A Cross Sectional Study of Socioeconomic Trends in the Colorectal Cancer Screening Population in the United States

by
Ifeyinwa Okoye

Dissertation Submitted to Rutgers University
Rutgers School of Health Professions
In Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Biomedical Informatics

Department of Health Informatics
August 2016
Final Dissertation Approval Form

A Cross Sectional Study of Socioeconomic Trends in the Colorectal Cancer Screening Population in the United States

by
Ifeyinwa Okoye

Dissertation Committee:
Syed Haque PhD
Shankar Srinivasan PhD
Masayuki Shibata PhD

Approved by the Dissertation Committee:

________________________________________________________________________ Date: ____________

________________________________________________________________________ Date: ____________

________________________________________________________________________ Date: ____________

________________________________________________________________________ Date: ____________

________________________________________________________________________ Date: ____________

________________________________________________________________________ Date: ____________
ABSTRACT

Colorectal cancer is the 3rd most prevalent type of non-cancer, and the 2nd leading cause of cancer-related death in both men and women in the United States. This is despite being one of the most preventable and curable cancer types, when detected early. While the incidence and mortality from the disease has been declining over the past decade, its decline can be further accelerated by improving screening rates in order to identify the disease at the earliest stage, when the cure rate is at its highest.

This goal of this study was to identify the socioeconomic attributes of the colorectal cancer screening population, and to assess if any socioeconomic attribute statistically increases the likelihood that a patient tests positive at screening. This was achieved by stratifying socioeconomic attributes, and the colorectal cancer screening population data from a national clinical diagnostics lab for patients screened in a 6 year period, beginning in 2012 through 2015. Some of the key findings from the study are outlined below:

- There was a significant increase in CRC screening volume over the 6 years studied. This was an encouraging trend that shows a possible increase in CRC screening benefits awareness, which is pivotal in the ongoing effort to reduce the incidence and mortality rates from the disease.
- The population median income showed a decline over the 6 year studied, even as the CRC screening volume grew over the same period. This was an encouraging finding because it was indicative of the less affluent becoming more aware of CRC screening, as well as having better access to screening.
- There was an inverse relationship between the population’s median income and the positivity rate; as the population’s median income increased, the positivity rate declined.
Poverty rate in the CRC screening population grew over the study period, in support of the trend observed with the population median income attribute. Another indication that CRC screening was becoming more accessible to the poorer population.

The screening population’s bachelor’s degree attainment rate remained relatively stable of the 6 year study period, even as the CRC screening volume grew. Bachelor’s degree attainment had an inverse relationship with positivity rate.
# TABLE OF CONTENTS

ABSTRACT ........................................................................................................................................... iii

TABLE OF CONTENTS ......................................................................................................................... v

LIST OF FIGURES ............................................................................................................................... viii

LIST OF TABLES ................................................................................................................................... xi

CHAPTER 1 INTRODUCTION ................................................................................................................ 1

1.1. Introduction .................................................................................................................................. 1

1.2. Causes and Risk Factors for Colorectal Cancer ........................................................................... 2

1.3. Types of Colorectal Cancer ......................................................................................................... 9

1.4. Impact of Colorectal Cancer ....................................................................................................... 10

1.5. Improving the 5 Year Survival Rates .......................................................................................... 12

1.6. Goals and Objectives .................................................................................................................. 13

1.7. Significance of the Problem ......................................................................................................... 14

1.8. Relevance to Biomedical Informatics ......................................................................................... 17

1.9. Research Questions and Hypothesis ......................................................................................... 17

CHAPTER 2 LITERATURE REVIEW ..................................................................................................... 19

2.1 Colorectal Cancer and Population Socioeconomic Attributes .................................................. 19

2.2 Colorectal Cancer Screening ...................................................................................................... 23

2.2.1 Adult Colorectal Cancer Risk Assessment ............................................................................. 25

CHAPTER 3 METHODOLOGY ............................................................................................................. 31
3.1 Available Data Variables

3.1.1 LIS Clinical Diagnostics Available Data Variables

3.1.2 Zip Code to ZIP Code Tabulation Area (ZCTA) Crosswalk File

3.1.3 US Census Bureau’s Demographic Dataset

3.1.4 Age-Adjustment Weight Data Table

3.2 Analyses Tools

3.3 Research Design

3.3.1 Clinical Diagnostics Dataset

3.3.2 US Census Demographic Data

3.3.3 Zip Code to ZIP Code Tabulation Area (ZCTA) Crosswalk File

3.4 Study Variables

3.4.1 Core Clinical Diagnostics Study Variables

3.4.2 Core Socioeconomic Variables

3.4.3 Socioeconomic Variable Derivations

3.5 Statistical Tests

3.4 Data Preparation

CHAPTER 4 RESULTS

4.1 Clinical Diagnostics Data Distribution by Year

4.2 Trend and Descriptive Analysis Results for the Screening Population’s Median Income
4.3. Trend and Descriptive Analysis Results for the Screening Population’s Poverty Rates 56

4.3.1. Poverty Rate Attributes Distribution by Year...........................................58

4.3.2. Poverty Rate Summaries for the Combined 6 Years Studied.......................65

4.4. Trend and Descriptive Analysis Results for the Screening Population’s High School Graduation Rates .......................................................................................................................68

4.4.1. High School Graduation Rate Attributes Distribution by Year.................70

4.4.2. High School Graduation Summaries for the Combined 6 Years Studied ......77

4.5. Trend and Descriptive Analysis Results for Bachelor’s Degree Attainment Rate...80

4.5.1. Bachelor’s Degree Attainment Rate Attributes Distribution by Year.........82

4.5.2. Bachelor’s Attainment Rate Summaries for the Combined 6 Years Studied ...89

4.6. Correlation Analysis of the Screening Positivity Rates and the Socioeconomic Attributes .....................................................................................................................................92

4.7. Multiple Regression Analysis of the Four Socioeconomic Attributes as Predictors of a Population’s CRC Screening Positivity Rate.....................................................................92

4.8. ANOVA Analysis on Positivity Rates by Year .................................................95

CHAPTER 5 CONCLUSIONS, STUDY LIMITATIONS AND NEXT STEPS ..........97

5.1. Conclusions ..................................................................................................................97

5.2. Study Limitations ........................................................................................................99

5.3. Next Steps...................................................................................................................101

REFERENCES.......................................................................................................................102
LIST OF FIGURES

Figure 2.1. SEER 9 Incidence & U.S. Mortality 1992-2013, All Races, Both Sexes. Rates are Age-Adjusted. .................................20

Figure 2.2. Incidence per 100,000 Persons by Race/Ethnicity & Gender: Colorectal Cancer. SEER 18 2009-2013, Age-Adjusted. .................................21

Figure 2.3. Deaths per 100,000 Persons by Race/Ethnicity & Gender: Colorectal Cancer. SEER 18 2009-2013, Age-Adjusted. .................................22

Figure 2.4. Percentage Distributions of New Cases by Age Group. SEER 18 2009-2013, All Races, Both Sexes. .................................23

Figure 2.5. Side-by-side comparison CRC screening types by estimated life years gained. .................................29

Figure 2.6. Side-by-side comparison CRC screening types by estimated per 1,000 deaths averted. .................................30

Figure 4.1. CRC screening volume distribution histogram by year. ..............................................43

Figure 4.2. Combined 6 Year (2010 – 2015) Study Period Median Income Distribution and Probability Plot. ..............................................45

Figure 4.3. 2010 Median Income Distribution and Probability Plot. ..............................................46

Figure 4.4. 2011 Median Income Distribution and Probability Plot. ..............................................47

Figure 4.5. 2012 Median Income Distribution and Probability Plot. ..............................................48

Figure 4.6. 2013 Median Income Distribution and Probability Plot. ..............................................49

Figure 4.7. 2014 Median Income Distribution and Probability Plot. ..............................................50

Figure 4.8. 2015 Median Income Distribution and Probability Plot. ..............................................51

Figure 4.9. Graphical overlay of the population’s, mean and median income over the screening population by year. ..............................................52
Figure 4.10. US heat map above shows the geographical heat distribution by median income bucket, of the screened patients included in the study 6 year cross sectional study. ..................53

Figure 4.11. Graphical representation of screening population’s median income by size in relationship to positivity rate ......................................................................................................................54

Figure 4.12. Combined 6 Year (2010 – 2015) Study Period Poverty Rate Distribution and Probability Plot ........................................................................................................................................................................................................56

Figure 4.13. 2010 Poverty Rate Distribution and Probability Plot ........................................58

Figure 4.14. 2011 Poverty Rate Distribution and Probability Plot ........................................59

Figure 4.15. 2012 Poverty Rate Distribution and Probability Plot ........................................60

Figure 4.16. 2013 Poverty Rate Distribution and Probability Plot ........................................61

Figure 4.17. 2014 Poverty Rate Distribution and Probability Plot ........................................62

Figure 4.18. 2015 Poverty Rate Distribution and Probability Plot ........................................63

Figure 4.19. Graphical overlay of the population’s, poverty rate over the screening population by year ...........................................................................................................................................................................64

Figure 4.20. US heat map above shows the geographical heat distribution by poverty rate bucket, of the screened patients included in the study 6 year cross sectional study. ..................65

Figure 4.21. Graphical representation of screening population’s poverty rate by size in relationship to positivity rate ...........................................................................................................................................................................66

Figure 4.22. Combined 6 Year (2010 – 2015) Study Period High School Graduation Rate Distribution and Probability Plot ...........................................................................................................................................................................68

Figure 4.23. 2010 High School Graduation Rate Distribution and Probability Plot.............70

Figure 4.24. 2011 High School Graduation Rate Distribution and Probability Plot.............71

Figure 4.25. 2012 High School Graduation Rate Distribution and Probability Plot.............72

Figure 4.26. 2013 High School Graduation Rate Distribution and Probability Plot.............73
Figure 4.27. 2014 High School Graduation Rate Distribution and Probability Plot.............74
Figure 4.28. 2015 High School Graduation Rate Distribution and Probability Plot.............75
Figure 4.29. Graphical overlay of the population’s, high school graduation rate over the screening population by year .............................................................76
Figure 4.30. US heat map above shows the geographical heat distribution by high school graduation rate bucket, of the screened patients that met the criteria for inclusion in the 6 year cross sectional study. .............................................................77
Figure 4.31. Graphical representation of screening population’s high school graduation rate by size in relationship to positivity rate.............................................................78
Figure 4.32. Combined 6 Year (2010 – 2015) Study Period Bachelor’s Degree Attainment and Probability Plot Rate Distribution .............................................................80
Figure 4.33. 2010 Bachelor’s Degree Attainment Rate Distribution and Probability Plot......82
Figure 4.34. 2011 Bachelor’s Degree Attainment Rate Distribution and Probability Plot......83
Figure 4.35. 2012 Bachelor’s Degree Attainment Rate Distribution and Probability Plot......84
Figure 4.36. 2013 Bachelor’s Degree Attainment Rate Distribution and Probability Plot......85
Figure 4.37. 2014 Bachelor’s Degree Attainment Rate Distribution and Probability Plot......86
Figure 4.38. 2015 Bachelor’s Degree Attainment Rate Distribution and Probability Plot......87
Figure 4.39. Graphical overlay of the population’s, bachelor’s degree attainment rate over the screening population by year .............................................................88
Figure 4.40. US heat map above shows the geographical heat distribution by bachelor’s degree attainment rate bucket, of the screened patients that met the criteria for inclusion in the 6 year cross sectional study .............................................................89
Figure 4.41. Graphical representation of screening population’s bachelor’s degree attainment rate by size in relationship to positivity rate.............................................................90
LIST OF TABLES

Table 1.1 Cancer Survival Rates by Stage at Diagnosis ..........................................................13
Table 2.1. US Preventive Services Task Force Colorectal Cancer Screening Recommendation
29 ..................................................................................................................................................27
Table 2.2. Summary of Colorectal Cancer Screen Types 43 42 ..................................................28
Table 3.1. ZCTA file select variables with description ...............................................................34
Table 3.2. Age adjustment factors by age group .......................................................................35
Table 4.1. CRC screening with positivity rate summary by year .............................................43
Table 4.2. Distribution of CRC screens with a resolved median income attribute by year ....44
Table 4.3. Tabular distribution of the screening rate by median income bucket .....................55
Table 4.4. Tabular distribution of the positivity rate by median income bucket .....................55
Table 4.5. Distribution of CRC screens with a resolved poverty rate attribute by year ..........57
Table 4.6. Tabular distribution of the screening rate by poverty rate bucket .........................67
Table 4.7. Tabular distribution of the positivity rate by poverty rate bucket ............................67
Table 4.8. Distribution of CRC screens with a resolved high school graduation rate attribute
by year ...........................................................................................................................................69
Table 4.9. Tabular distribution of the screening rate by high school graduation rate bucket ..79
Table 4.10. Tabular distribution of the positivity rate by high school graduation rate bucket 79
Table 4.11. Distribution of CRC screens with a resolved bachelor’s degree attainment rate
attribute by year ..............................................................................................................................81
Table 4.12. Tabular distribution of the screening rate by bachelor’s degree attainment bucket
.....................................................................................................................................................91
Table 4.13. Tabular distribution of the positivity rate by bachelor’s degree attainment rate
bucket .............................................................................................................................................91
Table 4.14. Correlation analysis summaries between each of the four socioeconomic attributes and positivity rate.................................................................92

Table 4.15. Stepwise multiple regression summary by the 4 socioeconomic attributes studied, and the age attribute........................................................................................................93

Table 4.16. Stepwise multiple regression parameter estimates by the 4 socioeconomic attributes studied, and the age attribute ........................................................................................................93

Table 4.17. Waller-Duncan positivity rate mean grouping by year.................................................................95

Table 4.18. ANOVA analysis paired years means comparisons .................................................................96
CHAPTER 1 INTRODUCTION

1.1. Introduction

Colorectal cancer is used to refer to any neoplastic growth occurring anywhere along the luminal surface of the large bowel. About 41% of cases originate in the colon, 32% in the rectosigmoid and rectum, 27% in the cecum and ascending colon, and 12% in the transverse colon. Colorectal cancer is the 4th most common newly diagnosed type of non-skin cancer in the United States (US) after prostate, breast and lung cancers, and represents 9% of all newly diagnosed cases for both men and women, and the 3rd most prevalent type of cancer. About 5% of the US population will develop invasive colon or rectal cancer at some point in their lifetime; about 40% of these cases will be diagnosed at the localized stage, at which time the 5 year survival rate is approximately 90%, about 36% will be diagnosed when the disease is still regionalized and the 5 year survival rate is at 70%, and about 20% of cases will be diagnosed when the disease has already reached the metastatic stage at which time their survival prognosis is extremely poor and their survival years are typically discussed in terms of months. The patient’s age and tumor stage at the time of diagnosis are the strongest predictors for cure.

On an annual basis, over 160,000 new cases are reported, with about 55,000 deaths being directly attributed to colorectal cancer. That said, because the disease develops slowly, it has a significantly better chance of being prevented or detected early, in comparison to other types of cancer.
1.2. Causes and Risk Factors for Colorectal Cancer

The exact trigger for the development of the polyps that lead to colorectal cancer are not very well understood, but it is believed to develop as a result of changes to the DNA, that can be attributed to environmental factors. Some cases occur spontaneously, and others as a result of inherited genes. Most colorectal cancers develop slowly over several years. Prior to the onset of cancer, a growth of tissue or tumor usually begins as a non-cancerous adenomatous polyp on the inner lining of the colon or rectum, and then slowly increases in size, followed by dysplasia and finally cancer. A tumor is abnormal tissue and can be benign (not cancerous) or malignant (cancerous). A polyp is a benign, non-cancerous tumor. Some polyps can develop into cancer but not all do. The odds of developing into a cancer depend on the type of polyp. There are two types of polyps; adenomatous and hyperplastic polyps.

- Adenomatous polyps (adenomas) are polyps with the ability to develop into cancer, so as such are referred to as a pre-cancerous tumor.

