risk of developing CRC. Depression was listed as a common issue with high prevalence in women and a common comorbid condition in patients with CRC in the United States. Various studies like CONCORD and CONCORD -2 were done to evaluate the CRC survival across the globe and the United States, respectively. Further study was done by White et al. (2017) on CRC patients in the United States and did a state wise comparison. Study by Štor et al. (2019) investigated to understand the trend of CRC survival over the period. Age was considered as one of the prognostic factors in study by Itatani, Kawada, and Sakai (2018). Liu et al. (2019) did a study and reported that mortality rates increased with age. Like age, higher initial stage at diagnosis of CRC showed more risk (Joachim et al., 2019).

Study by White et al. (2017) showed poor survival in black individuals than whites. Various studies mentioned about the effect of MS on CRC survival (Wang, Wilson, Stewart, & Hollenbeak, 2011; Ching-Chieh et al., 2019). Socioeconomic status also had association with CRC mortality and incidence (Singh, & Jemal, 2017). There had been a gender difference in CRC survival as women had better survival per various studies (Afshar et al., 2018; Quirt et al., 2017). Meta-analysis by Yang et al. (2017) and as well as study by Kotake et al., (2016) concluded gender as predictor in CRC survival. As mentioned above, there were various factors that affected CRC survival. CRC patients did have many comorbid conditions among which preexisting depression was one of them. Depression affected cancer survival and cancer related mortality (Pinquart and Duberstein, 2010; Das-Manushi et al., 2017). Similarly, a study by Cunningham et al. (2015) also concluded that survival was better in cancer patients with no mental illness. In addition, study by Yang et al. (2018) concluded that the prognosis was not good when patients had pre-existing mental illness. Several other studies had similar findings. More attention should be given to patients with serious psychiatric illness with CRC as it has been seen that the stage of diagnosis is higher in patients with preexisting illness compared to CRC patients without preexisting mental illness (Kaerlev et al., 2018). There have been less studies on gender difference in CRC patients with PD. It would be interesting to see how gender and depression influence survival of CRC patients as survival of CRC is better in women. In addition, I can also see if race, marital status, socioeconomic status, initial stage of diagnosis affects the CRC survival between men and women in CRC patients with PD.

Next, chapter 3 will include comprehensive information regarding the study design and methods that were employed to complete the research. Aims, hypothesis along with the statistical test that will be used to get the answer will be stated in an elaborate way. The study population will be defined along with the information for secondary data source, data dictionary, data analysis plan, ways to collect data, direction of statistical test, significance level as well as the statistical analysis software to be used.

Chapter 3: Research Method

In this chapter, I detail the research methodology for the study. I also will include information on the design and the rationale for choosing the study design. The datasets to be used from SEER-Medicare database and the justification are also explained in this section. I also identify the independent variables, covariates, and the dependent variable in the study. The details in the code book and location of the needed variables are also discussed. Details on study population, inclusion and exclusion criteria, power analysis are mentioned. The use of SPSS software and the justification for using the software is presented. The statistical methods I used are also explained in detail. Moreover, the internal, external, construct, and statistical validity, and ethical practices of using SEER-Medicare database are presented.

Research Design and Rationale

This was a retrospective quantitative study to investigate if there was gender disparity in 5-YS in elderly CRC patients with PD. The dependent variable was the 5-YS. The independent variables were gender, initial diagnosis stage, and MS at diagnosis. The covariates were race, ethnicity, socioeconomic status, and primary site of tumor. Secondary data analysis was done using data from SEER-Medicare database. Since this was retrospective cohort study, manipulation was not done in the study. As I analyzed an already existing database, it took less time compared to collecting data on my own. The SEER-ML database is a large database with cancer cases of patients older than 65 years old and the study can be generalized to an elderly population. All the research questions

were addressed using quantitative data either in continuous or categorical form. Therefore, quantitative study was an appropriate study design for this study.

Time and Resource Constraints

As I used the SEER-ML database and I did not collect the data, the time constraint was less compared to study design where primary data needed to be collected. Even if the analysis was done on archival data, some time constraints still existed as data was requested via an application. SEER-Medicare Cancer Data file and Chronic condition file were requested. Data files were requested for CRC cases diagnosed with CRC during the year 2006 to 2012. The chronic condition file had data up to 2017. I evaluated if a patient was dead or alive after 5 years of diagnosis. Since the study evaluated patients' 5-YS, data up to 2012 was requested. After the proposal was completed and approved, I got institutional review board (IRB) approval from Walden University. Then I sent an application to the SEER Medicare oversight committee after approval and completion of the data user agreement from the dissertation chair to request data. It took approximately 4 weeks to get approval from SEER-Medicare oversight committee and I got data in 3 weeks. Training for data extraction and compilation process was completed during the data acquisition time.

Methodology

In this chapter, I define the target population for the study and the procedure that will be used to sample the population. Secondary data analysis was done SEER- ML database so no new data was collected.

Population

The primary research question of the study was to see if there was gender difference in 5-YS rate in elderly CRC patients diagnosed with PD. The population included elderly patients diagnosed with CRC between January 2006 and December 2012. SEER-Medicare database had patients 65 years of age and older. This study included CRC patients with PD of age greater than or equal to 67 years. Both men and women were included in this study. The patients who were diagnosed with depression after diagnosis of CRC were excluded from the study as this study included elderly CRC patients with PD prior to diagnosis of cancer.

Sampling and Sampling Procedures

Data from CRC patients (SEER-Medicare Cancer Data file) and chronic condition (CC) file from 2006 to 2012 was requested from SEER-Medicare. Only patients who had diagnosis of depression prior to CRC diagnosis were included in the study. After evaluating the inclusion and exclusion criteria, the final subset of patients was identified.

Sample Size Estimation

As per Majek et al. (2013), 5-YS was 64.5% in women and 61.9% in men.

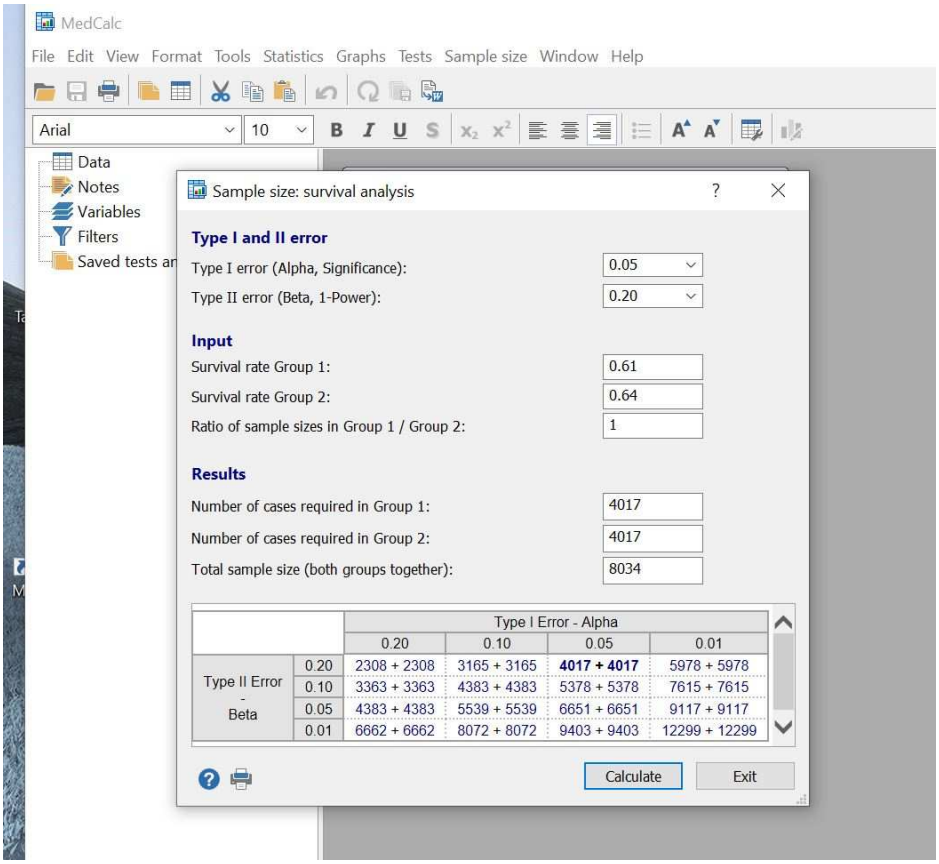
- Probability of Type I error (α)= 0.05
- Power (1 - β) = 0.8
- Ratio (R) = Sample size for men / sample size for women= 1 (equal size in two groups)
- Time= 5
- 5-YS rate in men (SR1) = 0.61

- 5-YS rate in women (SR2) =0.64

I used MedCalc software to calculate the total sample size needed as seen in Figure 5.

Figure 5

MedCalc Calculation for Total Sample Size



The total sample size estimated for alpha=0.05, 80 % power was 8,034 patients.

Research Instruments

As secondary data analysis was completed using the SEER- ML database, no instrument was used for this study. The data for this study came from the cancer registry and Medicare chronic condition flag. Data from cancer registries and mortality data were

linked appropriately and de-identified. SEER- ML database had diverse race/ethnicities in the United States.

Gaining Access to the Data

There were two main files that were required for this study. The SEER data was a part of SEER-Medicare. SEER-Medicare This file had one record per individual which matched Medicare enrollment records (NCI, 2020-f). Demographics information was also available in this file. There were separate files available for geographically-based socioeconomic information as well. The chronic condition file was required for the study. This file contained yearly, mid-year and ever flags for 27 conditions starting in 1999 (NCI, 2020-d). Depression was one of the chronic conditions listed (CMS, 2021-a). Based on a claim-based algorithm (except for HMOs), the flag was generated in SEER-ML database (NCI, 2020-d). As per the cost calculator in NCI website, the estimated cost for the cancer data file and Chronic condition flag file was $375 (NCI, 2021-d). To request data for restricted variables, I applied, and my application was reviewed by the NCI (NCI, 2021-b). The restricted variable data files were stored separately than other SEER-Medicare. IRB approval was obtained before obtaining SEER-Medicare data. I complied with Health Insurance Portability and Accountability Act (HIPAA) regulations and signed a data user agreement before obtaining the data (NCI, 2019-g). I obtained a SEER-Medicare database when I signed the data user agreement (DUA). The DUA accompanied a detailed data storage plan. The data was stored protected with password. Special permission was needed if the data was to be stored in removable media. If the data was stolen or its data security had been compromised, SEER - Medicare had to be

informed within 24 hours of data breach (NCI, 2019-d). The SEER- ML data delivered was stored securely in a locked file cabinet where only I have access. If I plan to submit the paper for publication, the manuscript will be submitted to NCI for approval and only after approval can it be submitted for publication (NCI, 2019-i). The contact person for SEER-ML database was as follows.

Elaine Yanisko

Information Management Services, Inc.

SEER- ML database is not a public use database so I got approval for the specific research question. The entire dataset was not requested. Therefore, when applying, the years of data were specified in DUA. Data was encrypted in a thumb drive with a protected password. The files were compressed using GUNZIP compression utility. Data could only be obtained for the same project three times (NCI, 2020-e). If more data was needed, a new application would have to be submitted. The data requested has to be retained for 5 years with 1 year of extension. If the data was needed for more than a 10-year period, a new application would have to be submitted. Once the data extends the retention period, it will be destroyed . Once the study has been completed, the data should be destroyed . All the raw files along with the analytic files, back up files and original period would be destroyed. Then SEER - Medicare contact would be informed about the destruction of files.

SEER-Medicare database cannot not be released outside of the United States (NCI, 2021-h). I did not request all the data available. There were five important documents to request data. The first file was the application form which needed

information like hypothesis and research question, description of study, explanation on how the key variables were identified, list of files to be used in the study, timeline for completion and references. I received a mock application as well. The DUA consisted of a list of all staff who had access to data. I submitted the application. The third important document was IRB approval from Walden University. The fourth document was the request for restricted variables. The fifth document was the funder letter. Since this study did not have any funding, the fifth document was not needed. I submitted all the documents for one request in one email. SEER-Medicare reviewed all the requests and approved the data purchase. It took 6- 8 weeks for the approval process. Once approved, it took 2- 4 weeks for processing of data. The total predicted estimated time for getting data for this study was 10 weeks.

Data Analysis

The following section describes the data analysis for the study. Detailed information on the software package used, the methods to obtain data in appropriate format, data cleaning procedure, and statistical methods to be used are provided.

Software

To make the dataset ready for analysis, I compiled the variables that are needed and created a database. The cohort in the study were elderly CRC patients with PD. The cancer file was linked with a CCW file to find the CRC patient with PD. I did a query, selected the cases based on inclusion and exclusion criteria. IBM SPSS Statistics 24 was used for descriptive and inferential statistics. Survival analysis and Cox regression analysis were the inferential statistics completed in SPSS.

Study Data Source

Secondary data analysis was done for the study. SEER is a large population-based data source (NCI, 2019-b). The SEER data is linked to another large population-based data source, Medicare. The SEER-ML database includes information about clinicals, demographics, and death information of cancer patients along with the Medicare claims of eligible patients (NCI, 2019-b). With the collaborative effort of NCI, SEER registries and CMS, SEER data was linked with Medicare claims (NCI, 2020-c). The linkage was not done from investigator side. The individual identifiers from each file were matched with Medicare master enrollment file. The linkage is done every 2 years. The method of linkage was based on social security number (SSN) and date of birth. If SSN was missing or incomplete the linkage completed based in first name, last name, and full date of birth.

Information regarding the patient's diagnosis and related information was available up to year 2017 (NCI, 2021-c). Claim files are available up to 2018. CC flag files are one of the important Medicare files for this study. CC flag files are available from 1999 – 2018. SEER-Medicare has cancer cases diagnosed in patients 65 years and older.

From 1991 to 2017, there were 21,041-21,052 cases of CRC in patients 65 years and older at time of diagnosis (NCI, 2021-f). Since I requested from year 2006 to 2012, the approximate number of cases of CRC was 85,705. Out of these patients only the patients who were older than 67 years old and diagnosed with PD were used in analysis. The information related to hospitals, providers and patients was encrypted (NCI, 2021-b).

Some other information like oncology diagnosis and Alaska native tumor registration data were also restricted variables in the SEER-Medicare database.

Setting and Sample

The SEER- ML database contains patients who are eligible (65 years old) for Medicare and diagnosed with cancer. The study population included individuals older than 67 years of age diagnosed with CRC who had PD. To identify patients with PD, I subset the CRC patients from CC flag documentation maintained by the CMS. There was a special claim-based algorithm to determine if a patient had a chronic condition or not. Then, if any cancer patients had treatment for that condition and had claims for Part A/B-FFS coverage during a special period then the patient was flagged as having that CC flag file. From the CC file, I got the date of first occurrence of chronic condition. This date was always after the Medicare eligibility date. A patient is eligible for Medicare at the age of 65 years (Commonwealth of Massachusetts, 2023). But to have a CC flag, there is some reference period after the patient's enrolled date. Therefore, the patients diagnosed with CRC at 65 years would not have an accurate CC flag for depression. To avoid this situation and have a correct CC flag, I only included patients of age 67 years and older. Data was requested from SEER-ML database from the period of 2006 to 2012. The most updated SEER-ML database has data up to the year 2017. I requested data up to 2012 as I was evaluating the 5-YS.

