Predictive Modeling For In-Hospital Mortality In Pancreatic Cancer Patients

BY

Fang Luo

A Dissertation Submitted to

Rutgers, the State University of New Jersey

School of Health Related Professions

In partial fulfillment of the Requirements for

The Degree of Doctor of Philosophy

In Biomedical Informatics

Department of Health Informatics

June, 2015

RUTGERS
School of Health Related Professions

Final Dissertation Defense Approval Form

Predictive Modeling For In-Hospital Mortality In Pancreatic Cancer Patients

BY

Fang Luo

Dissertation Committee:

Dinesh P. Mital, PhD. Committee Chair

Shankar Srinivasan, PhD. Committee Member

Frederick Coffman, PhD. Committee Member

Approved by the Dissertation Committee:

_____ Date: _____

_____ Date: _____

_____ Date: _____

_____ Date: _____

_____ Date: _____

Table of Contents

ABSTRACT

Background: Pancreatic cancer is very aggressive with few symptoms before the cancer can diagnosed and usually the cancer is advanced when it was found. It is the most deadly type of cancer, and it is listed as the fourth leading cause of cancer-related death with a poor prognosis because of the late finding of the disease. To patients, the pancreatic cancer diagnosis is a life-changing disaster; however, to maximally extend the pancreatic cancer patients' life is most important task after diagnosis. In this research study, we found certain statistical significant relationships between the death of pancreatic cancer patients and the comorbidities & demographics. Furthermore, patients with some comorbidities or demographics can make their life longer or shorter.

Objectives: To develop a mathematic model to predict the death of pancreatic cancer patients using certain patients' comorbidities and demographics. Also, pancreatic cancer patients with certain comorbidities or demographics can predict their rest of life will be longer or shorter.

Methods: The study uses HCUP NIS year 2005-2009 data files as research source database. With retrieving pancreatic cancer patient from NIS data files, the database used in this study includes pancreatic cancer patients' information with their comorbidities and some demographics. The algorithm used in this

research study in 1) Logistic regression ROC curve calculation 2) Logistic regression Odds Ratio calculation.

Results: 1) In output ROC curve, the area under the ROC curve (AUC) is 0.725 which is >0.5. It means when randomly pick one patient from the disease group (diagnosed as pancreatic cancer) and one from the no-disease group (not diagnosed as pancreatic cancer) and do the test on both. The patient with the more abnormal test result (Died) should be the one from the disease group (diagnosed as pancreatic cancer). 2) 11 comorbidities can significantly influence on life of pancreatic cancer patients, both LCL and UCL of odds ratio are either <1 or >1 in all 11 comorbidities, which means if patients were diagnosed as one of 11 comorbidities, they might either live longer or live shorter.

Conclusions: The study results were cross verified by 3 types of cancer patients, which is pancreatic cancer, breast cancer and stomach cancer. Those 11 comorbidities have the same statistical significance effect on influencing patients' life. So far no literatures can be found on such kind of study, this study can be considered as a new research field in the future.

ACKNOWLEDGEMENT

I’d like give my thinks to Dr. Dinsesh Mital, my dissertation advisor, for guiding me to finish this research project. Without his advise, it would be much harder for me to finish this project and receive good results. His solid knowledge background and sharp mind always can provide me some good ideas when I hit barriers in my research work. Also, I would like to thank Dr. Syed Haque, Chairman of Department of Health Informatics, Dr. Shankar Srinvasan, associate professor of Department of Health Informatics, for their advise and support in my research project.

List of Tables

List of Figures

Chapter I

INTRODUCTION

1.1 Background of the disease

The pancreas is an organ in the abdomen that sits in front of the spine above the level of the belly button. It performs two main functions. First, it makes insulin, a hormone that regulates blood sugar levels; and second, it makes enzymes which help break down proteins. The enzymes help digestion by chopping proteins into smaller parts so that they can be more easily absorbed by the body and used for energy. Enzymes leave the pancreas via a system of tubes called "ducts" that connect the pancreas to the intestines. The pancreas sits deep in the belly and is in close proximity to many important structures such as the small intestine (the duodenum) and the bile ducts, as well as important blood vessels and nerves [2].

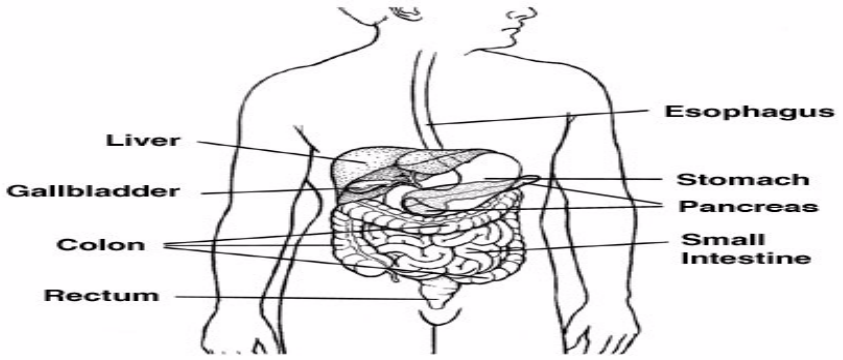


Figure 1: Pancreas organ in the body

Cancer is the uncontrolled growth of abnormal cells in the body. Cancer develops when the body's normal control mechanism stops working. Old cells do not die and cells grow out of control, forming new, abnormal cells.

According the definition from National Cancer Institute, "Cancer is the name given to a collection of related diseases. In all types of cancer, some of the body's cells begin to divide without stopping and spread into surrounding tissues." [1] And a disease in which malignant (cancer) cells are found in the tissues of the pancreas called exocrine cancer. Cancers that develop within the pancreas fall into two major categories: (1) cancers of the endocrine pancreas (the part that makes insulin and other hormones) are called "islet cell" or "pancreatic neuroendocrine" cancers and (2) cancers of the exocrine pancreas (the part that makes enzymes). Islet cell cancers are rare and typically grow slowly compared to exocrine pancreatic cancers. Islet cell tumors often release hormones into the bloodstream and are further characterized by the hormones they produce (insulin, glucagon, gastrin, and other hormones). Cancers of the exocrine pancreas develop from the cells that line the system of ducts that deliver enzymes to the small intestine and are commonly referred to as pancreatic adenocarcinomas. Adenocarcinoma of the pancreas comprises most of all pancreatic ductal cancers and is the subject of this review [3].

