Analyses of Colon Cancer Inpatients in the United States

By

Abdullah Alqahtani

A Dissertation Submitted In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Biomedical Informatics

Department of Health Informatics

School of Health Related Professions

Rutgers, The State University of New Jersey

October 2014

RUTGERS
School of Health Related Professions

Final Dissertation Approval Form

Analysis of Colon Cancer Inpatients in the United States

BY

Abdullah Alqahtani

Dissertation Committee:

Shankar Srinivasan PhD

Syed Haque PhD

Dinesh Mital PhD

Approved by the Dissertation Committee:

_______________________________________ Date ____________

_______________________________________ Date ____________

_______________________________________ Date ____________

_______________________________________ Date ____________

_______________________________________ Date ____________

_______________________________________ Date ____________

TABLE OF CONTENTS

Chapter III - Methodology

Chapter IV- Results & Discussion

Chapter V CONCLUSIONS & FUTURE RESEARCH

ABSTRACT

The overall goal of the project was to identify the factors and costs associated with Colon Cancer patients in terms of mortality, length of stay and costs in different types of clinical settings across the United States. Accordingly this research study utitlized the datasets for 2008 to 2010 available from the Nationwide Inpatient Sample (NIS) database with hospitalization characteristics of patients admitted with Colon Cancer as the principal diagnosis. Some of the important results found in this study were:

- Between 2008 and 2010 the age and population adjusted incidences and the hospital discharges both decreased significantly which is a promising trend speaking well of the state of health care in the United States as also possibly due to the effectiveness of nutritional counseling, patient education, screening for men aged 50 and above.
- It was found that while the total number of colon cancer patient discharges decreased significantly between 2008 to 2010, the Total Charges however significantly risen up between 2008 and 2010. The mean charges increased by nearly 8 %.
- The number of discharges across the various hospital types and their locations across the United States as shown above revealed that those large hospitals in metropolitan regions and those that are private not-for-profit have more discharges compared to the other types.
- Patients who are uninsured and those on Medicaid (low income) are more in number over the years 2008 to 2010 as compared to those on Medicare and Private Insurance which have decreasing trends.
- It was found that the mean and median length of stay of colon cancer patient discharges remained more or less the same between 2008 and 2010.
- It was found that the number of in-hospital mortality or deaths significantly reduced between 2008 and 2010. Alongside Home Health Care increased while discharges to another hospital also decreased (with a

smaller decrease in discharge to another institution such as rehab facility and nursing home).

- The number of in-hospital deaths has a decreasing trend in the number of deaths over the years 2008 to 2010.
- Southern United States has more (nearly 2 times) in-hospital deaths compared to the other regions in all the 3 years.
- This study seems to indicate that mortality is positively correlated with the total costs and this may be due to a significant admission source is from emergency department.
- The in-hospital mortality prediction model above revealed significant risk for patients with hypertension (nearly 100%) and with obesity (62 % more)
- Patients with white ethnicity have lower risk of dying in the hospital compared to the other ethnicities which all have similar odds ratio intervals.

This research study was limited to the datasets available from the Nationwide Inpatient Sample (NIS) database with hospitalization characteristics of patients admitted with Colon Cancer as the principal diagnosis. A similar large scale dataset based future study is indeed warranted to analyze demographic and hospital based outcomes for a wider variety of Colon Cancer patients admitted for screening, cancer management, clinical trials and education.

ACKNOWLEDGEMENTS

First and foremost I would like to express my gratitude to my advisor Dr. Shankar Srinivasan, for essential guidance and insight throughout my graduate career. His influence has contributed to my development as a scientist. Sincere appreciation is due to Dr. Syed Haque, Dr. Dinesh Mital and Dr. Masayuki Shibata who have enlarged my sphere of understanding of Biomedical Informatics with their encyclopedic knowledge and fresh perspectives.

List of Figures

List of Tables

Chapter I

Introduction

1.1. Background of the Disease:

Cancer is a disease in which cells in a particular part of the body proliferate out of control and if such an uncontrolled cell growth were to occur in the colon, it is called Colon Cancer (CC)[1]. As shown in Figure 1 the colon is the large bowel leading to the rectum which is the passageway that connects the colon to the anus. The colon is the primary target of many pathophysiological disorders and which could be of acute or chronic quality. Colon cancer is prevalent worldwide though it has been reported that there is a difference in the disease pattern across the developing and developed nations and this is true specifically of the developing nations with tropical and subtropical areas[2].

Over the years several evidence-based recommendations have been made for the colon screening of adults 50 years or older who may pose a possible risk for CC. It has been shown that the early detection and prompt removal of adenomous polyps can diminish both the incidence of CC and improve the survival rate. Unfortunately however it is reported that nearly half the population at risk in the United States has not been screened at all for this otherwise preventable disease[3].

Besides the guidelines for screening there has also been significant improvement in the last decade in the design and development of effective diagnostic tools for colon diseases specifically the colonoscopy procedure[4]. Colonoscopy examinations are deemed to be very effective surveillance techniques for detection of colonic lesions[5].

Figure 1: Illustration of the Colon Region in the body (Microsoft ClipArt).

It has been reported that in the United States, Colon Cancer (CC) killed more than 50,000 people in the year 2010[6]. According to the American Cancer Society CC is the third most common type of cancer and second most common cause of cancer death in the U.S[4]. It represented the second cause of death as colorn cancer accounted for 60000 deaths per year[1]. Although colon cancer is one of the leading causes of cancer-related deaths in the United States, in almost all cases, early diagnosis of colon cancer can lead to a complete cure. This is because many colon cancers are thought to arise from adenomatous polyps in the colon. These polyps are usually benign, but some can develop into cancer over a period of time[7].

The majority of the time, the diagnosis of localized colon cancer is through colonoscopy. Among the various screening methods used to detect the cancer, the role of colonoscopy in the prevention of CC has been recommended by both the medical community and the federal government[8]. The ACS, the American Gastroenterological Association and various health organizations have advised that routine screening for CC should begin at age 50 for individuals of average risk and be repeated at 10 year intervals. Without screening, 40% of all CC cancers were assumed to result in death within 5 years[9]. The Morbidity and Mortality Weekly Report in 2001, stated that the percentage of persons aged > 50 years who received screening colonoscopy had steadily increased from 30% in 1997 to 33% in 1999 to almost 40% by 2001. However improving screening rates is still of concern, particularly in the state of Connecticut where 42% of adults aged 50 or older have not been screened and an estimated 1,710 new cases are diagnosed and 660 people die of CC each year[10].

1.2. Goals and Objectives:

The overall goal of the project is to identify the factors and costs associated with Colon Cancer patients in terms of mortality, length of stay and costs in different types of clinical settings across the United States. Specifically the objectives are to determine:

1) what clinical factors (such as number and types of comorbidities and procedures) influence the mortality, costs and length of stay
2) whether mortality, costs and length of stay differ with race, age, or socio-economic status
3) whether there are differences in the mortality, costs and length of stay across the various regions of the US
4) whether there are differences in the mortality, costs and length of stay amongst the different types of hospital settings – rural/urban/hospital with and without teaching.

1.3. Research Hypotheses of the Project:

- Are there statistically significant associations between the number and types of comorbidities and procedures and mortality, costs and length of stay of Colon Cancer patients

- Are there statistically significant differences in mortality, costs and length of stay of Colon Cancer patients with race, age, or socio-economic status
- Are there statistically significant differences in the mortality, costs and length of stay of Colon Cancer patients across the various regions of the US
- Are there statistically significant differences in the mortality, costs and length of stay of Colon Cancer patients amongst the different types of hospital settings – rural/urban/hospital with and without teaching.

1.4. Data & Methods:

In this project we plan to utilize the datasets obtained from the Nationwide Inpatient Sample (NIS) database towards our analyses of Colon Cancer patients. The NIS is the largest all-payer inpatient care database in the United States containing data from 1998 to 2011. It contains data from approximately 8 million hospital stays each year accruing from all discharge data from 1,050 hospitals located in 44 States, approximating a 20-percent stratified sample of U.S. community hospitals. The sampling frame for the 2011 NIS is a sample of hospitals that comprises approximately 95 percent of all hospital discharges in the United States. The NIS includes more than 100 clinical and nonclinical data elements for each hospital stay. These include:

- Primary and secondary diagnoses
- Primary and secondary procedures
- Admission and discharge status
- Patient demographics (e.g., gender, age, race, median income for ZIP Code)
- Expected payment source
- Total charges
- Length of stay
- Hospital characteristics (e.g., ownership, size, teaching status).

Furthermore, the NIS is the only national hospital database containing charge information on all patients, regardless of payer, including persons covered by Medicare, Medicaid, private insurance, and the uninsured.

The outcomes of interest as indicated in the goals and hypotheses above are the mortality, the length of stay and the costs involved. Using the datasets obtained from the NIS database appropriate descriptive and inferential statistics will be effected. To relate the factors associated with the research outcome, the length of stay and the costs a multiple regression model will be setup and validated. Predictive models such as logistic regression will be employed to determine the risks and ratios for the various factors influencing mortality such as race, age groups, number and types of procedures and comorbidities. Details as to the state of art knowledge and research into Colon Cancer and its management are provided in the next chapter.

Chapter II

Literature Review

2.1. Introduction:

It has been reported that in the United States, Colon Cancer (CC) killed more than 50,000 people in the year 2010[6]. According to the American Cancer Society CC is the third most common type of cancer and second most common cause of cancer death in the U.S[4]. Although colon cancer is one of the leading causes of cancer-related deaths in the United States, in almost all cases, early diagnosis of colon cancer can lead to a complete cure. This is because many colon cancers are thought to arise from adenomatous polyps in the colon. These polyps are usually benign, but some can develop into cancer over a period of time[7].

There has been in increase in the incidence of CC worldwide in the last couple of decades in both developing and developed nations. For example in Scandinavia it has been reported that over a period of nearly 50 years there has been an increase of over 50% in the annual incidence rate of colon cancer. It has been estimated that for developed nations, the average risk is 5% for adults over 50 year's old[11].

Jorgensen et al conducted outcomes study in New South Wales, Australia wherein they tried to identify the predictors for colon cancer patient outcomes. They found that increased age, colon resection procedures after emergency admission and higher stages of the disease contributed to higher mortality. They however concluded that studies need to be conducted to figure how certain hospitals and patient groups are at risk of poor care[12].

Several authors have reported improvement in survival with early screening and detection of CC and they conclude that it can reduce mortality. The 5 year survival probability is more than 90% if the disease is diagnosed at an early stage but it decreases to only about 60% if the patient has tumors in the lymph node and further down to 10% if any metastases are present. Since unfortunately clinical symptoms develop somewhat late in the progress of the disease, an early detection is often not possible in symptomatic patients[13].

2.2. Colon Cancer Stages and Treatment:

The CC stage indicates the extent of the cancer in the body and how far the cancer has invaded the large intestine and how far it has gone beyond its original location[14]. Also, the staging of colon cancer is based on the result of tests and exams such as the physical, endoscopic biopsies, blood tests and imaging test results[15]. The advantage of staging is that it allows for a more consistent treatment regimen, make it possible to compare results and treatments from other clinical studies. Accordingly the treatment protocol for the cancer will vary with the stage, the age and the health condition of the patient. Similar to other cancers the most common options for colon cancer are surgery, radiation therapy and chemotherapy[13].

Odermatt et al from a study of colon resection and its effects on mortality and recurrence concluded that complications arising from elective colon resection leads to poorer long term survival[17].

Halpern et al analyzed data from 1998-2004 obtained from the National Cancer Data Base to determine relationships between types of colon cancer and patient and treatment facility characteristics. They found that the odds of advanced stage cancer incidence were higher for those from the black population, of female gender and those from poor socioeconomic status. Also uninsured and patients on Medicaid were also found to have greater chances of presenting themselves with an advanced stage of colon cancer. They concluded that programs to improve access to colon screening amongst the poorer populations will remedy this disparity[18].

Kao et al from a study of 297 patients with colon cancer (CC) in a veterans hospital in Taiwan found that those with first degree relatives with CC had better survival chances than those without a family history of CC[19].

As shown in the figure below colon cancer is a progressive disease and is demarcated by five stages of the disease progression. In the lowest stage (Stage 0) of the staging system the cancer is found only in the innermost lining of the colon and/or rectum and as the disease progresses it moves to distant sites in the body and the Stage number increases accordingly from Stage 0 all the way to Stage 5 which indicates that metastases are now found in distant sites in the body[20].

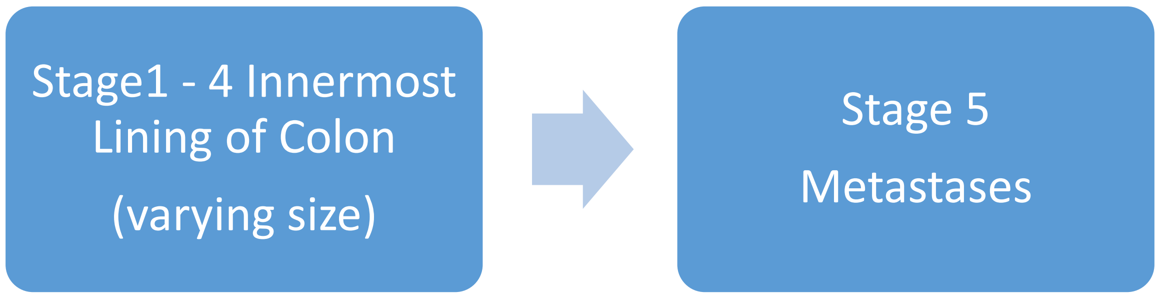


Figure 2: Stages of colon cancer.

2.3. Colon Carcinoma:

As mentioned earlier colon carcinoma (CC) is a very common neoplasm and affects lots of lives. It has been estimated that risk of developing CC during one's life over the age of 50 is approximately 5% with a 2.5% chance of mortality[21]. Most such neoplasms develop from adenomas with the possibility of malignant transformation within 10 years. Also the prevalence increases for people with ages more than 60 years upto 40%[22]**.** Once the cancer has been detected, even the 5 year survival declines significantly[23]**.**

Colon cancers are the third most frequent cancers in men and second in women. The lifetime probability of developing colon cancer in developed countries is 4.3% in men and 3.2% in women. In developed countries, rates of colon cancer are approximately four times higher than in developing countries[24]**.** Accordingly, together with serious morbidity and mortality rates, and that effective treatment can be offered in case of early detection, attention should be drawn to secondary prevention or screening[14]**.** Such screening tests should be sufficiently accurate for detecting colon cancers, acceptable to "patients", feasible in clinical practice, and cost-effective. Large scale detection and subsequent removal of lesions in an early stage, especially the advanced adenomas[23]**,** may lead to a significant decrease in the incidence of CC and a reduced cancer related morbidity and mortality[25].

Colon cancers arise through complex interactions between genetic and environmental influences. The relative contribution of each varies. Genetic factors predominate in defined hereditary syndromes such as familial adenomatous polyposis and hereditary non-polyposis colon cancer[26]. Sporadic colon cancers develop over longer periods of time as environmental influences produce genotoxic events eventually leading to cancer. In both types, tumors do not develop all at once but rather evolve from progressive identifiable changes in the colonic mucosa e.g., dysplasia, adenoma[27].

2.4. Risk Factors for Colon Cancer:

2.4.1. Age:

The risk of developing CC increases after age 40 years in the general population, with 90% of cancers occurring in persons aged 50 years and older; a 50-year-old person has approximately a 5% chance of developing CC if he or she survives to age 80 and a 2.5% risk of dying from the disease[17]. This has important implications for screening. Sporadic CCs do arise in other age groups as well), however, and this diagnosis must be considered in younger persons with signs and symptoms characteristic of this disease, especially if they have a family history of colon neoplasia[28].

The tables below (obtained from SEER) shows the age adjusted incidence rates of colon cancer for both men and women. Today the typical cancer patient is > 60 years of age has comorbidities and taking medication concurrently[29].

Table 1: Age Adjusted Incidence Rates for Males for different age groups 2007-2011.