Inclusion Criteria

The following criteria was used to determine whether a patient was included in my study:

- Patients 67 years or older diagnosed with CRC listed in the SEER-Medicare database
- CRC patients had PD
- CRC cases first diagnosed between 2006 to 2012

Both men and women were included in the study. The data was limited to those living in the United States, but I did not limit the scope to any particular state or region of the country.

Exclusion Criteria

The following criteria were used to exclude files from the data analysis:

- CRC patients in SEER-Medicare database younger than 67 years were excluded from the study.
- CRC patients diagnosed with depression after diagnosis of cancer

Process of Determining the Population for Analysis

I used the following process to determine the necessary population for data analysis.

- Obtained data from SEER-Medicare of CRC patients diagnosed between 2006-2001
- Subset the patients who were older than 67
- Subset patients whose date of first occurrence of depression was before date of diagnosis.
- Checked for inclusion and exclusion criteria

The SEER-Medicare cancer file contained the CRC patients who are age 65 or older. The month of diagnosis (Variable name: modx1-modx10) and year of diagnosis (Variable name: yrdx1-yrdx10) were available in the file. I imputed the midday of month for the day of diagnosis of CRC for these patients. For example, if the month of CRC diagnosis for a patient was March and the year of diagnosis available was 2010, the date of diagnosis for that patient was considered as 15MAR2010. The CC file contained information on first occurrence of depression (Variable name: DEPRESSION_EVER) with date shown in YYYYMMDD format. For the CRC patients in Medicare file, I included only those patients in study whose date of first occurrence of depression was before the imputed date of diagnosis.

Research Questions and Hypotheses

RQ1: What is the association between gender and 5-YS rates of elderly CRC patients with PD?

$H_0$1: There is no association between gender and 5-YS rates of elderly CRC patients with PD.

H_a1: There is an association between gender and 5-YS rates of elderly CRC patients with PD.

RQ2: What is the association between initial diagnostic stage (Stage I, Stage II, etc.) and 5-YS rates of elderly CRC patients with PD?

$H_0$2: There is no association between initial diagnostic stage (Stage I, Stage II, etc.) and 5-YS rates of elderly CRC patients with PD?

H_a2: There is an association between initial diagnostic stage (Stage I, Stage II, etc.) and 5-YS rates of elderly CRC patients with PD?

RQ3: What is the association between MS at diagnosis and 5-YS rates of elderly CRC patients with PD?

$H_0 3$: There is no association between MS at diagnosis and 5-YS rates of elderly CRC patients with PD?

$H_a 3$: There is an association between MS at diagnosis and 5-YS rates of elderly CRC patients with PD?

RQ4: What is the association between gender, initial diagnosis stage and MS at diagnosis and 5-YS rates of elderly CRC patients with PD after adjusting for race, ethnicity, socioeconomic status, primary site of tumor?

$H_0 4$: There is no association between gender, initial diagnosis stage and 5-YS rates of elderly CRC patients with PD after adjusting for race, ethnicity, socioeconomic status, primary site of tumor.

$H_a 4$: There is an association between gender, initial diagnosis stage and 5-YS rates of elderly CRC patients with PD after adjusting for race, ethnicity, socioeconomic status, primary site of tumor.

Study Variables

The independent, covariates and dependent variables for the study are mentioned in this section below. All the variables were available in SEER-Medicare cancer file in SEER-Medicare database. SEER-Medicare cancer file is a file that includes information of patients in SEER database that has Medicare enrollment records (NCI, 2020-f). Since

the SEER database and Medicare database is linked, the patients 67 and older was only included in this study.This file had 1 record per patient (NCI, 2020-b). These files had unique SEER case ID number with a cancer diagnosis which was merged with another file. These independent, dependent and covariates were used to answer the research question.

Independent Variables

The first research question in the study is to see if there was gender disparity in 5-YS in elderly CRC patients with PD. Gender was the first independent variable and was identified from the SEER-Medicare cancer data file (variable name is "sex"). In this Cancer file, Male was coded as 1 and Female as 2. The initial stage of diagnosis was another independent variable. The third independent variable was marital stage at diagnosis. The details of independent variables are listed in Appendix A.

Confounder Variables

The covariates in this study were socio economic factors, primary site of tumor (laterality) and race. More than one variable was needed from SEER-Medicare database and some restricted variables were requested for socioeconomic condition. Socioeconomic condition was derived, and the detail is described in "Data cleaning" section below. The primary site of tumor for CRC will be either left or right side of the origin of tumor. In SEER- ML database the primary site where the tumor originated is coded with International Classification of Diseases for Oncology, Third Edition (ICD-O-3) for topography codes.

Dependent Variables

Five-years survival was the dependent variable for the study. It was a categorical variable with values 1 or 0. SEER-Medicare Cancer file had variables of year of diagnosis (Year_of_diagnosis) and month of diagnosis (Month_of_diagnosis). The year and month of diagnosis variable were combined, and the day was imputed. The new variable for date of diagnosis was in format YYYYMMDD. There was another Medicare Date of Death (SEER_DateofDeath_Month, SEER_DateofDeath_Year) that provided the information of the date of death of an individual. I calculated the difference in years between death date and date of diagnosis variable. If a patient was alive from 5 years of diagnosis, then coded as "1" and if a patient was not alive at 5 years of diagnosis, then coded as "0".

Data Preparation

SEER-Medicare cancer data file and Chronic condition file were used for analysis for this study. SEER-Medicare cancer data file that contained all the information on demographics and information on cancer. Only the cancer patients whose date of first occurrence of Depression on chronic condition file was before the date of Diagnosis of CRC were included. For the survival analysis, the event of interest is death. If a patient was not alive at time point 5 years from date of diagnosis, then the patient was coded as 0. If the patient was alive at 5 years of CRC diagnosis it was coded as 1 and was censored when doing survival analysis. The time the patient was alive from the date of diagnosis was calculated. For example: if a patient had CRC and died after 2 years of diagnosis then the survival time was 2 years. The survival time was censored at 5 years of CRC

diagnosis. For example: If patient A had CRC in 2007 and was still alive then the survival time for the patient would be 5 years. This survival time was important to calculate the median survival time. Appropriate coding was done for the variables.

Data Cleaning Procedure

For analysis single and unmarried were grouped together. Similarly, separated, divorced, widowed were grouped together, blank and unknown were grouped together. For descriptive statistics, the initial stage of diagnosis, I grouped stages like Occult, Unknown, Stage 0 and Blank stages together. Later for inferential analysis only stages I to IV were included. For Primary site of tumor ICD-O03 codes were followed. If the ICD-O-3 site codes were 18.0–Cecum, 18.2–Ascending colon, 18.3–Hepatic flexure of colon, and 18.4–Transverse colon then it was coded as right sided. If the ICD-O-3 site codes were 18.5–Splenic flexure of colon, 18.6–Descending colon, 18.7–Sigmoid colon, and 19.9–Rectosigmoid, and C20.9-Rectum, NOS then it was coded as left sided. Socioeconomic status (SES) variable was not directly available in the SEER-Medicare dataset. Based on census tract data- per capita income and median household income within each state, neighborhood poverty level and patient's enrollment in Medicaid or state buy program were used to group a patient among one of the 4 Socioeconomic classes. Poor-SES, Near-Poor-SES, Middle SES, and High SES were the 4 groups. If a patient was enrolled in Medicaid regardless of census tract residence was categorized into poor SES. If a patient was not enrolled in Medicaid but lived in census with lowest quartile of PCI and HHI and higher quartile of poverty, then the patients was categorized into Near-poor SES. If a patient was not enrolled in Medicaid, does not live in the lowest

quartile of PCI and HHI and does not live in highest quartile of poverty then categorized into Middle SES. If a patient was not enrolled in Medicaid, lives in the upper quartile of both PCI and HHI and lives in lowest quartile of poverty then grouped into High SES.

For the survival analysis, recoding was done to get the time variable, censor variable. The time the patient was alive from date of diagnosis was calculated. If a patient was alive greater than 5 years, then the patient was censored at 5 years. If the patient was alive from 5 years of diagnosis, then coded as 1. If a patient was not alive at 5 years of diagnosis, then coded as 0. Please refer to Appendix B for more detail on data cleaning procedure.

Statistical Test

The statistical test section will describe the statistical tests used for the research question of the study. The interpretation of statistical tests will be discussed as well.

Descriptive Statistics

Descriptive statistics were used to understand and summarize the data in a meaningful way for better interpretation. SPSS was for descriptive statistics. For continuous variables measures of central tendency like mean were presented. For categorical variables, frequency and percentages were calculated. Gender, race, socioeconomic status, primary site of tumor, MS at diagnosis was presented in frequency and percentage as shown below. Age would be presented as mean, median, standard deviation, minimum and maximum. Please refer to Appendix C for more detail.

Inferential Statistics

To measure the 5-YS l of subjects after CRC diagnosis in elderly patients with PD, KM (KM) estimates were used. KM estimates the fraction of subjects living for 5 years after CRC diagnosis (Goel, Khanna, & Kishore, 2010). The starting time for this study was the CRC diagnosis. Death was the event in this study. KM analysis considers scenarios when some subject may not experience death in 5 years of diagnosis. Those observations were labelled as censored observations.

The KM estimates are also called "product limit estimates". Please refer to Appendix D for more detail on KM survival analysis.

Inferential Statistics for RQ1

RQ1: What is the association between gender and 5-YS rates of elderly CRC patients with PD?

$H_0$1: There is no association between gender and 5-YS rates of elderly CRC patients with PD.

H_a1: There is an association between gender and 5-YS rates of elderly CRC patients with PD.

Cox proportion hazard model was used to see if there was association between gender and 5-YS for this research question. Additionally, KM estimates, LR test was done to see if there is difference in survival between men and women. The KM estimates would show the difference in survival at different time points between 2 groups. If the p value of LR test was less than 0.05, then there was a significant difference in survival in 2 groups. Please refer to Appendix E for more detail for table layout.

Inferential Statistics for RQ2

RQ2: What is the association between initial diagnostic stage (Stage I, Stage II, etc.) and 5-YS rates of elderly CRC patients with PD?

$H_0$2: There is no association between initial diagnostic stage (Stage I, Stage II, etc.) and 5-YS rates of elderly CRC patients with PD?

H_a2: There is an association between initial diagnostic stage (Stage I, Stage II, etc.) and 5-YS rates of elderly CRC patients with PD?

Cox proportion hazard model was used to see if there was association between initial stage of diagnosis and 5-YS for this research question. Additionally, KM estimates, LR test was done to see if there is difference in survival between men and women at each initial stage of diagnosis. Please refer to Appendix E for more detail for table layout.

Inferential Statistics for RQ3

RQ3: What is the association between MS at diagnosis and 5-YS rates of elderly CRC patients with PD?

$H_0$3: There is no association between MS at diagnosis and 5-YS rates of elderly CRC patients with PD?

H_a3: There is an association between MS at diagnosis and 5-YS rates of elderly CRC patients with PD?

Cox proportion hazard model was used to see if there was association between marital stage of diagnosis and 5-YS for this research question. Additionally, KM estimates, LR test was done to see if there is difference in survival between men and

women at each marital stage of diagnosis. Please refer to Appendix E for more detail for table layout.

Inferential Statistics for RQ4

RQ4: What is the association between gender, initial diagnosis stage and MS at diagnosis and 5-YS rates of elderly CRC patients with PD after adjusting for race, ethnicity, socioeconomic status, primary site of tumor?

$H_0$4: There is no association between gender, initial diagnosis stage and 5-YS rates of elderly CRC patients with PD after adjusting for race, ethnicity, socioeconomic status, primary site of tumor.

H_a4: There is an association between gender, initial diagnosis stage and 5-YS rates of elderly CRC patients with PD after adjusting for race, ethnicity, socioeconomic status, primary site of tumor.

CP hazard model was used for this research question. At first a bivariate analysis was done. Then the significant variables were included in the multivariate Cox regression model. Whatever model comes insignificant is removed from the multivariate model and reran again. The final model contains only the significant variables, and the HR and confidence interval were interpreted.

Result Interpretation

Interpretation of Kaplan-Meier's Estimates

The probability of survival when the patient diagnosed with CRC was 1.0 or 100 percent alive. At 1 year, the probability survival would be x.xx or X%. Similarly, it

would be interpreted for up to 5 years for both groups (men and women). The median survival time would also be reported as years for both men and women.

Interpretation of LR Test

If the p value of the LR test was less than 0.05 then I rejected the null hypothesis and concluded that there was a difference in 5-YS between men and women in elderly CRC patients with PD.

Interpretation of Cox Proportional Hazard Model

The multivariate CP hazard model was used to see the association between the independent variable gender, initial stage at diagnosis, race, socioeconomic status, primary site of tumor, MS on 5-YS. If there was a positive regression coefficient it meant higher hazard and if there was negative coefficient it would indicate lower hazard.
For Example: If HR=2 and the reference group was men, then women have 2 times higher death 2 times compared to men in elderly CRC patients with PD.

For those parameters if the p value was significant, then I would conclude that there was an association between the independent variables/ covariates and 5-YS in elderly CRC depressed patients.

Threats to validity

The SEER-Medicare database had been extensively used for survival of cancer research and tracked cancer over time. SEER-Medicare could be generalized to elderly population above 65 years old. There might be some possibility of error in data collection and SEER-Medicare was aware of it. The SEER-Medicare data was protected, and access was not provided to unauthorized people. Threats for internal validity could be caused by

confounding variables. So, for this study, the Cox regression model was used to see the association between independent and dependent variables. All the cases of CRC between 2006 and 2012 that fitted the inclusion criteria were included.

Ethical Procedure

IRB approval was taken from IRB at Walden University. Data Use Agreement (DUA) and approval letter for use of SEER-Medicare database was provided. DUA was signed by me. This data application was approved by the SEER-Medicare (NCI) committee. There was no conflict of interest. Since, I used secondary data analysis using the SEER-Medicare database, DUA was signed so there was not any informed consent taken directly from patients for this study. The patient information on SEER-Medicare data was protected. All possible measures were taken to protect the SEER-Medicare data. The data was on a password protected server. No data related to hospital and physician information were needed for the study. Data was not shared with anyone not associated with the study. Once the research was complete, the data including the original data, analysis data, and back up data would be destroyed.

Summary

To see the gender disparity in 5-YS of elderly CRC patients with PD, secondary data analysis was planned. This was a retrospective study which also evaluated if there was an association between 5-YS and initial stage of diagnosis, MS at diagnosis and various covariates like race, socioeconomic status, primary site of tumor. The study took the data from SEER- ML database. The data was requested for the years 2006 to 2012. In this chapter, I described methodology for the research study which included study design,

study population, method for data collection, statistical method, interpretation of results, issues of reliability and validity.

Chapter 4: Results

In this study, I investigated the difference in 5-YS rates for elderly CRC patients with PD based on gender. I also analyzed the gender disparity in 5-YS for initial stage of diagnosis and MS at diagnosis. In this chapter, I outline of selection of CRC subjects with PD. I present the demographics and report the descriptive studies for the independent variables and covariates. The inferential statistics KM estimates, LR test and Cox regression model are shown in this section as well. Below are the research questions along with null and alternate hypotheses.