- Hyperplastic polyps also known as inflammatory polyps are typically not pre-cancerous, though some researchers believe it could be a potential indicator for a higher risk of developing an adenomatous polyp and cancer down the line. This is especially so, if the hyperplastic polyp develops in the ascending colon.

Carcinoid tumors originate in hormone producing cells (neuroendocrine system) in the intestine. The digestive system has more neuroendocrine cells than any other part of the body, and is also known to be the most frequent origination site for most carcinoid tumors. Gastrointestinal stromal tumors are very rare, and originate from special cells found in the wall of the GI tract, known as interstitial cells of Cajal (ICCs). These cells are part of the autonomic nervous system, responsible for digestion.
90% of all colorectal cases occur after age 50, and the peak incidence age is about 70. The disease is more predominant in men than women worldwide, but in the US, the distribution appears to equal for both genders. In addition to age and gender, there are a myriad of other factors that are known to predispose individuals to developing colorectal cancer. Broadly speaking, the causes of colorectal cancer can be classified as genetic and non-genetic. The genetic factors typically occur as a result of germline mutations that can be traced back to familial lineage. The non-genetic colorectal cancer incidents are considered to be sporadic in nature and influenced by factors such as the environment, diet, personal history of adenoma or carcinoma, existence of predisposing diseases, such as irritable bowel disease (IBD), and compromised immune system. These predisposing factors will be explored in the ensuing section.

1.2.1. Genetic Predisposing Factors to Colorectal Cancer

Research has shown that 5% to 6% of all colorectal cancer cases arise due to germline mutations. Familial colorectal cancer occurs as a result of an altered genome at birth, which with the right environmental trigger metamorphoses into a malignant phenotype.

1.2.1.1. Hereditary Polyposis Syndrome

Hereditary polyposis syndrome is an example of a genetically predisposing factor to colorectal cancer, and accounts for about 1% of all colorectal cancer cases. It is an autosomal dominant trait that manifests as colonic adenomas in the hundreds to thousands, and sometimes occurs with extracolonic tumors. These adenomas typically precede the onset of cancer by about 10 years, and the eventual development of cancer for the carrier of this trait is almost always certain.
1.2.1.2. Lynch Syndrome

Lynch syndrome is the most common \textsuperscript{12} genetically predisposing factor to colorectal cancer, and accounts for about 6\% of all cases. Lynch syndrome related colorectal cancer begins to manifest between the ages of 40 and 50 – about two decades prior to the non-lynch syndrome related colorectal cancers. It is also an autosomal dominant condition, for which 80\% of cases are due to germline mutations of genes that are responsible for repairing DNA mismatches during DNA replication \textsuperscript{4}.

Lynch syndrome is caused by a mutation in one of four mismatch repair (MMR) genes, MLH1, MSH2, MSH6, and PMS2. Deletions in the EPCAM gene which is located in the MSH2 can also give rise to Lynch syndrome due to its muting of MSH2 functions. Greater than 80\% of Lynch syndrome cases are as a result of mutations in the MLH1 and MSH2 germlines, and 10\% of cases are due to mutations in the MSH6 germline. Mutations in the PMS2 germline account for 2\% to 3\% of all Lynch syndrome cases \textsuperscript{12}.

Lynch syndrome patients typical have discrete polyps, as opposed to the hundreds to thousands of polyps seen in patients with polyposis syndrome. There are two types of lynch syndrome, namely Lynch’s syndrome I, in which colorectal cancer manifestation is limited to the colon and rectum areas and Lynch’s syndrome II, in which female members of the family tend to develop cancer in the female genital tract, as well as other sites.

1.2.2. Other Predisposing Factors

While it is best to assume that all cancers have a genetic component that may have either been inherited or acquired to some degree \textsuperscript{4,12}, certain “non-genetic” factors are also known to act as triggers for the inherited cancer traits. Some of these “non-genetic” factors include age, personal or family history of sporadic colorectal cancer, irritable bowel syndrome,
including ulcerative colitis and Crohn’s disease, diet, frequent use of antibiotics, and female history of endometrial or ovarian cancer. The likelihood of developing colorectal cancer as a result of the “non-genetic” factors is higher in the older population of ages 50 and over, more so than the younger population.

1.2.2.1. Age

Research has shown that 90% of colorectal cancer incidents occur in patients over 50 years of age, with the peak age at about 70 years in the general population. A 50 year old who lives to age 80 has a 5% chance of developing the disease, and a 2.5% chance of dying from it.

1.2.2.2. Personal or Family History of Sporadic Colorectal cancer

An individual’s odds of developing colorectal cancer increases two folds if one first degree relative is diagnosed with the disease, and three to four folds if two or more first degree relative are diagnosed with the disease or a first degree relative acquires the disease at an early age.

A previous personal history of colorectal cancer also predisposes one to developing the disease again. The incidence of one person developing more than one carcinoma is stated to be between 2% and 6%. This can occur in one of two forms – the synchronous carcinoma, in which the second carcinoma develops almost concurrently as the first, or metachronous carcinoma, in which the second carcinoma clearly develops after a previously diagnosed carcinoma. The time lapse from the onset of carcinomas for metachronous cases is about 5 to 7 years, and the reoccurrence site is usually different from the site of the initial diagnosis. The typical pattern for synchronous carcinoma is the existence of one lesion each in the distal and proximal colon regions, with differing degrees of invasion. The most advanced lesion is used to advise on the patient’s prognosis.
1.2.2.3. Diet and Lifestyle

Diet is believed to play some role in triggering the development of colorectal cancer in some patients. Specifically, a Western diet which is typically rich in animal fat, red and processed meat, processed grain, high in sugar and poor in fibers is strong risk factor for colorectal cancer. Research suggests that dietary patterns rich in red meat, processed meats and sugar promote the development of colorectal cancer.

Diets rich in red meat, as well as the excessive consumption of alcohol, smoking, and an inactive lifestyle are all associated with increased risk of developing colorectal cancer.

A study sponsored by the World Cancer Research Fund (WCRF) and the American Institute for Cancer Research (AICR) concluded red meat consumption significantly increased the risk of developing colorectal cancer. It put the relative risk of developing colorectal cancer at 1.17 for every 100g of red meat consumed per day, and an even higher relative risk of 1.24 for every 50g of processed meat consumed a day. It is believed that the oncogenic effects of heme iron, heterocyclic amines and polycyclic aromatic hydrocarbons found on the surface of well done and charred meat may be responsible for the higher than average relative risk of consuming red meat, and the nitrates typically used in preserving processed meat which subsequently converts to carcinogenic nitrosamines is responsible for the carcinogenic tendencies of processed meat. More studies are ongoing to better understand these theories.

A year 2000 study concluded that people with a family history of Colorectal Cancer, who consumed a predominantly Western diet, were at an increased risk of developing Colorectal Cancer themselves. Statistically significant differences were observed in how the effect of Western diet is influenced by age. Study participants age 56 and older with a family history of colorectal cancer showed a higher likelihood of developing the disease if their Western diet
included a high component of eggs, refined grains and red meat. For patients with a family history of colorectal cancer, a Western diet rich in high sugar content was the influencer.\textsuperscript{18}

1.2.2.4. Frequent Use of Antibiotics

A retrospective study using insurance claims over a five year period showed that there was an increased likelihood of developing colorectal cancer with an increase in the frequency of using antibiotics.\textsuperscript{14} The thought here is that the frequent use of antibiotics creates a microbial imbalance in the colon, which can lead to inflammatory conditions that increase the predisposition to developing colorectal cancer.\textsuperscript{19}

Imbalances in the colonic microbiota disrupt the digestive process, which includes the fermentation of undigested carbohydrates from fibers into short-chain fatty acids. These fatty acids are the preferred energy source of the colon mucosa, and have anti-inflammatory, anti-carcinogenic, and anti-proliferative properties.\textsuperscript{14}

1.2.2.5. Female History of Endometrial or Ovarian Cancer

Endometrial cancer is the most common gynecologic cancer in the United States, with upwards of 40,000 cases diagnosed on an annual basis. It is the most common type of extra-colonic tumor in patients with Lynch syndrome II/hereditary non-polyposis colorectal cancer, occurring at a rate of up to 43\% in women from affected families by the time they are 70 years of age.\textsuperscript{19,20}

Links have been found to exist between the development of colorectal cancer and endometrial cancer. Though no genetic disorder is found in majority of the cases, 10\% of cases in which there is a history of both endometrial and colorectal are believed to be attributable to Lynch syndrome II/hereditary non-polyposis colorectal cancer. On the other hand, just 2\% to 3\% of all colorectal cancers are attributable to hereditary non-polyposis. A
A retrospective study conducted using the electronic medical records of 267 women with a history of endometrial cancer, in order to understand their colorectal screening compliance pattern found that fewer than 50% of these patients followed up with screening for colorectal cancer; and of the 127 who did, 13.6% had either adenomatous polyps or tubulovillous polyps, and 2.4% had colorectal cancer.

The good news is that colorectal cancer that develops as a result of hereditary non-polyposis is less likely to metastasize, and has better survival rate in comparison to the sporadic type.

1.2.2.6. **Inflammatory Bowel Disease (IBD)**

Inflammatory bowel disease (IBD) is a chronic inflammatory condition of all or part of the digestive system. The term is used to broadly describe about four specific gastrointestinal conditions, including Crohn’s disease, ulcerative colitis, collagenous colitis and lymphocytic colitis — though less so for the latter two. IBD is the most prevalent functional disorder of the gastrointestinal system. The condition most commonly manifests by signs of abdominal pain, discomfort, bloating and changes in bowel habits with no known diagnosis of a structural or biochemical abnormality within the gastrointestinal system. It is believed to have a 5% to 24% prevalence rate in Western nations, depending on the classification criteria used, and accounts for about 30% of all visits to gastroenterologists. There have been various studies that attempt to quantify the incidence rate of developing colorectal cancer following a diagnosis for irritable bowel syndrome, with one of the reviewed studies putting the incidence rate of developing colorectal cancer at 5.8% within the first year of diagnosis.

Following a diagnosis with IBD, patients undergo follow up colonoscopies which is used to assess the patient’s risk for developing colorectal cancer based on the gastrointestinal
histologic inflammation. A colonoscopy schedule is then recommended for the patient based on the assessed risk findings and the most current American Gastroenterological Association guidelines, with changes in colonoscopy intervals being made if necessary, based on the histologic findings. 25 26.

1.3. Types of Colorectal Cancer

Colorectal cancers fall into five different types of cancers, including adenocarcinomas, lymphomas, sarcomas, carcinoid tumors and gastrointestinal stromal tumors, with adenocarcinomas making up the vast majority of all colorectal cancer cases. Adenocarcinomas account for 95% of all colorectal cancers. They typically originate in cells responsible for forming glands that create the colon and rectum inner lining lubricating mucus. 27.

Lymphomas are cancers that typically originate in cells known as lymphocytes found in lymphoid tissues, which are part of the immune system, but could also originate in the colon, rectum, and other parts of the body. 28.

Sarcomas, which are the rarest colorectal cancer type typically originates in the muscle and connective tissues of the walls of the colon and rectum, but can also originate in the blood vessels. 29.

Colorectal cancer is a preventable disease that still persists, though both the new incidence and mortality rate have been on a decline since 1975, albeit at a slower rate in the last decade, though populations with lower socioeconomic attributes continue to have a disproportionally higher incidence and mortality rate as a result of the disease. 30 31 32.

Historically, low-income, less educated, and poor populations are more likely than their counterparts not to adhere to the recommended screening schedule, and more likely to be
diagnosed at a more advanced stage which limits their odds for survival. Understanding the reason behind why there continues to be new incidents of the disease, as well as mortality associated with the disease requires a population-based research to help develop insights into why the disease continues to persist. A good understanding of who develops the disease, where they are, at what stage they were diagnosed, and whether they followed the screening recommendations, etc., can help healthcare policy makers better understand the types of barriers that stand in the way of the disease being prevented or diagnosed at the earliest stage when the survival rate is at its highest, and enable them to create policies to address those barriers.

Colorectal cancer awareness, including the importance of understanding one’s risk or disposition to developing the disease, and following through with the screening regimen and schedule for one’s risk status is a verified way of reducing the incidence and mortality rates within populations.

1.4. Impact of Colorectal Cancer

Colorectal cancer is the 3rd most prevalent type of cancer, and the 3rd leading cause of cancer-related death in the United States ². This is despite being one of the most preventable and curable cancer types, when caught early. Americans have a 5% lifetime risk of developing colorectal cancer ⁸.

Significant incidence variations have been observed in countries around the world ⁸. Some of the high incidence countries include North America, countries in Western and Eastern Europe, Japan, Israel, Singapore, Australia, New Zealand, Brazil, Argentina, Hong Kong, and the affluent parts of China and Malaysia ³³. In addition, marked correlation between the incidence of colorectal cancer and the more developed and affluent parts of the
world have been observed. CRC incidence variations have also been shown to exist across
ethnic and racial groups; Hispanics and Asians/Pacific Islanders have the lowest incidence
and mortality rates, while African Americans American and Indians/Alaska Natives have the
highest mortality and incidence rates respectively 1.

Incidence of mortality from the disease is directly related to its prevalence, as well as the
availability of diagnostics services, adequate therapy, co-morbidity 33 and financial resources.
In the United States, disparities in mortality rates have also been observed across
populations, with minority groups such as African American and Alaska Natives showing a
disproportionate mortality to incidence rate in comparison to other groups 34. A number of
developed countries are beginning to show some stabilization and in some cases, even a
decline in both the incidence and related mortality rate for colorectal cancer. Early detection
as a result of increased awareness and screening has been credited with the recent declines.
One such example is the United States, which has shown a 2% to 3% rate of decline in
incidence rate over the past 15 years 8. There was an estimated 1.22 million people living
with colorectal cancer in 2010, with an estimated 1.38 million projected to be living with the
disease after adjusting for the US population increase by 2020, if the current incidence and
survival trends continue. This represents a 13% decline in the number of people that will be
living in the United States with colorectal cancer in 2020(1.52 million), if the incidence and
survival remained unchanged at the 2010 levels 34. Worthy of note is that even with a
declining colorectal cancer related mortality rate in the United States, published research still
puts its 2010 estimated health burden for colorectal cancer at $14 billion, and it is expected
to rise to about $17 billion by 2020 35 36.
This mission to increase the colorectal cancer screen rate is particularly important because when colorectal cancer is identified at a localized stage, the 5-year survival rate is typically at 90%, and then declines to 40% at the early cancer stages and further declines to a mere 12% when detected at the late stage. Put differently, screening leads to a reduction in the number of cases found at the point when the cancer has metastasized to distant regions of the body. In addition, screening also helps in the identification and prompt removal of polyps, which if left unaddressed, will almost certainly lead to the development of colorectal cancer. Research has shown that if all adults 50 and older underwent colorectal cancer screening, its associated mortality rate could be cut in half, saving approximately 25,000 lives per year in the United States.

1.5. Improving the 5 Year Survival Rates

A population’s cancer prognosis is usually described using a 5 year survival rate measure. It describes the percentage of a patient population with similar cancer presentation surviving up to 5 years following their initial cancer diagnosis. The earlier the disease is diagnosed and treated following its development, the better the 5 year survival rate.

Colorectal cancer statistics show that since the mid-1980s, the colorectal cancer 5 year survival rate has been increasing. Nevertheless, more still needs to be done in order to improve upon the current 5 year survival rates. The most recent 5 year survival rate statistics compiled from the National Cancer Institute’s SEER database on patients diagnosed with colorectal cancer between 2004 and 2010, sheds some additional insights into how much of a difference the 5 year survival rate can be depending on how early in the colorectal cancer development stage that a diagnosis is made, and indirectly the undeniable importance of
adhering to the colorectal screening guidelines in order to identify the disease at the earliest stage possible.

The 5 year survival rates for colorectal cancer has been split by colon cancer and rectum cancer, and then by stage at the time of initial diagnosis. The table below outlines the 5 year survival distribution.