	All Sites					Colon excluding Rectum				
	All ages	00-19	20-54	55-64	65+	All ages	00-19	20-54	55-64	65+
All races	536.1	18.2	173.6	1,181.50	2,699.10	33.6	0.1	10.3	60.3	183
White	540.3	19.4	175.8	1,165.40	2,731.90	32.9	0.1	9.9	56.1	182.4
Black	626	13.9	203.5	1,646.60	2,980.60	43.1	0.1	12.9	95.9	223.5

Table 2: Age Adjusted Incidence Rates for Females for different age groups 2007-2011.

	All Sites					Colon excluding Rectum				
	All ages	00-19	20-54	55-64	65+	All ages	00-19	20-54	55-64	65+
All races	420.1	16.8	249.7	880.3	1,691.70	28.2	0	9.4	44.7	154.7
White	433.3	17.9	258.5	898.1	1,745.80	27.6	0	8.7	40.6	156.1
Black	411.3	13.4	225.9	922.1	1,694.70	35.1	0	13.5	69.7	176.4

2.4.2. Gender:

With regard to the difference in the occurrence of colon cancer between the two sexes, as can be seen in the tables above, colon cancer has predominance in men over and this has been reported in several other studies[30]**.** The group most impacted by colon cancer in the United States are men of African descent. However, incidence rates in women are catching up to those in men. The Center of Disease Control reports the incidence rate by State for colon cancer in the U.S. showing the different percentages from State to State. The rates are per 100,000 based on the 2000 U.S. population[31].

Table 3: Rates are per 100,000 and are age-adjusted to the 2000 U.S. standard population (Year 2011)[32].

31.5 to 37.3	Arizona, Colorado, Florida, Idaho, Maryland, Massachusetts, New Hampshire, New Mexico, Oregon, Utah, Vermont, Virginia, and Wisconsin
37.4 to 40.8	California, Connecticut, Georgia, Maine, Michigan, Nevada, North Carolina, Ohio, Rhode Island, South Carolina, Texas, and Washington
40.9 to 43.6	Alabama, Delaware, Indiana, Kansas, Missouri, Montana, New Jersey, New York, Oklahoma, Pennsylvania, Tennessee, and Wyoming
43.7 to 51.3	Alaska, District of Columbia, Hawaii, Illinois, Iowa, Kentucky, Louisiana, Mississippi, Nebraska, North Dakota, South Dakota, and West Virginia
Data not available	Arkansas and Minnesota

2.4.3. Environmental Influences:

Several pieces of evidence suggest that the environment plays a role in the development of colon cancers. The frequency of colon cancer varies remarkably worldwide, with the highest rates in North America, Australia, and Europe, and much lower rates in regions of Asia, South America, and sub-Saharan Africa[33]. The risk of cancer rises rapidly in populations migrating from low-risk to high-risk areas, again suggesting that environmental factors, “especially: dietary differences, lifestyle, tobacco exposure, physical activity, and occupational exposures” have important implications in its development[34]**.**

2.5. Physical Activity:

It has been reported that approximately 50% reduction in incidence of colon cancer was found among individuals with the highest level of physical activity[26]**.**

2.6. Diet:

Many epidemiologic studies have shown an association between the intake of a diet high in fruits and vegetables and protection from colon cancer. Whether this protective effect is due to the fiber, antioxidant vitamins, folic acid, minerals such as selenium, other micronutrients in vegetables or is due to some other constituent(s) is not known[35]. Some reports suggest that a diet low in red meat, animal fat and/or cholesterol may also be protective[36]**.**

Ning et al performed a systematic review of the association between body mass index (BMI) and the risk of colon cancer and from their meta analytical study they determined that patients with BMI of greater than 30 kg per m^2 had a 41 % increased risk compared to those with BMI less than 23 kg per m^2[37].

2.7. Tobacco:

Although tobacco has not been clearly implicated as a cause of colon malignancies, a higher risk of colon cancers has been consistently observed among smokers in numerous studies[38]. Tobacco smoke is an initiator of colon carcinogenesis, yet the induction period is very long, possibly up to four decades[39].

2.8. Alcohol Consumption:

Alcohol significantly increases the risk of developing rectal cancer. It increases the risk of adenoma and carcinoma due to abnormal DNA methylation[40]. Excessive alcohol use, including underage drinking and binge drinking can lead to increased risk of health problems such as injuries, liver diseases, and colon cancer[41].

2.9. Drugs:

2.9.1. Hormone Replacement Therapy (HRT):

HRT has been shown to be associated with a decreased risk of colon cancer in the majority of studies[42].

2.9.2. Non-Steroidal Anti-inflammatory Drugs (NSAIDs):

Patients with rheumatoid arthritis, who generally have higher use of NSAIDs, have lower incidence and mortality rates of gastrointestinal malignancies. Also a statistically significant reduction of colon cancer is noted in individuals after 20 years of consistent (two or more tablets per week) aspirin use[43].

2.10. Genetic Influences:

Although it convenient to categorize colon cancers as hereditary (familial) and nonhereditary (sporadic), it is more appropriate to consider all colon cancers as having genetic components that are inherited or acquired to varying degrees.

Individuals with familial adenomatous polyposis, hereditary non-polyposis colon cancer, and other familial syndromes are born with genetic alterations that make them susceptible to the development of colonic neoplasia. Environmental factors then contribute additional "hits," leading to malignant transformation[44]**.**

Spence et al did a systematic review of studies pertaining to determining a causal relationship between physical activity (PA) and colon cancer (CC). They concluded that there being significant heterogeneity in the results and the there was no clear evidence to claim a convincing relationship exists between physical activity and the risk of CC[45].

Genetic changes that may lead to the development of colon cancer can be organized into three major classes: alterations in proto-oncogenes, loss of tumor suppressor gene activity, and abnormalities in genes involved in DNA repair. While much of what is known about the molecular genetics of colon cancer has come from the study of familial syndromes, similar changes are associated with the development of sporadic cancers[46]**.**

2.11. Length of Hospital Stay:

The length of stay for inpatient hospitalization or number of days is completely variable and depends on the specific stage of cancer. Some patients are hospitalized and are released quickly, for example, the patient who administration of certain chemotherapy regimens that require one night because the long infusions and close monitoring[47]. Other patients who are suffering from leukemia require up to a week or more of continuous chemotherapy[34].

Huebner et al conducted a retrospective study of rectal cancer patients going in for elective colon resection between 2005 and 2011 in a single hospital. They found that an early diagnosis of post-operative complications was related to a shorter length of stay for the patients and a faster recovery even in the presence of comorbidities[48].

2.12. Cost

According to a National Institutes of Health analysis the U.S. population, medical costs for cancer in the year 2010 reach at least $158 billion. That covered the cost-effectiveness of various treatment and procedures for advanced colon cancer chemotherapy, radiation, clinical trial and palliative care. According to Dr Croyle of the National Cancer Institute the advancing costs of cancer care makes it imminent to further the task of cancer prevention and treatment and to ensure that the most effective approaches are being employed. This is corroborated by the reports that costs for late stages of CC are rising exponentially with little or no improvement in survival[49].

Perng et al in a retrospective study of colon patients in Taiwan found that during the period of 2004 to 2009 the average hospitalization cost and the average length of stay decreased from by nearly 1100 dollars and by 5 days respectively[50].

2.13. Mortality:

With improvement in the use of appropriate medical equipment, accurate diagnoses and prompt treatment the mortality risks are still high for CC. It is estimated that the mortality numbers are increasing in the United States and the estimated mortality for the year 2012 were 52,857 deaths from colon cancer alone[51].

2.14. Race:

CC incidence has been reported to differ with ethnicity. It has been reported that Black males have the highest incidence rate for CC followed by white males and black females. Furthermore black males and black females have the lowest survival rates followed by white males. Also those blacks with family history of CC are at a higher risk than whites especially if they also happen to be smokers and obese[52].

Zullig et al did a comprehensive analysis of approximately 40,000 incident cancer cases as reported in the Veterans Affairs Central Cancer Registry (VACCR). They found that roughly 78.5% of newly diagnosed cases were of White population, 19.0% Black, and 2.5% were another race. They also found that the median age at diagnosis was 66 years and of the total around 8.8% had colon or rectal cancer. This they believe resembles those observed among U.S. men[53].

2.15. Research Gap in Literature:

Although as detailed above much has been researched into the etiology and associations of factors with incidence and progression of colon cancer there is no study yet looking at a comprehensive view of the cofactors and comorbidities that might influence the incidence and the outcomes of colon cancer. Also it is not yet known how the nature of care as determined by hospitals (and their types), insurance availability and types of insurances will affect the outcomes as well. It is therefore proposed in this research project to evaluate the effects of demographic, hospital characteristics, comorbidities (smoking, alcohol usage, diabetes and the like) on three outcomes namely, the length of stay, mortality and total charges. The next chapter will provide details on the data source employed for this research undertaking as also the various analytical procedures to be used.

Chapter III

Research Methodology

3.1. Nationwide Inpatient Sample Data:

The sample data consist of inpatient hospital stay file from the HCUP Nationwide Inpatient Sample (NIS). The NIS is nationwide database of community hospital inpatient stays. Research and policymakers use NIS data to identify, track and analyze trends in health care utilization, access, charges, quality and outcome. The NIS is nationally representative of all community hospitals (i.e. short-term, non-federal, non-rehabilitation hospitals). The NIS is a sample of hospitals and includes all patients from each hospital, regardless of payer including uninsured. It is drawn from a sampling frame that contains hospitals comprising about 95 percent of all discharges in the United States. The sampling frame for the 2011 NIS is a sample of hospitals that comprises approximately 95 percent of all hospital discharges in the United States. The NIS includes more than 100 clinical and nonclinical data elements for each hospital stay.

Table. 4 Data Variables Used for Analysis:

Study Variables	Original Variable Name in the NIS Data Set	Variable Description
Age	AGE	Age in years, Numerical Variable
Mortality	DIED	Patient did not die during hospitalization (DIED=0); Patient died during hospitalization (DIED=1), Categorical (binary) Variable
GENDER	FEMALE	Gender of patient FEMALE = 1 is Male; FEMALE=0 is female, Categorical (binary) Variable
TOTAL CHARGE	TOTCHG	Total charges , Numerical Variable
RACE	RACE	1 = White, 2 = Black, 3 = Hispanic, 4 = Asian/Pacific, 5 = Native Am., 6 = Other, Categorical Variable
INSURANCE TYPE	PAY1	1=Medicare, 2=Medicaid, 3=Private insurance,4=Self-pay,5=No charge,6=Other, Categorical Variable
NUMBER OF PROCEDURES	NPR	The number of procedures performed while patient was hospitalized, Numerical Variable
SOCIO_ECONOMIC STATUS (SES)	ZIPINC	Median household income for patient's ZIP Code, 1=$1-24,999, 2=$25,000-34,999, 3=$35,000-44,999, 4=45,000 or more, Categorical Variable
COMORBIDITIES	CM_DRUG, CM_ALCOHOL, CM_OBESE, CM_ULCER, CM_DM, CM_HTN	Comorbidities (drug abuse, alcohol abuse, obesity, ulcer, diabetes, hypertension), Categorical (binary) Variables
LENGTH OF STAY	LOS	The number of days patient was hospitalized, Numerical Variable
NUMBER OF DIAGNOSES	NDX	The number of diagnoses on the patient record, Numerical Variable
REGION	REGION	Four regions are included Northeast = 1, Midwest =2, South = 3, west =4 , Categorical Variable

Table 5. Study Hypotheses and Corresponding Statistical Tests:

Research Question	Hypothesis	Independent Variables	Outcome Variables	Inferential Analyses	Descriptive Analyses & Predictive Models
Do Type of Comorbidities & Number of Procedures significantly affect LOS, TOTCHG & DIED?	Hypothesis1	CM_ALCOHOL, CM_DRUG, CM_DM, CM_HTN, CM_OBESE	LOS, TOTCHG, DIED	ANOVA (LOS vs CM_ALCOHOL, vs CM_DRUG, vs CM_DM, vs CM_HTN, vs CM_OBESE) ANOVA (TOTCHG vs CM_ALCOHOL, vs CM_DRUG, vs CM_DM, vs CM_HTN, vs CM_OBESE) CHI-SQ (DIED vs vs CM_ALCOHOL, vs CM_DRUG, vs CM_DM, vs CM_HTN, vs CM_OBESE)	Means & SDs for LOS and TOTCHG
Do Race, Age & Socioeconomic Status significantly affect LOS, TOTCHG & DIED?	Hypothesis2	RACE, AGEGRP, ZIPINC	LOS, TOTCHG, DIED	ANOVA (LOS vs RACE, LOS vs ZIPINC, LOC vs AGEGRP) ANOVA (TOTCHG vs RACE, TOTCHG vs ZIPINC, TOTCHG vs AGEGRP) CHI-SQ (DIED vs RACE, DIED vs AGEGRP, DIED vs ZIPINC)	Frequency Distributions for DIED
Do LOS, TOTCHG & DIED significantly differ with the Geographical Regions?	Hypothesis3	REGION	LOS, TOTCHG, DIED	ANOVA (LOS vs REGION) ANOVA (TOTCHG vs REGION) CHI_SQ (DIED vs REGION)	Multiple Linear Regression for LOS and TOTCHG with respect to the Independent Variables
Do LOS, TOTCHG & DIED significantly differ with the Type of Hospital?	Hypothesis4	HOSPTYPE	LOS, TOTCHG, DIED	ANOVA (LOS vs HOSPTYPE) ANOVA (TOTCHG vs HOSPTYPE) CHI_SQ (DIED vs HOSPTYPE)	Logistic Regression for DIED with respect to the Independent Variables

3.2. Research Design & Methods:

In this project we plan to utilize the datasets obtained from the Nationwide Inpatient Sample (NIS) database towards our analyses of Colon Cancer patients. The NIS is the largest all-payer inpatient care database in the United States containing data from 1998 to 2011. It contains data from approximately 8 million hospital stays each year accruing from all discharge data from 1,050 hospitals located in 44 States, approximating a 20-percent stratified sample of U.S. community hospitals. The sampling frame for the 2011 NIS is a sample of hospitals that comprises approximately 95 percent of all hospital discharges in the United States. The NIS includes more than 100 clinical and nonclinical data elements for each hospital stay. These include:

- Primary and secondary diagnoses
- Primary and secondary procedures
- Admission and discharge status
- Patient demographics (e.g., gender, age, race, median income for ZIP Code)
- Expected payment source
- Total charges
- Length of stay

- Hospital characteristics (e.g., ownership, size, teaching status).

Furthermore, the NIS is the only national hospital database containing charge information on all patients, regardless of payer, including persons covered by Medicare, Medicaid, private insurance, and the uninsured.

We plan to acquire all NIS data for 2008 to 2010 and the statistical analysis software SAS 9.2 will be employed to extract the datasets and perform the analyses. The figure below illustrates the conceptual model employed in this research project.