RQ1: What is the association between gender and 5-YS rates of elderly CRC patients with PD?

$H_0$1: There is no association between gender and 5-YS rates of elderly CRC patients with PD.

H_a1: There is an association between gender and 5-YS rates of elderly CRC patients with PD.

RQ2: What is the association between initial diagnostic stage (Stage I, Stage II, etc.) and 5-YS rates of elderly CRC patients with PD?

$H_0$2: There is no association between initial diagnostic stage (Stage I, Stage II, etc.) and 5-YS rates of elderly CRC patients with PD?

H_a2: There is an association between initial diagnostic stage (Stage I, Stage II, etc.) and 5-YS rates of elderly CRC patients with PD?

RQ3: What is the association between MS at diagnosis and 5-YS rates of elderly CRC patients with PD?

$H_0$3: There is no association between MS at diagnosis and 5-YS rates of elderly CRC patients with PD?

H_a3: There is an association between MS at diagnosis and 5-YS rates of elderly CRC patients with PD?

RQ4: What is the association between gender, initial diagnosis stage and MS at diagnosis and 5-YS rates of elderly CRC patients with PD after adjusting for race, ethnicity, socioeconomic status, primary site of tumor?

$H_0$4: There is no association between gender, initial diagnosis stage and 5-YS rates of elderly CRC patients with PD after adjusting for race, ethnicity, socioeconomic status, primary site of tumor.

H_a4: There is an association between gender, initial diagnosis stage and 5-YS rates of elderly CRC patients with PD after adjusting for race, ethnicity, socioeconomic status, primary site of tumor.

Data Collection

I did secondary data analysis for the study. NCI's SEER-Medicare database was used for the study. For this study data for SEER-Medicare was requested for CRC (CRC) patients from year 2006 to 2012. The SEER-Medicare Cancer file, Chronic condition file and census tract file was requested from SEER-Medicare. The application to SEER-Medicare was submitted in December 2021. I received comments from the protocol reviewers in January 2022. The reviewer's comments were addressed, and the application was again sent to SEER- Medicare. The final application was approved, and I received data in March 2022. The raw datafiles were securely stored in secure server.

Cohort Selection

SEER-Medicare file had one record per subject per primary cancer site. Demographics information was available in this file from the year 2006 to 2012. The second file CC file contained the yearly, mid-year, and ever flag for 27 conditions including for depression. I received the CC files from the year 2006 to 2012. The raw data was read into SPSS. Once the files were read in SPSS, all the records from these CC files were stacked and I subset the subjects who had data available for yearly, midyear, and ever flag for depression available. Thereafter, I selected the subjects whose date of diagnosis of first ever depression was before the date of diagnosis of cancer to get the subjects who had PD. Total of 28,278 CRC patients were available who had PD and were 67 years old or older. Then the cancer information of these subjects was taken from the SEER-Medicare cancer file. Descriptive and inferential statistics were done on a total of 28,278 CRC patients with PD.

Results

Descriptive Statistics

The study cohort included 28,278 elderly (67+) CRC patients with PD diagnosed between 2006 to 2012. The demographics of this cohort are described in detail in Tables 2- 7. The mean age in this dataset was 80.16 years with range 68 years to 100 years. In this study cohort, about 68% are women and only 32% are men. As shown in Table 2, most patients (89.1%) were White. A total of 59 patients did not have known race which are excluded from the inferential analysis. Table 3 below shows the ethnicity of the study cohort.

Table 2

Demographics of Study Cohort by Gender and Race

	Women		Men		Total	
Race	*n*	%	*n*	%	*n*	%
White	17,058	88.7	8133	89.8	25,191	89.1
Black	1,561	8.1	603	6.7	2164	7.6
American Indian/Alaska Native	52	0.3	27	0.3	79	0.3
Asian or Pacific Islander	511	2.7	274	3.0	785	2.8
Unknown	41	0.2	18	0.2	59	0.2
Total (*n* %)	19,223	100	9,055	100	28,278	100

Table 3

Demographics of Study Cohort by Gender-Ethnicity

	Women		Men		Total	
Ethnicity	*n*	%	*n*	%	*n*	%
Non-Spanish-Hispanic-Latino	17,985	93.6	8456	93.4	26,441	93.5
Spanish-Hispanic-Latino	1,238	6.4	599	6.6	1837	6.5
Total (*n* %)	19,223	100	9,055	100	28,278	100

Table 4 describes the MS of the patients at diagnosis. For the inferential analysis separated, divorced, and widowed were grouped together. Most patients (35.4%) were under this category. Thirty-one-point-five percent of patients who did not have known and blank MS at diagnosis were excluded from the inferential statistics.

Table 4

Demographics of Study Cohort by Gender-MS at Diagnosis

	Women		Men		Total	
MS at diagnosis	*n*	%	*n*	%	*n*	%
Single/Unmarried	1,091	5.7	695	7.7	1,786	6.3
Married	3,310	17.2	3,235	35.7	6,545	23.1
Separated/Divorced/Widowed	8,160	42.4	1,852	20.5	10,012	35.4
Blank or Unknown	6,662	34.7	3,273	36.1	9,935	35.1
Total (*n* %)	19,223	100	9,055	100	28,278	100

Table 5 shows the descriptive statistics for the initial stage of diagnosis. Most were stage NA, Stage Occult, Unknown, Blank, or Stage 0. These records were not included in the inferential statistics. The percentage of patients were similar for Stages I-III. The Stage IV patients were less compared to the rest three stages.

Table 5

Demographics of Study Cohort by Gender-Initial State of Diagnosis

	Women		Men		Total	
Initial stage of diagnosis	*n*	%	*n*	%	*n*	%
Stage I	2,849	14.8	1,335	14.7	4,184	14.8
Stage II	3,135	16.3	1,457	16.1	4,592	16.2
Stage III	2,533	13.2	1,098	12.1	3,631	12.8
Stage IV	1,930	10	871	9.6	2,801	9.9
NA/Stage Occult/Unknown/Blank/Stage 0	8,776	45.7	4,296	47.4	13,070	46.2
Total (n %)	19,223	100	9,055	100	28,278	100

Table 6 shows that there were higher number of patients with right sided tumor (13,368) as primary site compared to left sided tumor (10,683) as primary site. 14.9% of cases were other than colon and rectum which were included in the analysis.

Table 6

Demographics of Study Cohort by Gender- Primary Site

	Women		Men		Total	
Primary Site	*n*	%	*n*	%	*n*	%
Right sided	9,689	50.4	3,679	40.6	13,368	47.3
Left Sided	6,727	35.0	3,956	43.7	10,683	37.8
Other than colon and rectum	2,807	14.6	1,420	15.7	4,227	14.9
Total (*n* %)	19,223	100	9,055	100	28,278	100

Out of 28,278 patients, 1,497 (5.3%) of patients did not meet the criteria for SES. These patients were excluded from inferential statistics. Out of 26,781 patients with SES data available, most of the patients were of near-poor SES. The percentages are similar when analyzed separately for men and women.

Table 7

Demographics of Study Cohort by Gender- Socioeconomic Status

	Women		Men		Total	
Socioeconomic Status (SES)	*n*	%	*n*	%	*n*	%
Poor SES	3,196	17.5	1,491	17.4	4,687	17.5
Near-Poor SES	7,471	41.0	3,534	41.3	11,005	41.1
Middle SES	1,747	9.6	821	9.6	2,568	9.6
High SES	5,802	31.9	2,719	31.7	852	31.8
Total (*n* %)	18,216	100	8,565	100	26,781	100

Total number of patients with missing SES was 1,497

Inferential Statistics

For the four research questions in this study, LR test and CP hazard were done.

RQ1

RQ1: What is the association between gender and 5-YS rates of elderly CRC patients with PD?

$H_0$1: There is no association between gender and 5-YS rates of elderly CRC patients with PD.

H_a1: There is an association between gender and 5-YS rates of elderly CRC patients with PD.

The 5-YS rate for women was 35.8% whereas 5-YS was 30.8% in men. I completed the LR test to determine if there was a difference in 5-YS between women and men. The survival distribution was statistically significant. Since the *p* value from Cox regression model is less than 0.05 (Table 8), I rejected the null hypothesis that there is association between gender and 5-YS rates for elderly CRC patients with PD. I concluded that there was an association between gender and 5-YS in elderly CRC patients with PD. The hazard ratio (HR) from Cox regression model showed the HR as 0.891(0.865, 0.919). Therefore, women who were diagnosed with CRC and PD had 10.9% less death within 5 years when compared to men who were diagnosed with CRC and PD.

Table 8

Inferential Statistics for RQ1

Variable	Women	Men	Total
Median Survival (estimate in yrs)	2.505	2.160	2.415

KM Estimates			
1 year	54.2%	51.2%	53.2%
2 years	46.9%	42.5%	45.5%
3 years	41.2%	36.2%	39.6%
4 years	35.8%	30.8%	34.2%
5 years	35.8%	30.8%	34.2%
Log- rank p value, Women Vs Men	**<0.001***		
HR (95% CI)[Men as reference	**0.891 (0.865, 0.919) ***		

Table 9

Inferential Statistics for RQ1 (Gender)- Cox Regression

	OR	*SE*	Wald	*df*	Sig.	HR	95% Lower CI	95% Upper CI
Gender (Men as reference)	-0.115	0.016	54.210	1	**<0.001***	**0.891***	**0.865 ***	**0.919 ***

RQ2

RQ2: What is the association between initial diagnostic stage (Stage I, Stage II, etc.) and 5-YS rates of elderly CRC patients with PD?

$H_0$2: There is no association between initial diagnostic stage (Stage I, Stage II, etc.) and 5-YS rates of elderly CRC patients with PD?

H_a2: There is an association between initial diagnostic stage (Stage I, Stage II, etc.) and 5-YS rates of elderly CRC patients with PD?

As the *p* value for the LR test was <0.001, which is less than level of significance alpha 0.05, I concluded that there is a significant difference in 5-YS rate among the four initial stages of diagnosis (Table 10).

Table 10 shows the HR and *p* value for Stage II, Stage III and Stage IV initial stage of diagnosis, taking Stage I as reference. At level of significance alpha=0.05, the *p* value for the overall Cox-regression model was less than 0.05, which is significant. Therefore, I rejected the null hypothesis that there was no link between initial diagnosis stage and 5-YS rate for elderly CRC patients with PD. Since the overall Cox regression model is significant, there is a link between the initial stage of diagnosis and 5-YS in elderly CRC patients with PD. When I compared the 5 years survival between Stage I and Stage II, the *p* value was greater than 0.05, so I rejected the null hypothesis. There was no significant difference in 5-YS when Stage I and Stage II were compared. But the *p* value when I compared 5-YS for Stage III and Stage IV with Stage I. The *p* value was less than 0.05, so I rejected the null hypothesis. Therefore, there was significant difference in 5-YS in Stage III and Stage IV compared to Stage I. The HR from Cox regression model showed the HR as 1.634 (1.539, 1.734). Therefore, Stage III CRC patients with PD had 63.4% higher death compared to Stage I CRC patients with PD. The HR was much higher when Stage IV is compared to Stage I initial stage of Diagnosis.

Table 10

Inferential Statistics for RQ2- Cox Regression

	OR	*SE*	Wald	*df*	Sig.	HR	95% Lower CI	95% Upper CI
Initial stage of diagnosis (Stage I as reference)								
Stage II	0.045	0.054	0.698	1	0.404	1.046	0.941	1.163
Stage III	0.491	0.030	260.431	1	**<0.001***	**1.634***	**1.539***	**1.734***
Stage IV	1.724	0.031	3182.275	1	**<0.001***	**5.605***	**5.279***	**5.950***

The 5-YS rates were compared between men and women at each initial stage of diagnosis (Table 11). For Stage I initial stage of diagnosis, the 5-YS for men was 45.8% and 55% for women. For Stage II, 5-YS was 41.4% and 48.4%. For Stage III, 5-YS was 29.1% for men and 37.6% for women. For stage IV patients, women (4.1%) have lesser 5-YS compared to men (4.6%). The LR test p value when I compared 5-YS between male and female at each initial stage of diagnosis was less than 0.05 except for Stage IV. I rejected the null hypothesis that there was no significant difference in 5-YS between men and women for initial stage of diagnosis I, II and III. Therefore, I conclude that there was significant difference in 5-YS between men and women for initial stage of diagnosis I, II and III. For Stage IV initial stage of diagnosis LR test P value = 0.429 which was greater than 0.05 so I do not reject the null hypothesis and concluded that there was no

significant difference in 5-YS between men and women at initial stage of diagnosis IV. For Initial Stage of Diagnosis, Stage I, Stage II and Stage, HR is less than 1 for all these stages when I had Male as reference. Therefore, I can conclude that women had better 5-YS compared to men for Stage I, Stage II and Stage III initial stage of diagnosis.

Table 11

Inferential Statistics for RQ2- KM

Variable	Women	Men	Total
Initial stage of Diagnosis=1 (N)	2849	1335	4184
Median Survival (estimate in yrs)	NE	4.252	NE
KM Estimates			
1 year	75.6%	68.3%	73.3%
2 years	68.8%	59.3%	65.8%
3 years	62.1%	52.3%	58.9%
4 years	55.0%	45.8%	52.0%
5 years	55.0%	45.8%	52.0%
Log- rank p value, Women Vs Men	**<0.001***		
HR(95% CI)[Men as reference	**0.754 (0.688, 0.826) ***		
Variable	Women	Men	Total
Initial stage of Diagnosis=II (N)	3135	1457	4592
Median Survival (estimate in yrs)	4.747	3.748	4.419

KM Estimates			
1 year	71.1%	65.5%	69.3%
2 years	62.9%	56.2%	60.8%
3 years	55.5%	48.2%	53.2%
4 years	48.4%	41.4%	46.2%
5 years	48.4%	41.4%	46.2%
Log- rank p value, Women Vs Men	**<0.001***		
HR(95% CI)[Men as reference	**0.823 (0.758, 0.894) ***		
Variable	Women	Men	Total
Initial stage of Diagnosis=III (N)	2533	1098	3631
Median Survival (estimate in yrs)	2.834	2.335	2.664
KM Estimates			
1 year	58.5%	54.8%	57.4%
2 years	48.8%	42.9%	47.0%
3 years	42.7%	35.8%	40.6%
4 years	37.6%	29.1%	35.0%

5 years	37.6%	29.1%	35.0%
Log- rank p value, Women Vs Men	**<0.001***		
HR (95% CI)[Men as reference	**0.827 (0.759, 0.902) ***		
Variable	Women	Men	Total
Initial stage of Diagnosis=IV (N)	1930	871	2801
Median Survival (estimate in yrs)	0.334	0.334	0.334
KM Estimates			
1 year	13.6%	14.7%	14.0%
2 years	8.2%	9.3%	8.6%
3 years	5.5%	6.4%	5.8%
4 years	4.1%	4.6%	4.2%
5 years	4.1%	4.6%	4.2%
Log- rank p value, Women Vs Men	0.429		
HR (95% CI)[Men as reference	1.033 (0.951, 1.121)		

RQ3

RQ3: What is the association between MS at diagnosis and 5-YS rates of elderly CRC patients with PD?