<table>
<thead>
<tr>
<th>Cancer Stage at Diagnosis</th>
<th>Colon Cancer 5 Year Survival Rate</th>
<th>Rectum Cancer 5 Year Survival Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage I</td>
<td>92%</td>
<td>87%</td>
</tr>
<tr>
<td>Stage IIA</td>
<td>87%</td>
<td>80%</td>
</tr>
<tr>
<td>Stage IIB</td>
<td>63%</td>
<td>49%</td>
</tr>
<tr>
<td>Stage IIIA</td>
<td>89%</td>
<td>84%</td>
</tr>
<tr>
<td>Stage IIIB</td>
<td>69%</td>
<td>71%</td>
</tr>
<tr>
<td>Stage IIIC</td>
<td>53%</td>
<td>58%</td>
</tr>
<tr>
<td>Stage IV</td>
<td>11%</td>
<td>12%</td>
</tr>
</tbody>
</table>

Table 1.1 Cancer Survival Rates by Stage at Diagnosis

1.6. Goals and Objectives

The purpose of this research is to study the screening and positivity rates for colorectal cancer across various socioeconomic groups, in order to identify if socioeconomic attributes play a role in the likelihood of a patient getting screened for colorectal cancer, and the likelihood of a screened patient having a positive colorectal cancer screen result. Some of the questions that we will attempt to gain insights into are as follows:

- Does a population's rate of graduation from high school or Bachelors program have a statistically significant impact on the population's CRC screening rate?
- Does a population's median income have a statistically significant influence on the population's CRC screening rate?
Does a population's poverty rate have a statistically significant influence on the population's CRC screening rate?

- Do the screening results change statistically from year to year?

This will be achieved by analyzing de-identified records for approximately 4 million distinct diagnostics testing encounters for patients who have at least one colorectal cancer related diagnosis on file between January 2010 and December 2015 for screening and incidence rate trends. Given that colorectal cancer is one of the most treatable cancers that there is, if caught at its earliest stages, it is clear that any successful initiative that drives up colorectal cancer screening rates will conversely bring down its incidence and mortality rates.

In order to efficiently execute well-designed colorectal screening awareness programs, identifying target populations and where they are, is critical.

The findings from this research will help health care policy makers, providers and other professionals better understand regions where more colorectal cancer awareness campaigns need to be done, in order to increase colorectal cancer screening rates, thereby accelerating the declining incidence and mortality rates for the disease.

1.7. Significance of the Problem

While the incidence and mortality rates have been on a decline due to growing awareness about colorectal cancer and the benefits of screening for the disease, there is still a need to do all that is possible to eradicate its incidence and mortality from the disease. The issue can be addressed via many facets, from increased awareness, to the identification and removal of barriers to screening, to increased precision in the identification of target audience for more aggressive outreach campaigns, amongst other facets.
Research has shown that variations in age, gender, race, environment, genetics and region, just to name a few are factors that predispose populations to colorectal cancer. Identifying and understanding the contributory effects of these factors will go a long way in helping to target audiences for awareness campaigns. This research will focus on understanding the socioeconomic attributes of the screening population in order to understand socioeconomic populations that are likely to get screened and those likely to have a positive screen result. Findings from this study will serve as a guide in when targeting populations for colorectal cancer screening educational and/or outreach campaigns, thereby maximizing the desired results of such campaigns.

The National Cancer Institute (NCI) estimated that new incidents of colorectal cancer in 2015 made up 8% of all cancers, and ranked at 4th for all cancers/genders. It also placed the mortality rate for colorectal cancer 2nd for all cancers/genders. While there has been a steady decline in both the incidence and mortality rates, there is still a lot of research to be done to understand why this preventable disease continues to rank so high amongst all cancers.

One of the factors that make colorectal cancer both preventable and relatively more curable in comparison to other types of cancer when diagnosed at the earliest stages is that it has a long detectable, preclinical phase; it can take up to 10 years to metamorphose from the preclinical stage to the pathological stage. In addition, there are several safe and accurate diagnostics tests, which if utilized according to the specific test’s screening schedule can prevent colorectal cancer and/or allow for the treatment of colorectal cancer while it is still localized and in its earliest stage when the cure and survival rate is at the most favorable level.
For instance, an individual with an average risk for colorectal cancer, who opts for the non-invasive fecal occult blood screening test, and tests positive for human heme, will typically be recommended for additional screening, typically the more invasive colonoscopy, which uses a scope to screen for the presence of polyps and/or any unusual presentation in the lining of the colon. Patient screening and adherence to the screening guidelines based on the preferred screen test type and by risk level, increases the odds that any abnormal cell growth will be diagnosed at the earliest stage when the survival rate is at its highest.

According to statistics published by the Colon Cancer Alliance, only 40% of all colon colorectal cancers are diagnosed at the localized stage, when it is still confined to the colon or rectum area. 36% of cases are diagnosed at the regionalized stage, when it has begun to spread to adjoining tissues of the colon or rectum where it originated. 20% of cases are diagnosed at the metastatic stage, at which time it has spread to distant organs, and has the poorest prognosis for survival. Increasing awareness in order to increase the percentage of cases diagnosed at the localized stage, and reduce to the barest minimum the percentage of cases that are diagnosed at the regionalized and metastatic stages is pivotal to the goal of improving colorectal cancer incidence and mortality rates. A solid understanding of colorectal cancer trends, including populations at risk, colorectal cancer stage at time of diagnosis, and colorectal cancer screening disparities across socioeconomic populations is pivotal to creating programs designed to increase screening rates, thereby lowering the incidence and mortality rates for the disease.

The results from this cross-sectional study of the impact of socioeconomic attributes on colorectal cancer screening rates in the United States will shed some light on socioeconomic populations that are less likely to get screened and enable additional studies on
understanding some of the screening barriers faced by these socioeconomic groups, thereby
enabling the design of colorectal cancer screening awareness programs that are most likely to
resonate with the specific at-risk socioeconomic group. This targeted awareness messaging
approach will be more likely to increase screening rates.

1.8. Relevance to Biomedical Informatics
The relevance of this study to the field of biomedical informatics is that it advances the body
of knowledge around colorectal cancer, by utilizing health care analytics techniques to better
understand patient screening outcomes and population health trends.

1.9. Research Questions and Hypothesis

Research Question 1 - Does a population's rate of graduation from high school or
Bachelors program have a statistically significant impact on the population's CRC screening
rate?

Hypothesis I:
- Null Hypothesis (H₀: \( \mu_1 = \mu_2 \)) - A population’s rate of graduation from high
  school or Bachelors program will have no statistically significant effect on
colorectal cancer screening using the fecal immunochemistry test method

- Alternative Hypothesis (H₁: \( \mu_1 \neq \mu_2 \)) - A population’s rate of graduation from
  high school or Bachelors program will have a statistically significant effect on
  screening for colorectal cancer

Research Question 2 - Does a population's median income have a statistically significant
influence on the population's CRC screening rate?

Hypothesis II:
- **Null Hypothesis (H₀: μ₁ = μ₂)** - Median income level of a population will have no statistically significant impact on the population’s colorectal cancer screening rate

- **Alternative Hypothesis (H₁: μ₁ ≠ μ₂)** - Median income level of a population will have a statistically significant impact on the population’s colorectal cancer screening rate

**Research Question 3** - Does a population's poverty rate have a statistically significant influence on the population's CRC screening rate?

**Hypothesis III:**

- **Null Hypothesis (H₀: μ₁ = μ₂)** – The poverty rate of a population will have no statistically significant impact on the population’s colorectal cancer screening rate

- **Alternative Hypothesis (H₁: μ₁ ≠ μ₂)** – The poverty rate of a population will have a statistically significant impact on the population’s colorectal cancer screening rate

**Research Question 4** – Do the results change statistically from year to year?

**Hypothesis IV:**

- **Null Hypothesis (H₀: μ₁ = μ₂)** - There will be no statistical difference in CRC screening results from one year to the next

- **Alternative Hypothesis (H₁: μ₁ ≠ μ₂)** - There will be a statistical difference in CRC screening results from one year to the next
CHAPTER 2 LITERATURE REVIEW

The literature search process involved the search of a vast number of scientific databases that Rutgers University students have access privileges to, the websites of authoritative bodies on the subject of colorectal cancer, such as the National Institute of Health (NIH), National Cancer Institute (NCI), and the American Cancer Society. Other sources of information include textbooks on the subject of colorectal cancer.

To ensure the relevance of the research, and given the rate at which new research on the subject has grown in recent years, a conscious effort was made to rely more heavily on peer reviewed articles published within the last 11 years.

Some sample search strings used include “Colorectal Cancer Screening and Socioeconomic Attributes”, “Colorectal Cancer Screening Disparities”, “Colorectal Cancer Incidence and Poverty”, “Colorectal Cancer Prevention”, and “Colorectal Cancer and Education”, just to name a few. The search results were limited to articles published on or after 2005. The article abstracts were then reviewed for inclusion to the list for detailed review based on direct relevance to my research.

2.1 Colorectal Cancer and Population Socioeconomic Attributes

2.1.1 Colorectal Cancer Trends

Colorectal cancer is the 3rd most prevalent type of non-skin cancer, the second leading cause of cancer-related death in both men and women in the United States, and the 4th most newly diagnosed types of cancer. The survival rate for the disease from 2006 through 2012 (the most current 5 years for which data is available) is pegged at 65.1%, an improvement from the previous rate of 64.9% for the period 2005 through 2011; the most differentiating factor for survival being the cancer stage at the time of diagnosis. Cancer
registry data for 2006 through 2012 shows that 39% of cases are diagnosed while the cancer is still localized; the stage at with the 5 year survival rate is at its highest (90.1%) \(^{30}\). A lot of work has been done around increasing awareness for the disease, and publicizing preventive measures that can be taken by at risk patients to reduce the likelihood of developing the disease; more still needs to be done to get this very preventable and curable disease under control. While colorectal cancer disease trends show that there have been remarkable decline in both the incidence and mortality rates for colorectal cancer in the last two decades, some populations still remain at a high risk for developing the disease. According to SEER estimates \(^{30}\), 134,490 or 8% of all new cancer diagnosis in 2016 and 49,490 or 8.3% of cancer related deaths in 2016 will be attributable to cancer.

The table below shows yearly trend for both incidence and mortality from 1992 through 2013.

**New Cases, Deaths and 5-Year Relative Survival Graph**

![Figure 2.1. SEER 9 Incidence & U.S. Mortality 1992-2013, All Races, Both Sexes. Rates are Age-Adjusted \(^{30}\).](image-url)
2.1.2 Colorectal Cancer Prevalence in the United States

2.1.2.1 Race/Ethnicity and Gender

According to various research literatures, including the NIH’s SEER Cancer Statistics Factsheets, men have a higher incidence of colorectal cancer, in comparison to women, at a ratio of 48.9 per 100,000 to 37.1 per 100,000. The mortality rates also mirror the same pattern, with the death per 100,000 at 16.6 and 13.1 for men and women respectively.

As a race, Blacks have the highest incidence for both genders, and are also more likely than other races, to die of the disease. Hispanic women are the least likely to develop the disease, while Asian Pacific women are the least likely to die from it.

The info graphics below provides a snapshot of the gender to race statistics on the incidence and mortality rates per 100,000 persons.

Figure 2.2. Incidence per 100,000 Persons by Race/Ethnicity & Gender: Colorectal Cancer. SEER 18 2009-2013, Age-Adjusted 30
2.1.2.2 Age

Current colorectal cancer screening recommendations for patients at an average risk of developing the disease is to begin routine screening at age 50 and continue through age 75, except for patients with a higher than average risk, for who the current recommendation is to begin at about age 40 or following the incident that created the greater than average risk – whichever occurs first.

The NIH’s SEER Cancer Statistics Factsheets shows that the peak age group for the diagnosing colorectal cancer is between ages 65 and 74; about 23.9% of all cases are diagnosed for patients who fall in this age bracket, with patients age 68 at the median of the distribution. See the graph below for the age group distribution for all newly diagnosed cases of colorectal cancer.
2.2 Colorectal Cancer Screening

The benefits of colorectal cancer screening as a tool to identify the disease in its earliest stage, as well as reduce the mortality rate has been well documented in many peer reviewed literatures, and yet the colorectal cancer screening rate lag behind what is considered an effective rate in the United States. It enables the early detection and removal of premalignant adenomatous polyp, as well as localized cancer, resulting in higher survival rates.

Various authoritative bodies on the subject of colorectal cancer screening including the American Cancer Society, the U.S. Preventive Services Task Force (USPSTF), the U.S. Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology amongst others, have developed screening guidelines to help inform both patients and healthcare practitioners, with a goal of increasing the screen rates for at risk individuals, reducing the incidence and increasing the survival rates.
Specifically, the USPSTF, an independent volunteer panel of national health experts in prevention and evidence-based medicine has been working on initiatives to increase the CRC screening rate to 80% of the eligible population with a view to reducing the incidence and mortality rates, and have made recommendations for colorectal cancer screening designed to achieve this goal 41.

Broadly speaking, their recommendation target population includes asymptomatic patients, who have an average risk for developing disease, and are age 50 or older. Average risk here is defined as not having conditions that predispose the individual to a higher likelihood of developing the disease as a result of familial history of related genetic disorders such as Lynch syndrome, a personal history of the disease, inflammatory bowel disease or previous adenomatous polyp.

It is recommended that all adults that have an average risk of developing colorectal cancer begin periodic screening at age 50, and continue through age 75. This screening can be yearly, every 3 years, every 5 years, or every 10 years, depending on the screening test that they have elected to use. While routine screening is definitively recommended for patients age 50 to 75 who have an average risk for developing the disease, it is selectively recommended for patients age 76 to 86 if the patient has never been previously screened for the disease, and are in a position to get treatment for the disease in the event that it is detected. It is also not recommended for patients in the 76 to 85 age group if they have any comorbid conditions that could also significantly impact their life expectancy 41. This translates to a decline in the benefits associated with routine screening in comparison to the risks from developing the disease beyond age 75. In addition, there is also a long average
time (about 10 years) between adenoma development and cancer diagnosis; as such, screening routinely without specific cause is not recommended. Screening beyond age 85 is not recommended.

2.2.1 Adult Colorectal Cancer Risk Assessment

The guideline recommends that all adults be assessed for their risk of developing colorectal cancer, in order to set a more individualized screen start date. The physician assesses the patient’s risk for developing colorectal cancer by reviewing the patient’s personal and family health history (including history of diagnosis of colorectal cancer and other types of cancer, history of polyps, history of inflammatory bowel disease such as ulcerative colitis, genetic predispositions such as lynch syndrome, etc.), race, diet, obesity, and so on.

Patients who are considered high risk for predisposing conditions are usually advised to begin routine screening earlier than their peers who have average risks. For example, patients who have a one-degree relative with a history of colorectal cancer before age 50 are usually advised to begin routine screening 10 years before the age at which the youngest one-degree relative was diagnosed with the disease. A personal history of the condition may require ongoing routine screening in the event that there is a relapse. Patients of African American decent have a higher incidence of colorectal cancer in comparison to all other races. The physician does a complete assessment of the patient, and then decides the best course of action on an individual basis.

2.2.2 Types of Colorectal Cancer Screening Tests

There are various methods for administering colorectal cancer screen, some more invasive than others, and all having well defined advantages and disadvantages. The colorectal cancer
screening tests fall into two broad categories; the less invasive stool based tests and the generally more invasive endoscopic and radiologic tests\(^8\). The non-invasive stool based tests include the guaiac-based fecal occult blood test (gFOBT) or FIT, immunochemical-based fecal occult blood test (iFOBT), and the stool DNA panel (sDNA). The more invasive tests include the flexible sigmoidoscopy, optical colonoscopy, double-contrast barium enema (DCBE), and virtual colonoscopy.

According to a guidance statement research paper published in 2012, of all the available colorectal cancer screening tests, only the guaiac-based fecal occult blood test and flexible sigmoidoscopy have been evaluated in randomized, controlled trials, and the results showed an encouraging decrease in both colorectal cancer incidence and its associated mortality rates\(^8\).