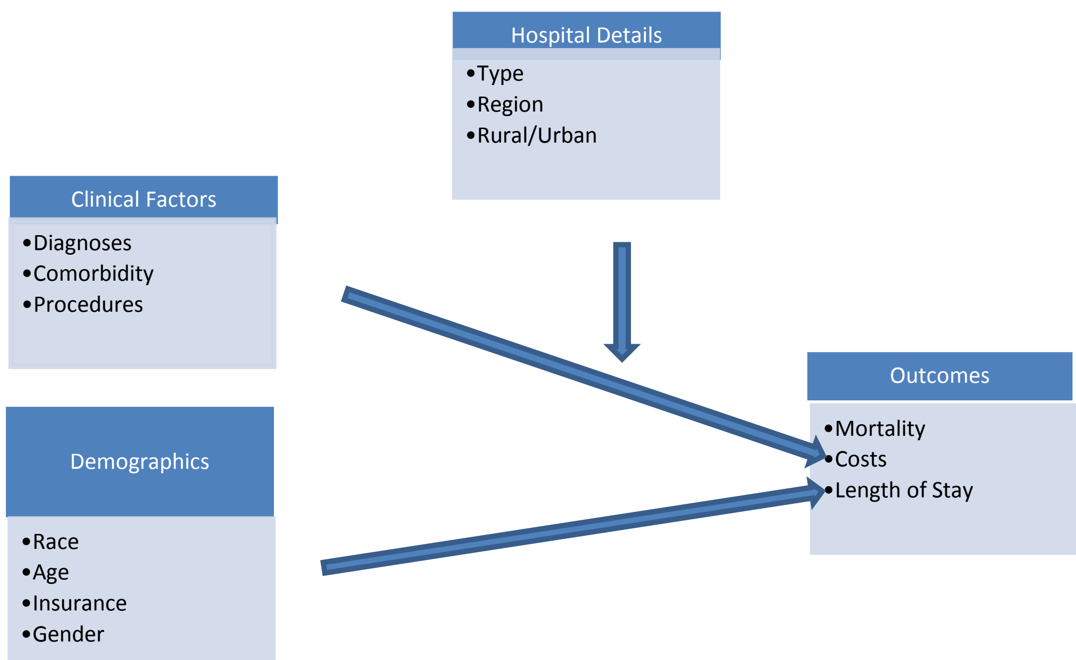


Figure 3: Conceptual model for identifying the factors determining Mortality, Length of Stay and Costs**.**

Essentially it shows the factors that are hypothesized to affect the research outcomes, the mortality, and length of stay as also the hospitalization costs for the colon cancer patient data acquired from the NIS database. These factors are categorized as being: clinical such as the type of diagnoses, the number and type of comorbidities and the number of procedures; demographics group delineates the race and age of the patient as also the type of insurance (Private, Medicare and the like) while the Hospital Details category consists of the type of the hospital (teaching/non-teaching), its location in the US, and whether it is in the rural, metro or urban areas. The outcomes of interest as identified in this proposal are the mortality, the length of stay and the costs involved. Using the datasets obtained from the NIS database appropriate descriptive and inferential statistics (such as frequency distributions, chi-square analysis and ANOVAs) will be effected. To relate the factors associated with the research outcome, the length of stay and the costs a multiple regression model will be setup and validated. Predictive models such as logistic regression will be employed to determine the risks and ratios for the various factors influencing mortality such as race, age groups, number and types of procedures and comorbidities.

3.2.1. Statistical Methodology:

The following methods will be used to analyze the data as appropriate.

Parametric methods such as:

- Linear regression models
- Correlation analysis: Pearson correlation
- Paired and unpaired t-test
- One-way ANOVA and
- Mean, SD for descriptive analyses

will be used to analyze continuous data that are normally distributed.

Non-parametric methods such as:

- Wilcoxon Rank sum test and Mann-Whitney test
- Correlation analysis: Spearman correlation
- Kruskal-Wallis test and
- Median, interquartile range for descriptive analyses

will be used to analyze variables that all not normally distributed and are ranks or scores.

For binary and categorical variables the following methods will be use where appropriate.

- Logistic regression models.
- Chi-square test or Fisher exact test.
- Contingency coefficients (Cochran-Mantel-Haenszel tests).
- McNemar's test and.
- Proportion for descriptive analyses.

These will be used to analyze data with binary or categorical outcome.

3.2.2. Statistical Analyses:

The NIS data is completely de-identified secondary data ready for analyses. Data will be categorized as appropriate to investigate research questions. All computations will be performed with SAS® Release 9.3 running on a Windows operating system. All invalid data will be reported and a reason given for why the data is considered invalid (example –missing value). Where outlying data are observed, analyses will be performed with and without the outlying data. Sound statistical evidence that the data are outlying (i.e. outlying data is more than 4 standard deviations beyond the mean of comparable data) will be documented. Outlying data can be removed from an analysis if it can be shown to improve the power of the statistical tests or if not removing it would skew the result.

Descriptive and distribution analyses will be performed for all appropriate variables. Continuous variables will be assessed for normality. If the data is normally distributed, parametric methods will be used to analyze data otherwise non-parametric methods will be used. Non-parametric methods will be used to analyze score data. Categorical analyses with the appropriate methods will be used to compare categorical variables. Cochran-Mantel-Haenszel tests (for categorical variables) or linear models (for continuous variables) will be used to compare the baseline clinical characteristics. Relationships between outcome and clinical characteristics will be tested by using Pearson correlations. When data are not normally distributed, nonparametric tests such as Spearman correlation will be used. Also nonparametric tests (Wilcoxon rank sum test) will be used where appropriate. Categorical variables will be analyzed using the chi-square test or Fisher exact test where appropriate to compare groups. A two-sample student's t test will be used to compare difference in scores between clinical groups. Pearson correlation or the Spearman rank correlation coefficient will be used where appropriate to test independence between variables. Spearman rank correlation coefficients between categorical factors and the continuous research outcomes (length of stay, total charges and mortality) will be calculated. All means will be provided with the standard deviation (SD). For comparison of means, the Student t-test will be used, and where appropriate, a paired t-test will be performed.

The following SAS procedures will be used to perform the analyses: The Proc CORR, the PROC FREQ Procedure, PROC GLM Procedure, PROC LOGISTIC Procedure and PROC UNIVARIATE. These are standard procedures common across statistical softwares.

3.3. Modeling Techniques:

The following is a brief summary of two major statistical modeling techniques such as linear regression and Logistic regression as is used in this thesis.

3.4. Linear Regression:

Linear regression is used to model the relationship between two variables using the least error squares approach. The linear regression technique tries to fit a straight line for the data obtained from a single or from multiple predictor variables and an output or response variable. If it is a single input or predictor variable it is called Simple Linear Regression but if however there are several predictor variables it is called Multiple Linear Regression.

In this dissertation we will specifically employ the Multiple regression and to do so in SAS 9.2 the Stepwise Regression technique which incrementally adds variables or factors into the regression model based on maximizing R^2 will be used. Those factors which are a result of models with higher R^2 values will be used for the simulation. The different models will be compared for their inclusion of data elements as independent variables. In general statistical models use variable statistic regression methods, which utilize variable selection. Selection of the correct variables is critical to achieving “good” results. Variables are either defined as quantitative because they are continuous or as qualitative because they are discrete (e.g., categories).

A linear model usually assumes that the dependent variable is continuous. If least squares estimation is used, then if it is assumed that the error component is normally distributed, the model is fully parametric. If it is not assumed that the data are normally distributed, the model is semi-parametric. If the data are not normally distributed, there are often better approaches to fitting than least squares. In particular, if the data contain outliers then regression might be better.

If two or more independent variables are correlated, then those variables are multicollinear. Multicollinearity results in parameter estimates that are unbiased and consistent, but inefficient. If the regression error is not normally distributed but is assumed to come from an exponential family, generalized linear models should be used. For example, if the response variable can take only binary values, such as a Boolean or Yes/No variable, then logistic regression is preferred. The outcome of this type of regression is a function which describes how the probability of a given event or achieving a “yes” or “no” varies with the predictors. In this thesis, the outcome variable could be “lived” vs. “died” or “hypertension” vs. “non-hypertension” and so on.

Simple linear regression and multiple regression are related statistical methods for modeling the relationship between two or more random variables using a linear equation[2]. Simple linear regression refers to a regression on two variables while multiple regression refers to a regression on more than two variables. A simple linear regression equation is represented below.

Simple Linear Regression: $$y = \alpha + \beta_1 X_1 \quad \ldots (3\text{-}1)$$

A Multiple Linear Regression equation is represented below.

Multiple Linear Regression: $$y = \alpha + \beta_1 X_1 + \beta_2 X_2 + .. + \beta_n X_n + \varepsilon \quad \ldots (3\text{-}2)$$

In both the simple linear and multiple regression models α is the constant. β represents the coefficient for the independent variable(s) 1 through n. Where n is the number for the subscript for the last independent variable. In the case of the simple linear model there is only one independent variable. ε is the error value in the multiple regression model.

A related method to determine if a model is a “good” fit is called the Akaike information criterion (AIC). The AIC was developed by Hirotsugu Akaike in 1971 and is a measure of the goodness of fit for an estimated statistical model[33]. The AIC trades the complexity of an estimated model against how well the model fits the data.

Simple variable selection algorithms are ad hoc or method based. A common method is the greedy hill climbing approach. This approach evaluates a variable subset and then modifies that subset to determine if an improved subset exists. Thus, this greedy algorithm adds or deletes the respective best or worst variable. The stepwise regression method is a popular choice, which demonstrates a greedy algorithm.

Stepwise regression is an automatic procedure for statistical model selection where there are a large number variables added or dropped[54]. New variables are added at each stage in the process and variables are checked to see if some can be deleted without increasing the Residual Sum of Squares (RSS). The stepwise regression process stops when a selection is maximized or when it can no longer be improved. Backward regression and forward regression are variations of the technique.

A confidence of 95% will be used unless there is a need to increase or decrease the percentage. Thus, a p value < 0.05 for a 95% confidence would be significant. In many cases all the data records will be used in the analysis unless stated.

3.5. Logistic Regression:

Logistic regression is used to predict the outcome of a categorical dependent variable based on one or more independent variables. Logistic regression can be used for binomial or multinomial dependent variables, i.e. dead or alive, incidence or no incidence, true or false and so on. Typically, the outcome is denoted as “0” and “1”. This technique was developed by Boyd et al in 1987[55].

Logistic regression predicts therefore probabilities of the outcome of a categorical dependent variable based on one or more predictor variables being present or not. The probability of the outcome can be modeled as a function of predictor variables, using a logistic function which is a sigmoid function, given its name in 1844 by Pierre Francois Verhulst who used the logistic function in studying population growth. A simple logistic function is defined as P (t) = 1/ $(1+e^{-t})$, where the probability function P is a function of time (t).

The Logistic Regression equation is typically denoted as $F(x) = \beta_0 + \beta_{1*X} + e$ on the horizontal axis where F(x) is the Logit Function. The Logit function is useful in this context because it can take as an input any value from negative infinity to positive infinity, whereas the output is constrained to be between 0 and 1. β_0 is the intercept in the regression equation, β_1 is the regression coefficient multiplied by the predictor variable ‘x’ and *e* denotes the residual function.

SAS 9.2 can perform the logistic regression modelling given the output and the input variables to be included and the results include Model Fit Statistics describes and tests such as AIC, Schwarz Criterion and Wilks Statistic. AIC or the Akaike information criterion (AIC) provides a relative measure of the information lost when a given model is used to describe reality. It describes the tradeoff between bias and variance in a model, i.e. accuracy and complexity of the model. The formula for AIC is given by 2k – 2ln (M), where k is the number of variables and M is the maximum likelihood value for the estimated model. Besides AIC SAS also provides other results for the Logistic Regression modelling such as the Schwarz criterion or Bayesian information criterion (BIC) developed by Gideon Schwarz in 1978 to predict the maximum likelihood function by adding a penalty term of the number of variables in the model when the results are over-fitting[29].

Wilks' statistic was constructed by Samuel Wilks in 1938 to compute the likelihood 'M' for the outcome and it compares -2Log (M) to the chi-square value corresponding to a desired statistical significance as an approximate statistical test[13]. The next chapter presents the results arising from conducting these statistical tests and the discussions on the results is also presented.

Chapter IV

RESULTS & DISCUSSIONS

4.1. Colon Cancer Incidences and Discharges:

Incidence Rates across Ethnic Groups – Trend between 2008 and 2010

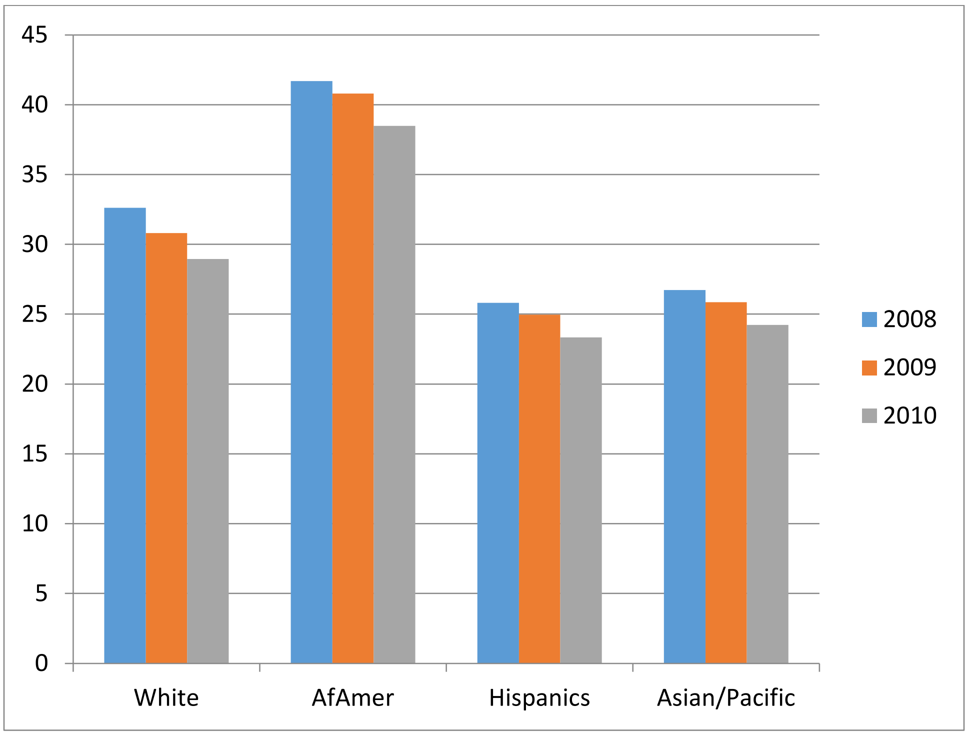
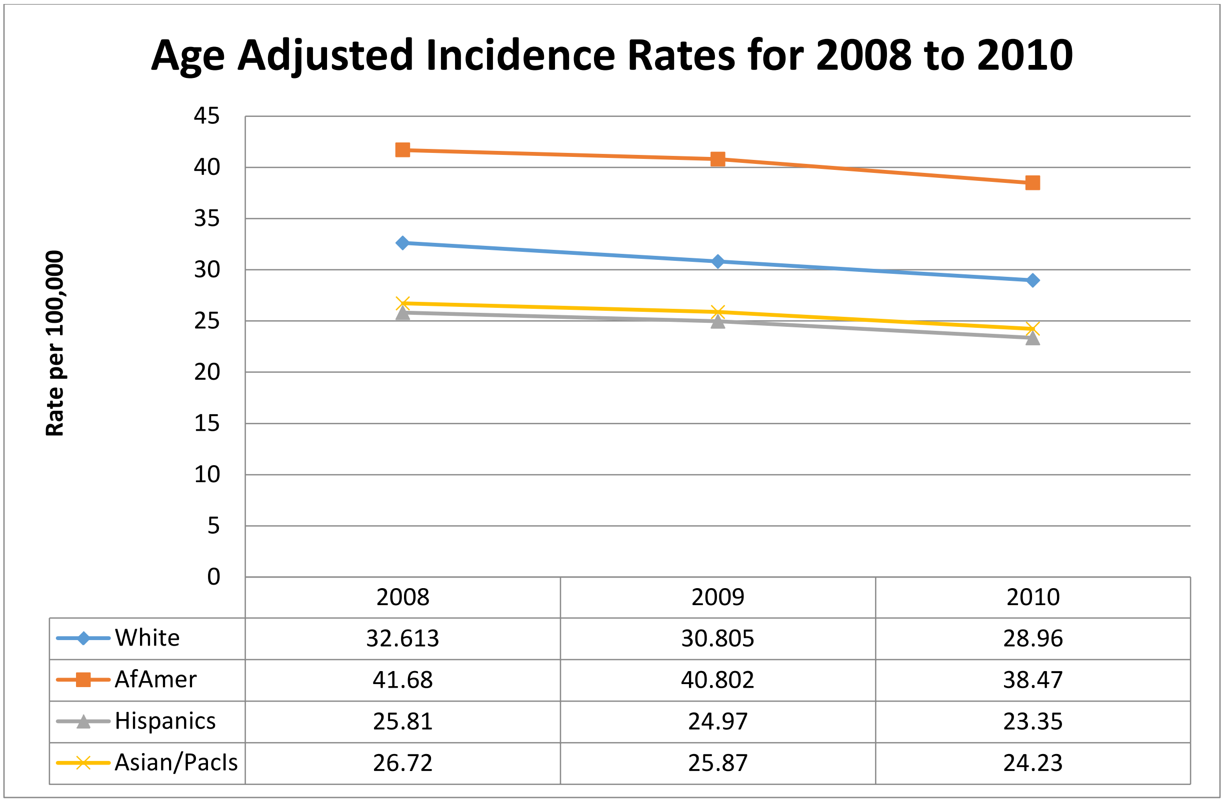
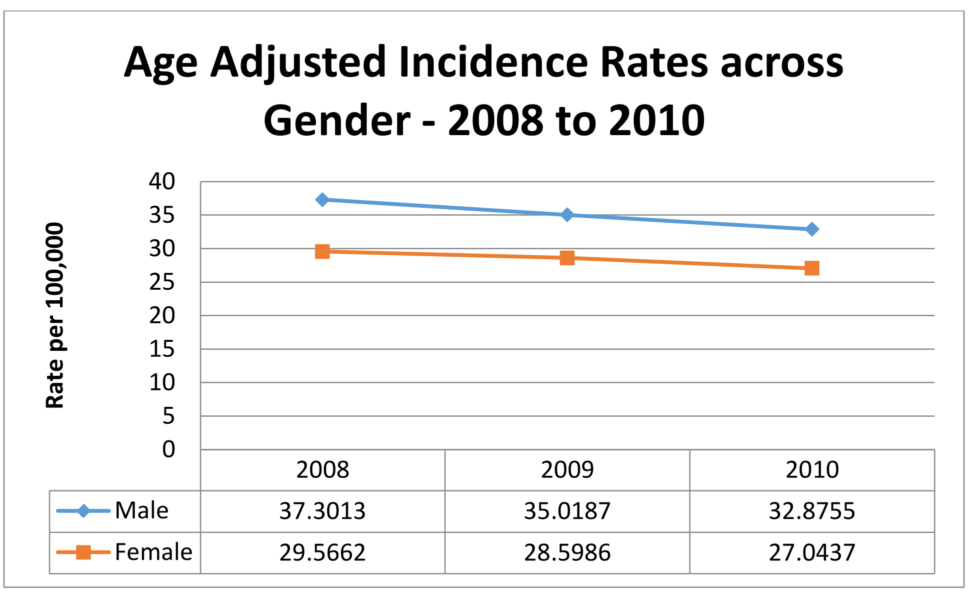


Figure 4: Age Adjusted Incidence Rates (rate per 100,000) of Colon Cancer Cases from 2008 to 2010.