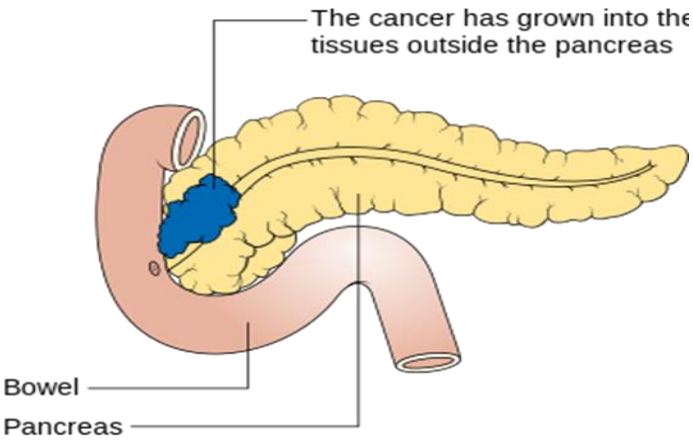


Figure 2: Pancreatic cancer

According to the American Cancer Society, in 2015 about 48,960 people will be diagnosed with pancreatic cancer and about 40,560 people will die of pancreatic cancer. Cancer is the second leading cause of death in the United States following heart disease. Pancreatic cancer is the fourth leading cause of cancer-related death in the United States [41-43] with lung cancer as #1. Because pancreatic cancer progresses rapidly, and no method of early detection has been discovered, it is one of the most dangerous types of cancer. In the case of pancreatic cancer, more than 85% of diagnoses have metastasized, which means they are unsectable [35-37]. The one-year survival rate is 25 percent, and the five-year survival rate sits at only 6 percent. The median survival rate is less than six months and at least 95$ of those diagnosed with pancreatic cancer will expire to the disease within 5 year of diagnosis [38-40]. In other words, pancreatic cancer is the most aggressive and the most lethal form of cancer [28-32]. Nevertheless, pancreatic cancer has the worst prognosis in all of medicine, it is a deadly disease with a dismal prognosis [45]. It has really poor record on

prognosis [46], one important reason is lacking of symptoms in the early stages of the disease [47, 48].

Furthermore, even if a patient was diagnosed as pancreatic cancer before metastasis, still, the full recovery chance is small. The American Cancer Society prognosis figures have indicated that the twelve month all-stage survival for pancreatic cancer is estimated to be 20%. And that the five year rate is considered to be about 4%. The treatment options for pancreatic cancer are either surgery if it is still localized or therapy if it has metastasized. Pancreatic cancer can be controlled only if it is found before it has spread, when it can be completely removed by surgery. If the cancer has spread, palliative treatment can improve the patient's quality of life by controlling the symptoms and complications of this disease. Although, surgery presents the best treatment option with the highest survival rate, only about 15% of those diagnosed with pancreatic cancer fall under this category [49]. Even then, not all of these cancers turn out to be truly resectable once the surgery is started. Sometimes once the surgeon starts the operation it becomes clear that the cancer has grown too far to be removed completely. If this happens, the operation may be stopped, or the surgeon might continue with a smaller operation with a goal of relieving or preventing symptoms [50]. For those patients who could not be treated by surgery, undergoing treatments will follow a) Ablation or embolization treatments b) Radiation therapy and c) Chemotherapy and other drugs. However, the survival rate for those patients who could not do surgery is really low, most patients will succumb to the disease eventually. Some researchers believe that one of the

reasons why pancreatic cancer is hard to cure is the location of the pancreas because the pancreas is a 6 to 10 inch (18 to 25 cm) long organ located behind the stomach in the back of the abdomen. And the tumors are often surrounded by dense tissues which might be a problem for drug delivery [51]. This location makes the diagnosis of pancreatic cancer more difficult than diagnosing other digestive tract cancers.

Obviously, early diagnosis of pancreatic cancer is the most efficient way to cure the disease. Because of traditional diagnose methods usually can only find cancer cell in advanced stage, nowadays, researchers are working on finding the highly sensitive and specific biomarkers for pancreatic malignancies, and it has become one of the hottest research fields. Using molecular markers thought to identify the presence of pancreatic cancer is not currently recommended because researchers are yet to identify specific biomarkers and the reliability of the currently used biomarkers is yet to be proven [53,54,55]. However, recently the researchers, from H. Lee Moffitt Cancer Center in Tampa, FL, report their findings in the journal PLOS ONE. The Moffitt team focused on a small group of molecules called microRNAs, they looked specifically at microRNAs linked to intraductal papillary mucinous neoplasms (IPMNs). IPMNs are a type of pancreatic cyst or lesion that can lead to pancreatic cancer in the same way that precancerous polyps can lead to colon cancer. MicroRNAs are small molecules that act as "master regulators" that control many cancer processes in the body. They can be found in tumor tissue and in blood and other body fluids. There is growing evidence that microRNAs could serve as biomarkers of early pancreatic

cancer. For example, in October 2014, Medical News Today reported how researchers from Indiana University School of Medicine found a panel of microRNAs that could be used as a blood test for pancreatic cancer. In that study, the team suggested three microRNAs - miR-10b, miR-155 and miR-106b - might serve as highly accurate early indicators of pancreatic ductal adenocarcinoma (PDAC) - the most common type of pancreatic cancer. In this new study, the Moffitt team looked for microRNAs that might be linked with high-risk IPMN lesions that should probably be removed as soon as possible to avoid full-blown pancreatic cancer. First author Dr. Jennifer Permuth-Wey, applied research scientist and molecular epidemiologist in Moffitt's Cancer Epidemiology Program, says: “The hope is that this line of research may eventually lead to a microRNA-based blood test that could be used in conjunction with imaging features and other factors to aid the medical team and patient in accurately predicting disease severity at the time of IPMN diagnosis or follow-up."

Another recent published research study shows that miR-21, miR-192 and miR-200 could be used as new diagnostic markers for pancreatic cancer. A dynamics of these miRNAs could serve as a prognostic marker in patients after cancer removal. Further prospective studies with newly-onset diabetic patients with no signs of malignancy will be needed to validate if suggested miRNAs could be used as early markers as well [67].

According to research statistics, the incidence of pancreatic cancer has been rising in the past several decades [33, 34]. In August 2012, the Pancreatic Cancer Action Network prepared a report outlining predictions for changes in cancer

incidence and deaths in the coming years. The main message of the Cancer Research paper is that pancreatic cancer is predicted to become the second leading cause of cancer-related death in the United States around 2020. In addition, the paper reports that the projected top four cancer killers by 2030 will be lung, pancreatic, liver and colorectal.

The earlier report, entitled The Alarming Rise of Pancreatic Cancer Deaths in the United States: Why We Need to Stem the Tide Today, has been archived. This report, printed and distributed by the organization, was based on older data, and the methodology has been changed slightly to provide the most accurate predictions [52].