Grounded in further research into the effectiveness of screening which found no conclusive study that showed any one screening method to be more effective than the others, as well as an ongoing effort to boost CRC screening rates, the USPSTF recently updated their 2008 colorectal cancer screening recommendation statement in June 2016 to show that the use of any of the industry screening methods for colorectal cancer screening has a higher net benefit than not screening at all; as such they are not recommending any specific method over the other, but rather, they are stipulating that the best screening method is the one that the patient decides to have\(^42\). A summary of their most current recommendation is shown in the table below.
Table 2.1. US Preventive Services Task Force Colorectal Cancer Screening Recommendation

<table>
<thead>
<tr>
<th>Population</th>
<th>Adults aged 50 to 75 y</th>
<th>Adults aged 76 to 85 y</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Recommendation</strong></td>
<td>Screen for colorectal cancer starting at age 50 y.</td>
<td>The decision to screen for colorectal cancer is an individual one.</td>
</tr>
<tr>
<td><strong>Grade</strong></td>
<td>A</td>
<td>C</td>
</tr>
</tbody>
</table>

For a summary of the evidence systematically reviewed in making this recommendation, the full recommendation statement, and supporting documents, please go to [US Preventive Services Task Force](http://www.uspreventiveservicestaskforce.org).

**Risk Assessment**
- For the vast majority of adults, the most important risk factor for colorectal cancer is older age. Other associated risk factors include family history of colorectal cancer, male sex, and black race.

**Screening Tests**
- There are numerous screening tests to detect early-stage colorectal cancer, including stool-based tests (gFOBT, FIT, and FIT-DNA), direct visualization tests (flexible sigmoidoscopy, alone or combined with FIT; colonoscopy; and CT colonography), and serology tests (SEPT9 DNA test). The USPSTF found no head-to-head studies demonstrating that any of these screening strategies are more effective than others, although they have varying levels of evidence supporting their effectiveness, as well as different strengths and limitations.

**Starting and Stopping Ages**
- The USPSTF concluded that the evidence best supports a starting age of 50 y for the general population. The age at which the balance of benefits and harms of colorectal cancer screening becomes less favorable varies based on a patient’s life expectancy, health status, comorbid conditions, and prior screening status. The USPSTF does not recommend routine screening for colorectal cancer in adults 85 y and older.

**Treatment and Interventions**
- Treatment of early-stage colorectal cancer generally consists of local excision or simple polypectomy for tumors limited to the colonic mucosa or surgical resection (via laparoscopy or open approach) with anastomosis for larger, localized lesions.

**Balance of Benefits and Harms**
- The USPSTF concludes with high certainty that the net benefit of screening for colorectal cancer is substantial. The USPSTF concludes with moderate certainty that the net benefit of screening for colorectal cancer in adults aged 76 to 85 y who have been previously screened is small. Adults who have never been screened are more likely to benefit. Screening is most appropriate for those healthy enough to undergo treatment and those without comorbid conditions that significantly limit their life expectancy.

**Other Relevant USPSTF Recommendations**
- The USPSTF has made a recommendation on aspirin use for the primary prevention of cardiovascular disease and colorectal cancer in average-risk adults. This recommendation is available on the USPSTF website (http://www.uspreventiveservicestaskforce.org).

A physician committee at the American College of Physicians, identified and reviewed the most critical guidelines that are in use today, and summarized their findings to come up with a single list of guidelines that captures the essence, as well as meet the goals of increasing the screen rates for at-risk individuals, reducing the incidence and increasing the survival rates. A summary of the available tests, along with their recommended schedule is shown in the table below.
<table>
<thead>
<tr>
<th>Screening Test</th>
<th>Testing Frequency</th>
<th>Methodology</th>
<th>Polyps and Cancer</th>
<th>Cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flexible Sigmoidoscopy</td>
<td>5 Years</td>
<td>Examination of the rectum and lower colon with a rigid or flexible lighted instrument (Lab Tests Online, 2015)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Double Contrast Barium Enema</td>
<td>5 Years</td>
<td>Series of x-rays of the colon and rectum; the patient is given an enema with a white, chalky solution that outlines the colon and rectum on the x-rays (Lab Tests Online, 2015)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Colonoscopy</td>
<td>10 Years</td>
<td>Examination of the rectum and entire colon with a flexible lighted instrument (Lab Tests Online, 2015)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Virtual Colonoscopy</td>
<td>5 Years</td>
<td>CT scan of the entire colon</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Guaiac-Based Fecal Occult Blood Test</td>
<td>1 Year</td>
<td>Uses a chemical indicator that shows a color change in the presence of blood in stool (Lab Tests Online, 2015)</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Stool DNA</td>
<td>3 Years</td>
<td>The stool is analyzed for certain mutations in the DNA that are associated with colon cancer (Lab Tests Online, 2014)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immunochemical-Based Fecal Occult</td>
<td>1 Year</td>
<td>Uses antibodies directed against human hemoglobin to detect blood in stool (Lab Tests Online, 2015)</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Colonoscopy</td>
<td>10 Years</td>
<td>Examination of the rectum and entire colon with a flexible lighted instrument (Lab Tests Online, 2015)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Flexible Sigmoidoscopy</td>
<td>5 Years</td>
<td>Examination of the rectum and lower colon with a rigid or flexible lighted instrument (Lab Tests Online, 2015)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Double Contrast Barium Enema</td>
<td>5 Years</td>
<td>Series of x-rays of the colon and rectum; the patient is given an enema with a white, chalky solution that outlines the colon and rectum on the x-rays (Lab Tests Online, 2015)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Colonoscopy</td>
<td>10 Years</td>
<td>Examination of the rectum and entire colon with a flexible lighted instrument (Lab Tests Online, 2015)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Virtual Colonoscopy</td>
<td>5 Years</td>
<td>CT scan of the entire colon</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Guaiac-Based Fecal Occult Blood Test</td>
<td>1 Year</td>
<td>Uses a chemical indicator that shows a color change in the presence of blood in stool (Lab Tests Online, 2015)</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Stool DNA</td>
<td>3 Years</td>
<td>The stool is analyzed for certain mutations in the DNA that are associated with colon cancer (Lab Tests Online, 2014)</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Immunochemical-Based Fecal Occult</td>
<td>1 Year</td>
<td>Uses antibodies directed against human hemoglobin to detect blood in stool (Lab Tests Online, 2015)</td>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>

Table 2.2. Summary of Colorectal Cancer Screen Types 43 42
2.2.3 Overview of Benefits and Burdens of Select Recommended Screening Strategies

The most recent publication by USPSTF with the update on colorectal cancer screening included graphical summaries of models showing the life years gained per 1,000 patients screened and deaths averted per 1,000 patients screened. These graphical representations have been reproduced below.

![Graph showing life years gained per 1,000 patients screened](image)

**Figure 2.5. Side-by-side comparison CRC screening types by estimated life years gained**

The USPSTF model shows the number of life years gained per 1,000 patients screened for patients who begin their routine screening at age 50 and follow the recommended guideline for their preferred tests through age 75. The FIT which is recommended annually, and is non-invasive shows the least number of life years gained at an average of 244 life years gained that could range from 231 to 260 life years gained. Colonoscopy on the other hand, is the most invasive, is recommended every 10 years, and has the most life years gained.
Figure 2.6. Side-by-side comparison CRC screening types by estimated per 1,000 deaths averted

The graph above shows the number of potential deaths averted as a result of adopting each respective screening approach per 1,000 screens beginning at age 50 and continuing through age 75 on the recommended schedules. Both FIT and gFOBT show similar death aversion rates per 1,000 screens, but once again colonoscopy captures the highest death aversion rate in comparison to the other tests.
CHAPTER 3 METHODOLOGY

A cross sectional study was conducted using approximately 4 million de-identified in-care patient population dataset for CRC screening from a national clinical diagnostics labs, who had their colorectal cancer screening test completed via the fecal immunochemical test (FIT) method. The dataset spans screenings that were completed between January 2010 through December 2015, and the population studied was captured by identifying tests that had any one of CPT codes 82270, 82274, G0328 captured for reimbursement purposes, as well as had Fecal Globin Immunochemistry test as the CRC screening test utilized.

The fecal globin immunochemistry test was the preferred colorectal cancer screening type for this research for the following reasons:

- Unlike other colorectal cancer screening tests such as colonoscopy which is conducted at a doctor’s office by a physician, the fecal globin immunochemistry screening test is typically ordered by a physician; the patient collects the sample, and then sends it off to the lab for testing. This makes it easier for a clinical diagnostics lab to collect data on this type of CRC screening test

- The fecal globin immunochemistry screening test only reacts to human hemoglobin as opposed to the guaiac-based tests, which also react to heme peroxidase activity and is non-specific to human hemoglobin. This was important for the study, since positivity rate was also being studied; this approach will help improve positivity rater accuracy by reducing the incidence of false positives

The test data was retrieved from the Lab Information system (LIS) database of the national clinical lab, parsed, cleaned and then linked to the US Census Bureau’s socioeconomic data
in order to assess the impact of specific socioeconomic attributes on screening and positivity rates.

### 3.1 Available Data Variables

#### 3.1.1 LIS Clinical Diagnostics Available Data Variables

The clinical diagnostics dataset retrieved from the LIS of a national clinical diagnostics lab comprised of de-identified testing administrative details for patients who were screened for colorectal cancer using the fecal immunochemical test method. The data is used for billing, trending and a myriad of operational improvement purposes, as well as by researchers, pharmaceutical companies, and healthcare policy makers to advance the clinical body of knowledge in pursuit of evidence-based healthcare policies, life sciences innovations and medical practice delivery. The initial run of the CRC screening dataset was retrieved from the LIS database with the following filters in place:

- All patients who had a fecal immunochemical test done for colorectal cancer screening
- Fecal immunochemical test order and/or CPT code ‘82270’, ‘82274’ or ‘G0328’
- Patients with dates of service beginning January 2010 through December 2015

The LIS houses clinical and administrative data elements for all clinical diagnostics visits. Listed below, is a non-exhaustive sample of variables included in the LIS database:

- Patient Unique Identifier
- Patient Visit ID
- Date of Service
- Patient Demographics (age, gender, date of birth, address, state, zip code, etc.)
- Test(s) Ordered
- Diagnosis Code(s)
- CPT Code(s)
- Test Result
- Test Result Unit
- Result Lower Limit (for tests results that fall within a range)
- Result Upper Limit (for tests results that fall within a range)
- Abnormal Result Flag (indicator for result that are out of possible range)
- Physician NPI
- Physician Specialty
- Payer Type (commercial/private insurance, Medicare, Medicaid, self-pay, etc.)
- Performing Lab

3.1.2. Zip Code to ZIP Code Tabulation Area (ZCTA) Crosswalk File

ZCTAs data is a generalized area representations of United States Postal Service (USPS) ZIP Codes, put together by the US Census Bureau, which it uses for various demographic, socioeconomic, congressional and other aggregations 45.

A zip code to ZCTA crosswalk file was downloaded from UDS Mapper, a non-profit organization that is funded by the US Department of Health & Human Services’ Health Resources and Services Administration (HRSA). They specialize in building standardized geographic mapping and decision support tools used by researchers and other healthcare
practitioners to better understand the geographic reach of various U.S. federal (Section 330) Health Center Program (HCP) initiatives.

The ZCTA crosswalk file was downloaded to serve as a link to connect the zip codes captured in the clinical diagnostics dataset to the US Census Bureau’s socioeconomic datasets using the common field of ‘GEO.id’. The ZCTA data fields needed for the analyses are shown below:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zip Code</td>
<td>US postal geographic code</td>
<td>To link the clinical dataset zip code to the ZCTA GEO.id</td>
</tr>
<tr>
<td>ZCTA GEO.id</td>
<td>Zip code tabulation area ID</td>
<td>To link the socioeconomic dataset ZCTA GEO.id to GEO.id in the US Census Bureau’s socioeconomic datasets</td>
</tr>
</tbody>
</table>

Table 3.1. ZCTA file select variables with description

3.1.3. **US Census Bureau’s Demographic Dataset**

The US census demographic dataset was generated and downloaded from the Facts Finder web portal. This dataset includes US demographic information such as age, gender, education, median income, poverty rate, amongst other demographic variables. This file will be linked to the clinical diagnostics dataset via the UDS mapper ZCTA file.

3.1.4. **Age-Adjustment Weight Data Table**

The age adjustment weight tables published by the Department of Health and Human Services (DHHS) in the Healthy People 2010 Statistical Notes, and based on the 2000 projected US population will be used to adjust the populations’ age distribution. This will be done by applying the respective weights to observed age distribution in order to even out variations that may occur in age distributions across populations and/or points in time. This
will ensure a more equitable comparison of age distributions across groups. The age adjustment table that will be used is shown below.

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Adjustment Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Under 18 years</td>
<td>0.257736</td>
</tr>
<tr>
<td>18 - 44 years</td>
<td>0.393797</td>
</tr>
<tr>
<td>45 - 54 years</td>
<td>0.134834</td>
</tr>
<tr>
<td>55 - 64 years</td>
<td>0.087247</td>
</tr>
<tr>
<td>65 - 74 years</td>
<td>0.066037</td>
</tr>
<tr>
<td>75 years and over</td>
<td>0.060349</td>
</tr>
</tbody>
</table>

Table 3.2. Age adjustment factors by age group

3.2. Analyses Tools

Data analyses will be completed primarily using SAS 9.4. Other tools that will be used in varying capacities include SAS JMP 11, Tableau 9.3 and Excel 2010

3.3. Research Design

3.3.1. Clinical Diagnostics Dataset

The goal of this research is to analyze colorectal cancer screening test results which were performed using the fecal immunochemical test (FIT) method. The test data was retrieved from the LIS database of a national clinical lab, parsed, cleaned and then linked to the US Census Bureau’s socioeconomic data in order to assess if socioeconomic attributes have an influence on receiving a positive result on the test. 6 years’ worth of testing de-identified patients’ data beginning in 2010 through 2015 was analyzed.

The initial run of the CRC screening dataset was retrieved from the LIS database with the following filters in place:

- All patients who had a fecal immunochemical test done for colorectal cancer screening
- Fecal immunochemical test order and/or CPT code ‘82270’, ‘82274’ or ‘G0328’
- Visited beginning January 2010 through December 2015

3.3.1.1. Clinical Diagnostics Test Dataset Cleaning and Preparation

Using the dataset extracted from the LIS, a base dataset was created in order to provide some insights into the general testing population. The diagnostics lab’s database assigns unique numbers to each patient, as well as a unique number to each patient’s visit. The base number of fecal immunochemical tests for the study was generated using the number of patient visits during the testing period, and then tests that had inconclusive or missing results, as well as records that were missing vital patient information such as age and zip code, which were considered necessary for the analysis were excluded. Other administrative and test processing quality control records were also deleted as part of the cleaning process.

A summary of the initial data pull by year showed the following distribution table.

<table>
<thead>
<tr>
<th>Year</th>
<th>Distinct Visits</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>407,244</td>
</tr>
<tr>
<td>2011</td>
<td>615,317</td>
</tr>
<tr>
<td>2012</td>
<td>675,105</td>
</tr>
<tr>
<td>2013</td>
<td>736,987</td>
</tr>
<tr>
<td>2014</td>
<td>814,170</td>
</tr>
<tr>
<td>2015</td>
<td>980,882</td>
</tr>
</tbody>
</table>

Table 3.3. Clinical diagnostics screening study raw data summary by year

In preparation for the core part of our analyses, the research population was further narrowed down to only patients ages 50 to 85. This was to ensure that we capture the core average risk population recommended to undergo routine screening by the USPSTF. This resulted in a further 15% reduction of the dataset.

The data was then cleaned, and the necessary exclusions and data transformations performed. A non-exhaustive list of the criteria for exclusion is outlined below:
\[\begin{itemize}
\item Inconclusive test result (positivity rate is at the core of our analysis)
\item Missing data in any of the fields considered vital to our analysis, such as the patient age or zip code
\item Quality control records that are randomly captured alongside tests to help validate the test results
\item Inconsistent field values
\item Duplicate values
\end{itemize}\]

The clinical diagnostics testing dataset was then linked to the zip code to ZCTA crosswalk file, in order to pull in the GEO.id required to pull in the US Census Bureau’s socioeconomic attributes dataset, and then using the GEO.id code, the details of the US Census Bureau’s socioeconomic demographic file was linked to the clinical diagnostics dataset.