	2008	2009	2010
White	32.613	30.805	28.96
AfAmer	41.68	40.802	38.47
Hispanics	25.81	24.97	23.35
Asian/PacIs	26.72	25.87	24.23

Figure 5: Age Adjusted Incidence Rates (rate per 100,000) of Colon Cancer Cases from 2008 to 2010.



	2008	2009	2010
Male	37.3013	35.0187	32.8755
Female	29.5662	28.5986	27.0437

Figure 6: Age Adjusted Incidence Rates (rate per 100,000) of Colon Cancer Cases from 2008 to 2010 across Gender.

Between 2008 and 2010 the age and population adjusted incidences (both genders) both decreased significantly which is a promising trend speaking well of the state of health care in the United States as also possibly due to the effectiveness of nutritional counselling, patient education, screening for men aged 50 and above.

Colon Cancer Discharges based on Hospital Types – 2010

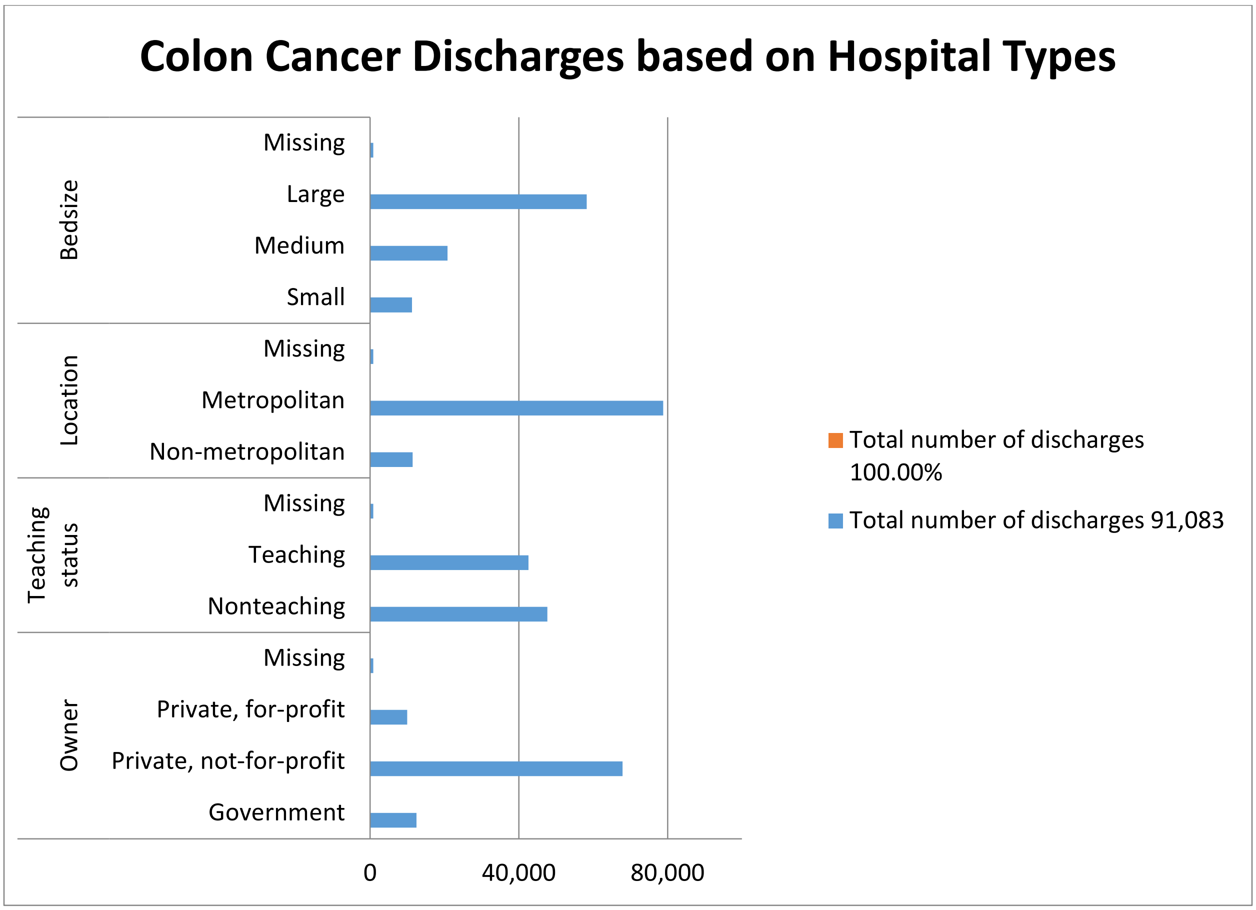


Figure 7: Illustration of Colon Cancer Discharges based on Hospital Types as determined for 2010.

The number of discharges across the various hospital types and their locations across the United States as shown above reveals those large hospitals in metropolitan regions and those that are private not-for-profit have more discharges compared to the other types. This could probably be attributed to the large population that the hospitals serve being in metropolitan regions.

Table 6: Data Table of Colon Cancer Discharges based on Hospital Types as determined for 2010.

		Total number of discharges	
All discharges		91,083	100.00%
Owner	Government	12,431	13.65%
	Private, not-for-profit	67,916	74.57%
	Private, for-profit	9,938	10.91%
	Missing	799	0.88%
Teaching status	Nonteaching	47,698	52.37%
	Teaching	42,586	46.76%
	Missing	799	0.88%
Location	Non-metropolitan	11,457	12.58%
	Metropolitan	78,827	86.54%
	Missing	799	0.88%
Bedsize	Small	11,217	12.31%
	Medium	20,830	22.87%
	Large	58,238	63.94%
	Missing	799	0.88%

Table 7: Table of Example Bedsize Categories based on Hospital Types for Northeast based Hospitals

Region	**Location and Teaching Status**	**Hospital Bedsize**		
		Small	**Medium**	**Large**
Northeast	Rural	Jan-49	50-99	100+
	Urban, nonteaching	1-124	125-199	200+
	Urban, teaching	1-249	250-424	425+

Regional Differences in Colon Cancer Patient Discharges – 2008 to 2010

Table 8: Data Table of Colon Cancer Discharges based on US Regions 2008 to 2010

		2008	2009	2010
All discharges		103,143 (100.00%)	93,936 (100.00%)	91,083 (100.00%)
Region	Northeast	19,085	18,579	18,483
	Midwest	25,178	22,595	22,448
	South	39,290	35,413	32,780
	West	19,591	17,349	17,371

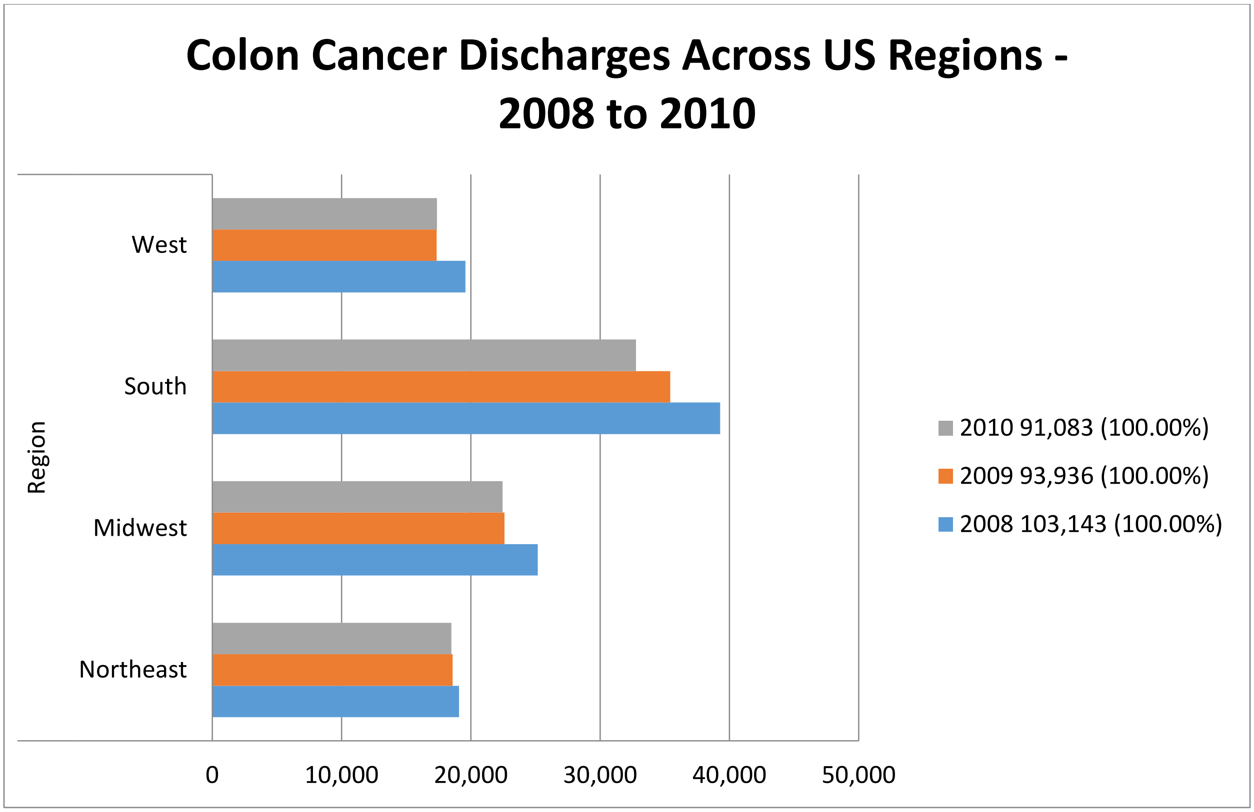
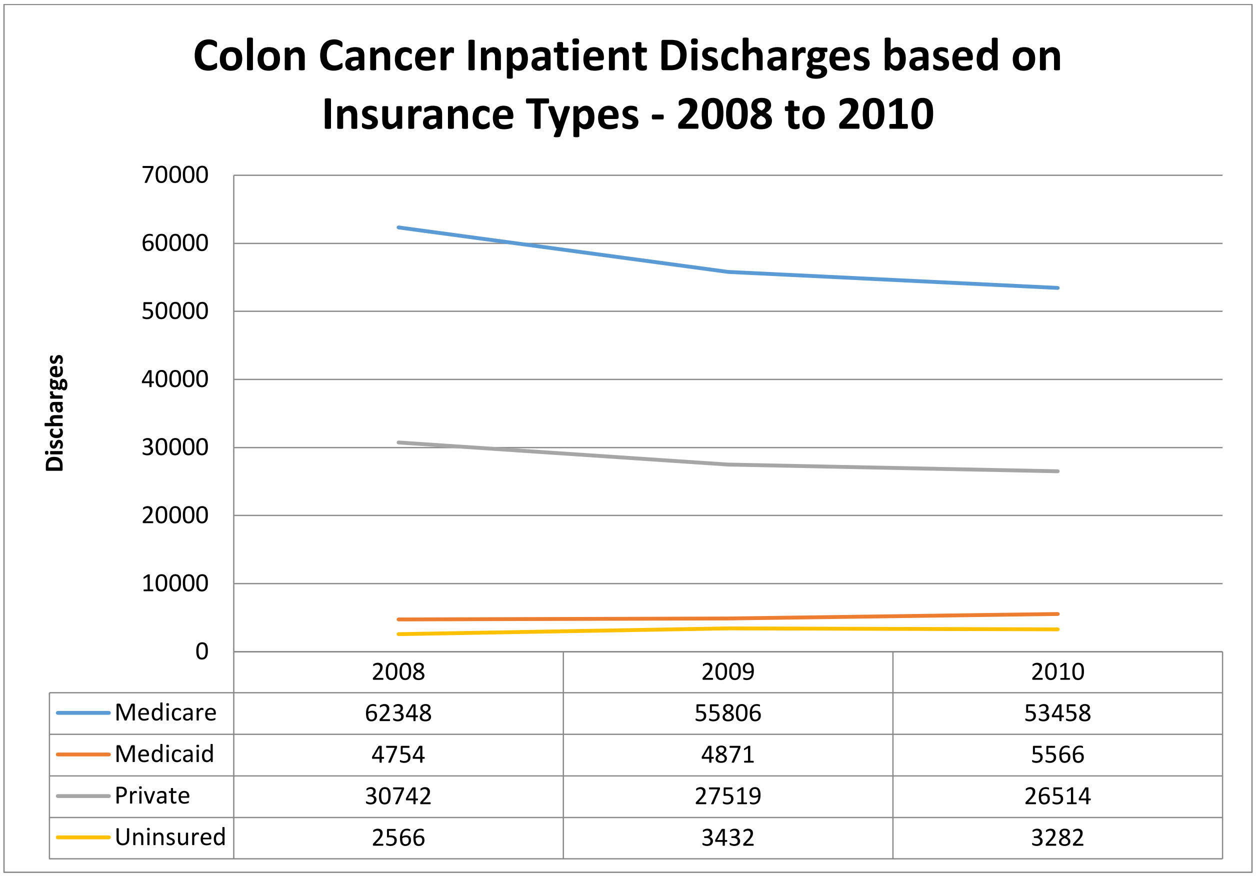


Figure 8: Illustration of Colon Cancer Discharges based on Hospital Types as determined for 2010.

The number of discharges across the various regions of the United States as shown above reveals the South has more (nearly 2 times) discharges compared to the other regions in all the 3 years. This could probably be attributed to the dietary aspects of the South as also there is a greater African American population in the South as compared to the other regions.



	2008	2009	2010
Medicare	62348	55806	53458
Medicaid	4754	4871	5566
Private	30742	27519	26514
Uninsured	2566	3432	3282

Figure 9: Illustration of Colon Cancer Discharges based on Insurance Types 2008 to 2010

The results above pertaining to the discharges across the various age groups reveal that the patients who are uninsured and those on Medicaid (low income) are more in number over the years as compared to those on Medicare and Private Insurance which have decreasing trends. The reasons are not very apparent but it could be due to the severe economic downturn during this period with the chances that there are more patients who were uninsured and/or relying on Medicaid otherwise.