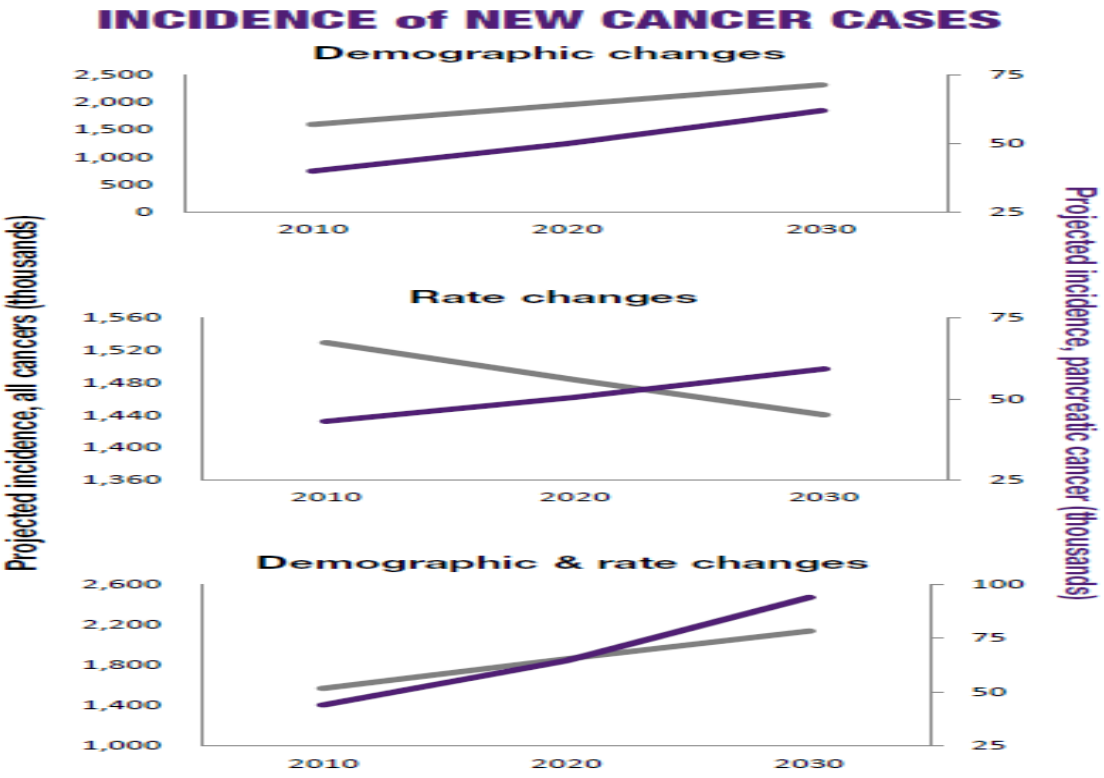


Figure 3: Projected new pancreatic cancer cases in the future [52]

The figure 3 above is the prediction of the new pancreatic cancer cases from 2010 to 2030. The figure shows that the incidence of pancreatic cancer will rise quickly in the near future.

Top: The number of new cases of cancer (grey line) is projected to increase as a result of demographic changes, namely the growing number of older adults and minorities in the population. The number of new cases of pancreatic cancer (purple line) is growing at a slightly faster rate than cancer in general.

Middle: The average number of new cases of cancer/100,000 population is decreasing for men (0.6%/year) but not changing for women, resulting in a slight decline in the projected incidence of cancer overall (grey line). In contrast, the average number of new cases of pancreatic cancer/100,000 population (purple line) is increasing in both men and women, resulting in a steady increase in the projected incidence of pancreatic cancer.

Bottom: The combination of changing demographics and the increase in the average annual incidence rate for pancreatic cancer work together to result in a more rapid increase in the projected total number of new cases of pancreatic cancer (purple line) compared to cancer in general (grey line).

There are two types of pancreatic cancer, that of the exocrine gland and that of the endocrine gland. About 95 percent of pancreatic cancers begin in the exocrine (enzyme-producing) cells of the pancreas.

- Exocrine tumors: Most tumors affecting the exocrine gland are called adenocarcinomas. This type of cancer forms in the pancreas ducts. Treatment for these tumors is based on stage of growth.
- Endocrine tumors: These tumors are less common and are most often benign. Though rare, cancer stemming from a pancreatic endocrine tumor (PET) affects the hormone-producing cells. These tumors are also called islet cell tumors or neuroendocrine tumors [8].

In the United States, pancreatic cancer is the fourth leading cause of cancer-related death in both men and women. In 2013, more than 45,000 people in the United States will be diagnosed with pancreatic cancer and more than 38,000 will die of this disease. Because pancreatic cancer usually is diagnosed at an advanced stage, the survival rate is extremely low compared with those of many other cancer types. The pancreatic cancer incidence rate has increased since 1999, and the mortality rate also has increased slightly since 2000.

African Americans have higher rates of pancreatic cancer incidence and mortality than whites or other racial/ethnic groups. The incidence of pancreatic cancer in African Americans is 50% to 90% higher than the incidence in other racial groups. African Americans also have the worst prognosis. This is an evidence-based review of pancreatic cancer in African Americans with particular emphasis on baseline characteristics, treatment, and survival [44]. Pancreatic cancer incidence and mortality rates also are higher in men than in women.

Estimated new cases and deaths from pancreatic cancer in the United States in 2014: [9]