$H_0$3: There is no association between MS at diagnosis and 5-YS rates of elderly CRC patients with PD?

H_a3: There is an association between MS at diagnosis and 5-YS rates of elderly CRC patients with PD?

For this analysis, I grouped single and unmarried together and separated, divorced, widowed as one group.35.1% of observations were Blank and Unknown MS at diagnosis. Blank and Unknown MS at diagnosis were excluded from inferential statistics. Therefore, there were 3 groups for MS at diagnosis. CP hazards regression model and relate an indicator of MS at diagnosis and 5-YS was estimated and related p values are shown in table below (Table 11). Since the p value for the model was less than 0.05, Irejected the null hypothesis. I concluded that there was an association between MS at diagnosis and 5-YS in elderly CRC patients with PD. Table 12 further showed a significant difference in 5-YS between single/ unmarried group and married group (p value=0.000). The HR and 95% CI for married group was 0.739 (0.692, 0.788). Therefore, there was 26.1% better 5-YS in married group compared to unmarried/single group in elderly CRC patients with PD. However, there was no significant difference in 5-YS in unmarried/single group versus separated/divorced/widowed group (p value=0.830).

Table 12

Inferential Statistics for RQ3 (MS at Diagnosis)- Cox Regression

	OR	SE	Wald	df	Sig.	HR	95% Lower CI	95% Upper CI
MS at Diagnosis (Single/Unmarried (N) as reference)								
Married	-0.303	0.033	83.571	1	**<0.001***	**0.739***	**0.692***	**0.788***
Separated/Divorced/Widowed	0.007	0.031	0.046	1	0.830	1.007	0.947	1.070

Further, a LR test also showed that there was difference in the 5-YS distributions for various MS at diagnosis. The log-rank test p vlaue was <0.001 which is less than 0.05, therefore I rejected the null hypothesis that there was no significant difference in 5 years surval rates at various marital stage at diagnosis for elderly CRC patients with PD.

For additional analysis, KM survival analysis was conducted to compare the difference in 5-YS in men and women at each different MS at diagnosis (Table 13). Since the p values was less than 0.05 for all the three groups, I rejected the null hypothesis that there was no significant difference in 5-YS between men and women at each level of MS at diagnosis. There was significant difference in 5-YS between men and women in all three groups. For Single/Unmarried group, women had19 % better 5-YS compared to men. For Married group, women had 26% better 5-YS in women compared to men. Like Single/Unmarried group, women had 19% better survival in women compared to men in

Separated/Divorced/Widowed group. For all three groups, women had higher median survival time compared to men.

Table 13

Inferential Statistics for RQ3- KM

Variable	Women	Men	Total
Marital Status= Single/Unmarried (N)	1091	695	1786
Median Survival (estimate in yrs)	2.338	1.580	2.001
KM Estimates			
1 year	53.0%	45.8%	50.2%
2 years	46.6%	38.0%	43.2%
3 years	40.8%	32.4%	37.5%
4 years	35.2%	28.0%	32.4%
5 years	35.2%	28.0%	32.4%
Log- rank p value, Women Vs Men	**<0.001***		
HR (95% CI) [Men as reference	**0.811 (0.723, 0.909) ***		
Marital Status= Married (N)	3310	3235	6545
Median Survival (estimate in yrs)	4.583	2.916	3.663
KM Estimates			
1 year	54.2%	51.2%	61.4%
2 years	46.9%	42.5%	54.0%

3 years	41.2%	36.2%	48.3%
4 years	35.8%	30.8%	42.8%
5 years	35.8%	30.8%	42.8%
Log- rank p value, Women Vs Men	**<0.001***		
HR (95% CI) [Men as reference	**0.748 (0.701, 0.797) ***		
Marital Status= Separated/Divorced/Widowed (N)	8160	1852	10012
Median Survival (estimate in yrs)	2.248	1.580	2.084
KM Estimates			
1 year	65.2%	57.5%	50.9%
2 years	58.6%	49.4%	43.1%
3 years	53.2%	43.4%	37.3%
4 years	48.3%	37.2%	32.0%
5 years	48.3%	37.2%	32.0%
Log- rank p value, Women Vs Men	**<0.001***		
HR (95% CI) [Men as reference	**0.818 (0.771, 0.868) ***		

RQ4

RQ4: What is the association between gender, initial diagnosis stage and MS at diagnosis and 5-YS rates of elderly CRC patients with PD after adjusting for race, ethnicity, socioeconomic status, primary site of tumor?

$H_0$4: There is no association between gender, initial diagnosis stage and 5-YS rates of elderly CRC patients with PD after adjusting for race, ethnicity, socioeconomic status, primary site of tumor.

H_a4: There is an association between gender, initial diagnosis stage and 5-YS rates of elderly CRC patients with PD after adjusting for race, ethnicity, socioeconomic status, primary site of tumor.

To see if there is association between gender, initial stage of diagnosis, MS at diagnosis and 5-YS after adjusting for race, ethnicity, socioeconomic status, and primary site Cox regression model was done. At first the bivariate Cox regression model was done. Table 14 below shows the result of bivariate Cox regression model.

Table 14

Inferential Statistics for RQ4- Cox Regression Model Bivariate Analysis

	OR	SE	Wald	*df*	Sig.	HR	95% Lower CI	95% Upper CI
Gender (Male as reference)	-0.115	0.016	54.210	1	**0.000**	0.891	0.865	0.919
Initial stage of diagnosis (Stage I as reference)								
Stage II	0.167	0.030	30.924	1	**<0.001***	**1.182***	**1.114***	**1.254***
*Stage III	0.491	0.030	260.431	1	**<0.001***	**1.634***	**1.539***	**1.734***
Stage IV	1.724	0.031	3182.275	1	**<0.001***	**5.605***	**5.279***	**5.950***
MS at diagnosis (Single/Unmarried as reference)								
Married	-0.303	0.033	83.571	1	<0.001*	**0.739***	**0.692***	**0.788***
Separated/Divorced/Widowed	0.007	0.031	0.046	1	0.830	1.007	0.947	1.070
Race (White as reference)								
Black	0.090	0.027	10.983	1	**0.001**	**1.094***	**1.038***	**1.154***
American Indian/Alaska Native	0.150	0.136	1.212	1	0.271	1.162	0.890	1.518
Asian or Pacific Islander	-0.123	0.046	6.985	1	**0.008**	**0.884***	**0.808***	**0.969***

*Unknown	-0.955	0.243	15.480	1	**<0.001***	**0.385***	**0.239***	**0.619***
Ethnicity (Non-Spanish-Hispanic-Latino as reference)	-0.175	0.031	30.855	1	<0.001*	**0.840***	**0.790***	**0.893***
SES (Poor SES as reference)								
Near Poor SES	-0.055	0.021	6.447	1	**0.011***	**0.947***	**0.908***	**0.988***
Middle SES	0.019	0.030	0.419	1	0.517	1.020	0.962	1.081
High SES	-0.061	0.022	7.269	1	**0.007***	**0.941***	**0.901***	**0.984***
Analysis Primary Site (Right sided as reference)	0.061	0.016	14.416	1	**<0.001***	**1.063***	**1.030***	**1.098***

All the 7 variables had p value less than 0.05 and were found to be significant. And further included in multivariate initial Cox regression model. Stepwise Cox regression was done to remove the variable with highest p value. For the step I Cox regression model, primary site had the highest p value (0.457). Therefore, for the step II Cox regression model, the primary site was excluded from the Cox regression model. Step II Cox regression model had highest p value for SES (p value=0.169), so SES removed for step III Cox regression model. The Step III Cox regression model showed all the variables had p value less than 0.05. Gender, Initial stage at diagnosis, MS at diagnosis, Race, Ethnicity were included in the final model. The values related to the final model is shown in Table 16 below.

Table 15 below is the results for final multivariateCox regression model. The variables included in this multivariate model are gender, initial stage of diagnosis, marital stage at diagnosis, race, ethnicity. At level of significance alpha =0.05, the p value was less than 0.05 for gender, initial stage of diagnosis, MS at diagnosis after adjusting for race, ethnicity. Therefore, I concluded that there was association between gender, initial stage of diagnosis and MS at diagnosis in elderly CRC patients with PD. Women have better 21.4% better 5-YS compared to men. Stage II patients have 17.7% higher death compared to stage I patients. Stage III patients have 64.1% higher death compared to stage I. Stage IV patients have 5.6 times more death compared to stage I patients. There was a significant difference in 5-YS between single/unmarried group vs married. Married patients had 26.5% better 5-YS compared to single/unmarried patients. However, there was not significant difference in 5-YS between single/unmarried group vs separated/divorced/widowed group as the p value was greater than 0.05 and I rejected the null hypothesis.

Table 15

Inferential Statistics for RQ4- Cox Regression Model Multivariate Analysis (Final Model)

	OR	SE	Wald	*df*	Sig.	HR	95% Lower CI	95% Upper CI
Gender (Male as reference)	-0.240	0.023	104.731	1	**<0.001***	**0.786***	**0.751***	**0.823***
Initial stage of diagnosis (Stage I as reference)								
Stage II	0.163	0.031	27.786	1	**<0.001***	**1.177***	**1.108***	**1.251***
Stage III	0.495	0.031	250.324	1	**<0.001***	**1.641***	**1.543***	**1.745***
Stage IV	1.727	0.031	3008.399	1	**<0.001***	**5.621***	**5.285***	**5.979***
MS at diagnosis (Single/Unmarried as reference)								
Married	-0.308	0.038	65.935	1	**<0.001***	**0.735***	**0.683***	**0.792***
Separated/Divorced/Widowed	0.013	0.036	0.133	1	0.715	1.013	0.944	1.088
Race (White as reference)								
Black	0.059	0.038	2.444	1	0.118	1.061	0.985	1.143
American Indian/Alaska Native	0.151	0.175	0.750	1	0.386	1.163	0.826	1.638
Asian or Pacific Islander	-0.148	0.060	6.024	1	**0.014***	**0.862***	**0.766***	**0.971***
Unknown	-0.258	0.409	0.400	1	0.527	0.772	0.347	1.720
Ethnicity (Non-Spanish-Hispanic-Latino as reference)	-0.157	0.044	12.504	1	**<0.001***	**0.855***	**0.783***	**0.932***

Summary

In this study secondary data analysis was done from SEER-Medicare database. The cohort for this analysis is the elderly patients diagnosed with CRC between 2006 to 2012 with pre-existing diagnosis. A total of 28,278 subjects with age greater than 67 were included for the analysis. Regular data cleaning procedures were done. Most of the patients were Whites and Non-Spanish- Hispanic Latino. Most patients were separated, divorced, and widowed. Most patients had the initial stage of diagnosis as occult, unknown, blank or Stage 0. These were excluded from the analysis. There was a similar distribution for Stages I to IV. Most of the patients had a right sided tumor. In terms of socioeconomic status (SES) most patients had Near-Poor status followed by High-SES. Result showed that there is association between gender and 5-YS rates in elderly CRC patients with PD. Women had better 5-YS rate compared to Men. There was an association between the initial stage of diagnosis and 5-YS rates. The risk of death in 5 years is highest in Stage IV initial stage of diagnosis compared to Stage I. In all stages of diagnosis except for Stage IV, women had better 5-YS compared to Men. There was an association between the marital stage of diagnosis and 5-YS. Married group had better 5-YS compared to single/unmarried group. There was significant difference in 5-YS between men and women for all three categories of MS of diagnosis. From the research question 4, I concluded that there was association between gender, initial stage of diagnosis and MS at diagnosis with 5-YS rate in elderly CRC patients with PD after adjusting for race, ethnicity, socioeconomic status, and primary site

Chapter 5: Discussion, Conclusions, and Recommendations

CRC is one of the most common cancers affecting both men and women (ACS, 2021a). The prevalence of CRC is significantly higher for older groups. Also, OS and CSS were comparatively higher in female patients. Depression is one of the comorbidities that affect cancer patients (Saracino et al., 2015). Studies showed that there was an association between depression and worst cancer prognosis. Studies suggested that there is gender disparity in survival of CRC (Yang et al., 2017). But few studies have investigated gender differences in survival of elderly CRC patients with PD. The purpose of this retrospective cohort study was to evaluate the gender disparity in 5-YS in this population. Secondary data analysis was done using SEER-Medicare database. I report significant findings and discuss the limitations, recommendations, and implications of this study in Chapter 5.

Interpretation of the Findings

Finding 1: Gender and 5-YS in elderly CRC patients with PD

Few studies have looked for an association between gender and 5-YS in CRC patients. A few studies, like Maajani et al. (2019) and Afshar et al. (2018), looked for survival rate in CRC patients. Results from Maajani et al. study showed 1-, 3-, and 5-YS as 84%, 64% and 54%, respectively. In an extension of CONCORD 2, White et al. (2017) evaluated the 5-YS rate of those in 37 states in the United States diagnosed with CRC between 2001 and 2009. The 5-YS rate for CRC patients diagnosed with CRC for 2001-2003 was 63.7% and the 5-YS rate for CRC for year 2004- 2009 was 64.4%. In this study I evaluated 5-YS in elderly CRC patients with PD. The 5-YS rate for elderly CRC

patients with PD was much less when compared to that found by Maajani et al. The overall 5-YS rate for my study was only 34.2%. The 5-YS rate for elderly CRC patients with PD is better in women compared to men, which is consistent with the result from Maajani et al.

Quirt et al. (2017) and Afshar et al. (2018) showed that the incidence of CRC was higher in men than women but there were more women than men elderly CRC patients with PD. This inconsistency of higher incidence of CRC in women in this subgroup may have been due to higher incidence of depression in females. Women are twice as vulnerable for depression when compared to men in all age groups (Brody et al., 2018).

My results showed that there was an association between gender and 5-YS in elderly CRC patients with PD. Gender was one of the factors that affected 5-YS in elderly CRC patients with PD. This finding was consistent with studies by Maajani et al., (2019) and Yang et al., (2017). The median 5-YS time was lower in men than in women for elderly CRC patients with depression, which is consistent with my findings. A Japanese population-based study reported a significant difference in survival between men and women compared (Kotake et al., 2016).

Finding 2: Initial Stage of Diagnosis and 5-YS In Elderly CRC Patients With PD

According to ACS (2021-b), OS for CRC is dependent on stage of initial diagnosis. CRC patients with localized stage at diagnosis had higher survival rate compared to CRC patients whose cancer already spread to distal organs (ACS, 2021-b). Joachim et al. (2019) concluded that there was difference in OS by gender, age at diagnosis and stage of initial diagnosis. White et al. (2017) reported that 5-YS was higher

in patients with localized disease compared to distant stages of CRC. My studytudy on elderly CRC patients with PD showed that there was an association between the initial stage of diagnosis and 5-YS rate for elderly CRC patients with PD. Stage III CRC patients with PD had 63.4% higher death compared to Stage I CRC patients with PD. The HR was much higher when Stage IV is compared to Stage I initial stage of diagnosis. I also compared the difference in survival between men and women at each of initial stage of diagnosis. Results showed that there is significant difference in 5-YS between men and women at all stage of diagnosis except for Stage IV. For Stage IV, there was no significant difference between men and women in 5-YS for elderly CRC patients with PD. Similarly, women have higher survival compared to men except for Stage III.