3.3.2. US Census Demographic Data

The US census demographic dataset was generated and downloaded from the Facts Finder web portal. This dataset includes US demographic information such as age, gender, education, median income, poverty rate, amongst other socioeconomic attributes.

3.3.2.1. US Census Bureau’s Demographic Dataset Cleaning and Preparation

Population age buckets were transformed and combined in order to get the a single number for age 50 and over, and so were the high school and Bachelors program graduation figures in order to create the required flags for high school and Bachelors program graduation rates for the age 45 and over population.

Secondary fields for high school graduation rate quintile buckets and Bachelors program graduation rate quintile buckets were then created in the ranges of:
The poverty rates were also grouped into quintile buckets in the ranges of:

- 0.0% - 19.9%
- 20.0% - 39.9%
- 40.0% - 59.9%
- 60.0% - 79.9%
- 80.0% - 100%

Median income was grouped in the ranges:

- =<$19,999
- $20,000 - $29,999
- $30,000 - $49,999
- $50,000 - $69,999
- $70,000 - $99,999
- $100,000 - $149,999
- $150,000 - $199,999
- $200,000 - $249,999
- $250,000+

3.3.3. Zip Code to ZIP Code Tabulation Area (ZCTA) Crosswalk File

This data file was downloaded from UDS Mapper, a non-profit organization that is funded by the US Department of Health & Human Services’ Health Resources and Services Administration (HRSA). The organization specializes in building standardized geographic
mapping and decision support tools used by researchers and healthcare practitioners in order to make it easier to compare research results.

This file contains fields that are common to the clinical diagnostics dataset and the US Census Bureau’s demographic dataset (zip code and GEO.id code), which will enable the two files to be linked. This will be achieved by linking the zip code field from this ZCTA crosswalk file to the zip code field in the clinical diagnostics dataset, and then linking the GEO.id code field from the ZCTA crosswalk to the GEO.id code field in the US Census demographic file.

3.4. Study Variables

The study will primarily utilize core variables whose attributes have been hypothesized to have a statistically significant impact on CRC screening outcome, as well as other variables that shed additional insights on the core variables. The nature of these variables range from clinical and non-clinical/administrative variables extracted from the LIS, to socioeconomic variables extracted from the US Census Bureau’s socioeconomic dataset.

3.4.1. Core Clinical Diagnostics Study Variables

- Test Result (binary categorical variable)
- Positivity Rate (quantitative variable)
- Patient Gender (binary categorical variable)

3.4.2. Core Socioeconomic Variables

- High School Graduation Rate (quantitative variable)
- High School Graduation Rate Bucket (categorical variable)
- Year College Graduation Rate (quantitative variable)
- Year College Graduation Rate Bucket (categorical variable)
3.4.3. Socioeconomic Variable Derivations

The socioeconomic attributes data was extracted from the US Census Bureau’s data portal and is grouped by the administrative geographical areas known as ZCTA. The extracted data was linked to zip codes, and then the population estimates per variable were converted into percentage rates by using the following approaches.

- The poverty rates of the population were derived by dividing the estimated total population by the estimated population living at or below the poverty rate in a given zip code.
- The median income of the population was used “as is” at the zip code level.
- The high school graduation rates of the population were derived by summing up the estimates for the total age 45 and over population, and then summing up the estimates for age 45 and over with a high school diploma. The age 45 and over high school diploma population estimate was then divided by the total age 45 and over population, to arrive at the high school graduation rate for our study.
- The bachelor’s degree attainment rates of the population were derived by summing up the estimates for the total age 45 and over population, and then summing up the estimates for age 45 and over with a bachelor’s degree. The age 45 and over bachelor’s degree population estimate was then divided by the total age 45 and over population, to arrive at the bachelor’s degree attainment rate.
3.5. Statistical Tests

A variety of statistical tests were carried out to understand the distribution of the socioeconomic attributes of the screening population, relationship between each of the 4 selected socioeconomic attributes and positivity rates, and identify the socioeconomic attributes’ parameter estimates that can be used to predict a population’s colorectal cancer screening positivity rate, as well as verify if there are variations in the positivity rates over the years.

Descriptive analyses were carried out to understand the extracted data, as well as to gain some insights into the testing population distribution over time. Examples of descriptive statistics that will be used include Mean, Median, Mode, and Frequencies through the use of the SAS Univariate procedure.

Inferential and predictive analyses, including correlation analysis, multiple regression analysis and ANOVA were be carried out to understand how, and to what extent the socioeconomic attributes were related to the positivity rates, which socioeconomic attributes best predicts a positive screen test, and to what extent it/they can predict a positive screen tests.

The SAS ‘PROC CORR’ procedure was used to complete the correlation analysis in order to assess the nature of the relationship between each of the four socioeconomic attributes and the positivity rates. The SAS ‘PROC REG’ procedure was used to carry out a multiple regression analysis, in order to generate the parameter estimates for the socioeconomic variables that will enable us to predict the positivity rate for a given population. The ‘PROC ANOVA’ procedure was used to carry out the ANOVA analysis on the positivity rate over the years, in order to determine if there were any statistically significant variations in the
positivity rates over the 6 years of study. The test for statistical significance for all tests will be set at p<0.05.

3.4. Data Preparation

We will be analyzing de-identified CRC screening test data extracted from a LIS database of a national clinical diagnostics lab. Data for variables deemed to be relevant to answering our research questions will be generated and exported for analysis using SAS 9.4 statistical tool.

The data will be cleaned, transformed and then analyzed. The cleaning phase will include the identification and exclusion of the following:

- Incomplete records
- Inconsistent records
- Duplicate records
- Records with inconclusive results
- Cancelled tests
- Quality control records that are randomly captured alongside tests to help validate the test results
- Patients whose age do not fall between ages 50 – 85

The transformation phase will involve assigning more meaningful names to coded data, creating secondary fields that aggregate data up into groups, such as age groups, service month, city and state.

The analysis phase will involve the analyses of the data to understand its attributes, distribution and any inferences that can be drawn based on our hypothesis.
CHAPTER 4 RESULTS

4.1. Clinical Diagnostics Data Distribution by Year

The clinical diagnostic data was summarized by year after cleaning and transformation, and a weighted distribution of the data by year, as well as the relative deviation by year was generated. The weighted screening trend showed a marked increase, at a rate of 10% from the base year of 2010 through 2015. The summary results are shown below:

<table>
<thead>
<tr>
<th>Year</th>
<th>Negative Results</th>
<th>Positive Results</th>
<th>Distinct Screens</th>
<th>Percent</th>
<th>Cumulative Percent</th>
<th>Positivity Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>376,210</td>
<td>32,121</td>
<td>407,244</td>
<td>9.65%</td>
<td>9.65%</td>
<td>7.90%</td>
</tr>
<tr>
<td>2011</td>
<td>548,063</td>
<td>67,571</td>
<td>615,317</td>
<td>14.55%</td>
<td>24.20%</td>
<td>11.00%</td>
</tr>
<tr>
<td>2012</td>
<td>612,251</td>
<td>63,103</td>
<td>675,105</td>
<td>15.96%</td>
<td>40.15%</td>
<td>9.30%</td>
</tr>
<tr>
<td>2013</td>
<td>676,466</td>
<td>60,729</td>
<td>736,987</td>
<td>17.42%</td>
<td>57.57%</td>
<td>8.20%</td>
</tr>
<tr>
<td>2014</td>
<td>734,050</td>
<td>80,315</td>
<td>814,170</td>
<td>19.24%</td>
<td>76.82%</td>
<td>9.90%</td>
</tr>
<tr>
<td>2015</td>
<td>881,312</td>
<td>99,799</td>
<td>980,882</td>
<td>23.18%</td>
<td>100.00%</td>
<td>10.20%</td>
</tr>
<tr>
<td>Study Totals</td>
<td>3,828,352</td>
<td>403,638</td>
<td>4,229,705</td>
<td>100.00%</td>
<td>9.50%</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.1. CRC screening with positivity rate summary by year

Figure 4.1. CRC screening volume distribution histogram by year
Descriptive analyses of the core measures of interest were then carried out in order to understand their distribution, including the analysis of the high school graduation rates, Bachelors program graduation rates, median income levels, and poverty rates. These were also analyzed by screen rates and positivity rates. Summaries reports were then generated, including heat maps that show the population foot print across the US and bubble charts showing the descriptive relationship between the population, screen rate and positivity rate were then generated.

The SAS Univariate procedure was run on the variables of interest in order to gain insights into their descriptive attributes.

Some screening records did not have their median income resolved. Since the study dataset was limited to screen records that had a valid zip code, the unresolved records were as a result of being affiliated with geographical areas where the disclosure of some socioeconomic attributes could increase the likelihood of personally identifying people within the population.

For the purpose of analyzing the screening population's median income, screening records without a resolved median income attribute were excluded from this analysis. Shown below is the yearly distribution of records that were resolved for median income.

<table>
<thead>
<tr>
<th>Year</th>
<th>Total Distinct Screens</th>
<th>Distinct Screens with Median Income Attributes</th>
<th>Positive Results</th>
<th>% Year-Over-Year Distinct Screen Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>407,244</td>
<td>381,605</td>
<td>23,339</td>
<td>-</td>
</tr>
<tr>
<td>2011</td>
<td>615,317</td>
<td>528,124</td>
<td>47,476</td>
<td>38.40%</td>
</tr>
<tr>
<td>2012</td>
<td>675,105</td>
<td>599,176</td>
<td>44,067</td>
<td>13.45%</td>
</tr>
<tr>
<td>2013</td>
<td>736,987</td>
<td>651,918</td>
<td>42,733</td>
<td>8.80%</td>
</tr>
<tr>
<td>2014</td>
<td>814,170</td>
<td>706,991</td>
<td>56,774</td>
<td>8.45%</td>
</tr>
<tr>
<td>2015</td>
<td>980,882</td>
<td>762,450</td>
<td>69,875</td>
<td>7.84%</td>
</tr>
<tr>
<td>Totals</td>
<td>4,229,705</td>
<td>3,630,264</td>
<td>284,264</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.2. Distribution of CRC screens with a resolved median income attribute by year
4.2. Trend and Descriptive Analysis Results for the Screening Population's Median Income

Of the 4,229,705 distinct screens between 2010 and 2015, only 3,630,264 had a median income match in the US census socioeconomic data table.

The population median income for all patients who got screened in the period studied was positively skewed, with a range from $4,550 to $236,912, and had a mean and median of $59,525 and $54,666 respectively. The mode was $32,670.

Figure 4.2. Combined 6 Year (2010 – 2015) Study Period Median Income Distribution and Probability Plot
4.2.1. Median Income Attribute Distribution by Year

To better understand if the median income distribution for patients who got screened varied over the years, we performed some descriptive analysis on the median income variable by year of service; the results are presented in the ensuing sections.

2010 had 381,605 distinct screens that matched all the criteria for inclusion in the median income variable analysis. The median income distribution was positively skewed, and ranged from $11,413 to $235,172. The distribution had a mean and median income of $65,013 and $59,869 respectively. The mode was $47,226.

Figure 4.3. 2010 Median Income Distribution and Probability Plot
In 2011, there was a 38% increase over the previous year (see table 4.2). The median income distribution was positively skewed, and ranged from $10,625 to $235,172. The distribution had a mean and median income of $61,211 and $56,065 respectively. The mode was $47,226.
In comparison to the 2010 to 2011 change in screen volume, there was a more modest change in volume in 2012 over 2011 with a 13% increase over the previous (see table 4.2). Similar to the previous years, the median income distribution was positively skewed, and ranged from $4,550 to $235,172. The distribution had a mean and median income of $60,128 and $55,143 respectively. The mode was $22,329.

Figure 4.5. 2012 Median Income Distribution and Probability Plot
2013 experienced a slightly more modest volume growth of 9% in screens in comparison to the previous year (see table 4.2). The 2013 number of distinct screens captured was 651,918. The median income remained positively skewed as in previous years, and ranged from $4,550 to $235,172. The distribution had a mean and median income of $59,480 and $54,680 respectively. The mode was $32,670.

Figure 4.6. 2013 Median Income Distribution and Probability Plot
2014 continued growth in screening volume with an 8% increase over the previous year with 706,991 distinct screens (see table 4.2). The positively skewed median income trend continued, and ranged from $10,169 to $235,172. The distribution had a mean and median income of $57,779 and $52,761 respectively. The mode held steady $32,670.

Figure 4.7. 2014 Median Income Distribution and Probability Plot
2015, the last of 6 years included in the study captured the highest number of screens per year at 762,450 distinct screens, which was an 8% increase in screening volume over the previous year (see table 4.2). As in other years, the median income remained positively skewed. 2015 however had the widest income range, spanning populations with a median income of $4,550 to $236,912. The distribution had a mean and median income of $57,166 and $52,045 respectively. The mode held steady $32,670.

![Distribution and Probability Plot for Median Income](image)

**Figure 4.8.** 2015 Median Income Distribution and Probability Plot
Figure 4.9. Graphical overlay of the population's, mean and median income over the screening population by year

The median income trend graph above shows that, the colorectal cancer screening volume grew over the years, from 2010 through 2015. In the same period, both the population mean and median income showed some decline. The increase in screening volume over the years could imply a greater awareness of the disease and the need to screen in order to detect the disease early on, given that there was no remarkable change in the US geographical population during the same period; and the decline in median income for the screening population could be attributed to increased awareness and accessibility of screening to lower income populations over the years.
4.2.2. Median Income Summaries for the Combined 6 Years Studied

With the entire screening population across the 6 years of study, all screens were combined irrespective of screen year, and then analyzed based using the median income buckets previously described in the methodology section. The results were then summarized and have been provided below.