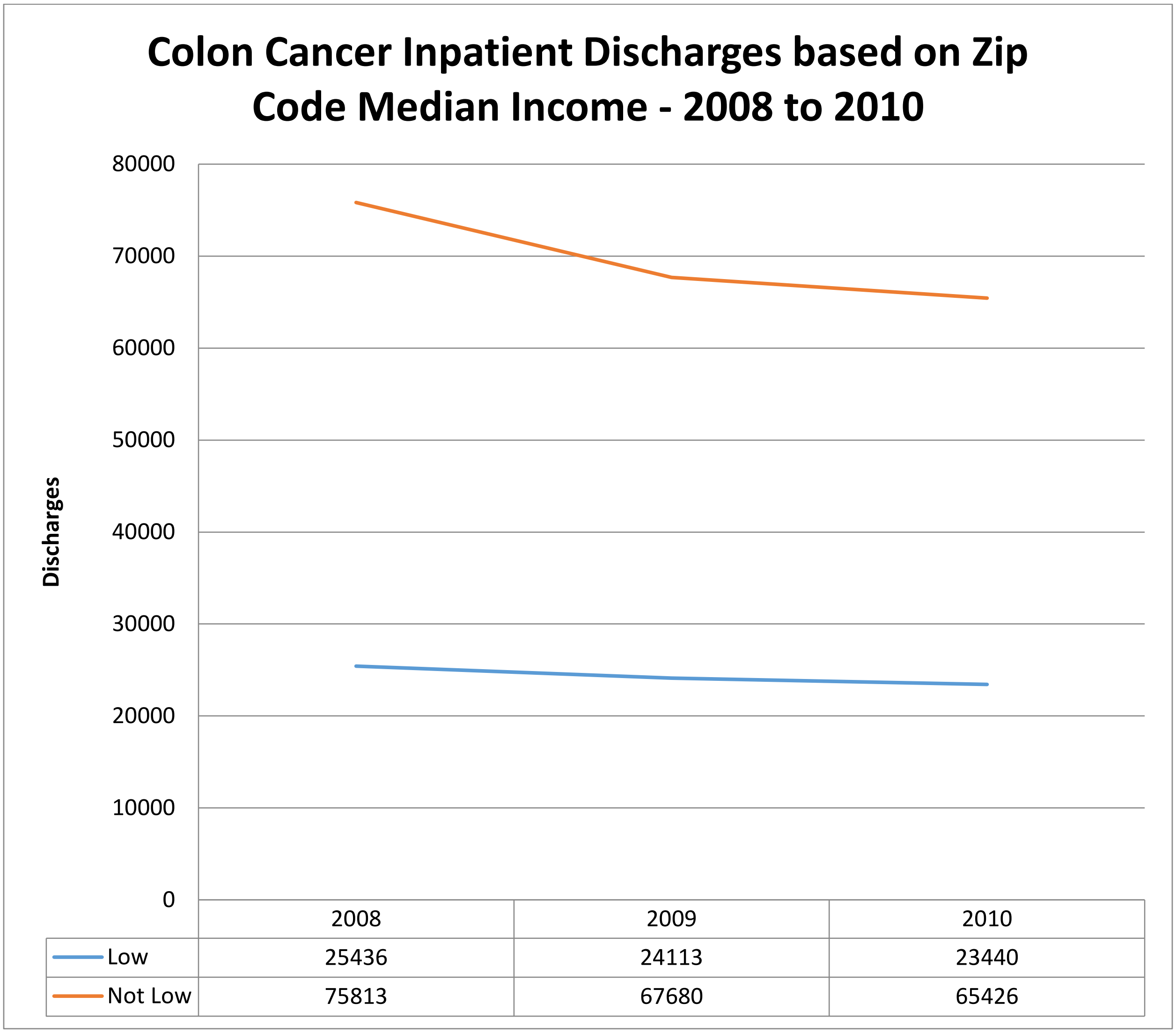


Figure 10: Illustration of Colon Cancer Discharges based on Patient Socioeconomic Status (as assumed by zip code median income) 2008 to 2010

ZIP Code Median income is the median household income of the patient's ZIP code of residence and allows one to compare patients that come from very low income ZIP Codes to patients who come from more affluent ZIP Codes. This is a proxy measure of a patient's socioeconomic status. The definition of very low income is the lowest quartile of income while the others are considered ‘Not Low’.

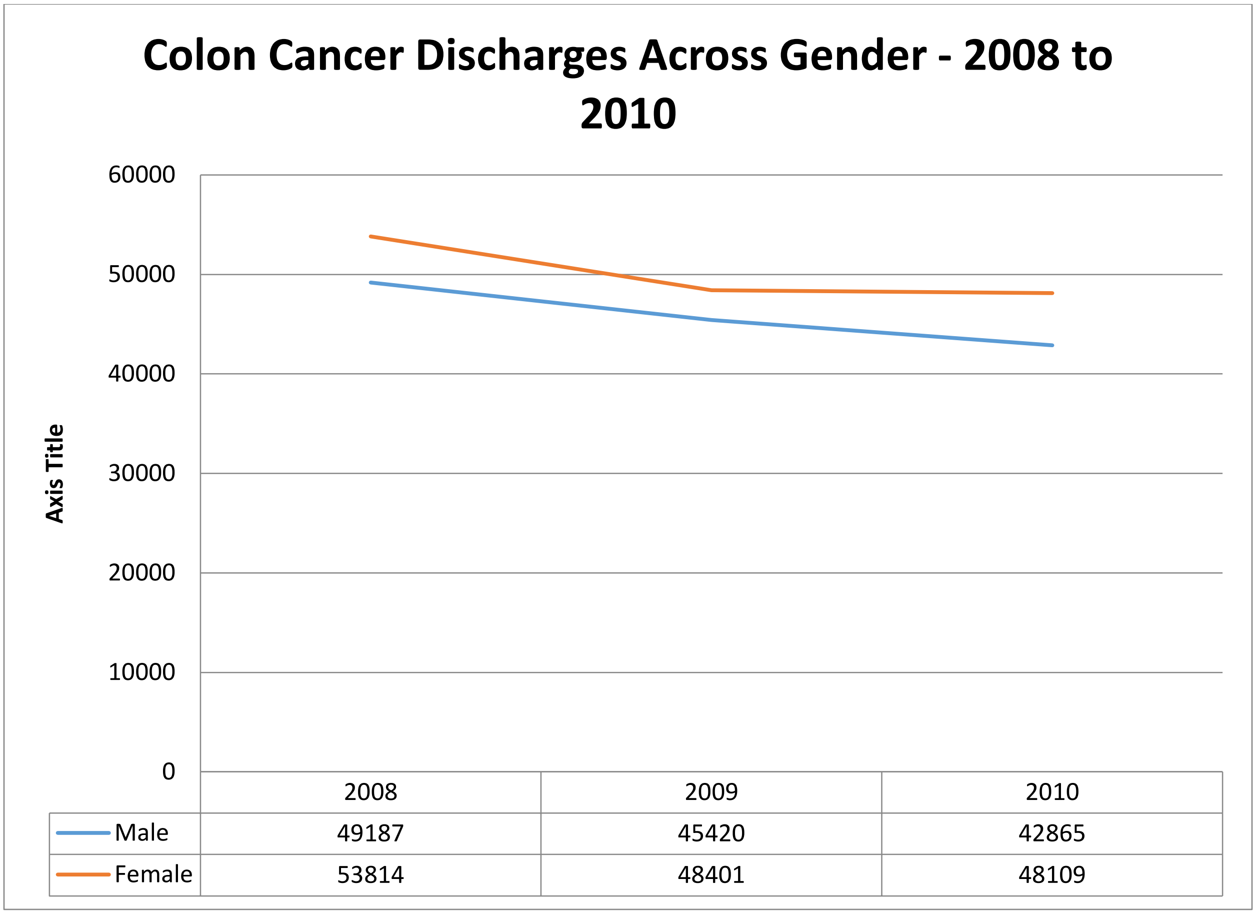
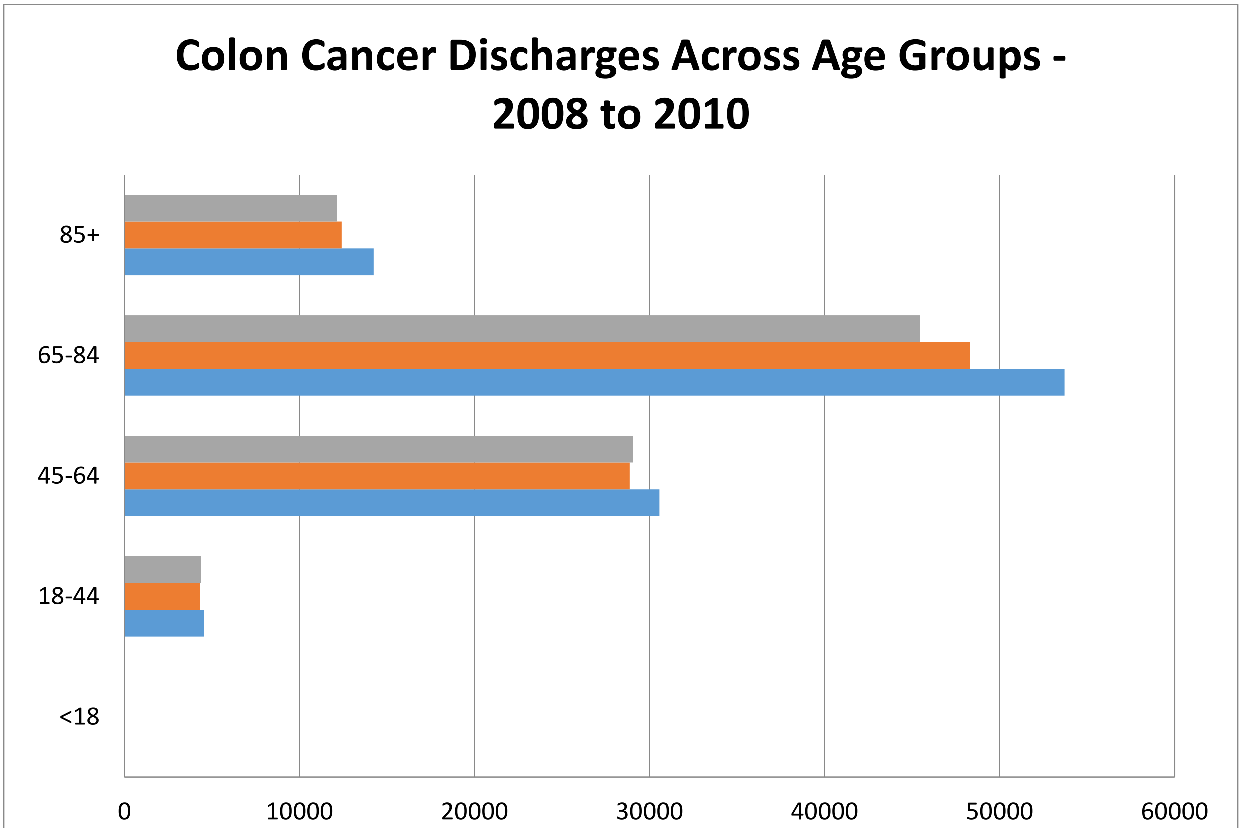


Figure 11: Illustration of Colon Cancer Discharges based on Gender - 2008 to 2010

The results above show that the discharges across the gender are decreasing for both Males and Females across the years and which is a promising trend speaking well of the state of health care in the United States as also possibly due to the effectiveness of nutritional counselling, patient education and preventative screening.



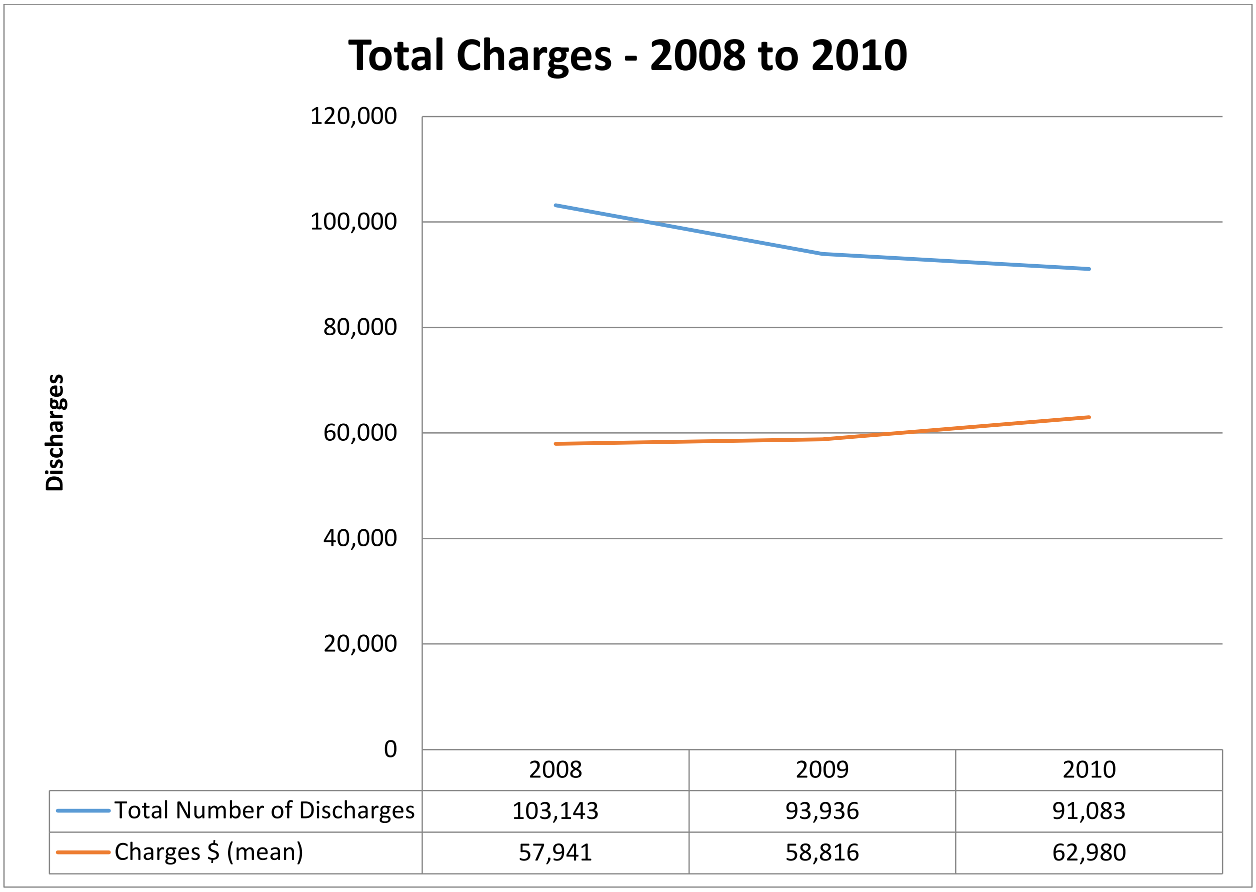
	<18	18-44	45-64	65-84	85+
2010		4390	29042	45447	12132
2009		4305	28858	48292	12411
2008		4548	30559	53715	14231

Figure 12: Illustration of Colon Cancer Discharges based on Age Groups - 2008 to 2010

The results above pertaining to the discharges across the various age groups reveal that they are more or less the same across the years for the age groups 18 to 44 while there is a decreasing trend for all ages 45 and above and this could possibly due to effectiveness of preventative screening for adults over 50 years old.

4.2. Colon Cancer In-Patent Hospital Total Charges (TOTCHG) Statistics:

Total Charges – Trend between 2008 and 2010



	2008	2009	2010
Total Number of Discharges	103,143	93,936	91,083
Charges $ (mean)	57,941	58,816	62,980

Figure 13: Trends in Total Charges over the years 2008 to 2010.