New cases: 46,420

Deaths: 39,590

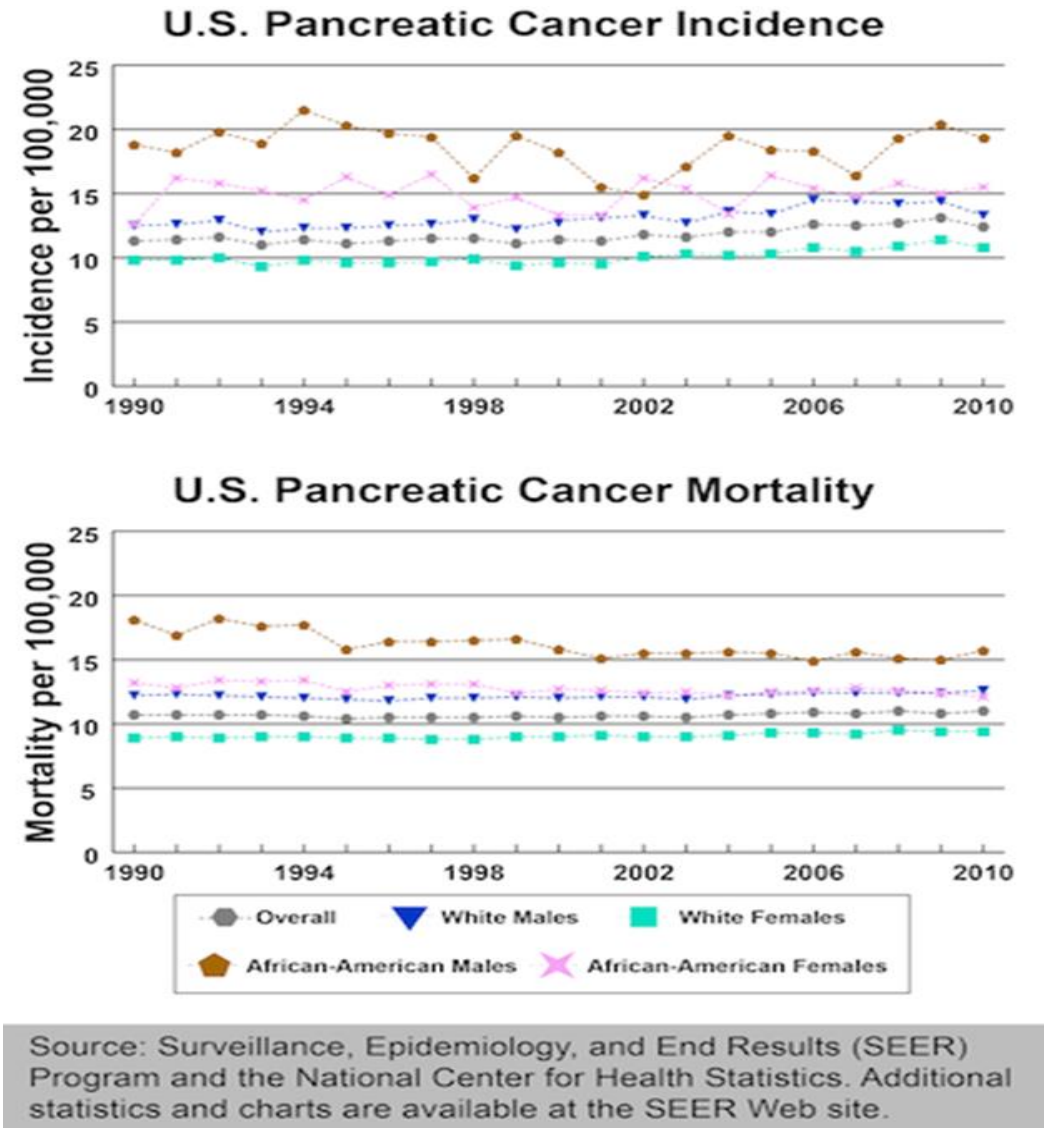


Figure 4: Pancreatic Cancer Incidence and Mortality in U.S

The American Cancer Society's most recent estimates for pancreatic cancer in the United States are for 2015:

- About 48,960 people (24,840 men and 24,120 women) will be diagnosed with pancreatic cancer.

- About 40,560 people (20,710 men and 19,850 women) will die of pancreatic cancer [13].

1.2 Goal and Objects

The overall goal of the project is to identify how the comorbidities affect the lifespan of a cancer patient. We target pancreatic cancer as a sample cancer in this research project. Specifically the objectives are to determine:

1) What the potential mortality possiblity of pancreatic cancer patient during their stay in hosptial
2) What clinical factors (such as diagnosis associated with pancreatic cancer) influence the life span of a pancreatic cancer patient
3) Whether mortality of pancreatic cancer patients differ with race, age or socio-economic status

1.3 Research Hypotheses of the project

- Are there statistically significant associations between mortality possibility and inpatient pancreatic cancer patients across the various region of the US
- Are there statistically significant difference in life span of pancreatic cancer patients and their comorbidities across the various region of the US
- Are there statistically significant differences in mortality of pancreatic cancer patients with race, age or socio-economic status

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 2005 to 2009. 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 2009 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 will show the relationship between the pancreatic cancer mortality and the commodities as well as how commodities of pancreatic cancer affect a patient life span. 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, a logistic regression model will be setup and validated includes Receiver Operating Characteristic (ROC) Curves and odd ratios statistic calculation. 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, and comorbidities. Details as to the state of art knowledge and research into Pancreatic Cancer and its management are provided in the next chapter.

Chapter Two

LITERATURE REVIEW

2.1 Epidemiology of Pancreatic Cancer

Worldwide, over 200000 people die annually of pancreatic cancer. The highest incidence and mortality rates of pancreatic cancer are found in developed countries. In the United States, pancreatic cancer is the 4th leading cause of cancer death, and in Europe it is the 6th. Because of high fatality rates, pancreatic cancer incidence rates are almost equal to mortality rates. Pancreatic cancer is diagnosed late in the natural history of the disease, given the few early indicators of illness, and the lack of screening tests for this disease. Treatment has not improved substantially over the past few decades and has little effect on prolonging survival time [56]. According to recent research conducted by the Pancreatic Cancer Action Network, pancreatic cancer will be the second leading cause of cancer-related death by the year 2030 [52, 58].

It is well known that the mortality rate of pancreatic cancer is one of the highest. Pancreatic cancer is an uncommon tumor, but because the mortality rate approaches 100%, this form of cancer has now become a common cause of cancer mortality [57].

According to National Cancer Institute (NCI), Each year in the United States, more than 43,000 people are diagnosed with cancer of the pancreas. Most are

over 65 years old. The annual incidence rate for all types of pancreatic cancer is approximately 9 new cases per 100,000 people, ranking it 11th among cancers. According to data from The Surveillance, Epidemiology, and End Results (SEER) Program, **number of New Cases and Deaths per 100,000:** The number of new cases of pancreas cancer was 12.3 per 100,000 men and women per year. The number of deaths was 10.9 per 100,000 men and women per year. These rates are age-adjusted and based on 2007-2011 cases and deaths.

Lifetime Risk of Developing Cancer: Approximately 1.5 percent of men and women will be diagnosed with pancreas cancer at some point during their lifetime, based on 2009-2011 data.

Prevalence of this cancer: In 2011, there were an estimated 43,538 people living with pancreas cancer in the United States [4].

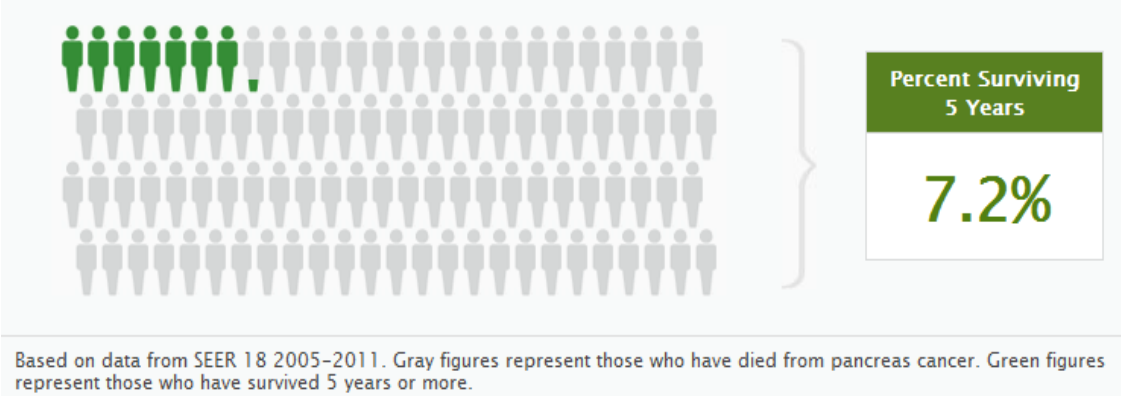


Figure 5: Pancreatic cancer patients' 5-year surviving rate (Data 2005-2011)

In above figure 5, it statistics compare the survival of patients diagnosed with cancer with the survival of people in the general population who are the same age, race, and sex and who have not been diagnosed with cancer. The number is low, only 7.2% of total pancreatic cancer patients can survive more than 5

years. However, it is the overall 5-year surviving rate without discrimination of age, race and gender and cancer stage. If cancer stage is counted, then in figure 6 below shows the difference.