**US Heat Map of CRC Screening Footprint by Median Income Bucket**

![US Heat Map of CRC Screening Footprint by Median Income Bucket](image)

Figure 4.10. US heat map above shows the geographical heat distribution by median income bucket, of the screened patients included in the study 6 year cross sectional study.
Screening and Positivity Rate by Median Income Bucket

Figure 4.11. Graphical representation of screening population’s median income by size in relationship to positivity rate

Bubble graph depicting the screening and positivity rates by median income for the 6 years of study. The bigger and further right along the horizontal axis a bubble is, the higher the screen rate. The higher along the vertical axis a bubble is, the higher the positivity rate. Each bubble has been labeled to show the median income bucket that it represents.
### CRC Screening Rate by Median Income

<table>
<thead>
<tr>
<th>Median Income Range</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>All 6 Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>=&lt; $19,999</td>
<td>0.07%</td>
<td>0.08%</td>
<td>0.12%</td>
<td>0.12%</td>
<td>0.25%</td>
<td>0.56%</td>
<td>1.04%</td>
</tr>
<tr>
<td>$20,000 - $29,999</td>
<td>3.18%</td>
<td>3.90%</td>
<td>4.09%</td>
<td>4.30%</td>
<td>4.85%</td>
<td>4.56%</td>
<td>4.74%</td>
</tr>
<tr>
<td>$30,000 - $49,999</td>
<td>29.92%</td>
<td>33.60%</td>
<td>34.60%</td>
<td>35.79%</td>
<td>38.29%</td>
<td>39.99%</td>
<td>38.59%</td>
</tr>
<tr>
<td>$50,000 - $69,999</td>
<td>30.78%</td>
<td>33.28%</td>
<td>33.62%</td>
<td>33.23%</td>
<td>32.66%</td>
<td>31.09%</td>
<td>30.98%</td>
</tr>
<tr>
<td>$70,000 - $99,999</td>
<td>25.62%</td>
<td>21.87%</td>
<td>20.84%</td>
<td>20.01%</td>
<td>18.33%</td>
<td>18.13%</td>
<td>18.53%</td>
</tr>
<tr>
<td>$100,000 - $149,999</td>
<td>9.53%</td>
<td>6.69%</td>
<td>6.27%</td>
<td>6.15%</td>
<td>5.24%</td>
<td>5.38%</td>
<td>5.58%</td>
</tr>
<tr>
<td>$150,000 - $199,999</td>
<td>0.87%</td>
<td>0.58%</td>
<td>0.46%</td>
<td>0.39%</td>
<td>0.38%</td>
<td>0.30%</td>
<td>0.50%</td>
</tr>
<tr>
<td>$200,000 - $249,999</td>
<td>0.01%</td>
<td>0.01%</td>
<td>0.01%</td>
<td>0.01%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.04%</td>
</tr>
<tr>
<td>$250,000+</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
</tbody>
</table>

Table 4.3. Tabular distribution of the screening rate by median income bucket

### CRC Screening Positivity Rate by Median Income

<table>
<thead>
<tr>
<th>Median Income Positivity Rate</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>=&lt; $19,999</td>
<td>5.00%</td>
<td>9.70%</td>
<td>10.60%</td>
<td>11.90%</td>
<td>16.30%</td>
<td>13.50%</td>
</tr>
<tr>
<td>$20,000 - $29,999</td>
<td>9.10%</td>
<td>12.30%</td>
<td>10.50%</td>
<td>9.20%</td>
<td>9.70%</td>
<td>9.60%</td>
</tr>
<tr>
<td>$30,000 - $49,999</td>
<td>7.90%</td>
<td>11.50%</td>
<td>9.40%</td>
<td>8.40%</td>
<td>10.10%</td>
<td>10.10%</td>
</tr>
<tr>
<td>$50,000 - $69,999</td>
<td>8.00%</td>
<td>11.00%</td>
<td>9.40%</td>
<td>8.30%</td>
<td>10.30%</td>
<td>10.90%</td>
</tr>
<tr>
<td>$70,000 - $99,999</td>
<td>8.30%</td>
<td>11.00%</td>
<td>9.60%</td>
<td>8.50%</td>
<td>9.80%</td>
<td>10.20%</td>
</tr>
<tr>
<td>$100,000 - $149,999</td>
<td>7.30%</td>
<td>9.90%</td>
<td>8.30%</td>
<td>7.80%</td>
<td>9.50%</td>
<td>9.90%</td>
</tr>
<tr>
<td>$150,000 - $199,999</td>
<td>5.70%</td>
<td>6.60%</td>
<td>7.10%</td>
<td>8.10%</td>
<td>9.40%</td>
<td>10.90%</td>
</tr>
<tr>
<td>$200,000 - $249,999</td>
<td>2.90%</td>
<td>10.00%</td>
<td>8.10%</td>
<td>7.10%</td>
<td>18.80%</td>
<td>19.40%</td>
</tr>
</tbody>
</table>

Table 4.4. Tabular distribution of the positivity rate by median income bucket
4.3. Trend and Descriptive Analysis Results for the Screening Population’s Poverty Rates

3,631,397 out of the 4,229,705 distinct screens captured between 2010 and 2015 had an associated poverty rate variable match in the US census socioeconomic data table. The population poverty rate for all patients who got screened in the period studied was positively skewed, and ranged from 0% to 100%. The mean population poverty rate was 15.12% and median poverty rate was 13.38%. The mode was 15.54%.

Figure 4.12. Combined 6 Year (2010 – 2015) Study Period Poverty Rate Distribution and Probability Plot
Some screening records were missing a poverty rate attribute data capture. The records with missing poverty rate attributes were most likely as a result of not being provided by the US Census Bureau, in order not to inadvertently disclose information that could personally identify individuals.

For the purpose of analyzing the screening population’s poverty rate, screening records without a resolved poverty rate attribute were excluded from this analysis. Shown below is the yearly distribution of records that were resolved for poverty rate.

<table>
<thead>
<tr>
<th>Year</th>
<th>Total Distinct Screens</th>
<th>Distinct Screens with Poverty Rate Attributes</th>
<th>Positive Results</th>
<th>±Δ Year-Over-Year Distinct Screen Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>407,244</td>
<td>381,933</td>
<td>23,355</td>
<td>-</td>
</tr>
<tr>
<td>2011</td>
<td>615,317</td>
<td>528,402</td>
<td>47,494</td>
<td>38.35%</td>
</tr>
<tr>
<td>2012</td>
<td>675,105</td>
<td>599,359</td>
<td>44,079</td>
<td>13.43%</td>
</tr>
<tr>
<td>2013</td>
<td>736,987</td>
<td>652,153</td>
<td>42,756</td>
<td>8.81%</td>
</tr>
<tr>
<td>2014</td>
<td>814,170</td>
<td>707,241</td>
<td>56,798</td>
<td>8.45%</td>
</tr>
<tr>
<td>2015</td>
<td>980,882</td>
<td>762,609</td>
<td>69,896</td>
<td>7.83%</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td><strong>4,229,705</strong></td>
<td><strong>3,631,697</strong></td>
<td><strong>284,378</strong></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.5. Distribution of CRC screens with a resolved poverty rate attribute by year
4.3.1. Poverty Rate Attributes Distribution by Year

Poverty rate was further analyzed by year of service to determine if there variations in the poverty levels over the study years. 2010 had 381,933 distinct screens that matched all the criteria for inclusion in the population poverty rate variable analysis. The poverty rate distribution was positively skewed, and ranged from 0.00% to 61.29%. The 2010 poverty rate distribution had a mean and median poverty rate of 13.97% and 12.18% respectively. The mode was 6.34%.

Figure 4.13. 2010 Poverty Rate Distribution and Probability Plot
2011 had 528,402 distinct screens that matched all the criteria for inclusion in the poverty rate variable analysis. The poverty rate distribution was positively skewed, and ranged from 0.00% to 100.00%. The 2011 poverty rate distribution had a mean and median poverty rate of 14.30% and 12.71% respectively. The mode was 6.32%.

Figure 4.14. 2011 Poverty Rate Distribution and Probability Plot
2012 had 599,359 distinct screens that matched all the criteria for inclusion in the poverty rate variable analysis. The poverty rate distribution was positively skewed, and ranged from 0.00% to 79.64%. The 2012 poverty rate distribution had a mean and median poverty rate of 14.84% and 13.10% respectively. The mode was 15.54%.

Figure 4.15. 2012 Poverty Rate Distribution and Probability Plot
2013 had 652,153 distinct screens that matched all the criteria for inclusion in the poverty rate variable analysis. The poverty rate distribution was positively skewed, and ranged from 0.00% to 100.00%. The 2013 poverty rate distribution had a mean and median poverty rate of 15.10% and 13.33% respectively. The mode was 15.54%.

Figure 4.16. 2013 Poverty Rate Distribution and Probability Plot
2014 had 707,241 distinct screens that matched all the criteria for inclusion in the poverty rate variable analysis. The poverty rate distribution was positively skewed, and ranged from 0.00% to 84.66%. The 2014 poverty rate distribution had a mean and median poverty rate of 15.63% and 13.78% respectively. The mode was 15.54%.

Figure 4.17. 2014 Poverty Rate Distribution and Probability Plot
2015 had 762,609 distinct screens that matched all the criteria for inclusion in the poverty rate variable analysis. The poverty rate distribution was positively skewed, and ranged from 0.00% to 100.00%. The 2015 poverty rate distribution had a mean and median poverty rate of 16.04% and 14.14% respectively. The mode was 22.93%.

Figure 4.18. 2015 Poverty Rate Distribution and Probability Plot
Figure 4.19. Graphical overlay of the population's, poverty rate over the screening population by year

The poverty rate trend graph above shows that, as the colorectal cancer screening volume grew over the years, from 2010 through 2015, the mean and median poverty rate for the screening population was also growing. This increase in the mean and median poverty rate could be seen as an additional indicator that in addition to the poorer population being more aware of the disease and the preventive power of screening, the test was also becoming more accessible to the poorer population, and allowing them to be a part of the screening population.
4.3.2. Poverty Rate Summaries for the Combined 6 Years Studied

To get an overall poverty rate view as it pertains to all screens that met our criteria for inclusion, all 6 years screen data were combined and then grouped into poverty rate buckets as outlined in the methodology section prior to being analyzed for screening and positivity trends as they relate to the population poverty rates. The results were then summarized and have been provided below.

**US Heat Map of CRC Screening Footprint by Poverty Rate Bucket**

Figure 4.20. US heat map above shows the geographical heat distribution by poverty rate bucket, of the screened patients included in the study 6 year cross sectional study.
Screening and Positivity Rate by Poverty Rate Bucket

Figure 4.21. Graphical representation of screening population’s poverty rate by size in relationship to positivity rate

The bubble graph above depicts the screening and positivity rates by poverty rate for the 6 years of study. The bigger and further right along the horizontal axis a bubble is, the higher the screen rate. The higher along the vertical axis a bubble is, the higher the positivity rate. Each bubble has been labeled to show the poverty rate bucket that it represents.
### CRC Screening Rate by Population Poverty Rate

<table>
<thead>
<tr>
<th>Population Poverty Rate Range</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>Demographic Poverty Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0% - 19.9%</td>
<td>79.42%</td>
<td>77.52%</td>
<td>75.92%</td>
<td>74.41%</td>
<td>72.82%</td>
<td>70.92%</td>
<td>74.41%</td>
</tr>
<tr>
<td>20.0% - 39.9%</td>
<td>19.86%</td>
<td>21.75%</td>
<td>23.28%</td>
<td>24.63%</td>
<td>26.05%</td>
<td>27.47%</td>
<td>23.17%</td>
</tr>
<tr>
<td>40.0% - 59.9%</td>
<td>0.71%</td>
<td>0.72%</td>
<td>0.79%</td>
<td>0.94%</td>
<td>1.01%</td>
<td>1.27%</td>
<td>2.33%</td>
</tr>
<tr>
<td>60.0% - 79.9%</td>
<td>0.01%</td>
<td>0.01%</td>
<td>0.01%</td>
<td>0.02%</td>
<td>0.12%</td>
<td>0.33%</td>
<td>0.08%</td>
</tr>
<tr>
<td>80.0% - 100.0%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
</tbody>
</table>

Table 4.6. Tabular distribution of the screening rate by poverty rate bucket

### CRC Screening Positivity Rate by Population Poverty Rate

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0% - 19.9%</td>
<td>7.94%</td>
<td>11.06%</td>
<td>9.29%</td>
<td>8.35%</td>
<td>10.11%</td>
<td>10.70%</td>
</tr>
<tr>
<td>20.0% - 39.9%</td>
<td>8.18%</td>
<td>11.35%</td>
<td>9.63%</td>
<td>8.49%</td>
<td>9.78%</td>
<td>9.47%</td>
</tr>
<tr>
<td>40.0% - 59.9%</td>
<td>8.78%</td>
<td>13.28%</td>
<td>13.23%</td>
<td>10.81%</td>
<td>10.43%</td>
<td>10.30%</td>
</tr>
<tr>
<td>60.0% - 79.9%</td>
<td>7.50%</td>
<td>13.64%</td>
<td>15.71%</td>
<td>12.26%</td>
<td>19.88%</td>
<td>13.83%</td>
</tr>
<tr>
<td>80.0% - 100.0%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
</tbody>
</table>

Table 4.7. Tabular distribution of the positivity rate by poverty rate bucket
4.4. Trend and Descriptive Analysis Results for the Screening Population’s High School Graduation Rates

3,635,032 distinct screens out of 4,229,705 screens conducted between 2010 and 2015, had an associated high school graduation attribute in the US census socioeconomic data table. The population high school graduation rate for all patients who got screened during the study period was negatively skewed, and ranged from 13.31% to 100.00%. The overall mean, median and mode high school graduation rate were 82.78%, 87.30% and 88.84% respectively.

Figure 4.22. Combined 6 Year (2010 – 2015) Study Period High School Graduation Rate Distribution and Probability Plot
Similar to the median income and poverty rate attributes, some screening records were missing a high school graduation rate attribute data capture. The reason behind this is also for the US Census Bureau to limit the chance of inadvertently disclosing information that could personally identify individuals within a population.

For the purpose of analyzing the screening population’s high school graduation rate, screening records without a resolved high school graduation rate attribute were excluded from this analysis. Shown below is the yearly distribution of records that were resolved for high school graduation rate.

<table>
<thead>
<tr>
<th>Year</th>
<th>Total Distinct Screens</th>
<th>Distinct Screens with High School Graduation Attributes</th>
<th>Positive Results</th>
<th>±△ Year-Over-Year Distinct Screen Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>407,244</td>
<td>382,648</td>
<td>23,405</td>
<td>-</td>
</tr>
<tr>
<td>2011</td>
<td>615,317</td>
<td>529,032</td>
<td>47,541</td>
<td>38.26%</td>
</tr>
<tr>
<td>2012</td>
<td>675,105</td>
<td>599,900</td>
<td>44,140</td>
<td>13.40%</td>
</tr>
<tr>
<td>2013</td>
<td>736,987</td>
<td>652,694</td>
<td>42,807</td>
<td>8.80%</td>
</tr>
<tr>
<td>2014</td>
<td>814,170</td>
<td>707,866</td>
<td>56,864</td>
<td>8.45%</td>
</tr>
<tr>
<td>2015</td>
<td>980,882</td>
<td>762,892</td>
<td>69,935</td>
<td>7.77%</td>
</tr>
<tr>
<td>Totals</td>
<td><strong>4,229,705</strong></td>
<td><strong>3,635,032</strong></td>
<td><strong>284,692</strong></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.8. Distribution of CRC screens with a resolved high school graduation rate attribute by year
4.4.1. High School Graduation Rate Attributes Distribution by Year

As with other study variables, a year by year descriptive analysis of the high school graduation rate of the screened population was also conducted. The results are shown in the sections below. 2010 had 382,648 distinct screens that matched all the criteria for inclusion in the high school graduation rate variable analysis. The high school graduation rate distribution was negatively skewed, and ranged from 19.75% to 100.00%. The 2010 high school graduation rate distribution had a mean and median high school graduation rate of 84.63% and 88.62% respectively. The mode was 92.89%.

Figure 4.23. 2010 High School Graduation Rate Distribution and Probability Plot
There were 529,032 distinct screens that matched all the criteria for inclusion in the high school graduation rate variable analysis in 2011. The high school graduation rate population distribution was negatively skewed, and ranged from 13.31% to 100.00%. The 2011 high school graduation rate distribution had a mean and median high school graduation rate of 84.02% and 88.01% respectively. The mode was 92.89%.

Figure 4.24. 2011 High School Graduation Rate Distribution and Probability Plot
There were 599,900 distinct screens that matched all the criteria for inclusion in the high school graduation rate variable analysis in 2012. The high school graduation rate population distribution was negatively skewed, and ranged from 18.23% to 100.00%. The 2012 high school graduation rate distribution had a mean and median high school graduation rate of 83.31% and 87.72% respectively. The mode was 88.84%.

Figure 4.25. 2012 High School Graduation Rate Distribution and Probability Plot
There were 652,694 distinct screens that matched all the criteria for inclusion in the high school graduation rate variable analysis in 2013. The high school graduation rate population distribution was negatively skewed, and ranged from 18.23% to 100.00%. The 2013 high school graduation rate distribution had a mean and median high school graduation rate of 82.73% and 87.27% respectively. The mode was 88.84%.

Figure 4.26. 2013 High School Graduation Rate Distribution and Probability Plot
There were 707,866 distinct screens that matched all the criteria for inclusion in the high school graduation rate variable analysis in 2014. The high school graduation rate population distribution was negatively skewed, and ranged from 18.23% to 100.00%. The 2014 high school graduation rate distribution had a mean and median high school graduation rate of 81.89% and 86.72% respectively. The mode was 88.84%.

Figure 4.27. 2014 High School Graduation Rate Distribution and Probability Plot
There were 762,892 distinct screens that matched all the criteria for inclusion in the high school graduation rate variable analysis in 2015. The high school graduation rate population distribution was negatively skewed, and ranged from 18.23% to 100.00%. The 2015 high school graduation rate distribution had a mean and median high school graduation rate of 81.43% and 86.25% respectively. The mode was 61.03%.