Finding 3: MS at Diagnosis and 5-YS in elderly CRC Patients with PD

Wang et al. (2011) identified MS as one of the factors associated with better outcome on CRC patients. When married and unmarried patients were compared, survival was better in married group for both genders (Wang et al., 2011). Ching-Chieh et al. (2019) evaluated if there is impact of MS on survival of patients who treated surgically for colon cancer. Like White et al., Ching-Chieh et al. concluded that better survival in married individuals compared to single individuals. The results were similar in elderly CRC patients with PD. My study showed that there was an association between MS at diagnosis and 5-YS in elderly CRC patients with PD. Married groups had better survival than single/unmarried groups. When I compared the single/unmarried group with the separated/divorced/widowed group, there was no significant difference in 5-YS. Gomez et al. (2016) also showed better OS in married when compared with unmarried/

single. The results from Gomez et al. showed that the reason married groups have better survival may be due to increased social support, early diagnosis, and early treatment. These may be the reasons why survival was better in the married group compared to unmarried/ single. Wang et al. (2011) also concluded that survival was better in married patients for both genders.

Further, I evaluated the data to see if there was difference in survival distribution between men and women. The results showed that there was significant difference in 5-YS between men and women in all three marital statuses of diagnosis for this study. In all three subgroups (married, single/unmarried and separated/divorced/widowed) women had better 5-YS. Median 5-YS time was better in women. This result is also consistent with Gomez et al.'s (2016) study on California CRC patients. Gomez et al. showed better survival in women compared to men and that the protective effect decreases with age.

Finding 4: Gender, Initial Stage of Diagnosis and MS at Diagnosis

I performed bivariate and multivariate Cox regression models. A stepwise Cox regression model was done. Finally, gender, initial stage of diagnosis, race, and ethnicity were included in the final model and was interpreted in Chapter 4. There was an association between gender, initial stage, and MS at diagnosis after adjusting for race, ethnicity, socioeconomic status, and primary site. Women had better 5-YS compared to men. This result echo Kotake et al.'s (2016) findings where gender as considered as prognostic factor. My finding was also consistent with studies by Maajani et al. (2019), Yang et al. (2017), and Afshar et al. (2018).

Elderly CRC patients with depression and Stage II, III and IV initial stages of cancer at diagnosis have higher death compared to Stage I patients after adjusting for other covariates. This result from multivariate analysis was consistent with the results from bivariate analysis. There was significant difference in 5-YS between single/unmarried group versus married, whereas there was no significant difference in 5-YS single/unmarried group versus separated/divorced/widowed group for elderly CRC patients with PD. This result from multivariate analysis persisted after controlling for race, ethnicity, socioeconomic status, and primary site so the conclusion from bivariate analysis was consistent with the multivariate analysis.

Many studies have evaluated the difference in CRC survival in race. Coleman et al. (2018) evaluated the racial difference in cancer survival. In the United States, the cancer survival was less in Black patients when compared to White patients (Coleman et al., 2018). Coleman et al.'s results were not consistent with my results. For elderly CRC patients with PD, there was significant difference in survival between White patients and Black patients for bivariate analysis but not for multivariate analysis where other covariates were considered. Since race is one of the significant factors in the multivariate Cox regression model, race is one of the factors that needs to be considered in survival in elderly CRC patients with PD.

Zhang et al. (2017) showed that SES is one of the prognostic factors for CRC. Also, Singh and Jemal (2017) showed that cancer survival was lower in the most disadvantaged group. Zhang et al. investigated SEER database for CRC patients diagnosed between 2007 to 2013 to evaluate the impact of SES on OS of CRC. Result

showed that lowest education group had poor prognosis (Zhang et al., 2017). Van den Berg et al. (2019) studied CRC patients in the Netherlands and found that there was increase in OS if there was increase in SES. In elderly CRC patients with PD, there was an association between SES and 5-YS, but SES was not significant when other covariates like race, ethnicity, or primary site were added in a multivariate Cox regression model.

Like SES, the primary side of cancer was significant when bivariate analysis was done for elderly CRC patients with PD. But with multivariate analysis, the primary site of tumor was not significant and was not included in the final multivariate model. In the literature there have been contradictory findings. Some studies stated that the right side had better survival and some studies stated that left side had better survival. Contradictory to this, Warschkow et al. (2016) concluded that those with a right-sided tumor had the worst survival rate which is like the findings of Petrelli et al. (2017) who found that patient death was less likely when the tumor originated on the left side. However, a study by Wang et al. (2017) showed higher survival in patients with right sided tumor. More research needs to be done for elderly CRC patients with PD.

Limitations of the Study

Despite various contributions listed above from this study, there are few limitations of this study. The author likes to acknowledge that this study relies on the secondary data, there is a chance of misclassification bias due to error in data entry (Frankfort- Nachmias & Nachmias, 2008). Another limitation of such retrospective cohort study always poises the risk of accuracy of the exposure and outcome variables as they were collected beforehand (Setia, 2016). Because of misclassification and accuracy

of data in secondary data there may be threat of external validity so temporal relationship could not be assessed (Frankfort- Nachmias & Nachmias, 2008). The cause and effect of the relationship may not be assessed as well.

SEER-Medicare had many elderly CRC patients with PD, so there was no issue of small sample size in this study. Although withdrawal from Medicare is rare as individuals enrolled are followed up to death, and SEER-Medicare is a high-quality data, there are some limitations of SEER-Medicare database. One such limitation can be the patients with only Medicare claims were included in the data. If any subjects are enrolled in managed care, veteran's hospitals and cancer information from hospitals are not included in the database. Similarly, the results will not be able to generalize to younger population as this study includes CRC patients with PD greater than 67 years old. Also, MS being variables of interest, MS at diagnosis is self-reported by patients so the accuracy of data is questionable. In addition, there was some missing data and some of the observations were removed from analysis.

This study is mainly focused on quantitative factors only. Qualitative factors like behavioral characteristics which includes diet and exercise that affect cancer were not evaluated. Another limitation of this study is that other pre-existing conditions were not evaluated. Finally, this study did not evaluate the causal mechanism for gender difference in 5-YS of elderly CRC patients with PD.

Recommendations

This study evaluated the gender disparity in 5-YS in elderly CRC patients with PD. This study advocated a strong message that survival of terminal disease such as

cancer (CRC) should always be accomplished with awareness of mental diseases. This study serves as good foundation for future to evaluate longer survival time for such patients. This study strictly focused on depression as an example of mental diseases, future studies should also evaluate the complication and influence of other forms of mental disease to understand the gender disparity of CRC patients. Not only cancer management, but mental care should also be emphasized on the management of other terminal diseases. Some specifics from this study, author recommends research should be done to see the effect of treatment on survival of 5-YS in this population of interest. Also, the other factors that affect the 5-YS for elderly CRC patients with depression should be evaluated. This study clearly pointed out the difference in 5-YS in elderly CRC patients with depression and women have better 5-YS. Further evaluation can be done to find out the factors or cause for better survival in women compared to men for the elderly CRC patients with PD. It would be interesting to see if there is difference in 5-YS in CRC patients with PD in the younger age group category.

The author also recommends evaluating the difference in 5-YS in elderly CRC patients with or without PD. Future studies related to impact of gender specific approach especially with PD for cancer care and its impact on 5-YS will be a monumental aid to this study.

Various studies have shown survival for CRC greater than 60%. But for this specific population of elderly CRC patients with PD the 5-YS was only 34.2%. Future studies can also be done to evaluate the factors affecting such lower 5-YS rate for this group.

Implications

This study demonstrated that there is a strong difference among genders for 5-YS of elderly CRC patients with PD. Elderly CRC women tends to have better 5-YS compared to their counterpart which was also supported by lesser death compared to men. As per literature review, there are numerous studies done to see gender difference in 5-YS between men and women but very few studies evaluated the influence of mental health in cancer patients especially in CRC. So, this study attempted to fill the gap in literature by evaluating the 5-YS for elderly CRC patients with PD. This study should serve as a cornerstone for future studies for better understanding of CRC patients with PD. This study should point out practitioners the importance of PD for 5-YS outcomes for CRC patients. This knowledge can be used in clinical settings to implement better care to the population of interest. A collaborative, interdisciplinary approach could be used where oncologist and psychiatrist worked together to deliver profound care for cancer patients. This study also pointed out that there might be numerous cancer patients suffering with PD. So, the success of cancer treatment could be aided by quality mental healthcare for cancer patients. There should be inclusion of mental health care during the training of medical professionals. More funding should be allocated to understand the mental health of cancer patients.

Conclusion

CRC is among the leading cancer diagnosed in the US. Among the CRC patients, mental health status has a direct influence in the quality of life and the progression of the cancer. Mental health such as depression is often ignored in elderly cancer patients since

depression symptoms are easily overlooked as aging symptoms. In addition to this, literature reviews strongly suggested that there is a gender disparity in the survival rate of CRC. However, to the author's knowledge, there is a very limited understanding regarding the gender bias in 5-YS of elderly CRC patients with PD. So, this study focused on this existing gap in literature. This study showed that there is a strong association between gender, initial stage of diagnosis and MS of diagnosis and 5-YS in elderly CRC patients with PD. It further pointed out the difference in 5-YS among gender for all initial stages of diagnosis and MS at diagnosis. The results particularly pointed out the difference in survival between among gender for initial stages of diagnosis but is indifference to the late stage (stage IV). Strikingly, there was a significant difference in 5-YS between men and women for MS at diagnosis. All in all, this study concluded elderly women CRC patients initially diagnose with depression have better 5-YS compared to the counterparts.

References

Afshar, N., English, D. R., Thursfield, V., Mitchell, P. L., Te Marvelde, L., Farrugia, H., Giles, G.G. & Milne, R. L. (2018). Differences in cancer survival by sex: a population-based study using cancer registry data. *Cancer Causes Control 29*, 1059–1069 (2018). https://doi.org/10.1007/s10552-018-1079-z

Allemani, C., Weir, H. K., Carreira, H., Harewood, R., Spika, D., Wang, X.-S., Bannon, F., Ahn, J. V., Johnson, C. J., Bonaventure, A., Marcos-Gragera, R., Stiller, C., Azevedo e Silva, G., Chen, W.-Q., Ogunbiyi, O. J., Rachet, B., Soeberg, M. J., You, H., Matsuda, T., …Coleman, M. P. (2015). Global surveillance of cancer survival 1995–2009: Analysis of individual data for 25,676,887 patients from 279 population-based registries in 67 countries (CONCORD-2). *Lancet*, *385*(9972), 977–1010. https://doi.org/10.1016/S0140-6736(14)62038-Al-Husseini, M. J., Saad, A. M., Gaad, M.M., Rachid, A., Salama, A.H., Hilaal, T. (2018). Racial disparities in CRC survival: A population-based cohort. *Journal of Clinical Oncology. 36*. https://ascopubs.org/doi/abs/10.1200/JCO.2018.36.15_suppl.e15643

American Cancer Society. (2017). *CRC Facts & Figures 2017-2019*. https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/colorectal-cancer-facts-and-figures/colorectal-cancer-facts-and-figures-2017-2019.pdf

American Cancer Society. (2020a). *Cancer facts & figures 2020*. https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2020/cancer-facts-and-figures-2020.pdf

American Cancer Society. (2020b). *What is CRC?* https://www.cancer.org/cancer/colon-rectal-cancer/about/what-is-colorectal-cancer.html

American Cancer Society. (2021a). *Key statistics for CRC*. https://www.cancer.org/cancer/colon-rectal-cancer/about/key-statistics.html

American Cancer Society. (2021b). *Survival Rates for CRC.* https://www.cancer.org/cancer/colon-rectal-cancer/detection-diagnosis-staging/survival-rates.html

American Cancer Society. (2023a). *Colorectal cancer stages.* https://www.cancer.org/cancer/colon-rectal-cancer/detection-diagnosis-staging/staged.html

American Cancer Society. (2023b). *Colorectal cancer signs and symptoms*. https://www.cancer.org/cancer/colon-rectal-cancer/detection-diagnosis-staging/signs-and-symptoms.html

Arnold, M., Sierra, M. S., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2017). Global patterns and trends in CRC incidence and mortality. *Gut*, (4), 683. https://doi.org/10.1136/gutjnl-2015-310912

Balcazar, H., Wise, S., Rosenthal, E. L., Ochoa, C., Rodriguez, J., Hastings, D., Flores, L., Hernandez, L. & Duarte-Gardea, M. (n.d.). An ecological model using Promotores de Salud to prevent cardiovascular disease on the US-Mexico border:

The HEART Project. *Preventing Chronic Disease*, *9*. http://dx.doi.org/10.5888/pcd9.110100

Baran, B., Mert Ozupek, N., Yerli Tetik, N., Acar, E., Bekcioglu, O., & Baskin, Y. (2018). Difference between left-sided and right-sided CRC: A focused review of literature. *Gastroenterology Research, 11*(4), 264–273. http://dx.doi.org/10.5888/pcd9.110100

Benedix, F., Kube, R., Meyer, F., Schmidt, U., Gastinger, I., Lippert, H., & Colon/Rectum Carcinomas (Primary Tumor) Study Group (2010). Comparison of 17,641 patients with right- and left-sided colon cancer: Differences in epidemiology, perioperative course, histology, and survival. *Diseases of the Colon and Rectum, 53*(1), 57–64. https://doi.org/10.1007/DCR.0b013e3181c703a4

Blakemore, E. (2019). *Race and ethnicity: How are they different?* National Geographic. https://www.nationalgeographic.com/culture/article/race-ethnicity#close

Bland, J.M. & Altman, D.G. (2004). The logrank test. *BMJ, 328*(7447), 1073. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC403858/

Bojer, A. S., & Roikjær, O. (2015). Elderly patients with CRC are oncologically undertreated. *European Journal of Surgical Oncology*, *41*(3), 421–425. https://doi.org/10.1016/j.ejso.2014.10.065

Brody, D.J., Pratt, L.A., & Hughes, J. P. (2018). Prevalence of depression among adults aged 20 and over: United States, 2013–2016. *NCHS Data Brief*. https://www.cdc.gov/nchs/data/databriefs/db303.pdf

Campbell-Thompson, M., Lynch, I. J., & Bhardwaj, B. (2001). Expression of estrogen receptor (ER) subtypes and ERbeta isoforms in colon cancer. *Cancer Research, 61*(2), 632–640. https://cancerres.aacrjournals.org/content/61/2/632.long

Cancer Care. (2021). CRC: *The importance of screening and early detection*. https://www.cancercare.org/publications/116-colorectal_cancer_the_importance_of_screening_and_early_detection

Centers for Disease Control and Prevention. (2020). *CRC statistics.* https://www.cdc.gov/cancer/colorectal/statistics/index.htm

Centers for Disease Control and Prevention. (2021). *CRC Control Program (CRCCP)*. https://www.cdc.gov/cancer/crccp/about.htm

Centers for Disease Control and Prevention. (2022). *What is colorectal cancer?*. https://www.cdc.gov/cancer/colorectal/basic_info/what-is-colorectal-cancer.htm#:~:text=Colorectal%20cancer%20is%20a%20disease,large%20intestine%20or%20large%20bowel.