The total charges (mean and median) were determined using the Colon Cancer Patient Data for the years 2008 to 2010. *Total Charges* is the amount the hospital charged for the entire hospital stay. It does not include the fees for the physicians. Charges are not necessarily how much was reimbursed. It was found that while the total number of colon cancer patient discharges decreased between 2008 to 2010 the Total Charges however significantly rose up between

2008 and 2010. The mean charges increased by 9 % during 2008 to 2010. This implies that the per patient costs have significantly increased over the years 2008 to 2010. Laudabum et al.[51] comment that total payments in colon screening patients are mostly due to payments for anesthesia and pathology services and in this case it is possible (though not testable) that increases in such payments over the period of 2008 to 2010 contribute significantly to the increased trend observed in our results. Stroupe et al[21] in their study of multivariable regression to examine health care costs and utilization between patients whose chemotherapy was administered before or after March 1, 2007 did find significant increases in other cancer-related costs and concluded that those increases were mainly in outpatient services and pharmacy. They suggested this may likely due to the adoption of new high-cost diagnostic approaches and therapeutic modalities.

4.3. Discharge Status From Hospital – Trend between 2008 and 2010

Discharge status refers to the disposition of the patient at discharge from the hospital, e.g., routine (home), to another short term hospital, to a nursing home, to home health care, or against medical advice. The discharge status (how the patient was discharged) was determined using the Colon Cancer Patient Data for the years 2008 to 2010. It was found that the number of in-hospital mortality or deaths significantly reduced between 2008 and 2010. Alongside Routine Discharge and Home Health Care increased while discharges to another hospital also decreased (with a smaller decrease in discharge to another institution such as rehab facility and nursing home). As can be observed here the in-hospital deaths have significantly decreased over the period 2008 to 2010 while home health care has increased simultaneously due perhaps to the better colon health screening and procedures performed at the hospital and/or due to the availability of better surgical and/or treatment regimens in the recent years.

Table 9: The Discharge Status of Colon Cancer Patient from 2008 to 2010.

	2008		**2009**		**2010**	
Total number of discharges	103,143		93,936		91,083	
In-hospital deaths	4,488	4.35%	3,954	4.21%	3,448	3.79%
Routine discharge	63,494	61.56%	57,851	61.59%	55,326	60.74%
Another short-term hospital	1,368	1.33%	1,070	1.14%	1,140	1.25%
Another institution (nursing home, rehab)	16,695	16.19%	15,274	16.26%	14,753	16.20%
Home health care	16,654	16.15%	15,422	16.42%	16,073	17.65%
Against medical advice	238	0.23%	208	0.22%	178	0.19%

4.4. Stepwise Linear Regression for Total Charges – 2008

A Multiple Linear Regression Model for predicting Total Charges based on several of the demographic and hospitalization related variables was performed using a Stepwise regression technique and the following were revealed to be the significant predictors.

Table 10: List of Variables Identified as Significant in Prediction of Total Charges – 2008

Summary of Stepwise Selection							
Step	Variable Entered	Label	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	LOS	Length of stay (cleaned)	0.5417	0.5417	425.171	2638.62	<.0001
2	NPR	Number of procedures on this record	0.0724	0.6142	7.4295	418.91	<.0001
3	AGE	Age in years at admission	0.0007	0.6149	5.2449	4.18	0.0410
4	NDX	Number of diagnoses on this record	0.0006	0.6155	3.6496	3.60	0.0580

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-2893.50859	4021.75848	727070580	0.52	0.4719
AGE	-129.74011	55.86476	7575856345	5.39	0.0203
NPR	7729.42370	400.62042	5.22862E11	372.24	<.0001
NDX	361.16167	190.41619	5053061346	3.60	0.0580
LOS	4599.48818	118.92262	2.101111E12	1495.86	<.0001

The multiple linear regression model as elicited by the stepwise regression is accordingly,

TOTCHG = -2893.51 – 129.74*AGE + 7729.42* NPR + 361.16*NDX + 4599.5*LOS …. (4.1)

In this model we find that Total Charges or Costs are positively related to NPR the number of procedures, NDX the number of diagnoses and LOS the length of Stay. Interestingly Total Charges are inversely related to AGE which will be discussed below shortly. A recent study by Fox et al[22] on hospital costs and outcomes of mortality and readmission rates in a large California based hospital found no significant relationship between them and concluded that more research is warranted. In that regard our study of the linear regression involving larger data records also seem to indicate that mortality is not correlated with the total costs but the number of procedures (besides length of stay) is related to total charges. It can be reasoned that in-hospital mortality in the case of cancer patients may have had them come in as emergency admissions with high risk of organ failure and such and treatment and surgeries to prolong their lives may have been more contributing accordingly to greater costs. To confirm this data was procured as to the source of admission for Colon Cancer Patients and this piece of information as to the nature of admission source was only available upto

2006 after which it has been removed. However looking at the trend between 2001 and 2006 (as shown below) it can be seen that a significant number of patients (27% of discharges) admitted to the hospitals with Colon Cancer as the principal diagnosis did indeed come in from the Emergency Department. Similar results were also seen for those admitted with Cancer of the Lung and Bronchus as the principal diagnosis whereas those with Breast Cancer were non-emergency department sourced. In the case of the latter it could be that the admissions were for routine examinations and/or procedures.

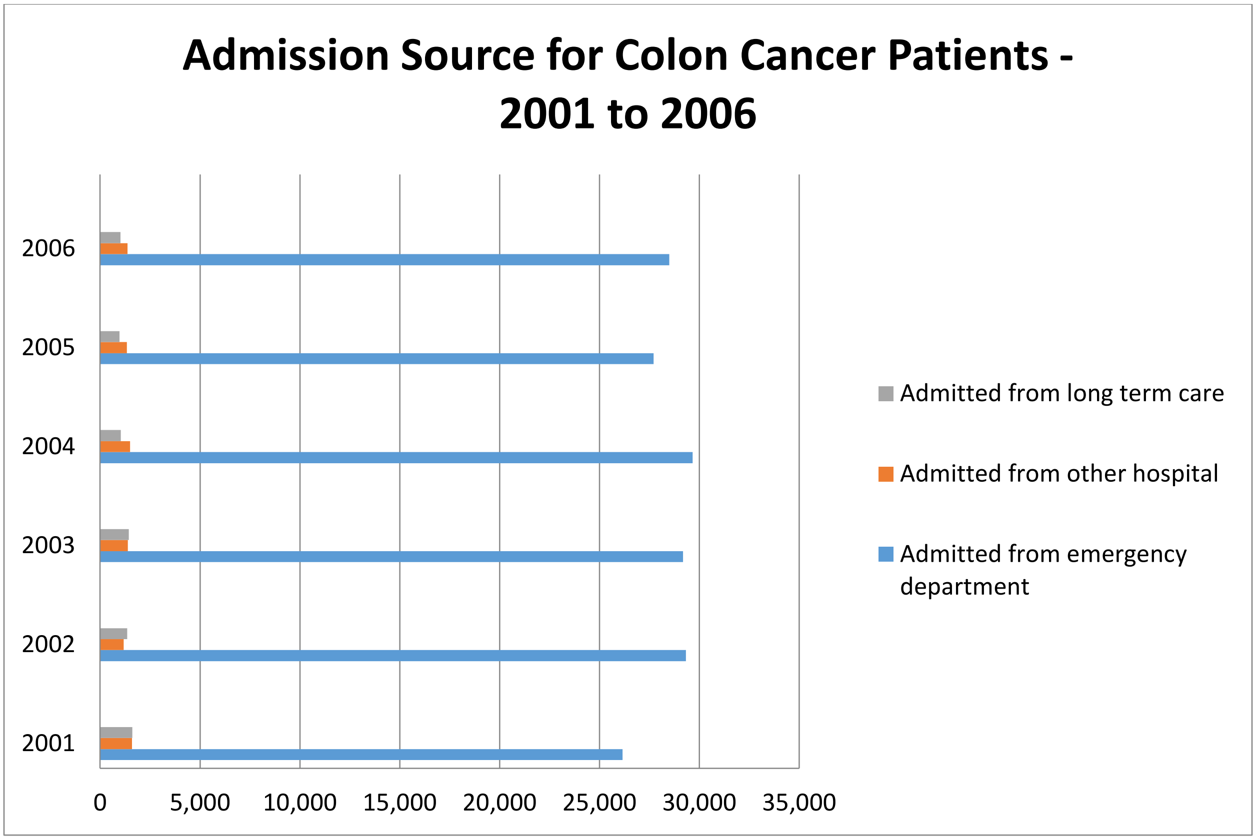
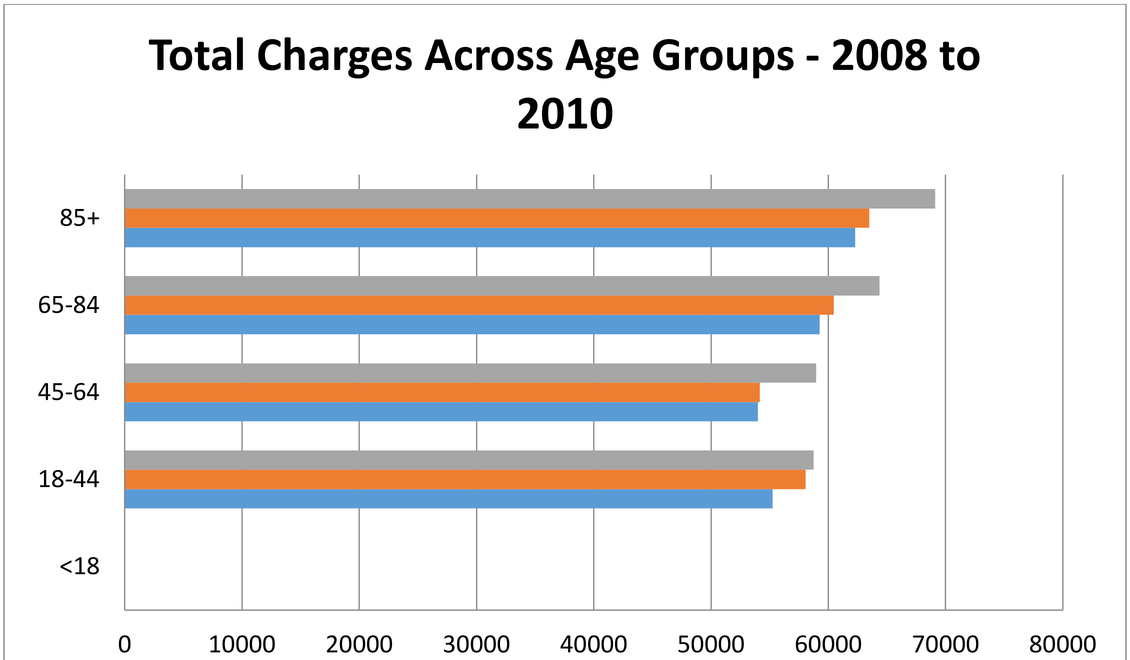


Figure 14: Illustration of Admission Sources for Colon Cancer In-Patients between 2001 and 2006 (no such data available after 2006)

In the multiple regression model above we found that Total Charges are inversely related to AGE and to investigate this further an analysis of AGE and TOTCHG was made and the chart and table below illustrates that indeed Per Discharge Costs are higher for the age ranges of 18 to 44 were higher than those aged 65 and above. Most likely the number of procedures might be more for these age groups and this was indeed found to be true by performing an analysis of variance of the number of procedures with the age groups listed above and the results are shown below. There were no significant variations in number of procedures with respect to ethnicity, gender or insurance types.



	<18	18-44	45-64	65-84	85+
2010		58732	58952	64357	69109
2009		58070	54154	60471	63493
2008		55258	54007	59273	62300

Figure 15: Illustration of Costs across Age Groups - 2010

Table 11: Results of ANOVA test for variation of NPR with age groups shown above (age group 1 is <18, 2 is between 18 and 45, 3 between 45 and 65, 4 between 65 and 85 and 5 is 85 years and above.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	79.94013	19.98503	3.79	0.0045
Error	2292	12080.92709	5.27091		
Corrected Total	2296	12160.86722			

Comparisons significant at the 0.05 level are indicated by ***.				
AGEGROUP Comparison	**Difference Between Means**	**Simultaneous 95% Confidence Limits**		
5 - 2	-0.76598	-1.42569	-0.10626	***
5 - 4	-0.49763	-0.89949	-0.09577	***
5 - 3	-0.46911	-0.89118	-0.04703	***
5 - 1	0.50962	-3.93646	4.95570	

4.5. Colon Cancer In-Patent Length of Stay (LOS) Statistics:

Length of Stay – Trend between 2008 and 2010

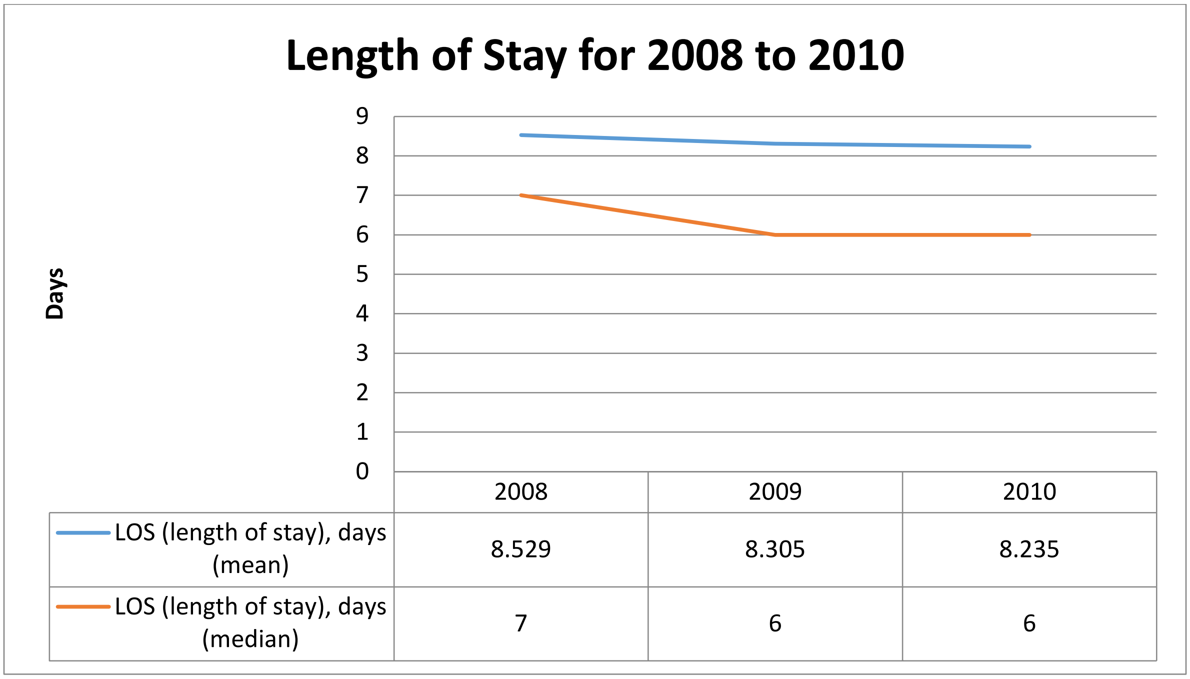


Figure 16: Trends in Length of Stay over the years 2008 to 2010.

The length of stay (mean and median) was determined using the Colon Cancer Patient Data for the years 2008 to 2010. *Length of stay* is the number of nights the patient remained in the hospital for this stay. A patient admitted and discharged on the same day has a length of stay = 0. Mean is the average while Median is the midpoint i.e. half of the cases fall below this value and half are above this value. The median is provided because the mean can be strongly influenced by extreme values. It was found that the mean length of stay of colon cancer patient discharges remained more or less the same while the median decreased by a day between 2008 to 2010. In the previous section we found that the Total Charges however had significantly risen up between 2008 and 2010.

The mean charges increase by 9 % implies that the per-day charges are significantly higher i.e. mean value by nearly 9% between 2008 and 2010.