Figure 4.28. 2015 High School Graduation Rate Distribution and Probability Plot
The high school graduation rate trend graph above shows a decline in the screening population’s high school graduation rate over the 6 years studied, even as the screening volume increased over the years. This is a potential indication that as the screening volume was increasing, it was more inclusive of people with limited high school education; an important inclusion, since the less educated are more vulnerable to not being aware of actions that help keep preventive diseases at bay.
4.4.2. High School Graduation Summaries for the Combined 6 Years Studied

Data from all 6 years that met our criteria for inclusion were later combined, in order to get an overall high school graduation rate view. The data was then grouped into high school graduation rate buckets as outlined in the methodology section, and analyzed for screening and positivity trends as they relate to the high school graduation rates. The results were summarized and have been provided below.

**US Heat Map of CRC Screening Footprint by High School Graduation Rate Bucket**

Figure 4.30. US heat map above shows the geographical heat distribution by high school graduation rate bucket, of the screened patients that met the criteria for inclusion in the 6 year cross sectional study.
Screening and Positivity Rate by High School Graduation Rate Bucket

Figure 4.31. Graphical representation of screening population’s high school graduation rate by size in relationship to positivity rate

The bubble graph above depicts the screening and positivity rates by high school graduation rate bucket for the 6 years of study. The bigger and further right along the horizontal axis a bubble is, the higher the screen rate. The higher along the vertical axis a bubble is, the higher the positivity rate. Each bubble has been labeled to show the poverty rate bucket that it represents.
<table>
<thead>
<tr>
<th>High School Graduation Rate</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0% - 19.9%</td>
<td>0.06%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.04%</td>
<td>0.02%</td>
<td>0.01%</td>
</tr>
<tr>
<td>20.0% - 39.9%</td>
<td>0.89%</td>
<td>1.84%</td>
<td>2.41%</td>
<td>3.19%</td>
<td>3.77%</td>
<td>3.87%</td>
</tr>
<tr>
<td>40.0% - 59.9%</td>
<td>3.93%</td>
<td>5.26%</td>
<td>5.59%</td>
<td>6.13%</td>
<td>7.61%</td>
<td>8.96%</td>
</tr>
<tr>
<td>60.0% - 79.9%</td>
<td>18.43%</td>
<td>27.96%</td>
<td>30.15%</td>
<td>32.44%</td>
<td>33.26%</td>
<td>34.74%</td>
</tr>
<tr>
<td>80.0% - 100.0%</td>
<td>76.70%</td>
<td>74.04%</td>
<td>72.38%</td>
<td>70.52%</td>
<td>69.13%</td>
<td>67.76%</td>
</tr>
</tbody>
</table>

Table 4.9 Tabular distribution of the screening rate by high school graduation rate bucket

<table>
<thead>
<tr>
<th>High School Graduate Positive Rate</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0% - 19.9%</td>
<td>6.94%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>5.30%</td>
<td>8.82%</td>
<td>2.94%</td>
</tr>
<tr>
<td>20.0% - 39.9%</td>
<td>6.37%</td>
<td>9.53%</td>
<td>7.46%</td>
<td>5.86%</td>
<td>5.50%</td>
<td>4.92%</td>
</tr>
<tr>
<td>40.0% - 59.9%</td>
<td>7.89%</td>
<td>10.50%</td>
<td>8.99%</td>
<td>7.77%</td>
<td>8.78%</td>
<td>7.87%</td>
</tr>
<tr>
<td>60.0% - 79.9%</td>
<td>7.93%</td>
<td>11.23%</td>
<td>10.02%</td>
<td>8.64%</td>
<td>9.68%</td>
<td>9.29%</td>
</tr>
<tr>
<td>80.0% - 100.0%</td>
<td>8.03%</td>
<td>11.17%</td>
<td>9.29%</td>
<td>8.46%</td>
<td>10.43%</td>
<td>11.18%</td>
</tr>
</tbody>
</table>

Table 4.10 Tabular distribution of the positivity rate by high school graduation rate bucket
4.5. Trend and Descriptive Analysis Results for Bachelor's Degree Attainment Rate

3,733,784 distinct screens out of 4,229,705 screens conducted between 2010 and 2015, had an associated bachelor’s degree attribute in the US census socioeconomic data table. The population bachelor’s degree attainment rate for all patients who got screened during the study period was positively skewed, and ranged from 0.00% to 100.00%. The overall mean, median and mode rates for bachelor’s degree attainment were 31.11%, 29.49% and 39.90% respectively.

![Distribution and Probability Plot for Bachelors_Over44](image)

Figure 4.32. Combined 6 Year (2010 – 2015) Study Period Bachelor’s Degree Attainment and Probability Plot Rate Distribution
The bachelor’s degree attainment attribute was not resolved for all screening records. This attribute is also one of the socioeconomic attribute of a population that the US Census Bureau excludes from the publicly available demographic data when the population is so small that disclosing the information could lead to the unintentional identification of individuals.

For the purpose of analyzing the screening population's bachelor's degree attainment rate, screening records without a resolved bachelor's degree attainment rate attribute were excluded from this analysis. Shown below is the yearly distribution of records that were resolved for bachelor’s degree attainment rate.

<table>
<thead>
<tr>
<th>Year</th>
<th>Total Distinct Screens</th>
<th>Distinct Screens with Bachelor’s Degree Attainment Rate Attributes</th>
<th>Positive Results</th>
<th>±∆ Year-Over-Year Distinct Screen Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>407,244</td>
<td>382,355</td>
<td>23,393</td>
<td>-</td>
</tr>
<tr>
<td>2011</td>
<td>615,317</td>
<td>528,834</td>
<td>47,533</td>
<td>38.31%</td>
</tr>
<tr>
<td>2012</td>
<td>675,105</td>
<td>599,833</td>
<td>44,136</td>
<td>13.43%</td>
</tr>
<tr>
<td>2013</td>
<td>736,987</td>
<td>652,634</td>
<td>42,800</td>
<td>8.80%</td>
</tr>
<tr>
<td>2014</td>
<td>814,170</td>
<td>707,773</td>
<td>56,860</td>
<td>8.45%</td>
</tr>
<tr>
<td>2015</td>
<td>980,882</td>
<td>762,850</td>
<td>69,929</td>
<td>7.78%</td>
</tr>
<tr>
<td>Totals</td>
<td>4,229,705</td>
<td>3,634,279</td>
<td>284,651</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.11. Distribution of CRC screens with a resolved bachelor’s degree attainment rate attribute by year
4.5.1. Bachelor’s Degree Attainment Rate Attributes Distribution by Year

The screened population that met our criteria for inclusion was further analyzed by year of service to determine if there were variations in the bachelor’s degree attainment rate over the years.

2010 had 382,355 distinct screens that matched all the criteria for inclusion in the study population for bachelor’s degree attainment rate variable analysis. The bachelor’s degree attainment distribution was positively skewed, and ranged from 0.00% to 100.00%. The 2010 bachelor’s degree attainment distribution had a mean and median attainment rate of 30.67% and 29.41% respectively. The mode was 28.23%.

Figure 4.33. 2010 Bachelor's Degree Attainment Rate Distribution and Probability Plot
In 2011, 528,834 distinct screens matched all the criteria for inclusion in the bachelor’s degree attainment study population variable analysis. The bachelor’s degree attainment distribution was positively skewed, and ranged from 0.00% to 100.00%. The 2011 bachelor’s degree attainment distribution had a mean and median poverty rate of 31.02% and 29.41% respectively. The mode was 28.23%.

Figure 4.34. 2011 Bachelor's Degree Attainment Rate Distribution and Probability Plot
In 2012, 599,833 screens matched all the criteria for inclusion in the bachelor’s degree attainment study population variable analysis. The bachelor’s degree attainment distribution was positively skewed, and ranged from 0.00% to 100.00%. The 2012 bachelor’s degree attainment distribution had a mean and median poverty rate of 31.14% and 29.35% respectively. The mode was 34.90%.

Figure 4.35. 2012 Bachelor’s Degree Attainment Rate Distribution and Probability Plot
In 2013, 652,634 screens matched all the criteria for inclusion in the bachelor’s degree attainment study population variable analysis. The bachelor’s degree attainment distribution was positively skewed, and ranged from 0.00% to 100.00%. The 2013 bachelor’s degree attainment distribution had a mean and median poverty rate of 31.06% and 29.34% respectively. The mode was 34.90%.

Figure 4.36.2013 Bachelor's Degree Attainment Rate Distribution and Probability Plot
In 2014, 707,773 screens matched all the criteria for inclusion in the bachelor’s degree attainment study population variable analysis. The bachelor’s degree attainment distribution was positively skewed, and ranged from 0.00% to 100.00%. The 2014 bachelor’s degree attainment distribution had a mean and median poverty rate of 31.20% and 29.60% respectively. The mode was 34.90%.

![Distribution and Probability Plot for Bachelors_Over44](image)

**Figure 4.37.2014 Bachelor's Degree Attainment Rate Distribution and Probability Plot**
In 2015, 762,850 screens matched all the criteria for inclusion in the bachelor’s degree attainment study population variable analysis. The bachelor’s degree attainment distribution was positively skewed, and ranged from 0.00% to 100.00%. The 2015 bachelor’s degree attainment distribution had a mean and median poverty rate of 31.36% and 29.73% respectively. The mode was 28.95%.

Figure 4.38. 2015 Bachelor’s Degree Attainment Rate Distribution and Probability Plot
The bachelor’s attainment rate trend graph above shows that, as the colorectal cancer screening volume grew over the years, from 2010 through 2015, the screening population’s bachelor’s attainment mean and median rates did not significantly change over the same period.
4.5.2. Bachelor’s Attainment Rate Summaries for the Combined 6 Years Studied

An analysis of the bachelor’s degree attainment rates all screens meeting the criteria for inclusion in the study period was also completed. The data from all 6 years were combined, and the then grouped into bachelor’s degree attainment rate buckets as outlined in the methodology section, and then analyzed for screening and positivity trends as they relate to the bachelor’s degree attainment rates. The results were then summarized and have been provided below.

**US Heat Map of CRC Screening Footprint by Bachelor's Degree Attainment Rate Bucket**

![US Heat Map](image)

Figure 4.40. US heat map above shows the geographical heat distribution by bachelor’s degree attainment rate bucket, of the screened patients that met the criteria for inclusion in the 6 year cross sectional study
The bubble graph above depicts the screening and positivity rates by bachelor's degree attainment rate bucket for the 6 years of study. The bigger and further right along the horizontal axis a bubble is, the higher the screen rate. The higher along the vertical axis a bubble is, the higher the positivity rate. Each bubble has been labeled to show the poverty rate bucket that it represents.

Figure 4.4: Graphical representation of screening population's bachelor's degree attainment rate by size in relationship to positivity rate
### Table 4.12. Tabular distribution of the screening rate by bachelor's degree attainment bucket

<table>
<thead>
<tr>
<th>Bachelors Graduation Rate</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0% - 19.9%</td>
<td>10.64%</td>
<td>11.10%</td>
<td>10.79%</td>
<td>11.35%</td>
<td>11.91%</td>
<td>11.23%</td>
</tr>
<tr>
<td>20.0% - 39.9%</td>
<td>77.90%</td>
<td>75.29%</td>
<td>75.35%</td>
<td>74.71%</td>
<td>72.94%</td>
<td>72.78%</td>
</tr>
<tr>
<td>40.0% - 59.9%</td>
<td>9.55%</td>
<td>10.99%</td>
<td>11.06%</td>
<td>11.17%</td>
<td>11.94%</td>
<td>12.52%</td>
</tr>
<tr>
<td>60.0% - 79.9%</td>
<td>1.64%</td>
<td>2.36%</td>
<td>2.53%</td>
<td>2.57%</td>
<td>2.98%</td>
<td>3.29%</td>
</tr>
<tr>
<td>80.0% - 100.0%</td>
<td>0.28%</td>
<td>0.25%</td>
<td>0.27%</td>
<td>0.20%</td>
<td>0.24%</td>
<td>0.18%</td>
</tr>
</tbody>
</table>

### Table 4.13. Tabular distribution of the positivity rate by bachelor's degree attainment rate bucket

<table>
<thead>
<tr>
<th>Bachelors Graduate Positive Rate</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0% - 19.9%</td>
<td>7.51%</td>
<td>10.41%</td>
<td>8.75%</td>
<td>7.93%</td>
<td>9.11%</td>
<td>9.38%</td>
</tr>
<tr>
<td>20.0% - 39.9%</td>
<td>8.05%</td>
<td>11.24%</td>
<td>9.53%</td>
<td>8.52%</td>
<td>9.96%</td>
<td>10.28%</td>
</tr>
<tr>
<td>40.0% - 59.9%</td>
<td>7.92%</td>
<td>11.21%</td>
<td>9.35%</td>
<td>8.30%</td>
<td>10.99%</td>
<td>11.43%</td>
</tr>
<tr>
<td>60.0% - 79.9%</td>
<td>8.96%</td>
<td>11.01%</td>
<td>8.92%</td>
<td>7.91%</td>
<td>11.54%</td>
<td>11.70%</td>
</tr>
<tr>
<td>80.0% - 100.0%</td>
<td>7.23%</td>
<td>10.20%</td>
<td>9.19%</td>
<td>7.94%</td>
<td>14.42%</td>
<td>12.62%</td>
</tr>
</tbody>
</table>
4.6. Correlation Analysis of the Screening Positivity Rates and the Socioeconomic Attributes

The Pearson’s correlation analysis was done to assess the nature of the relationship between positivity rates and each of the selected socioeconomic attributes used in our study. For this stage of the analysis, only records which had all four socioeconomic attributes resolved for were included in the analysis. The data was then summarized by zip code, and then the ‘PROC CORR’ procedure was then run on each of the socioeconomic and the positivity rate. Bachelor’s degree attainment, poverty rate and median income showed negative correlations in relationship with positivity rate. High school graduation rate had a positive correlation in relationship with positivity rate. The procedure results are shown in the tables below.

<table>
<thead>
<tr>
<th>Pearson Correlation Coefficients, N = 7291</th>
<th>Pearson Correlation Coefficients, N = 7291</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bachelors Over44</strong></td>
<td><strong>Poverty Rate</strong></td>
</tr>
<tr>
<td>Bachelors_Positivity Rate</td>
<td>Poverty Rate</td>
</tr>
<tr>
<td>1</td>
<td>-0.01081</td>
</tr>
<tr>
<td>-0.01081</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pearson Correlation Coefficients, N = 7291</th>
<th>Pearson Correlation Coefficients, N = 7291</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High School Graduation Rate</strong></td>
<td><strong>Median Income</strong></td>
</tr>
<tr>
<td>High School Graduation Rate</td>
<td>Median Income</td>
</tr>
<tr>
<td>1</td>
<td>-0.00475</td>
</tr>
<tr>
<td>0.02034</td>
<td>0.685</td>
</tr>
<tr>
<td>0.02034</td>
<td>1</td>
</tr>
<tr>
<td>0.0824</td>
<td>0.685</td>
</tr>
</tbody>
</table>

Table 4.14. Correlation analysis summaries between each of the four socioeconomic attributes and positivity rate

4.7. Multiple Regression Analysis of the Four Socioeconomic Attributes as Predictors of a Population’s CRC Screening Positivity Rate

We also ran a multiple regression analysis, using the four socioeconomic attributes being studied to generate a model for predicting a population’s positivity rate. We included the patient age in the model, given that the risk for developing colorectal cancer increases with
age, even for the average risk patients beginning at age 50. The thought here is that the inclusion of the patient age variable will improve the accuracy of the predictive model for the purpose of predicting a population’s positivity rate. The SAS ‘PROC REG’ procedure was used to run the multiple linear regression procedure.

The stepwise multiple linear regression approach was used to generate the predictive model for positivity rate. This approach showed each variable’s contributing effect on predicting the population’s positivity rate. The result showed a statistical significance of <0.0001 for all four socioeconomic attributes, as well as the age variable at each stepwise inclusion, an indication that all 5 variables were important and needed for building a model for predicting a population’s positivity rate. The age variable was the biggest predictor variable based on its R square value. The result tables are shown below.