Centers for Medicare & Medicaid Services. (2021). *Chronic conditions data warehouse*. https://www2.ccwdata.org/web/guest/condition-categories

Ching-Chieh, Y., Li-Chin, C., Yu-Wei, L., Shih-Chang, W., Te-Min, K., Chung-I., H., Yu-Chieh, S., & Ming-Hong, T. (2019). The impact of marital status on survival in patients with surgically treated colon cancer. *Medicine*, *11*, e14856. https://doi.org/10.1097/MD.0000000000014856 Coleman, M. P., Quaresma, M., Berrino, F., Lutz, J.-M., De Angelis, R., Capocaccia, R., Baili, P., Rachet, B., Gatta, G., Hakulinen, T., Micheli, A., Sant, M., Weir, H. K., Elwood, J. M.,

Tsukuma, H., Koifman, S., e Silva, G. A., Francisci, S., Santaquilani, M., … Young, J. L. (2008). Cancer survival in five continents: a worldwide population-based study (CONCORD). *Lancet Oncology*, *9*(8), 730–756. https://doi.org/10.1016/S1470-2045(08)70179-7

Cunningham, R., Sarfati, D., Stanley, J., Peterson, D., & Collings, S. (2015). Cancer survival in the context of mental illness: a national cohort study. *General Hospital Psychiatry, 37*(6), 501-506. https://doi.org/10.1016/j.genhosppsych.2015.06.003

Daly, M.C., & Paquette, I.M. (2019). Surveillance, epidemiology, and end results (SEER) and SEER-Medicare databases: Use in clinical research for improving CRC outcomes. *Clinics in Colon and Rectal Surgery, 32*(1), 61-68. https://www.thieme-connect.com/products/ejournals/abstract/10.1055/s-0038-1673355

Das-Munshi, J., Chang, C., Dutta, R., Morgan, C. Nazroo, J., Stewart, R., & Prince, M.J. (2017). Ethnicity and excess mortality in severe mental illness: a cohort study. *Lancet Psychiatry,4*(5), 389-399. https://doi.org/10.1016/S2215-0366(17)30097-4

Dekker, E., Tanis, P.J., Vleugels, J. L. A., Kasi, P. M., & Wallace, M.B. (2019). CRC. *Lancet, 39* (10207), 1467-1480. https://doi.org/10.1016/S0140-6736(19)32319-0

Drzayich, A, D., Waldman, C, A., Khoury, R, Michael, T., Renda, A., Hopson, S., Parikh, A., Stein, A., Costantino, M., Stemkowski, S., & Bunce, M. (2018). *Journal of Patient Reported Outcomes, 2*(29). https://doi.org/10.1186/s41687-018-0057-2

Feng, Y., Mo, S., Dai, W., Li, Q., Cai, G., & Cai, S. (2020). Increasing age-related survival gap among patients with CRC: a population-based retrospective study. *International Journal of Clinical Oncology*, *25*(1), 100–109. https://doi.org/10.1007/s10147-019-01538-3

Girgus, J. S., Yang, K., & Ferri, C.V (2017). The gender difference in depression: Are elderly women at greater risk for depression than elderly men? *Geriatrics*, *4*, 35. https://doi.org/10.3390/geriatrics2040035Goel, M. K., Khanna, P., & Kishore, J. (2010). Understanding survival analysis: KM estimate. *International Journal of Ayurveda Research*, *1*(4), 274–278. https://doi.org/10.4103/0974-7788.76794

Gomez, S. L., Hurley, S., Canchola, A. J., Keegan, T. H., Cheng, I., Murphy, J. D., Clarke, C. A., Glaser, S. L., & Martínez, M. E. (2016). Effects of marital status and economic resources on survival after cancer: A population-based study. *Cancer*, *122*(10), 1618–1625. https://doi.org/10.1002/cncr.29885

Itatani, Y., Kawada, K., & Sakai, Y. (2018). Treatment of elderly patients with CRC. *Biomedical Research International. 2018*, 2176056. https://doi.org/10.1155/2018/2176056

Jager, K. J., van Dijk, P. C., Zoccali, C., & Dekker, F. W. (2008). The analysis of survival data: The KM method. *Kidney international*, *74*(5), 560–565. https://doi.org/10.1038/ki.2008.217

Jia, Y., Li, F., Liu, Y. F., Zhao, J. P., Leng, M. M., & Chen, L. (2017). Depression and cancer risk: A systematic review and meta-analysis. *Public Health*, *149*, 138–148. https://doi.org/10.1016/j.puhe.2017.04.026

Joachim, C., Macni, J., Drame, M., Pomier, A., Escarmant, P., Veronique-Baudin, J., & Vinh-Hung, V. (2019). OS of CRC by stage at diagnosis: Data from the Martinique cancer registry. *Medicine*, (35), 16941. https://journals.lww.com/md-journal/fulltext/2019/08300/overall_survival_of_colorectal_cancer_by_stage_at.44.aspx

Kaerlev, L., Iachina, M., Trosko, O., Qvist, N., Ljungdalh, P. M., & Nørgård, B. M. (2018). Colon cancer patients with a serious psychiatric disorder present with a more advanced cancer stage and receive less adjuvant chemotherapy - A Nationwide Danish Cohort Study. *BMC cancer, 18*(1), 1050. https://doi.org/10.1186/s12885-018-4879-3

Kenner B. J. (2018). Early Detection of Pancreatic Cancer: The Role of Depression and Anxiety as a Precursor for Disease. *Pancreas*, *47*(4), 363–367. https://doi.org/10.1097/MPA.0000000000001024

Kim, S., Paik, H.Y., Yoon, H., Lee, J.E., Kim, N., & Sung, M. (2015). Sex- and gender-specific disparities in CRC risk. *World Journal of Gastroenterology, 21*(17), 5167-5175. https://dx.doi.org/10.3748/wjg.v21.i17.5167

Kilanowski, J.N. (2017). Breadth of the Socio-Ecological Model. *Journal of Agromedicine, 22*(4), 295-297. https://doi.org/10.1080/1059924X.2017.1358971

Ko, Z., Kim, K., Son, J.S., Park, H., & Park, S.M. (2019). Association of PD with all-cause, cancer related, and noncancer-related mortality among 5-year cancer survivors: a population-based cohort study. *Scientific Reports, 9* (18334). https://doi.org/10.1038/s41598-019-54677-y

Koch, M., McPherson, T. A., & Egedahl, R. D. (1982). Effect of sex and reproductive history on the survival of patients with CRC. *Journal Of Chronic Diseases*, *35*(1), 69–72. https://doi.org/10.1016/0021-9681(82)90032-7

Koo, J.H., & Leong R.W. (2010). Sex differences in epidemiological, clinical and pathological characteristics of CRC. *Journal of Gastroenterology Hepatology, 25*, 33–42. https://doi.org/10.1111/j.1440-1746.2009.05992.x

Kotake, K., Asano, M., Ozawa, H., Kobayashi, H., & Sugihara, K. (2016). Gender differences in CRC survival in Japan. *International Journal of Clinical Oncology*, *21*(1), 194–203. https://doi.org/10.1007/s10147-015-0868-6

Koyama, A., Matsuoka, H., Makimmura, C., Sakai, K., Sakamoto, R., & Murata, M. (2016). Gender differences in cancer-related distress in Japan: a retrospective observational study. *Biopsychosocial Medicine, 10*(10). https://doi.org/10.1186/s13030-016-0062-8

Lavnaditi, M., Barbas, G., Fratzana, A., Zyga, S. (n.d.). Evaluation of depression in colon cancer patients. *Health Science Journal*. https://www.hsj.gr/medicine/evaluation-of-depression-in-colon-cancer-patients.php?aid=3129

Lim, G.Y., Tam, W.W., Lu, Y., Ho, C.S., Zhang, M.W., & Ho, R.C. (2018). Prevalence of Depression in the Community from 30 Countries between 1994 and 2014. *Scientific Reports* 8,2861(2018). https://doi.org/10.1038/s41598-018-21243-x

Lin, J., McGlynn, K. A., Carter, C. A., Nations, J. A., Anderson, W. F., Shriver, C. D., & Zhu, K. (2016). The Impact of Preexisting Mental Health Disorders on the Diagnosis, Treatment, and Survival among Lung Cancer Patients in the U.S.

Military Health System. *Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, *25*(12), 1564–1571. https://doi.org/10.1158/1055-9965.EPI-16-0316

Liu, X., Bi, Y., Wang, H., Meng, R., Zhou, W., Zhang, G., Yu, C., & Zhang, Z. J. (2019). Different trends in CRC mortality between age groups in China: an age-period-cohort and joinpoint analysis. *Public health*, *166*, 45–52. https://doi.org/10.1016/j.puhe.2018.08.007

Luque-Fernandez, M. A., Gonçalves, K., Salamanca-Fernández, E., Redondo-Sanchez, D., Lee, S. F., Rodríguez-Barranco, M., Carmona-García, M. C., Marcos-Gragera, R., & Sánchez, M.-J. (2020). Multimorbidity and short-term overall mortality among CRC patients in Spain: A population-based cohort study. *European Journal of Cancer, 129*, 4–14. https://doi.org/10.1016/j.ejca.2020.01.021

Maajani, K., Khodadost, M., Fattahi, A., Shahrestanaki, E., Pirouzi, A., Khalili, F., & Fattahi, H. (2019). Survival Rate of CRC in Iran: A Systematic Review and Meta-Analysis. *Asian Pacific Journal of Cancer Prevention: APJCP*, *20*(1), 13–21. http://journal.waocp.org/article_81665_7169b22b93325ab4b3466631dbef364a.pdf

Meguid, R. A., Slidell, M. B., Wolfgang, C. L., Chang, D. C., & Ahuja, N. (2008). Is there a difference in survival between right- versus left-sided colon cancers? *Annals of surgical oncology, 15*(9), 2388–2394. https://doi.org/10.1245/s10434-008-0015-y

McLeroy, K. R., Bibeau, D, Steckler, A. & Glanz, K. (1988). The social ecology of health promotion interventions. *Health Education Quarterly,* 15(4):351-377. https://www.academia.edu/170661/An_Ecological_Perspective_on_Health_Promotion_Programs

Millan, M., Merino, S., Caro, A., Feliu, F., Escuder, J., & Francesch, T. (2015). Treatment of CRC in the elderly. *World journal of gastrointestinal oncology*, *7*(10), 204–220. https://doi.org/10.4251/wjgo.v7.i10.204

Mistry, S.A., Madala, H., Chacko, C., Jain, N., Gernand, J., Arora, M.L., Kakarala, R. (2014). Colon cancer: Characteristics and survival rates in young versus old patients. *Journal of Clinical Oncology, 32*(15_suppl), 3585-3585. https://doi.org/10.1200/jco.2014.32.15_suppl.3585

Moore, A. R., Buchanan, N. D., Fairley, T. L., & Lee Smith, J. (2015). Public Health Action Model for Cancer Survivorship. *American journal of preventive medicine*, *49*(6 Suppl 5), S470–S476. https://doi.org/10.1016/j.amepre.2015.09.001

Murphy, G., Devesa, S.S., Cross, A.J., Inskip, P.D., McGlyn, K.A., & Cook, M.B. (2011). Sex disparities in CRC incidence by anatomic subsites, race and age. *International Journal of Cancer, 128*(7), 1668-1675. https://doi.org/10.1002/ijc.25481

Musuuza, J., Sherman, M.E., Knudsen, K.J., Sweeney, H.A., Tyler, C.V., & Koroukian. (2013). Analyzing excess mortality from cancer among individuals with mental illness. *Cancer, 119* (13). https://doi.org/10.1002/cncr.28091

Myint, A.S., & Gérard, J.P. (2020). Role of radiotherapy in the treatment of rectal cancer in older patients. *European Journal of Surgical Oncology*. https://doi.org/10.1016/j.ejso.2019.12.017

National Cancer Institute. (n.d.-a). *Cancer stat facts: CRC*. https://seer.cancer.gov/statfacts/html/colorect.html

National Cancer Institute (n.d.-b). *Introduction to CRC*. https://training.seer.cancer.gov/colorectal/intro/

National Cancer Institute. (2019-a*). IRB Approval & HIPAA Regulations*. https://healthcaredelivery.cancer.gov/seermedicare/privacy/hipaa.html

National Cancer Institute. (2019-b). *SEER-Medicare: Brief description of the SEER-Medicare database*. https://healthcaredelivery.cancer.gov/seermedicare/overview/

National Cancer Institute. (2019-c*). SEER-Medicare: Overview of the process for obtaining data.* https://healthcaredelivery.cancer.gov/seermedicare/obtain/seerdua.pdf

National Cancer Institute. (2019-d). *SEER-Medicare Policy on Encryption & Data Security: Portable Devices & Removable Media*. https://healthcaredelivery.cancer.gov/seermedicare/privacy/portable.html

National Cancer Institute. (2019-e). *SEER-Medicare: Privacy & confidentiality Issues*. https://healthcaredelivery.cancer.gov/seermedicare/privacy/

National Cancer Institute. (2020-a). *Cancer stat facts: Common cancer sites*. https://seer.cancer.gov/statfacts/html/common.html

National Cancer Institute. (2020-b). *SEER-Medicare: About the data files*. https://healthcaredelivery.cancer.gov/seermedicare/aboutdata/

National Cancer Institute. (2020-c). *SEER-Medicare: How the SEER & Medicare data are linked*. https://healthcaredelivery.cancer.gov/seermedicare/overview/linked.html

National Cancer Institute. (2020-d). *SEER-Medicare: Medicare enrollment data*. https://healthcaredelivery.cancer.gov/seermedicare/medicare/enroll.html

National Cancer Institute. (2020-e). *SEER-Medicare policies*. https://healthcaredelivery.cancer.gov/seermedicare/aboutdata/policies.html

National Cancer Institute. (2020-f). *SEER-Medicare: SEER program & data*. https://healthcaredelivery.cancer.gov/seermedicare/aboutdata/program.htmlNational Cancer Institute. (2021-a). *Data user agreement*. https://healthcaredelivery.cancer.gov/seermedicare/obtain/seerdua.pdf

National Cancer Institute. (2021-b). *Encrypted or Restricted Variables*. https://healthcaredelivery.cancer.gov/seermedicare/privacy/variables.html

National Cancer Institute. (2021-c). *SEER-Medicare: Available Files by Year*. https://healthcaredelivery.cancer.gov/seermedicare/medicare/table.html

National Cancer Institute. (2021-d). *SEER-Medicare data cost calculator*. https://healthcaredelivery.cancer.gov/seermedicare/obtain/costcalc.html

National Cancer Institute. (2021-e). *SEER-Medicare: File sizes & distribution*. https://healthcaredelivery.cancer.gov/seermedicare/aboutdata/size.html

National Cancer Institute. (2021-f). *SEER-Medicare: Number of Cancer Cases for Selected Cancers Appearing in the Data*. https://healthcaredelivery.cancer.gov/seermedicare/aboutdata/cases.html