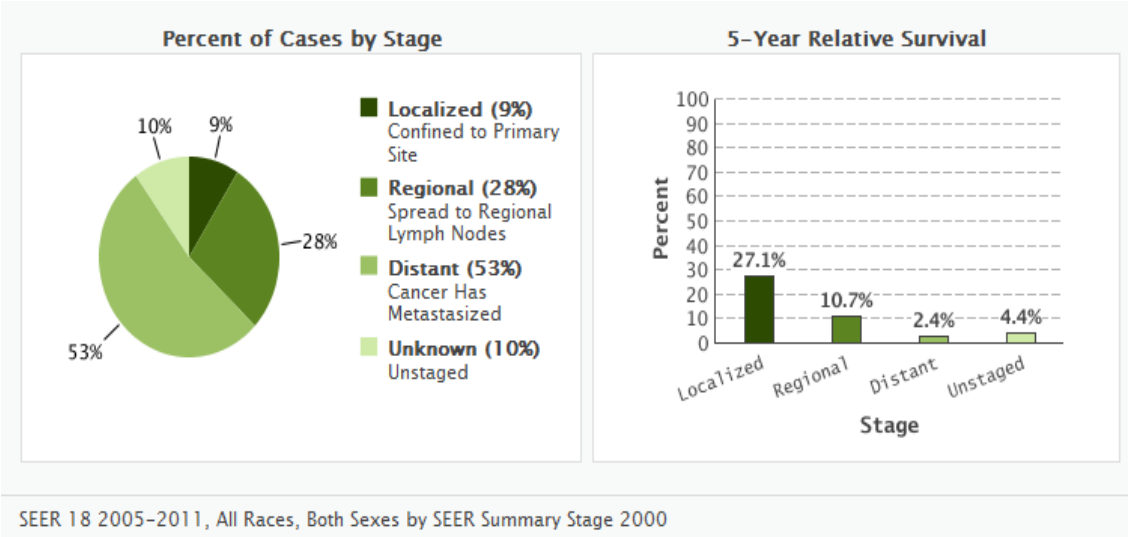


Figure 6: Percent of Cases & 5-Year Relative Survival by Stage at Diagnosis: Pancreas Cancer

Cancer stage at diagnosis, which refers to extent of a cancer in the body, determines treatment options and has a strong influence on the length of survival. In general, if the cancer is found only in the part of the body where it started it is localized (sometimes referred to as stage I). If it has spread to a different part of the body, the stage is regional or distant. The earlier pancreas cancer is caught, the better chance a person has of surviving five years after being diagnosed. For pancreas cancer, 9.0% are diagnosed at the local stage. The 5-year survival for localized pancreas cancer is 27.1% [4]. These low survival rates are attributable to the fact that fewer than 20% of patients' tumors are confined to the pancreas at the time of diagnosis; in most cases, the malignancy has already progressed to the point where surgical removal is impossible.

In those cases where resection can be performed, the average survival rate is 18 to 20 months. The overall five-year survival rate is about 10%, although this can rise as high as 20% to 25% if the tumor is removed completely and when cancer has not spread to lymph nodes [59].

	Common Types of Cancer	Estimated New Cases 2015	Estimated Deaths 2015
1.	Breast Cancer (Female)	231,840	40,290
2.	Lung and Bronchus Cancer	221,200	158,040
3.	Prostate Cancer	220,800	27,540
4.	Colon and Rectum Cancer	132,700	49,700
5.	Bladder Cancer	74,000	16,000
6.	Melanoma of the Skin	73,870	9,940
7.	Non-Hodgkin Lymphoma	71,850	19,790
8.	Thyroid Cancer	62,450	1,950
9.	Kidney and Renal Pelvis Cancer	61,560	14,080
10.	Endometrial Cancer	54,870	10,170
	-	-	-
12.	**Pancreas Cancer**	**48,960**	**40,560**

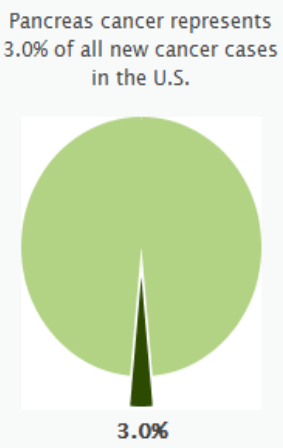


Figure 7: Estimated new cases and deaths of different cancers in 2015

In above figure 7, it shows that compare to other cancers, pancreas cancer is relatively rare, only represents 3.0% of all new cancer cases in the U.S. Pancreatic cancer is a relatively rare cancer, but it remains one of the most lethal. "It's a very complex disease; these people are hard to take care of, they have complicated medical issues," says Dr. Margaret Tempero, research director at UC San Francisco's Helen Diller Family Comprehensive Cancer Center. In 2015, it is estimated that there will be 48,960 new cases of pancreas cancer and

an estimated 40,560 people will die of this disease. From the estimation in figure 7, pancreatic cancer might become 12th instead of 11th in U.S.