Table 12: The Length of Stay for Different Hospital Types for 2008 to 2010

Year 2008		**Total number of discharges**	**LOS (length of stay), days (mean)**
All discharges		103,143 (100.00%)	8.6
Owner	Government	12,627 (12.24%)	8.7
	Private, not-for-profit	78,207 (75.82%)	8.6
	Private, for-profit	12,181 (11.81%)	8.5
	Missing	128 (0.12%)	6.3
Teaching status	Nonteaching	57,916 (56.15%)	8.4
	Teaching	45,099 (43.73%)	8.8
	Missing	128 (0.12%)	6.3
Location	Non-metropolitan	13,452 (13.04%)	7.9
	Metropolitan	89,563 (86.83%)	8.7
	Missing	128 (0.12%)	6.3
Bed size	Small	12,093 (11.72%)	8.2
	Medium	24,811 (24.05%)	8.6
	Large	66,111 (64.10%)	8.6
	Missing	128 (0.12%)	6.3

Year 2009		**Total number of discharges**	**LOS (length of stay), days (mean)**
All discharges		93,936 (100.00%)	8.3
Owner	Government	11,347 (12.08%)	8.5
	Private, not-for-profit	70,091 (74.62%)	8.3
	Private, for-profit	10,928 (11.63%)	8.6
	Missing	1,570 (1.67%)	7.7
Teaching status	Nonteaching	51,720 (55.06%)	8.2
	Teaching	40,647 (43.27%)	8.6
	Missing	1,570 (1.67%)	7.7
Location	Non-metropolitan	12,070 (12.85%)	7.5
	Metropolitan	80,296 (85.48%)	8.5
	Missing	1,570 (1.67%)	7.7
Bed size	Small	11,147 (11.87%)	7.7
	Medium	22,981 (24.46%)	8.3
	Large	58,239 (62.00%)	8.5
	Missing	1,570 (1.67%)	7.7

Year 2010		**Total number of discharges**	**LOS (length of stay), days (mean)**
All discharges		91,083 (100.00%)	8.3
Owner	Government	12,431 (13.65%)	8.6
	Private, not-for-profit	67,916 (74.57%)	8.2
	Private, for-profit	9,938 (10.91%)	8.3
	Missing	799 (0.88%)	8.1
Teaching status	Nonteaching	47,698 (52.37%)	8.1
	Teaching	42,586 (46.76%)	8.5
	Missing	799 (0.88%)	8.1
Location	Non-metropolitan	11,457 (12.58%)	7.4
	Metropolitan	78,827 (86.54%)	8.4
	Missing	799 (0.88%)	8.1
Bedsize	Small	11,217 (12.31%)	7.7
	Medium	20,830 (22.87%)	8.1
	Large	58,238 (63.94%)	8.4
	Missing	799 (0.88%)	8.1

The results above pertaining to the Length of Stay (LOS) across the various types of hospitals with different capacities are more or less the same across the years and there is not much of a variation across the various locations or types.

4.6. Stepwise Linear Regression for Length of Stay – 2008

A Multiple Linear Regression Model for predicting Length of Stay based on several of the demographic and hospitalization related variables was performed using a Stepwise regression technique. The model was not very robust with a R^2 of only 0.36 with the following (variables contributing more than 0.01 for the Partial R Square) as the significant predictor variables.

Table 13: List of Variables Identified as Significant in Prediction of Length of Stay – 2008

Summary of Stepwise Selection

Step	Variable Entered	Variable Removed	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	NPR		Number of procedures on this record	1	0.3252	0.3252	1184.55	10529.0	<.0001
2	NDX		Number of diagnoses on this record	2	0.0315	0.3566	112.794	1068.39	<.0001

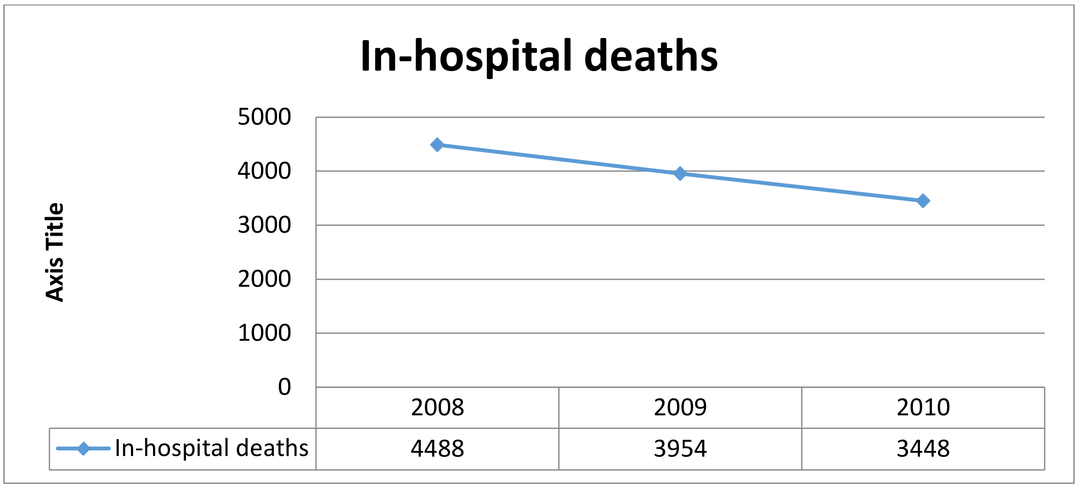
Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	1.04488	0.33158	336.66286	9.93	0.0016
NPR	1.40809	0.01800	207481	6119.87	<.0001
NDX	0.26871	0.00912	29410	867.47	<.0001

The multiple linear regression model as elicited by the stepwise regression (and using only those variables contributing more than 0.01 for the Partial R Square) is accordingly,

LOS = 1.045 + 1.41*NPR + 0.27 *NDX **…. (4.2)**

The multiple linear regression model above reveals that Length of Stay is dependent on Number of Procedures and Number of Diagnoses and doesn't relate significantly to race, gender

4.7. Colon Cancer In-Patient In-Hospital Mortality (DIED) Statistics:



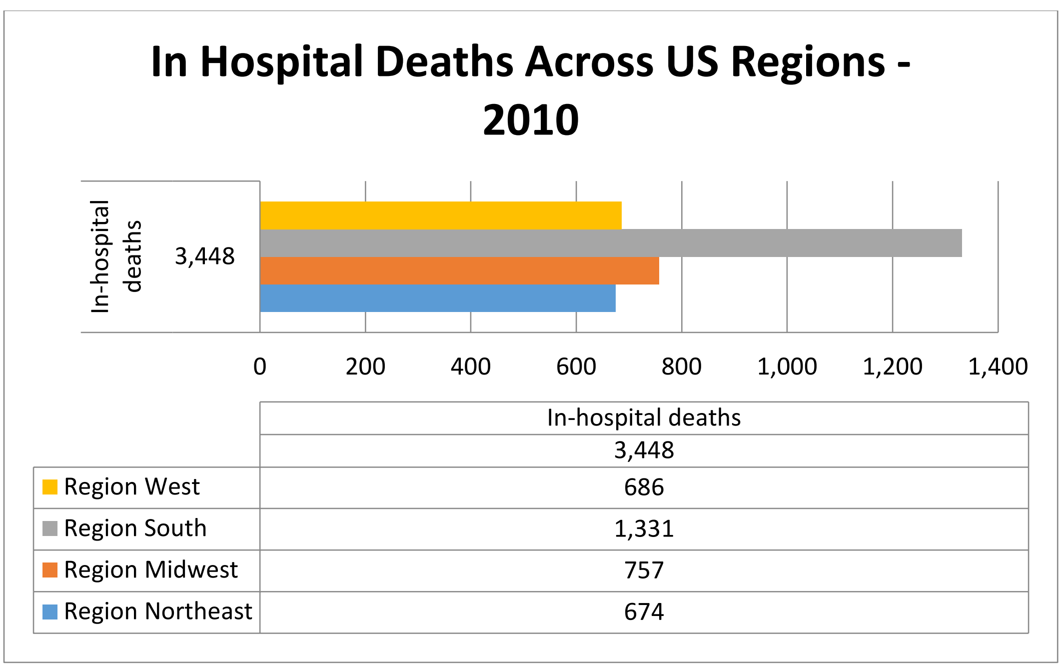


Figure 17: Illustration of Colon Cancer Inpatient In-Hospital Deaths based on US Regions 2008 to 2010

The number of in-hospital deaths across the various regions of the United States as shown above reveals that there is a decreasing trend in the number of deaths over the years 2008 to 2010. It also reveals that the South has more (nearly 2 times) in-hospital deaths compared to the other regions. This could probably be attributed to the dietary aspects of the South as also there is a greater African American population in the South as compared to the other regions.

Table 14. Mortality versus Age Groups for Colon Cancer In-Patients.

Table of AGE by DIED			
AGE('Age of Patient')	**DIED('Mortality Status')**		
Frequency	**0**	**1**	**Total**
'Less than 16 years'	7	0	7
'16 – 25 years'	56	4	60
'26 – 35 years'	337	5	342
'36 – 45 years'	1436	33	1469
'46 – 55 years'	4390	126	4516
'Over 55 years'	23734	1087	24821
Total	29960	1255	31215
Frequency Missing = 41			

The table above lists the variation of mortality statistics (of patients who died during their hospital stay) observed across the various age groupings as listed.

Table 15. Mortality versus Insurance Type for Colon Cancer In-Patients.

Table of PAY1 by DIED			
PAY1('Insurance type of Patient')	DIED('Mortality Status')		
Frequency Expected	0	1	Total
1	16893	780	17673
2	1677	72	1749
3	9790	296	10086
4	686	22	708
5	128	3	131
6	769	77	846
Total	29943	1250	31193
Frequency Missing = 63			

The table above lists the variation of mortality statistics (of patients who died during their hospital stay) observed across the various insurance groupings as listed. The PAY1 variable is the insurance used by the patient (values of PAY1 1 refers to Medicare, 2 to Medicaid, 3 to Private Insurance, 4 to Self-Pay, 5 is No Charge and 6 is Other). Here the results indicate that the majority of patients used Medicare across the years 2008 to 2010.

Table 16. Mortality versus Ethnic Groups for Colon Cancer In-Patients.

Table of RACE by DIED			
RACE('Ethnicity of Patient')	DIED('Mortality Status')		
Frequency Expected	0	1	Total
1	18139	752	18891
2	2598	125	2723
3	1667	71	1738
4	779	29	808
5	115	6	121
6	668	40	708
Total	23966	1023	24989
Frequency Missing = 6267			

The table above lists the variation of mortality statistics (of patients who died during their hospital stay) observed across the various ethnic groupings as listed. Race value of 1 indicates White population, 2 Blacks, 3 Hispanics, 4 Asian or Pacific Islanders, 5 Native American, and 6 others (not indicated by the patient). Most of the patients who died belong to the White population in proportion to the population in the country followed by African Americans, Hispanics and others. However looking at the percentages the White and the Black population in-hospital mortality due to Colon Cancer is very similar (around 4%).

Table 17. Mortality versus Diagnoses Types for Colon Cancer In-Patients.

Table of DX1 by DIED			
DX1('Diagnosis of Patient')	DIED('Mortality Status')		
Frequency	0	1	Total
Malignant neoplasm of colon	1110	24	1134
Transverse colon	1949	75	2024
Descending colon	1205	42	1247
Sigmoid colon	4577	125	4702
Cecum	3885	138	4023
Appendix	482	9	491
Ascending colon	4401	118	4519
Splenic flexure	658	23	681
Other specified sites of large intestine	702	52	754
Colon, unspecified	1941	353	2294
rectosigmoid junction	3223	134	3357
rectum	5054	137	5191
anal canal	124	3	127
anus, unspecified	192	10	202
other sites	471	12	483
Total	29974	1255	31229
Frequency Missing = 27			

The table above lists the variation of mortality statistics (of patients who died during their hospital stay) observed across the various diagnostic groupings as listed. Sigmoid Colon Cancer has the most mortality since sigmoidal region has the tendency to collect foreign substances that might trigger the cancer mechanism in that part of the colon.

Table 18. Mortality versus Diagnoses and Age Groups for Colon Cancer In-Patients

Table of AGE by DX1								
AGE('Age of patient')	DX1('Diagnosis of Patient')							
Frequency	Malignant neoplasm of colon	Transverse colon	Descending colon	Sigmoid colon	Cecum	Appendix	Ascending colon	Splenic flexure
'Less than 16 years'	0	0	1	0	1	2	0	1
'16 – 25 years'	0	4	6	8	2	15	3	0
'26 – 35 years'	8	17	12	53	20	21	19	11
'36 – 45 years'	39	52	56	232	121	63	114	25
'46 – 55 years'	99	174	203	755	367	138	399	100
'Over 55 years'	991	1776	970	3656	3513	251	3981	546
Total	1137	2023	1248	4704	4024	490	4516	683
Frequency Missing = 14								

The table above lists the variation of mortality statistics (of patients who died during their hospital stay) observed across the various diagnostic and age groupings as listed. As can be expected the majority of in-hospital deaths occurred for those over 55 years as has been reported in the literature earlier too.

4.8. Stepwise Logistic Regression for In-Hospital Mortality:

A Stepwise Logistic Regression technique was performed on the Colon Cancer data and all relevant variables were included into the technique which however identified the following as the only significant variables with predictive potential for In Hospital Mortality.

Table 19: Results of Stepwise Logistic Regression in Prediction of In-Hospital Mortality, their Odds Ratios and Model Validation

Summary of Stepwise Selection						
	Effect					
Step	**Entered**	**Removed**	**DF**	**Score Chi-Square**	**Pr > ChiSq**	**Variable Label**
1	**AGE**		1	80.3313	<.0001	Age in years at admission
2	**CM_HTN_C**		1	77.4726	<.0001	AHRQ comorbidity measure: Hypertension (combine uncomplicated and complicated)
3	**RACE**		5	17.1254	0.0043	Race (uniform)
4	**CM_OBESE**		1	6.6253	0.0101	AHRQ comorbidity measure: Obesity

Table 20: Wald Confidence Interval for Odds Ratios.

Wald Confidence Interval for Odds Ratios				
Effect	**Unit**	**Estimate**	**95% Confidence Limits**	
CM_HTN_C 0 vs 1	1.0000	1.950	1.675	2.271
CM_OBESE 0 vs 1	1.0000	1.624	1.118	2.360
RACE 1 vs 6	1.0000	0.604	0.416	0.875
RACE 2 vs 6	1.0000	0.861	0.567	1.308
RACE 3 vs 6	1.0000	0.760	0.480	1.205
RACE 4 vs 6	1.0000	0.697	0.401	1.214
RACE 5 vs 6	1.0000	0.994	0.373	2.649
AGE	1.0000	1.033	1.027	1.039

The validation of the Logistic Regression model is done using the **Hosmer and Lemeshow's Goodness of Fit Test**, which tests the null hypothesis that the data were generated by the model fitted by the researcher. The test divides subjects into deciles based on predicted probabilities, then computes a chi-square from observed and expected frequencies. Then a probability (p) value is computed from the chi-square distribution with 8 degrees of freedom to test the fit of the logistic model. If the Hosmer and Lemeshow (H-L) Goodness-of-Fit test statistic is .05 or less, we reject the null hypothesis that there is no difference between the observed and model-predicted values of the dependent. If the H-L goodness-of-fit test statistic is greater than .05, as we want for well-fitting models, we fail to reject the null hypothesis that there is no difference, implying that the model's estimates fit the data at an acceptable level. This does not mean that the model necessarily explains much of the variance in the dependent, only that however much or little it does explain is significant.

Table 21: Partition for the Hosmer and Lemeshow Test.

Partition for the Hosmer and Lemeshow Test					
		DIED = 1		DIED = 0	
Group	Total	Observed	Expected	Observed	Expected
1	1708	35	27.56	1673	1680.44
2	1788	43	42.32	1745	1745.68
3	1743	33	50.10	1710	1692.90
4	1711	72	56.80	1639	1654.20
5	1664	53	62.51	1611	1601.49
6	1807	75	76.60	1732	1730.40
7	1687	88	80.89	1599	1606.11
8	1724	100	95.77	1624	1628.23
9	1672	108	114.61	1564	1557.39
10	1588	149	148.88	1439	1439.12

The in-hospital mortality prediction model above reveals that the risks factors with significant odds ratios are hypertension comorbidity (patients with hypertension have nearly double the risk of mortality compared to those with no clinical hypertension), obesity (obese patients are 62 % more likely to die in the hospital compared to non-obese patients) and age (higher ages 0.03% more disposed to in-hospital death than those of lesser ages) and patients with white ethnicity have lower risk of dying in the hospital compared to the other ethnicities which all have similar odds ratio intervals.