National Cancer Institute. (2021-g). *SEER-Medicare: Privacy & confidentiality issues*.https://healthcaredelivery.cancer.gov/seermedicare/privacy/

National Cancer Institute. (2021-h). *SEER-Medicare: Required documents & instructions for submitting requests*. https://healthcaredelivery.cancer.gov/seermedicare/obtain/requests.html

National Cancer Institute. (2021-i). *Screening Tests to Detect Colorectal Cancer and Polyps*. https://www.cancer.gov/types/colorectal/screening-fact-sheet

National Institute of Mental Health. (n.d.). *Major depression*. https://www.nimh.nih.gov/health/statistics/major-depression.shtml

Peng, Y. N., Huang, M. L., & Kao, C. H. (2019). Prevalence of Depression and Anxiety in CRC Patients: A Literature Review. *International journal of environmental research and public health*, *16*(3), 411. https://doi.org/10.3390/ijerph16030411

Petrelli, F., Tomasello, G., Borgonovo, K., Ghidini, M., Turati, L., Dallera, P., Passalacqua, R., Sgroi, G. & Barni, S. (2017). Prognostic Survival Associated with Left-Sided vs Right-Sided Colon Cancer: A Systematic Review and Meta-analysis. JAMA Oncology, 3(2), 211–219. https://jamanetwork.com/journals/jamaoncology/fullarticle/2575468

Paulson, E.C., Wirtalla, C., Armstrong, K., & Mahmoud, N.N. (2009). Gender influences treatment and survival in CRC surgery. *Diseases of Colon Rectum, 52*(12), 1982–1991. https://pubmed.ncbi.nlm.nih.gov/19959975/

Quirt, J. S., Nanji, S., Wei, X., Flemming, J. A., & Booth, C. M. (2017). Is there a sex effect in colon cancer? Disease characteristics, management, and outcomes in routine clinical practice. *Current Oncology (Toronto, Ont.)*, *24*(1), e15–e23. https://doi.org/10.3747/co.24.3410

Rahman, M.S. (2017). The Advantages and Disadvantages of Using Qualitative and Quantitative Approaches and Methods in Language "Testing and Assessment" Research: A Literature Review. *Journal of Education and Learning, 6*(1). https://doi.org/10.5539/jel.v6n1p102

Riecher-Rössler, A. (2016). Sex and gender differences in mental disorders. *The Lancet Psychiatry, 4*(1), 8-9. https://doi.org/10.1016/S2215-0366(16)30348-0

Saracino, R. M., Rosenfeld, B., & Nelson, C. J. (2015). Towards a new conceptualization of depression in older adult cancer patients: a review of the literature. *Aging & Mental Health, 20*(12), 1230–1242. https://doi.org/10.1080/13607863.2015.1078278

Setia M. S. (2016). Methodology Series Module 1: Cohort Studies. Indian journal of dermatology, 61(1), 21–25. https://doi.org/10.4103/0019-5154.174011

Sherrill, C., Smith, M., Mascoe, C., Bigus, E., & Abbitt, D. (2017). Effect of Treating Depressive Disorders on Mortality of Cancer Patients. *Cureus*, *9*(10), e1740. https://doi.org/10.7759/cureus.1740

Siegel, R., Miller, K.D., & Jemal, A. (2020). Cancer statistics, 2020. *CA: A Cancer Journal for Clinicians, 70* (1), 7-30. https://doi.org/10.3322/caac.21590

Singh, G.K. & Jemal, A. (2017). Socioeconomic and Racial/Ethnic Disparities in Cancer Mortality, Incidence, and Survival in the United States, 1950–2014: Over Six Decades of Changing Patterns and Widening Inequalities. *Journal of Environmental and Public Health*. https://doi.org/10.1155/2017/2819372

Smith H. R. (2015). Depression in cancer patients: Pathogenesis, implications and treatment (Review). *Oncology letters*, *9*(4), 1509–1514. https://doi.org/10.3892/ol.2015.2944

Sotelo, J. L., Musselman, D., & Nemeroff, C. (2014). The biology of depression in cancer and the relationship between depression and cancer progression. *International review of psychiatry (Abingdon, England)*, *26*(1), 16–30. https://doi.org/10.3109/09540261.2013.875891

Stalpers, L.J.A. & Kaplan, E. L. (2018). Edward L. Kaplan and the KM Survival Curve, *BSHM Bulletin: Journal of the British Society for the History of Mathematics, 33*(2), 109-135. https://doi.org/10.1080/17498430.2018.1450055

Štor, Z., Blagus, R., Tropea, A., & Biondi, A. (2019). Net survival of patients with CRC: a comparison of two periods. *Updates in surgery*, *71*(4), 687–694. https://doi.org/10.1007/s13304-019-00662-z

Thanikachalam, K., & Khan, G. (2019). CRC and Nutrition. *Nutrients*, *11*(1). https://doi.org/10.3390/nu11010164

The Pennsylvania State University. (2021). *Lesson 13: Proportional Hazards Regression. Epidemiological Research Methods*. https://online.stat.psu.edu/stat507/lesson/13

Tsunoda, A., Nakao, K., Hiratsuka, K., Yasuda, N., Shibusawa, M., & Kusano, M. (2005). Anxiety, depression and quality of life in CRC patients. *International Journal of Clinical Oncology, 10*(6):411-417. https://doi.org/10.1007/s10147-005-0524-7

University of Washington. (2021). Evidence base for collaborative care in cancer treatment. https://aims.uw.edu/collaborative-care-cancer-treatment

U.S. Preventive Services Task Force recommendation statement. (2016). *JAMA, The Journal of the American Medical Association*, (23), 2564. https://doi.org/10.1001/jama.2016.5989

U.S. Preventive Services Task Force (2020). USPSTF A and B recommendations. https://www.uspreventiveservicestaskforce.org/uspstf/recommendation-topics/uspstf-and-b-recommendations

Van den Berg, I., Buettner, S., van den Braak, R., Ultee, K., Lingsma, H. F., van Vugt, J., & Ijzermans, J. (2020). Low Socioeconomic Status Is Associated with Worse Outcomes After Curative Surgery for CRC: Results from a Large, Multicenter Study. Journal of gastrointestinal surgery: official journal of the Society for Surgery of the Alimentary Tract, 24(11), 2628–2636. https://doi.org/10.1007/s11605-019-04435-2

Wang, C. B., Shahjehan, F., Merchea, A., Li, Z., Bekaii-Saab, T. S., Grothey, A., Colibaseanu, D. T., & Kasi, P. M. (2019). Impact of Tumor Location and

Variables Associated with OS in Patients With CRC: A Mayo Clinic Colon and Rectal Cancer Registry Study. *Frontiers in oncology*, *9*, 76. https://doi.org/10.3389/fonc.2019.00076

Wang, L., Wilson, S. E., Stewart, D. B., & Hollenbeak, C. S. (2011). Marital status and colon cancer outcomes in US Surveillance, Epidemiology and End Results registries: Does marriage affect cancer survival by gender and stage? *Cancer Epidemiology*, *35*(5), 417–422. https://doi.org/10.1016/j.canep.2011.02.004

Warschkow, R., Sulz, M. C., Marti, L., Tarantino, I., Schmied, B. M., Cerny, T., & Güller, U. (2016). Better survival in right-sided versus left-sided stage I - III colon cancer patients. *BMC cancer*, *16*, 554. https://doi.org/10.1186/s12885-016-2412-0

Weinberger, A. H., Gbedemah, M., Martinez, A. M., Nash, D., Galea, S., & Goodwin, R. D. (2018). Trends in depression prevalence in the USA from 2005 to 2015: widening disparities in vulnerable groups. *Psychological medicine*, *48*(8), 1308–1315. https://doi.org/10.1017/S0033291717002781

Wessels, H., de Graeff, A., Wynia, K., de Heus, M., Kruitwagen, C. L., Woltjer, G. T., Teunissen, S. C., & Voest, E. E. (2010). Gender-related needs and preferences in cancer care indicate the need for an individualized approach to cancer patients. *The oncologist*, *15*(6), 648–655. https://doi.org/10.1634/theoncologist.2009-0337

White, A., Ironmonger, L., Steele, R., Ormiston-Smith, N., Crawford, C., & Seims, A. (2018). A review of sex-related differences in CRC incidence, screening uptake,

routes to diagnosis, cancer stage and survival in the UK. *BMC cancer, 18*(1), 906. https://bmccancer.biomedcentral.com/articles/10.1186/s12885-018-4786-7

White, A., Joseph, D., Rim, S. H., Johnson, C. J., Coleman, M. P., & Allemani, C. (2017). Colon cancer survival in the United States by race and stage (2001-2009): Findings from the CONCORD-2 study. *Cancer*, *123 Suppl 24*(Suppl 24), 5014–5036. https://doi.org/10.1002/cncr.31076

World Cancer Research Fund (2018). Worldwide cancer data. https://www.wcrf.org/dietandcancer/cancer-trends/worldwide-cancer-data

Yang, L., Korhonen, K., Moustgaard, H., Silventoinen, K., & Martikainen, P. (2018). PD predicts survival in cardiovascular disease and cancer. *Journal of epidemiology and community health, 72*(7), 617–622. https://doi.org/10.1136/jech-2017-210206

Yang, Y., Wang, G., He, J., Ren, S., Wu, F., Zhang, J., & Wang, F. (2017). Gender differences in CRC survival: A meta□analysis. *International Journal of Cancer, 141*(10), 1942-1949. https://doi.org/10.1002/ijc.30827

Yu, B., Zhang, X., Wang, C., Sun, M., Jin, L. & Liu, X. (2020). Trends in depression among adults in the United States, NHANES 2005–2016. *Journal of Affective Disorders*. https://doi.org/10.1016/j.jad.2019.11.036

Yu, Y., Carey, M., Pollett, W., Green, J., Dicks, E., Parfrey, P., Yilmaz, Y. E., & Savas, S. (2019). The long-term survival characteristics of a cohort of CRC patients and baseline variables associated with survival outcomes with or without time-varying effects. *BMC medicine*, *17*(1), 150. https://doi.org/10.1186/s12916-019-1379-5

Zhang, Q., Wang, Y., Hu, H., Huang, R., Xie, L., Liu, E., Chen, Y. G., Wang, G., & Wang, X. (2017). Impact of socioeconomic status on survival of CRC patients. *Oncotarget*, *8*(62), 106121–106131. https://doi.org/10.18632/oncotarget.20859

Appendix A: Code Book

Table A1

Independent Variables

Variable	Nature	Type	SEER-Medicare file	Variable name	Coding
Gender	Independent	Categorical	SEER-Medicare cancer file	sex	1 =Male 2= Female 9= Not stated (unknown)
Initial Stage of Diagnosis	Independent	Categorical	SEER-Medicare cancer file	DerivedAJCCStageGroup6thed2004 2	00= Stage 0 01 =Stage 0a 02 =Stage 0is 10= Stage I 11 =Stage I NOS 12 =Stage IA 13= Stage IA1 14 =Stage IA2 15= Stage IB 16 =Stage IB1 17 =Stage IB2 18 =Stage IC 19 =Stage IS 30= Stage II 31 =Stage II NOS

					32 =Stage IIA 33 =Stage IIB 34 =Stage IIC 50= Stage III 51= Stage III NOS 52 =Stage IIIA 53 =Stage IIIB 54 =Stage IIIC 70 =Stage IV 71 =Stage IV NOS 72 =Stage IVA 73 =Stage IVB 74 =Stage IVC 88= Not applicable 90= Stage Occult 99 =Stage Unknown 126 =Blank
Marital status at diagnosis	Covariate	Categorical	SEER-Medicare cancer file	Marital_status_at_diagnosis (Marital Status at Diagnosis)	1= Single (never married) 2 =Married (including common law) 3

	=Separated 4= Divorced 5 =Widowed 6 =Unmarried or domestic partner (same sex or opposite sex or unregistered) 9= Unknown 14 =Blank

Table A2

Confounder Variables

Variable	Nature	Type	SEER-Medicare file	Variable name (label)	Codes and description
Race	Covariate	Categorical	SEER-Medicare cancer file	Race_recode_W_B_AI_API	1 =White 2 =Black 3 =American Indian/Alaska Native 4 =Asian or Pacific Islander 7 =Other unspecified (1991+) 9= Unknown
Ethnicity	Covariate	Categorical	SEER-Medicare cancer file	OriginrecodeNHIAHispanicNonHisp	0 =Non-Spanish-Hispanic-Latino 1 =Spanish-Hispanic-Latino
Socioeconomic Status	Covariate	Categorical	SEER-Medicare cancer file	1. Census Tract Poverty indicator-Census_Tract_Poverty_Indicator	1 =0% - <5% poverty 2 =5% - <10% poverty 3 =10% - <20% poverty 4 =20% - 100% poverty 9 =Unknown or not applicable

Socioeconomic Status	Covariate	Categorical	SEER-Medicare cancer file	Census_Tract_2010	Restricted variable
Socioeconomic Status	Covariate	Categorical	SEER-Medicare cancer file	Primary_Payer_at_DX	01= Not insured 02 =Not insured, self-pay 10 =Insurance, NOS 20= Private Insurance: Managed care, HMO, or PPO 21= Private Insurance: Fee-for-Service 31 =Medicaid 35 =Medicaid – Administered through a Managed Care plan 60 =Medicare/Medicare, NOS 61= Medicare with supplement, NOS 62 =Medicare – Administered through a Managed Care plan 63= Medicare with private supplement

					64= Medicare with Medicaid eligibility 65= TRICARE 66 =Military 67= Veterans Affairs 68= Indian/Public Health Service 99 =Insurance status unknown
Primary Site of tumor	Covariate	Categorical	SEER-Medicare cancer file	Primary_Site (Primary site)	ICD-O-3 codes are as follows: C18.0= Cecum C18.1= Appendix C18.2= Ascending colon; Right colon C18.3 =Hepatic flexure of colon C18.4 =Transverse colon C18.5= Splenic flexure of colon C18.6 =Descending colon; Left colon C18.7= Sigmoid colon

	C18.8= Overlapping lesion of colon C18.9= Colon, NOS C19.9= Rectosigmoid junction C20.9 =Rectum, NOS

Table A3

Summary of Dependent Variables in this Study

Variable	Nature	Type	SEER-Medicare file	Variable name
5-YS rate	Dependent	Nominal (Categorical-1/0)	SEER-Medicare Cancer file	The variable was derived. The 5-year survival rate was derived from variable name Year_of_diagnosis. This variable represented the year of diagnosis of cancer. Month_of_diagnosis is the month of diagnosis.

Appendix B: Data Cleaning Procedure

The coding for variables gender, race, ethnicity, marital status at diagnosis was used as it. Coding was done for the following variables.

Initial stage of diagnosis:

1. Stage 0:

   If the initial stage of diagnosis are 00= Stage 0, 01 =Stage 0a, 02 =Stage 0 then the new variable for new initial stage of diagnosis (Variable name: Stage) was coded as "0".