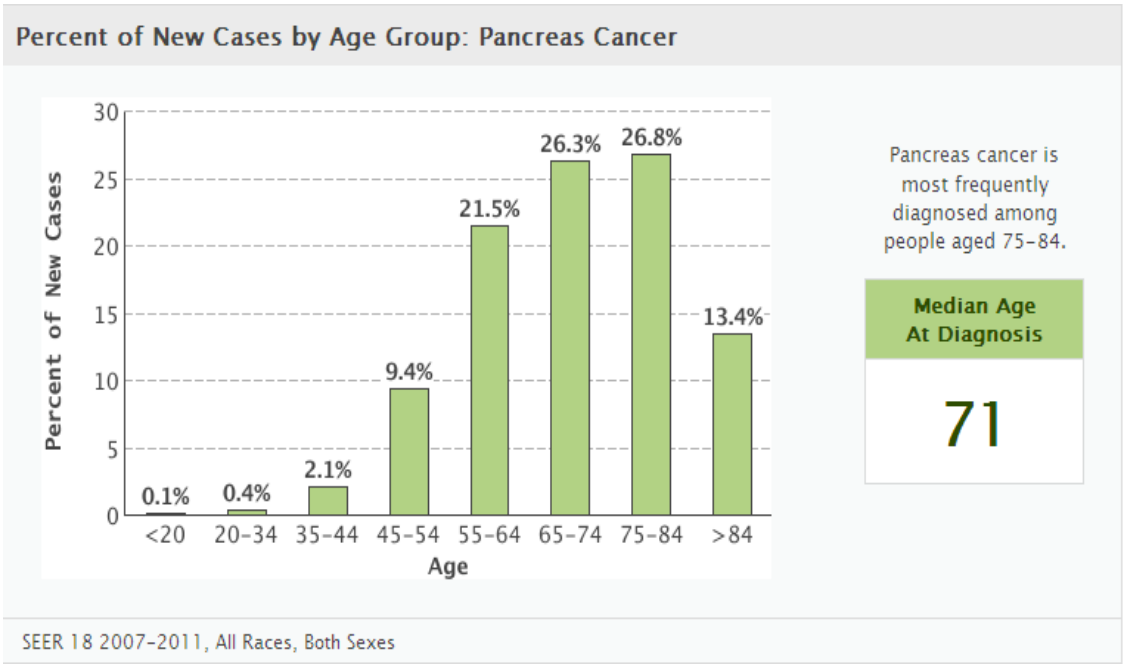


Figure 8: Percent of New Cases (Pancreatic Cancer) by Age Group

In figure 8, the risk of developing pancreatic cancer increases as people age. Almost all patients are older than 45. About two-thirds are at least 65 years old. The average age at the time of diagnosis is 71. Pancreas cancer is most frequently diagnosed among people aged 75-84.

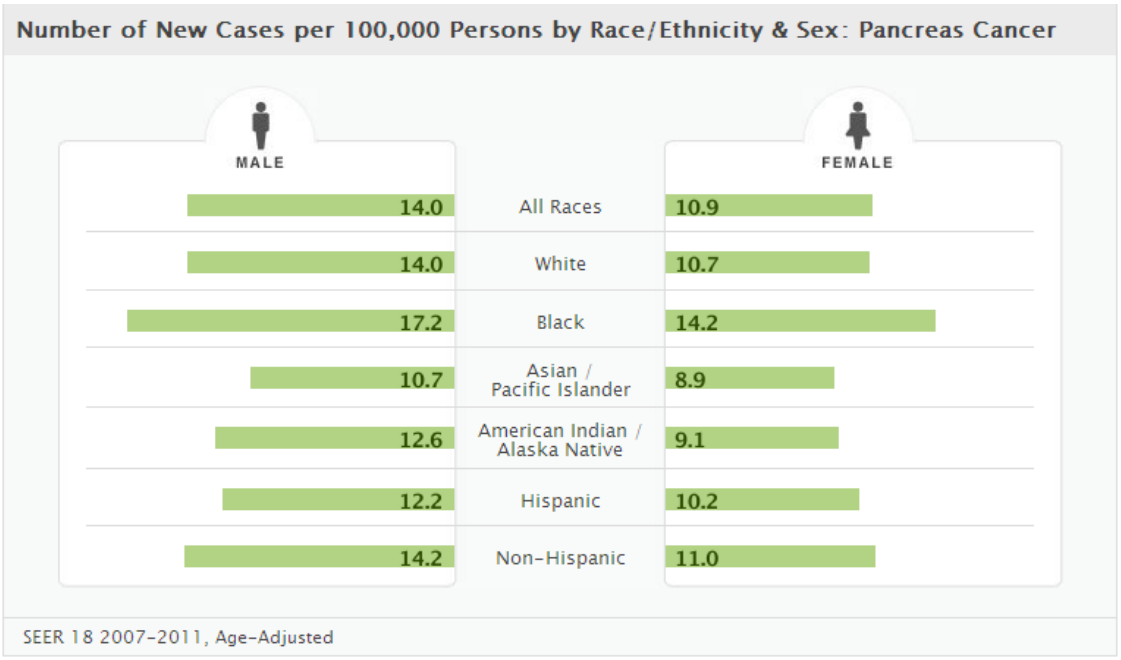


Figure 9: Number of New Cases per 100,000 Persons by Race/Ethnicity & Sex: Pancreas Cancer

Figure 9 presents that race and gender play significant role in pancreatic cancer incidence. Studies in the United States have shown that pancreatic cancer is more common in the African American population than it is in the white population. Some of this increased risk may be due to socioeconomic factors and to cigarette smoking. Cancer of the pancreas is more common in men than in women, one of the reasons is men are more likely to smoke than women. The age-adjusted incidence rates (AIR) of pancreatic cancer in the US were relatively stable between 1977 and 2001. The average age-adjusted yearly incidence of pancreatic cancer was 11.3 per 100 000 in 1977–1981 and 10.9 per 100 000 in 1997–2001. This stable temporal trend was noted across all gender and racial groups. In general, men had approximately 30% higher incidence rates than women with an overall average age-adjusted yearly incidence rate of 13.0 per

100 000 (95% CI: 12.9–13.2) among men, and 9.8 per 100 000 (95% CI: 9.7–9.9) among women. Among different racial groups, Blacks were approximately 50% more affected when compared with Whites and people of other races. The overall average age-adjusted yearly incidence rates were 16.4 per 100 000 (95% CI: 16.0–16.9) for Blacks, 10.8 per 100 000 (95% CI: 10.7–10.9) for Whites and 9.8 per 100 000 (95% CI: 9.4–10.1) for people of other races [60,61].