**Stepwise Multiple Regression Results for Positivity Rate**

<table>
<thead>
<tr>
<th>Variable Entered</th>
<th>Partial R-Square</th>
<th>Model R-Square</th>
<th>C(p)</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Age</td>
<td>0.0185</td>
<td>0.0185</td>
<td>504.66</td>
<td>3089.9</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>2 High_school_Over44</td>
<td>0.0005</td>
<td>0.019</td>
<td>426.39</td>
<td>80.06</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>3 Median_Income</td>
<td>0.001</td>
<td>0.02</td>
<td>253.7</td>
<td>174.42</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>4 Bachelors_Over44</td>
<td>0.0012</td>
<td>0.0212</td>
<td>57.353</td>
<td>198.28</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>5 Percentage_Poverty</td>
<td>0.0003</td>
<td>0.0215</td>
<td>6</td>
<td>53.35</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

**Summary of Stepwise Selection**

| Variable            | Parameter Estimate | Standard Error | t Value | Pr > |t| | Standard Estimate |
|---------------------|--------------------|----------------|--------|------| |-------------------|
| Intercept           | -0.04779           | 0.007          | -6.82  | <.0001 | 0 | |
| Age                 | 0.00217            | 3.90E-05       | 55.95  | <.0001 | 0.13685 | |
| High_school_Over44  | 0.07253            | 0.00627        | 11.57  | <.0001 | 0.04175 | |
| Percentage_Poverty  | -0.06614           | 0.00905        | -7.3   | <.0001 | -0.03188 | |
| Bachelors_Over44    | -0.06905           | 0.0049         | -14.08 | <.0001 | -0.03672 | |
| Median_Income       | -5.37E-07          | 2.94E-08       | -18.29 | <.0001 | -0.07018 | |

**Table 4.15. Stepwise multiple regression summary by the 4 socioeconomic attributes studied, and the age attribute**

**Table 4.16. Stepwise multiple regression parameter estimates by the 4 socioeconomic attributes studied, and the age attribute**
Using the parameter estimates output from the stepwise multiple linear regression, the positivity rates predictive model is given below.

\[
\text{Positivity Rate} = -0.04779 + (0.00217*\text{Age}) + (0.07253*\text{High\_school\_Over44}) + (-0.06614*\text{Percentage\_Poverty})*(-0.06905*\text{Bachelors\_Over44})*(-0.000000537304*\text{Median\_Income})
\]

The bachelor’s degree attainment rate and median income which both had inverse relationships with positivity rate can be seen as strong indicators of positive financial wellbeing of the population, which could predispose the population to being in a better position to afford healthier foods, as well as engaging in lifestyles to promote better health outcomes. As a result, their inverse relationship with positivity rate was not surprising, given the literatures that were reviewed indicated that diet and lifestyle choices were some of the factors that could influence the onset of colorectal cancer. The positive relationship between age and positivity rate was also expected because advancing age, beginning at age 50 is also known to predispose a patient with average risk to developing the disease.

Less easily explained were the positive relationship between high school graduation rate and positivity rate, and inverse relationship between poverty rate and positivity rate. It is not immediately clear why populations with a higher graduation rate will be more likely to test positive on the CRC screen test, or the poorer population will be less likely to have a positive rate. A potential reason for the positive relationship between the high school graduation rate and positivity rate could be the change in screening rate for the high school graduate populations. Reverting back to the high school graduate screening trends graph, we see that
as the number of patients screened grew over time from 2010 through 2015, the high school graduation rate mean and median rates for the screened patients was declining over the same period. When these are all taken into consideration, it could be that the positive relationship between the positivity rate and high school graduation rate is as a result of the declining high school graduation rate even as the overall screening volume increases over the years.

Also analyzed, was the positivity rate across the years to see if there were any significant differences in positivity rate from one year to the next.

4.8. ANOVA Analysis on Positivity Rates by Year

ANOVA analysis was carried out using the SAS ‘PROC ANOVA’ procedure with Waller-Duncan means option. The result showed that there were variations in the mean positivity rate over the years. 2015 showed the highest mean positivity rate, while 2010 showed the least. The yearly means appeared to be more sporadic, with the highest to lowest mean in the order 2015, 2011, 2014, 2012, 2013 and 2010 respectively.

<table>
<thead>
<tr>
<th>Waller Grouping</th>
<th>Mean</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.113588</td>
<td>2015</td>
</tr>
<tr>
<td>B</td>
<td>0.102659</td>
<td>2011</td>
</tr>
<tr>
<td>C</td>
<td>0.09704</td>
<td>2014</td>
</tr>
<tr>
<td>D</td>
<td>0.089033</td>
<td>2012</td>
</tr>
<tr>
<td>E</td>
<td>0.079855</td>
<td>2013</td>
</tr>
<tr>
<td>F</td>
<td>0.065174</td>
<td>2010</td>
</tr>
</tbody>
</table>

Table 4.17. Waller-Duncan positivity rate mean grouping by year

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Anova SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>5</td>
<td>122.949</td>
<td>24.5897235</td>
<td>534.26</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R-Square</th>
<th>Coeff Var</th>
<th>Root MSE</th>
<th>Positivity Rate Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.005011</td>
<td>232.153</td>
<td>0.21454</td>
<td>0.092411</td>
</tr>
<tr>
<td>Year Comparison</td>
<td>Difference Between Means</td>
<td>Simultaneous 95% Confidence Limits</td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>--------------------------</td>
<td>-----------------------------------</td>
<td></td>
</tr>
<tr>
<td>2015 - 2011</td>
<td>0.0109287</td>
<td>0.0080599 0.0138 ***</td>
<td></td>
</tr>
<tr>
<td>2015 - 2014</td>
<td>0.0165475</td>
<td>0.0137705 0.01932 ***</td>
<td></td>
</tr>
<tr>
<td>2015 - 2012</td>
<td>0.0245552</td>
<td>0.0217218 0.02739 ***</td>
<td></td>
</tr>
<tr>
<td>2015 - 2013</td>
<td>0.0337328</td>
<td>0.0309315 0.03653 ***</td>
<td></td>
</tr>
<tr>
<td>2015 - 2010</td>
<td>0.0484141</td>
<td>0.0453973 0.05143 ***</td>
<td></td>
</tr>
<tr>
<td>2011 - 2015</td>
<td>-0.0109287</td>
<td>-0.0137975 -0.0081 ***</td>
<td></td>
</tr>
<tr>
<td>2011 - 2014</td>
<td>0.0056188</td>
<td>0.0027346 0.0085 ***</td>
<td></td>
</tr>
<tr>
<td>2011 - 2012</td>
<td>0.0136265</td>
<td>0.010688 0.01657 ***</td>
<td></td>
</tr>
<tr>
<td>2011 - 2013</td>
<td>0.0228041</td>
<td>0.0198965 0.02571 ***</td>
<td></td>
</tr>
<tr>
<td>2011 - 2010</td>
<td>0.0374854</td>
<td>0.0343696 0.0406 ***</td>
<td></td>
</tr>
<tr>
<td>2014 - 2015</td>
<td>-0.0165475</td>
<td>-0.0193245 -0.0138 ***</td>
<td></td>
</tr>
<tr>
<td>2014 - 2011</td>
<td>-0.0056188</td>
<td>-0.008503 -0.0027 ***</td>
<td></td>
</tr>
<tr>
<td>2014 - 2012</td>
<td>0.0080078</td>
<td>0.0051588 0.01086 ***</td>
<td></td>
</tr>
<tr>
<td>2014 - 2013</td>
<td>0.0171853</td>
<td>0.0143683 0.02 ***</td>
<td></td>
</tr>
<tr>
<td>2014 - 2010</td>
<td>0.0318666</td>
<td>0.0288352 0.0349 ***</td>
<td></td>
</tr>
<tr>
<td>2012 - 2015</td>
<td>-0.0245552</td>
<td>-0.0273886 -0.0217 ***</td>
<td></td>
</tr>
<tr>
<td>2012 - 2011</td>
<td>-0.0136265</td>
<td>-0.016565 -0.0107 ***</td>
<td></td>
</tr>
<tr>
<td>2012 - 2014</td>
<td>-0.0080078</td>
<td>-0.0108567 -0.0052 ***</td>
<td></td>
</tr>
<tr>
<td>2012 - 2013</td>
<td>0.0091776</td>
<td>0.006305 0.01205 ***</td>
<td></td>
</tr>
<tr>
<td>2012 - 2010</td>
<td>0.0238589</td>
<td>0.0207757 0.02694 ***</td>
<td></td>
</tr>
<tr>
<td>2013 - 2015</td>
<td>-0.0337328</td>
<td>-0.0365341 -0.0309 ***</td>
<td></td>
</tr>
<tr>
<td>2013 - 2011</td>
<td>-0.0228041</td>
<td>-0.0257117 -0.0199 ***</td>
<td></td>
</tr>
<tr>
<td>2013 - 2014</td>
<td>-0.0171853</td>
<td>-0.0200023 -0.0144 ***</td>
<td></td>
</tr>
<tr>
<td>2013 - 2012</td>
<td>-0.0091776</td>
<td>-0.0120502 -0.0063 ***</td>
<td></td>
</tr>
<tr>
<td>2013 - 2010</td>
<td>0.0146813</td>
<td>0.0116276 0.01774 ***</td>
<td></td>
</tr>
<tr>
<td>2010 - 2015</td>
<td>-0.0484141</td>
<td>-0.0514309 -0.0454 ***</td>
<td></td>
</tr>
<tr>
<td>2010 - 2011</td>
<td>-0.0374854</td>
<td>-0.0406011 -0.0344 ***</td>
<td></td>
</tr>
<tr>
<td>2010 - 2014</td>
<td>-0.0318666</td>
<td>-0.034898 -0.0288 ***</td>
<td></td>
</tr>
<tr>
<td>2010 - 2012</td>
<td>-0.0238589</td>
<td>-0.026942 -0.0208 ***</td>
<td></td>
</tr>
<tr>
<td>2010 - 2013</td>
<td>-0.0146813</td>
<td>-0.017735 -0.0116 ***</td>
<td></td>
</tr>
</tbody>
</table>

Comparisons significant at the 0.05 level are indicated by ***.

Table 4.18. ANOVA analysis paired years means comparisons
CHAPTER 5 CONCLUSIONS, STUDY LIMITATIONS AND NEXT STEPS

5.1. Conclusions

This research had a primary goal of understanding the socioeconomic attributes of the screening population for colorectal cancer for patients aged 50 and over who utilized the services of a US national clinical diagnostic lab. The study spanned a 6 year period beginning in 2010 through 2015 and included distinct screen visits that met specific criteria. Some of the research questions that needed to be answered are outlined below:

**Research Question 1** - Does a population's rate of graduation from high school or Bachelors program have a statistically significant impact on the population's CRC screening rate?

**Research Question 2** - Does a population's median income have a statistically significant influence on the population's CRC screening rate?

**Research Question 3** - Does a population's poverty rate have a statistically significant influence on the population's CRC screening rate?

**Research Question 4** - Do the results change statistically from year to year?

Some of the major findings from the analyses of the clinical diagnostics data and review of literatures in support of my research are outlined below:

- There was a progressive increase in the colorectal cancer screening volume over the 6 years studied. A positive development in support of increasing the odds of diagnosing the colorectal cancer at the earliest stage when it is most curable, and conversely reducing the incidence and mortality rate of the disease United States
While there were differences in the positivity rate for colorectal cancer screening over the 6 years studied, there were no patterns or trends observed.

There were changes in the screening population’s median income over the years. As the volume of the screening population grew over the years, the population’s median income declined. Because the same socioeconomic figures (2010 US census) was used, which ensured there were no growths or decline in median income over the study period, this could be seen as an indication that over the years, the test was becoming more easily accessible and within the reach of the less affluent.

In relation to the median income, the study also found that there was an inverse relationship between the positivity rate for CRC screening and median income. If it is taken into consideration that the more affluent are usually in a better position to afford healthier foods as well as engaging in lifestyles activities that promote better health, then it is not a surprising finding.

The screening population’s poverty rate also showed an increase over the years, as the screening volume grew; another potential indication that CRC screening awareness and accessibility is increasingly reaching some of the poorer population. This is remarkable, because poverty is known to be one of the biggest barriers to wellness care, so any indication that they are getting more involved in receiving preventive care is move in the right direction.

Poverty rate had an inverse relationship with positivity rate.

The bachelor’s attainment rate trend remained relatively stable over the 6 years, even as the screening population grew. In addition, the bachelor’s attainment rate had an inverse relationship with positivity rate. This inverse relationship between
bachelor’s attainment rate and positivity rate could be attributable to a predisposition to higher income, which makes it easier for the population to afford healthier food and lifestyle.

- The high school graduation rate showed a decline over the years as the colorectal cancer screening volume increased, and there was a positive relationship between the high school graduation rate and positivity rate. Neither of the two analytical findings for high school graduation rate could be readily explained.

- The positivity rate by year was found to be sporadic at best. There were no trends observed over the years when all of the data from each year was combined and the positivity rates compared. This could be partially accounted for by the fact that the entire population from each year from the study population was combined, and that the study was done on a population with an average risk of developing the disease and no other stratifying attribute. This approach increases the likelihood that the odds of a positive screen result over time will be random in the absence of a sweeping change that affects all stratifications within the population in relation to the biggest factors that predispose an average risk individual to developing the disease.

5.2. Study Limitations

There were some limitations that were identified prior to, and during the course of the study that could not be adjusted for are outlined below:

- The clinical diagnostics dataset utilized for this study did not have the patient race variable, a very important socioeconomic variable. This would have been
very useful in gaining additional insights into the existence of colorectal cancer screening and positivity rates disparities across races, in relation to other socioeconomic attributes

- The colorectal cancer screening data used for the study is limited to tests conducted by a single national clinical diagnostics lab, so the data may under represent geographical areas where the national lab has limited footprint

- The study will not show the entire universe of all CRC screening by the fecal immunochemistry method in the US, just those carried out by the specific national clinical diagnostics lab

- The colorectal cancer screening test data being analyzed only represents screenings that were completed using the InSureFit® Fecal Immunochemical Test (FIT) method, so the results will not necessarily describe the entire universe of colorectal screening such as sigmoidoscopy, colonoscopy, and so on.

- As one of the least invasive screening tests recommended for patients of average risk, it’s sensitivity for colorectal cancer is 87% and specificity of 98%. As such, a positive result does not always translate to a positive diagnosis for colorectal cancer. A follow up with the more invasive colonoscopy is usually required

- The average risk population was identified “loosely” by capturing the entire population of patients age 50 and over. No adjustment was made for conditions such as family history, personal history, and so on, which could predispose the patient to developing colorectal cancer
5.3. Next Steps

It is recommended that this study be extended to address some outstanding socioeconomic questions related to colorectal cancer screening and more robust insights, using a more encompassing dataset. A dataset that spans across multiple clinical diagnostics labs, doctors’ offices and hospitals enhance this study by making the findings not just relevant to InSureFit® Fecal Immunochemical Test, but to all approved colorectal cancer screening types. In addition, this will also lend itself to a longitudinal study approach that will enable the measurement of adherence to recommended screening schedules by socioeconomic attributes, irrespective of the healthcare facility that the patients chooses for CRC screening and/or CRC screening follow up.

It will also be worthwhile to incorporate race into the study, which should shed some light on any racial disparities that may exist with regards to screening, as well as the effect of race in conjunction with other socioeconomic attributes on screening and positivity rates.

A study that goes on to identify truly positive screens by utilizing a more robust dataset with fields that make it possible to follow any given patient throughout their colorectal cancer screening journey, flag the patients for all predisposing factors to colorectal cancer, as well as identify patients who tests positive on the InSureFit® Fecal Immunochemical Test, and go on to have a colonoscopy and be conclusively confirmed to be positive for colorectal cancer or polyps – a predecessor to colorectal cancer will also be necessary for future steps.
REFERENCES


14. Dik VK, Van Oijen MGH, Smeets HM, Siersema PD. Frequent use of antibiotics is associated with colorectal cancer risk: Results of a nested case–control study. 2015;61.


18. Slattery ML, Potter JD, Ma KN, Cann BJ, Leppert M, Samowitz W. Western diet, Family history of colorectal cancer, NAT2, GSTM-1 and risk of colorectal cancer. 2000;11(1).