2. Stage I

   If the code are 10= Stage I, 11 =Stage I NOS, 12 =Stage IA, 13= Stage IA1, 14 =Stage IA2, 15= Stage IB, 16 =Stage IB1, 17 =Stage IB2, 18 =Stage IC, 19 =Stage IS then the new variable Stage was coded as "1".

3. Stage II:

   If the stage of diagnosis are 30= Stage II, 31 =Stage II NOS, 32 =Stage IIA, 33 =Stage IIB, 34 =Stage IIC then the new status variable was coded as "2".

4. Stage III:

   If the stage of diagnosis are 50= Stage III, 51= Stage III NOS, 52 =Stage IIIA, 53 =Stage IIIB, 54 =Stage IIIC then new stage variable was coded as "3".

5. Stage IV:

   If the code are 70 =Stage IV, 71 =Stage IV NOS, 72 =Stage IVA, 73 =Stage IVB, 74 =Stage IVC then new stage variable was coded as "4".

Note: Observations with unknown stage was removed from multivariate analysis. 88= Not applicable, 90= Stage Occult, 99 =Stage Unknown, 126 =Blank

Marital status at diagnosis

3 =Separated, 4= Divorced, 5 =Widowed were grouped together and coded as 7. 9= Unknown and 14 =Blank were grouped together and coded as 8. Marital status Blank and Unknown were removed from analysis.

Primary site of tumor:

To stratify CRC as right or left sides, I used a validated approach used by Meguid et al. (2008) and Benedix et al. (2010) on their published paper.

1. Right sided:

   If the ICD-O-3 site codes are 18.0–Cecum, 18.2–Ascending colon, 18.3–Hepatic flexure of colon, and 18.4–Transverse colon then it was right sided.

2. Left sided:

   If the ICD-O-3 site codes are 18.5–Splenic flexure of colon, 18.6–Descending colon, 18.7–Sigmoid colon, and 19.9–Rectosigmoid, and C20.9-Rectum, NOS then it was left sided.

Socioeconomic condition:

Socioeconomic conditions were divided into 4 categories as follows.

1. Poor
2. Near poor
3. Middle
4. High

These categories were based on patient enrollment information and geographic identifiers provided under census tract restricted data. The restricted geographic identifiers that were used were per capita income (PCI), median household income (HHI) of tracts within each state. Other nonrestrictive variables that used are neighborhood poverty level based on the census tract of diagnosis address and patient's enrollment in Medicaid or state buy in program.

1. Poor SES
    - Patient enrolled in Medicaid regardless of census tract residence. The variable in SEER-Medicare cancer file "Primary_Payer_at_DX" if have value of "31" and "35" then was categorized as "Poor" socioeconomic condition.
2. Near- poor SES
    - Patient not enrolled in Medicaid.
    - Patient lived in census with lowest quartile of PCI and HHI
    - Highest quartile of poverty
3. Middle SES
    - Patient not enrolled in Medicaid.
    - Patient did not live in the lowest quartile of PCI and HHI
    - Patients did not live in the highest quartile of poverty.
4. High SES
    - Patient not enrolled in Medicaid.
    - Patient lived in the upper quartile of both PCI and HHI

- Patient lived in the lowest quartile of poverty

Appendix C: Descriptive Statistics Table

For each variable, below format tables were generated.

Table C1

Descriptive Statistics

Variable	Women N (%)	Men N (%)	Total N (%)
Initial Stage at Diagnosis			
xx	xx (xx.x)	xx (xx.x)	xx (xx.x)
xx	xx (xx.x)	xx (xx.x)	xx (xx.x)
xx	xx (xx.x)	xx (xx.x)	xx (xx.x)
xx	xx (xx.x)	xx (xx.x)	xx (xx.x)
Marital Status at Diagnosis			
Single	xx (xx.x)	xx (xx.x)	xx (xx.x)
Married	xx (xx.x)	xx (xx.x)	xx (xx.x)
Divorced/Widowed/Separated	xx (xx.x)	xx (xx.x)	xx (xx.x)
Xxxxxxx	xx (xx.x)	xx (xx.x)	xx (xx.x)
Race			
White	xx (xx.x)	xx (xx.x)	xx (xx.x)
African American	xx (xx.x)	xx (xx.x)	xx (xx.x)

Asian	xx (xx.x)	xx (xx.x)	xx (xx.x)
xxxx	xx (xx.x)	xx (xx.x)	xx (xx.x)
xxxx	xx (xx.x)	xx (xx.x)	xx (xx.x)
xxxx	xx (xx.x)	xx (xx.x)	xx (xx.x)
Primary site of Tumor			
Right sided	xx (xx.x)	xx (xx.x)	xx (xx.x)
Left sided	xx (xx.x)	xx (xx.x)	xx (xx.x)
Socio economic condition			
Poor	xx (xx.x)	xx (xx.x)	xx (xx.x)
Near poor	xx (xx)	xx (xx.x)	xx (xx.x)
High	xx (xx.x)	xx (xx.x)	xx (xx.x)
Medium	xx (xx.x)	xx (xx.x)	xx (xx.x)

Appendix D: Kaplan-Meier Estimates Details

KM curve is defined as the probability of surviving in each length of time (Goel, Khanna, & Kishore, 2010).

The KM estimates are also called "product limit estimates".
The survival probability at certain time is defined as

$$S_t = \frac{\text{Number of subjects living at the start} - \text{Number of subjects died}}{\text{Number of subjects living at the start}}$$

The patients censored were counted in the denominator. The total probability of survival till the interval was calculated by multiplying all the probabilities of survival at all intervals preceding that time.

To compare if there was a significant difference between two survival curves, log rank test was used

Null hypothesis for the study:

H_0: There is no significant difference in survival among male and female elderly CRC patients with pre-existing depression.

Alternative Hypothesis for this study:

H_a: There is a significant difference in survival among male and female elderly CRC patients with pre-existing depression.

The formula for calculation of log-rank test:

$$\text{Log-rank test statistic} = \frac{(0_1 - E_1)^2}{E_1} + \frac{(0_2 - E_2)^2}{E_2}$$

E1 and E2 are the expected number of events in each group.

O1 and O2 are the total number of observed events in each group.

The calculated value was compared with the critical value (using chi square table) for a degree of freedom equal to one. If the test statistics was less than the critical value for degree of freedom equal to one, there was no significant difference between survival between 2 groups.

The log rank test will show that there is a difference in survival between 2 groups, but it will not estimate the size difference between 2 groups. The log rank test will not allow to test the effect of other independent variables on survival between 2 groups as well. So, Cox proportional hazard model (proportional hazard model) was used to test the effect of other covariate variables on survival of different groups. This test is like a multiple regression model. Univariate and multivariate models was performed in this test. Therefore, it will allow the simultaneous effect of various factors in survival.

The formula for Cox regression model is as follows.

$$h(t) = \log h_0(t) + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_p X_{pi}$$

Where:

- t is the survival time.
- $h(t)$ is the hazard function, determined by a set of p independent variables X_{1i}, X_{2i}, ..., Xpi for i subjects.
- $\beta_1, \beta_2, \ldots, \beta_p$ are the coefficients (also called parameters) which quantify the statistical relationship between the p covariates and the survival (regression coefficients).
- h_0 is the baseline hazard. It corresponds to the value of the hazard if all the X_i are equal to zero.

Hazard ratio was calculated from Cox proportional model. It is defined as the ratio of risk of hazard occurring at any given time when one group compared with another.

Appendix E: Table Layout for Inferential Statistics

Table E1

Table Layout RQ1

Variable	Women	Men
Median Survival	xx	xx
Kaplan Meier Estimates (95% CI) [No. at risk]		
1 year	x.xx (x.xx, x.xx)[N=x]	x.xx (x.xx, x.xx) [N=x]
2 years	x.xx (x.xx, x.xx)[N=x]	x.xx (x. xx, x.xx) [N=x]
3 years	x.xx (x.xx, x.xx) [N=x]	x.xx (x.xx, x.xx) [N=x]
4 years	x.xx (x.xx, x.xx) [N=x]	x.xx (x.xx, x.xx) [N=x]
5 years	x.xx (x.xx, x.xx) [N=x]	x.xx (x.xx, x.xx) [N=x]
Log- rank p value, female Vs Male	x.x	

Table E2

Table Layout RQ2

	Odds Ratio	SE	Wald	df	Sig.	HR	95% Lower CI	95% Upper CI
Initial stage of diagnosis (Stage I as reference)			xx	x	**x.xx**	x.xx	x.x	x.x
Stage II	xx	xx	xx	x	**x.xx**	x.xx	x.x	x.x
Stage III	xx	xx	xx	x	**x.xx**	x.xx	x.x	x.x
Stage IV	xx	xx	xx	x	**x.xx**	x.xx	x.x	x.x

Table E3

Table Layout RQ3

Variable	Women	Men
Initial Stage of Diagnosis- Stage I		
Median Survival	xx	xx
Kaplan Meier Estimates (95% CI) [No. at risk]		
1 year	x.xx (x. xx, x.xx) [N=x]	x.xx (x. xx, x.xx) [N=x]
2 years	x.xx (x. xx, x.xx) [N=x]	x.xx (x. xx, x.xx) [N=x]
3 years	x.xx (x.xx, x.xx)[N=x]	x.xx (x.xx, x.xx) [N=x]
4 years	x.xx (x.xx, x.xx) [N=x]	x.xx (x.xx, x.xx) [N=x]
5 years	x.xx (x.xx, x.xx) [N=x]	x.xx (x.xx, x.xx) [N=x]
Log- rank p value, female Vs Male	x.x	
Initial Stage of Diagnosis- Stage III		
Median Survival	xx	xx
Kaplan Meier Estimates (95% CI) [No. at risk]		
1 year	x.xx (x.xx, x.xx) [N=x]	x.xx (x.xx, x.xx) [N=x]
2 years	x.xx (x.xx, x.xx) [N=x]	x.xx (x.xx, x.xx)[N=x]
3 years	x.xx (x.xx, x.xx)[N=x]	x.xx (x.xx, x.xx)[N=x]
4 years	x.xx (x.xx, x.xx) [N=x]	x.xx (x.xx, x.xx) [N=x]
5 years	x.xx (x.xx, x.xx) [N=x]	x.xx (x.xx, x.xx) [N=x]
Log- rank p value, female Vs Male	x.x	
……		

Table E4

Table Layout for RQ3

Variable	Women	Men
Married		
Median Survival	xx	xx
Kaplan Meier Estimates (95% CI) [No. at risk]		
1 year	x.xx (x.xx, x.xx) [N=x]	x.xx (x.xx, x.xx) [N=x]
2 years	x.xx (x.xx, x.xx) [N=x]	x.xx (x.xx, x.xx) [N=x]
3 years	x.xx (x.xx, x.xx) [N=x]	x.xx (x.xx, x.xx) [N=x]
4 years	x.xx (x.xx, x.xx) [N=x]	x.xx (x.xx, x.xx) [N=x]
5 years	x.xx (x.xx, x.xx) [N=x]	x.xx (x.xx, x.xx) [N=x]
Log- rank p value, female Vs Male	x.x	
Single		
Median Survival	xx	xx
Kaplan Meier Estimates (95% CI) [No. at risk]		
1 year	x.xx (x.xx, x.xx) [N=x]	x.xx (x.xx, x.xx) [N=x]
2 years	x.xx (x.xx, x.xx) [N=x]	x.xx (x.xx, x.xx) [N=x]
3 years	x.xx (x.xx, x.xx) [N=x]	x.xx (x.xx, x.xx) [N=x]
4 years	x.xx (x.xx, x.xx) [N=x]	x.xx (x.xx, x.xx) [N=x]
5 years	x.xx (x.xx, x.xx) [N=x]	x.xx (x.xx, x.xx) [N=x]
Log- rank p value, female Vs Male	x.x	

Same for another group…

Table E5

Table Layout for RQ4

Independent variables/ covariates	Coefficient	Standard error	P value	HR	95 % CI(Lower)	95 % CI(Upper)
Gender (0: Male, 1: Female)	x.xxx	x.xxx	x.xxx	x.xxx	x.xxx	x.xxx
Initial Diagnostic Stage						
Stage 2	x.xxx	x.xxx	x.xxx	x.xxx	x.xxx	x.xxx
Stage 3	x.xxx	x.xxx	x.xxx	x.xxx	x.xxx	x.xxx
Stage4	x.xxx	x.xxx	x.xxx	x.xxx	x.xxx	x.xxx
Marital Status at Diagnosis (Married as reference)						
Single	x.xxx	x.xxx	x.xxx	x.xxx	x.xxx	x.xxx
Widow	x.xxx	x.xxx	x.xxx	x.xxx	x.xxx	x.xxx
XXX	x.xxx	x.xxx	x.xxx	x.xxx	x.xxx	x.xxx
Race (White as reference)						
African American	x.xxx	x.xxx	x.xxx	x.xxx	x.xxx	x.xxx
Asian	x.xxx	x.xxx	x.xxx	x.xxx	x.xxx	x.xxx

XXX	x.xxx	x.xxx	x.xxx	x.xxx	x.xxx	x.xxx
Ethnicity (Non-Spanish-Hispanic-Latino as reference)						
Spanish-Hispanic-Latino	x.xxx	x.xxx	x.xxx	x.xxx	x.xxx	x.xxx
Socioeconomic condition (low as reference)						
Medium	x.xxx	x.xxx	x.xxx	x.xxx	x.xxx	x.xxx
High	x.xxx	x.xxx	x.xxx	x.xxx	x.xxx	x.xxx
Primary Site of tumor (right sided as reference)						
Left sided	x.xxx	x.xxx	x.xxx	x.xxx	x.xxx	x.xxx

Appendix F: Detail on Cox Proportional Hazard Model

The formula for Cox regression model is as follows.

$$h(t) = \log h_0(t) + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_p X_{pi}$$

Where:

- t is the survival time.
- $h(t)$ is the hazard function, determined by a set of p independent variables and covariates $X_{1i}, X_{2i}, \ldots, X_{pi}$ for i subjects.
- X_{1i}=Initial Stage of Diagnosis
- X_{2}i= marital status at diagnosis
- X_{3i}=Ethnicity
- X_{4i}=Socioeconomic condition
- X_{5i}=primary site of tumor
- X_{6i}=Race
- $\beta 1, \beta 2, \ldots, \beta p$ are the coefficients (also called parameters) which quantify the statistical relationship between the p covariates and the survival (regression coefficients).
- $h0$ is the baseline hazard. It corresponds to the value of the hazard if all the Xi are equal to zero.

The positive regression coefficient denoted the higher hazard and negative coefficient indicates lower hazard. The coefficients were presented with standard error (SE) which is the measure of uncertainty of regression coefficient. Hazard ratio (HR) denotes the effect of predictors.

If the HR is equal to 1, it means both groups experienced an equal number of events. If HR is greater than 1, it means the number of times the group experienced the event compared to the reference group. If HR is less than 1 it means that the risk is less in the

group compared to the reference group. The statistical significance of the parameter and confidence interval also needs to be considered. 95 % CI means that if the estimated process is repeated an infinite number of times, 95% of time the interval will contain the parameter value. If the CI does not contain value 1, then the association between 5-YS and covariates as well as independent variables will be statistically significant at level of significance alpha is 0.05