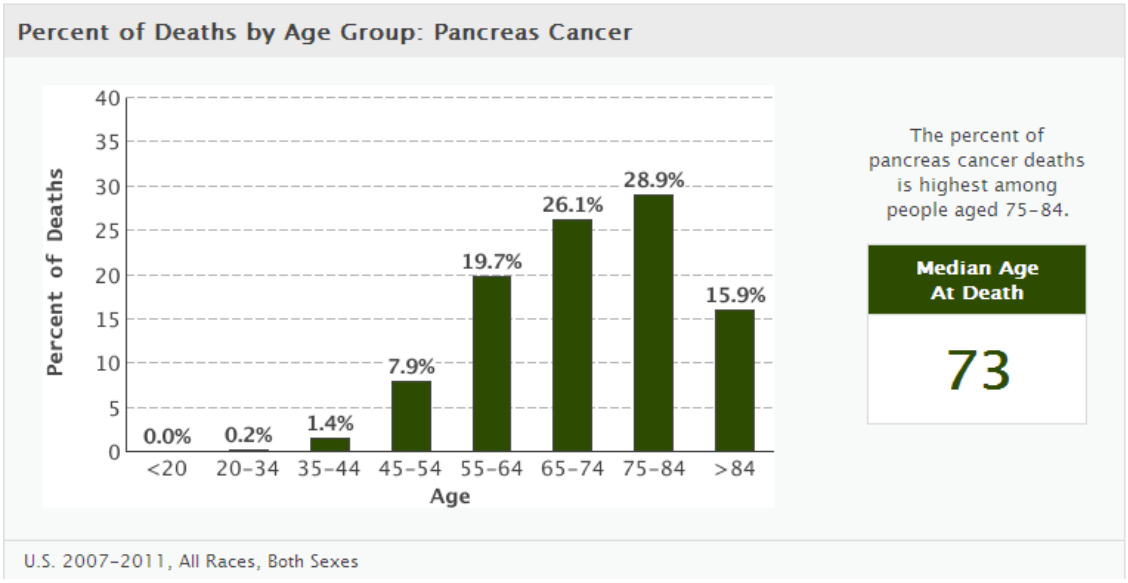


Figure 10: Percent of Deaths by Age Group: Pancreas Cancer

Figure 10 shows that median age at death is 73, and the number matches the median age of diagnosis, 71. Usually, a patient who was diagnosed as pancreatic cancer will expire within 18 months all cancer stages and the average life expectancy after diagnosis with metastatic disease is just three to six months [62]. Because survival is poor, the population distribution of people who die of pancreatic cancer is similar to that of people who are diagnosed with the disease. In part because it is difficult to detect early, the average survival time from

pancreatic cancer is low. Pancreas cancer is the third leading cause of cancer death in the United States. The number of deaths was 10.9 per 100,000 men and women per year based on 2008-2012 deaths [4].

2.2 Pancreas Cancer Stages and Treatment

The stage is evaluating the extent of the cancer in the body and how far the cancer grown into the large intestine and how far spread beyond its original location. Also, the staging of pancreas cancer is based on the result of many types of exam, which are physical exam, biopsy by endoscopy, blood test and imaging tests (CT or MRI scan, x-rays, PET scan, etc.). Staging allows for consistent treatment strategy, acceptability for entering clinical studies, and use of comparing new treatments for patients in a defined stage. The treatment of cancer depends on the type of stage, age and health status. The most common options for colon cancer are surgery, radiation therapy and chemotherapy. Stage is a term used in cancer treatment to describe the extent of the cancer's spread. The stages of pancreatic cancer are used to guide treatment and to classify patients for clinical trials. The stages of pancreatic cancer are:

- *Stage 0:* No spread. Pancreatic cancer is limited to a single layer of cells in the pancreas. The pancreatic cancer is not visible on imaging tests or even to the naked eye.
- *Stage I:* Local growth. Pancreatic cancer is limited to the pancreas, but has grown to less than 2 centimeters across (stage IA) or greater than 2 centimeters (stage IB).

- *Stage II:* Local spread. Pancreatic cancer has grown outside the pancreas, or has spread to nearby lymph nodes.
- *Stage III:* Wider spread. The tumor has expanded into nearby major blood vessels or nerves but has not metastasized.
- *Stage IV:* Confirmed spread. Pancreatic cancer has spread to distant organs.

Most of the time, the treatment of pancreatic cancer is based on its stage – how far it has spread in the body. But other factors, such as a person's overall health, can also affect treatment options. Ask doctors for any questions about the treatment plan he or she recommends. It is hard to stage pancreatic cancer accurately using imaging tests. Doctors do their best to decide before treatment whether there is a good chance the cancer is resecable – that is, if it can be removed completely. But sometimes cancers turn out to have spread farther than was first thought [5]. Although surgery remains the only curative treatment for pancreatic cancer, therapeutic strategies based on initial resection have not substantially improved the survival of patients with resectable disease over the past 25 years; presently, more than 80% of patients suffer disease relapse after resection [63].

Pancreatic cancer treatment is described as follows (Information from WebMD):

- Resectable: On imaging tests, pancreatic cancer hasn't spread (or at least not far), and a surgeon feels it might all be removable. About 10% of

pancreatic cancers are considered resectable when first diagnosed. If pancreatic cancer is resectable, surgery followed by chemotherapy or radiation or both may extend survival.

- Locally advanced (unresectable): Pancreatic cancer has grown into major blood vessels on imaging tests, so the tumor can't safely be removed by surgery.
- Metastatic: Pancreatic cancer has clearly spread to other organs, so surgery cannot remove the cancer [6].

2.3 Types of pancreatic cancer

Pancreatic cancer begins when abnormal cells within the pancreas grow out of control and form a tumor. There are two types of cells in the pancreas,

- Exocrine cells
- Endocrine cells [6]

2.3.1 *Neoplasms of the Exocrine Pancreas*

More than 95% of pancreatic cancers are classified as exocrine tumors. These tumors start in the exocrine cells that make pancreatic enzymes that help in digestion. Within this category, the vast majority of tumors are adenocarcinomas [64].

The following table2.1 is an overview of most common pancreatic exocrine cancers and their brief description.

Table 2.1: Most common pancreatic exocrine tumors [65]

TYPE	DESCRIPTION
Adenocarcinoma	Adenocarcinoma is the most common type of pancreatic cancer. It accounts for about 90% of all pancreatic cancers. It begins in the cells lining the pancreatic duct.
Intraductal Papillary-Mucinous Neoplasm (IPMN)	An IPMN is a cystic tumor that grows from the main pancreatic duct or from side branches of the duct. The tumor may appear as a finger-like (papillary) projection into the duct. An IPMN may be benign at the time of diagnosis. However, it has a risk of progressing to malignancy. This risk is high when the IPMN originates in the main pancreatic duct. An IPMN may therefore be a precursor for adenocarcinoma.
Mucinous Cystadenocarcinoma	Mucinous cystadenocarcinoma is a rare, malignant, cystic tumor. The cyst is filled with a thick fluid called mucin. It is similar to an IPMN but occurs in just one area of the pancreas, more commonly in the tail of the pancreas. These tumors are mostly seen in women.

2.3.2 *Neoplasms of the Endocrine Pancreas*

Pancreatic neuroendocrine tumors (pancreatic NETs or PNETs) account for less than 5% of all pancreatic tumors. They may be benign or malignant and they tend to grow slower than exocrine tumors [66]. These neoplasms are far less common than the exocrine neoplasms listed above. Neuroendocrine neoplasms (also known as endocrine or islet cell tumors) account for about 1-5% of pancreatic cancers. It is very important that endocrine neoplasms be distinguished from nonendocrine because the symptoms and the treatments for the two neoplasms are very different. The neuroendocrine neoplasms may produce highly active hormones and therefore have very dramatic symptoms [7]. Functional neuroendocrine tumors cause the pancreas to overproduce hormones consequently causing hormone-related symptoms. The majority of PNETs are

nonfunctional tumors. Nonfunctional tumors do not produce any hormones so they do not cause any hormone-related symptoms. As a result, these tumors are typically diagnosed once the tumor is advanced and is causing symptoms such as pain or jaundice [66].

In table 2.2, it lists the different types of pancreatic neuroendocrine tumors classified by the hormones they make.