Chapter V

CONCLUSIONS & FUTURE RESEARCH

5.1. Conclusion:

The overall goal of the project was to identify the factors and costs associated with Colon Cancer patients in terms of mortality, length of stay and costs in different types of clinical settings across the United States. Accordingly this research study utitlized the datasets for 2008 to 2010 available from the Nationwide Inpatient Sample (NIS) database with hospitalization characteristics of patients admitted with Colon Cancer as the principal diagnosis. Specifically the following were studied in this thesis:

1. what clinical factors (such as number and types of comorbidities and procedures) influenced the mortality, costs and length of stay
2. whether mortality, costs and length of stay differed with race, age, or socio-economic status
3. whether there were differences in the mortality, costs and length of stay across the various regions of the US
4. whether there were differences in the mortality, costs and length of stay amongst the different types of hospital settings – rural/urban/hospital with and without teaching.

Based on the analyses and studies done in this thesis the following were concluded:

- Between 2008 and 2010 the age and population adjusted incidences and the hospital discharges both decreased significantly which is a promising trend speaking well of the state of health care in the United States as also possibly due to the effectiveness of nutritional counselling, patient education, screening for men aged 50 and above.

- It was found that while the total number of colon cancer patient discharges decreased significantly between 2008 to 2010 the Total Charges however significantly risen up between 2008 and 2010. The mean charges increased by nearly 8 %. This implies that there must be greater costs associated with each patient over the years 2008 to 2010 due perhaps to the increased costs of new procedures as also more preventative screenings entailing greater costs per patient.
- The number of discharges across the various hospital types and their locations across the United States as shown above reveals those large hospitals in metropolitan regions and those that are private not-for-profit have more discharges compared to the other types. This could probably be attributed to the large population that the hospitals serve being in metropolitan regions and also that the private not-for-profit hospitals typically have more advanced procedures and newer treatment regimens available.
- The results pertaining to the discharges across the various insurance types reveal that the patients who are uninsured and those on Medicaid (low income) are more in number over the years 2008 to 2010 as compared to those on Medicare and Private Insurance which have decreasing trends. The reasons are not very apparent but it could be due

to the severe economic downturn during this period with the chances that there are more patients who were uninsured and/or relying on Medicaid otherwise.

- It was found that the mean and median length of stay of colon cancer patient discharges remained more or less the same between 2008 and 2010.
- The results pertaining to the Length of Stay (LOS) across the various types of hospitals with different capacities reveals that LOS is more or less the same across the years (2008 to 2010) and there is not much of a variation across the various locations or types of hospitals.
- It was found that the number of in-hospital mortality or deaths significantly reduced between 2008 and 2010. Alongside Home Health Care increased while discharges to another hospital also decreased (with a smaller decrease in discharge to another institution such as rehab facility and nursing home). The increase in home health care (between 2008 and 2010) is probably due to the better colon health screening and procedures performed at the hospital and/or due to the availability of better surgical and/or treatment regimens in the recent years.
- The number of in-hospital deaths across the various regions of the United States as shown above reveals that there is a decreasing trend in the number of deaths over the years 2008 to 2010. It also reveals that the South has more (nearly 2 times) in-hospital deaths compared to the other regions in all the 3 years. This could probably be attributed to the dietary

aspects of the South as also there is a greater African American population in the South as compared to the other regions.

- This study seems to indicate that mortality is positively correlated with the total costs though the reasons are not immediately apparent. However it can be reasoned that in-hospital mortality in the case of cancer patients may have had them come in as emergency admissions with high risk of organ failure and such and treatment and surgeries (hence number of procedures) to prolong their lives may have been more contributing accordingly to greater costs. To confirm this data was procured as to the source of admission for Colon Cancer Patients and this piece of information as to the nature of admission source was only available upto 2006 after which it has been removed. However looking at the trend between 2001 and 2006 it was seen that a significant number of patients admitted to the hospitals with Colon Cancer as the principal diagnosis did indeed come in from the Emergency Department. Similar results were also seen for those admitted with Cancer of the Lung and Bronchus as the principal diagnosis whereas those with Breast Cancer were non-emergency department sourced. In the case of the latter it could be that the admissions were for routine examinations and/or procedures.
- The in-hospital mortality prediction model above reveals that the risks factors with significant odds ratios are hypertension comorbidity (patients with hypertension have nearly double the risk of mortality compared to those with no clinical hypertension), obesity (obese patients are 62 %

more likely to die in the hospital compared to non-obese patients) and age (higher ages 0.03% more disposed to in-hospital death than those of lesser ages) and patients with white ethnicity have lower risk of dying in the hospital compared to the other ethnicities which all have similar odds ratio intervals.

5.2 Future Research Recommendations:

This research study was limited to the datasets available from the Nationwide Inpatient Sample (NIS) database with hospitalization characteristics of patients admitted with Colon Cancer as the principal diagnosis. As was shown most of these patients were admitted from the emergency department and therefore this study is relegated to this specific population group of Colon Cancer patients. A similar large scale dataset based future study is indeed warranted to analyze demographic and hospital based outcomes for a wider variety of Colon Cancer patients admitted for screening, cancer management, clinical trials and education. Also a comparative study can be made across different cancer patients to evaluate appropriate allocation of resources and education specific to different cancer types and the specific population sub-groups affected by such cancers. Large dataset studies like this are much more definitive in their results and conclusions can be drawn with a stronger statistical confidence.

References:

1. Center for disease control and preventive http://www.cdc.gov/cancer/colorectal/statistics/screening_rates.htm retrieved from internet on December 2, 2013.

2. Parkin DM, Whelan SL. et al. (1997), Cancer incidence in five continents vol. 7, IAKC Scientific Publication No.143, pp. 29.

3. Odze RD, Goldblum JR. (2009), Surgical Pathology of the GI Tract, Liver, Biliary Tract and Pancreas. Saunders, Elsevier; pp.506-507.

4. American Cancer Society: http://www.cancer.org/cancer/colonandrectumcancer/detailedguide/colorectal-cancer-stage retrieved from internet on February 2, 2014.

5. Hirotugu, A. (1974), "A new look at the statistical model identification", IEEE Transactions on Automatic Control, 19 (6): 716–723.

6. American Association for Cancer Research. “Genetic Differences Point to Ethnic and Racial Disparities in Colorectal Cancer Risk”. http://www.aacr.org/home/public--media/aacr-press-releases/press-releases-2007.aspx retrieved from internet on January 13,2014.

7. Rebecca S, Ahmedin J. (2008), “Behavioral Risk Factor Surveillance System Public Use Data Tapes 2006 and 2008”. National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention.

8. Hosmer DW, Lemeshow S. (1999), Applied survival analysis: regression modeling of time to event data. New York: Wiley.

9. Kaster S, Buckley S. et al. (1995), "Colonoscopy and barium enema in the detection of colorectal cancer."; Gastrointestinal Endoscopy, 41:379.

10. Helen AH, Maria M. et al. (2010), Chronic Disease Prevention and the New Public Health;32: 120-154.

11. Harrell FE. (2001), Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. New York.

12. Jorgensen ML, Young JM. et al. (2014), “Predictors of variation in colorectal cancer care and outcomes in New South Wales: a population-based health data linkage study”, Medical Journal of Australia. 21; 200 (7):403-407.

13. Wilks, SS. (1938), "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses". The Annals of Mathematical Statistics 9: pp.60–62.

14. Guillem JG, Forde KA et al. (1987), “The impact of colonoscopy on the early detection of colonic neoplasms in patients with rectal bleeding”, Annals of Surgery; 206:606.

15. Morris EJ, Whitehouse LE. et al. (2012), “A retrospective observational study examining the characteristics and outcomes of tumours diagnosed within and without of the English NHS Bowel Cancer Screening Programme.”; British Journal of Cancer, 107(5):757-64.

16. O'Brien MJ, Winawer SJ. et al. (1990), “The National Polyp Study: Patient and polyp characteristics associated with high-grade dysplasia in colorectal adenomas”; Gastroenterology, 98: 371.

17. Odermatt M, Miskovic D. et al.(2014), “Major postoperative complications following elective resection for colorectal cancer decrease long-term survival but not the time to recurrence”; Colorectal Diseases, [Epub ahead of print].

18. Halpern MT, Pavluck AL. et al. (2009), “Factors associated with colon cancer stage at diagnosis.”; Digestive Diseases and Sciences; 54(12): 2680-93.

19. Kao PS, Lin JK. et al. (2009), “The impact of family history on the outcome of patients with colorectal cancer in a veterans' hospital.”; International Journal of Colorectal Diseases, 24(11): 1249-54.

20. Ota Y, Matsui T. et al. (2003), “Value of virtual computed tomographic colonography for Crohn‘s colitis: comparison with endoscopy and barium enema”; Abdominal Imaging, 28(6): 778–783.

21. Anderson WF, Guyton KZ. et al.(2002), “Colorectal Cancer Screening for Persons at Average Risk”; Journal of National Cancer Institute, 94(15): 1126-33.

22. Wang A, Clouston SA. et al. (2012), “Fundamental Causes of Colorectal Cancer Mortality: The Implications of Informational Diffusion”; Milbank Quarterly, 90(3): 592-618.

23. El Batea H, Enaba M. et al. (2010), “Indications and Outcome of Colonoscopy in the Middle of Nile Delta of Egypt”; 10.1007.

24. Cancer.Net Editorial, Board, 9/2013 http://www.cancer.net/cancer-types/colorectal-cancer retrieved from internet on January 2,2014

25. Howard DH, Guy GP Jr. et al. (2014), “Eliminating cost-sharing requirements for colon cancer screening in Medicare”; Cancer, Oct 9. 10. [Epub ahead of print].

26. Boyd CR, Tolson MA. et al.(1987), “Evaluating trauma care: the TRISS method. Trauma Score and the Injury Severity Score”, Journal of Trauma; 27(4): 370–8.

27. Green, DM, Swets, JA. et al. (1966), Signal detection theory and psychophysics. New York, NY: John Wiley and Sons Inc.

28. Barish MA, Soto JA. et al. (2005), “Consensus on Current Clinical Practice of Virtual Colonoscopy”, American Journal of Radiology; 184:786-792.

29. Terry P, Giovannucci E. et al (2001), ”Fruit, vegetables, dietary fiber, and risk of colorectal cancer”; Journal of National Cancer Institute, 93(7): 525-33.

30. Schwarz, Gideon E. (1978), "Estimating the dimension of a model". Annals of Statistics; 6(2): 461–464.

31. Park SH, Lee SS. et al. (2007), “Flat colorectal neoplasms: definition, importance, and visualization on CT colonoscopy”, American Journal Roentgenology; 188: 953-9.

32. Hazra A, Chanock S. et al. (2008), Cancer Epidemiology Biomarkers and Prevention; 17: 311-319.

33. Centers for Disease Control and Prevention. “Colorectal Cancer Rates by State”. http://www.cdc.gov/cancer/colorectal/statistics/state.htm retrieved from internet on January 15, 2014.

34. Marchand L. (1999), “Combined influence of genetic and dietary factors on colorectal cancer incidence in Japanese Americans”; Journal of National Cancer Institute Monograph; (26): 101-5..

35. Endoscopic Classification Working Group. (2005), “Update on the paris classification of superficial neoplastic lesions in the digestive tract.”; Endoscopy, 37(6): 570-8.

36. National Cancer Institute http://www.cancer.gov/cancertopics/pdq/treatment/colon/Patient/page2 retrieved from internet on July 5, 2013.

37. Colon Cancer Screening and Prevention | CancerConnect.com http://news.cancerconnect.com/colon-cancer-screening-prevention/ retrieved from internet on July 5, 2013.

38. Haggar FA, Boushey R. (2009), “Colorectal Cancer Epidemiology: Incidence, Mortality, Survival, and Risk Factors”; Clinics in Colon and Rectal Surgery, 22(4):191-7.

39. Ning Y, Wang L. et al. (2010), “A quantitative analysis of body mass index and colorectal cancer: findings from 56 observational studies.”; Obesity Reviews; 11(1): 19-30.

40. Levin B, Lieberman DA et al. (2008), “Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology”; Gastroenterology, 134(5): 1570-95.

41. Martinez ME, McPherson RS. et al. (1995), “Aspirin and other non-steroidal anti-inflammatory drugs and risk of colorectal adenomatous polyps among endoscoped individuals”; Cancer Epidemiology, Biomarkers and Prevention, 7:703-7.

42. Bretthauer M., Kalagher M (2013), “Principles, effectiveness and caveats in screening for cancer.”; British Journal of Surgery, 100(1): 55-65.

43. Fox JP, Tyler JA. et al. (2014), “A variation in the value of colectomy for cancer across hospitals: Mortality, readmissions, and costs.”; Surgery: 156(4): 849-60.

44. Liang Z, Richards R. (2010), “Virtual colonoscopy vs optical colonoscopy”, Expert Opinion on Medical Diagnosis; 4(2): 159–169.

45. National Cancer Institute. Cancer costs project. http://www.nih.gov/news/health/jan2011/nci-12.htm retrieved from internet on December 2, 2013.

46. National Cancer Institute http://www.cancer.gov/cancertopics/factsheet/detection/colorectal-screening retrieved from internet on October 10, 2013.

47. Spence RR, Heesch KC. et al.(2009), “A systematic review of the association between physical activity and colorectal cancer risk.”, Scandinavian Journal of Medicine and Science in Sports; 6:764-81.

48. Howlader N, Noone AM. et al. (2012), “Use of imputed population-based cancer registry data as a method of accounting for missing information: application to estrogen receptor status for breast cancer”; American Journal of Epidemiology, 176(4): 347-56.

49. Gohil R, Rishi M. et al. (2013), “Pre-operative Serum Albumin and Neutrophil-Lymphocyte Ratio are Associated with Prolonged Hospital Stay Following Colorectal Cancer Surgery”, International Journal of Surgery, 11(8): 622.

50. Huebner M, Hübner M. et al.(2014), .“Timing of complications and length of stay after rectal cancer surgery.”; Journal of American College of Surgeons; 218(5):914-9.

51. Nurhee H, Seong H. (2014), “CT colonography in the diagnosis and management of colorectal cancer: Emphasis on pre- and post-surgical evaluation”; World Journal of Gastroenterology, 20(8):2014-22.

52. Perng DS, Lu IC. et al. (2014), “Incidence trends and predictors for cost and average lengths of stay in colorectal cancer surgery.”; World Journal of Gastroenterology, 20(2): 532-8.

53. O'Brien MJ, Winawer SJ. et al. (2004), “Flat adenomas in the National Polyp Study: Is there increased risk for high-grade dysplasia initially or during surveillance?”; Clinical Gastroenterology and Hepatology, 2(10): 905-11.

54. Ladabaum U et al. (2014). “Colorectal Testing Utilization and Payments in a Large Cohort of Commercially Insured US Adults”. American Journal of Gastroenterology, 109(10):1513-25.

55. Zullig LL, Jackson GL. et al. (2012),.“Cancer incidence among patients of the U.S. Veterans Affairs Health Care System.”; Military Medicine, 177(6):693-701.

56. O’Donovan AN, Somers S. et al. (1997), “A prospective blinded randomized trial comparing oral sodium phosphate and polyethylene glycol solutions for bowel preparation prior to barium enema”, Clinical Radiology, 52: 791-793.

57. Ladabaum U et al (2010), “Cost-effectiveness of colorectal cancer screening in high-risk Spanish patients: use of a validated model to inform public policy.”. Cancer Epidemiology, Biomarkers and Prevention, 19(11): 2765-76.

58. Stroupe K T. et al (2014), “Costs of care for lung and colon cancer patients receiving chemotherapy following FDA policy changes”. Supportive Care in Cancer: 2014 Jun 10 [epub ahead of print].