Table 2.2: Different types of pancreatic neuroendocrine tumors classified by the hormones that they produce

Types	Description
Gastrinoma (Zollinger-Ellison Syndrome)	Gastrinomas produce gastrin. When this tumor is inherited as part of a genetic syndrome called Multiple Endocrine Neoplasia Type 1 (MEN1) (see below), multiple tumors may be found in the head of the pancreas and/or the duodenum. They have a very high potential to become malignant.
Glucagonoma	Glucagonomas produce glucagon. They are commonly found in the body and tail of the pancreas. They are usually large, often metastasize and have a very high potential to become malignant.
Insulinoma	Insulinomas produce insulin. They are the most common type of functional pancreatic neuroendocrine tumor. They tend to be small and hard to diagnose. Most of them are benign.
Somatostatinoma	Somatostatinomas produce somatostatin. They are extremely rare and usually very large. They can occur anywhere in the pancreas and in the duodenum. They have a very high potential to become malignant.
VIPoma (Verner-Morrison Syndrome)	VIPomas produce vasoactive intestinal peptide (VIP). Two-thirds of VIPomas are found in women. The syndrome is also known as Watery Diarrhea and Hypokalemia Achlorhydria (WDHA) Syndrome. They have a high potential to become malignant.
Nonfunctional Islet Cell Tumor	Nonfunctional islet cell tumors are usually malignant. They are hard to detect.

2.4 Risk Factors for Pancreatic Cancer

A risk factor is anything that affects your chance of getting a disease such as cancer. Different cancers have different risk factors. Some risk factors, like

smoking, can be changed. Others, like a person's age or family history, can't be changed. But having a risk factor, or even several risk factors, does not mean that you will get the disease. And many people who get the disease may have few or no known risk factors. Researchers have found several factors that can affect a person's chance of getting cancer of the pancreas. Most of these are risk factors for exocrine pancreatic cancer.

Basically there are 3 types of risk factors for pancreatic Cancer

- Risk factors that can be changed
- Risk factors that can't be changed
- Risk factors with unclear effects

2.4.1 *Risk factors that can be changed*

- Tobacco use: Smoking is one of the most important risk factors for pancreatic cancer. The risk of getting pancreatic cancer is about twice as high among smokers compared to those who have never smoked. Scientists think this may be due to cancer-causing chemicals in cigarette smoke that enter the blood and damage the pancreas. About 20% to 30% of exocrine pancreatic cancer cases are thought to be caused by cigarette smoking. Cigar and pipe smoking also increase risk, as does the use of smokeless tobacco products [8].
- Overweight and obesity : Being overweight is a risk factor for pancreatic cancer. Very overweight (obese) people are about 20% more likely to develop pancreatic cancer.

- Carrying extra weight around the waistline may be a risk factor even in people who are not very overweight [8].
- Workplace exposure to certain chemicals: Heavy exposure at work to certain pesticides, dyes, and chemicals used in metal refining may increase the risk of developing pancreatic cancer [8].
- Religious background: Pancreatic cancer is proportionally more common in Ashkenazi Jews than the rest of the population. This may be because of a particular inherited mutation in the breast cancer gene (BRCA2) which runs in some Ashkenazi Jewish families [27].

2.4.2 Risk *factors that can't be changed*

- Age: The risk of developing pancreatic cancer increases as people age. Almost all patients are older than 45. About two-thirds are at least 65 years old. The average age at the time of diagnosis is 71 [8].

Table 2.3: Pancreatic Cancer, Incidence Rates per 100,000 Populations, by Age, Males, 2000-11 [12]

Age Range	2000 - 2002	2001 - 2003	2002 - 2004	2003 - 2005	2004 - 2006	2005 - 2007	2006 - 2008	2007 - 2009	2008 - 2010	2009 - 2011
0-49	0.9	0.9	0.8	0.8	0.8	0.8	0.8	0.9	0.9	0.9
50-59	12.8	12.6	12.9	13.4	13.6	13.7	13.5	13.8	13.8	13.6
60-69	33.1	34.1	34	34.5	34.3	34.8	35.6	35.8	35.8	35.2
70-79	63.8	61.6	64.1	64.1	65.3	64.1	64.2	65.2	66.2	66.7
80+	98.9	96.4	96.7	95.6	98.8	96.6	98.7	97.6	101.2	101.9

Table 2.4: Pancreatic Cancer, Inci. Rates per 100,000 Populations, by Age, Females, 2000-2011 [12]

Age Range	2000 - 2002	2001 - 2003	2002 - 2004	2003 - 2005	2004 - 2006	2005 - 2007	2006 - 2008	2007 - 2009	2008 - 2010	2009 - 2011
0-49	0.6	0.6	0.6	0.5	0.6	0.6	0.7	0.6	0.7	0.7
50-59	8.7	8.9	9.4	9.8	9.8	9.3	9.1	9.3	9.3	9.7
60-69	24	24.1	25.1	26.2	26.8	26.8	26.7	26.4	27.1	27.2
70-79	49.8	49.5	50	51.6	52.4	53.8	54	55.3	55.4	55.5
80+	81.8	83.6	83.2	84.5	85.6	85.7	86.2	86.6	87.5	87.5

- Gender: Men are about 30% more likely to develop pancreatic cancer than women. This may be due, at least in part, to higher tobacco use in men, which raises pancreatic cancer risk (see above). The difference in pancreatic cancer risk was more pronounced in the past (when tobacco use was much more common among men than women), but the gap has closed in recent years [8].
- Race: African Americans are more likely to develop pancreatic cancer than whites. The reasons for this are not clear, but it may be due in part to having higher rates of other risk factors for pancreatic cancer, such as diabetes, smoking in men, and being overweight in women [8].
- Family history: Pancreatic cancer seems to run in some families. In some of these families, the high risk is due to an inherited syndrome (explained below). In other families, the gene causing the increased risk is not known [8].
- Genetic syndromes: Inherited gene changes (mutations) can be passed from parent to child. These abnormal genes may cause as many as 10%

of pancreatic cancers and can cause other problems as well. Examples of the genetic syndromes that can cause exocrine pancreatic cancer include:

- Hereditary breast and ovarian cancer syndrome, caused by mutations in the gene *BRCA2*
- Familial melanoma, caused by mutations in the gene *p16/CDKN2A*
- Familial pancreatitis, caused by mutations in the gene *PRSS1*
- Hereditary non-polyposis colorectal cancer (HNPCC), also known as *Lynch syndrome*, most often caused by a defect in the genes *MLH1* or *MSH2*. Changes in other genes can also cause HNPCC, such as *MLH3, MSH6, TGBR2, PMS1*, and *PMS2.*
- Peutz-Jeghers syndrome (PJS), caused by defects in the gene *STK11*. This syndrome is also linked with polyps in the digestive tract and several other cancers.
- Von Hippel-Lindau syndrome, caused by mutations in the gene *VHL*. It can lead to an increased risk of pancreatic cancer and carcinoma of the ampulla of Vater [8].
- Pancreatic neuroendocrine tumors and cancers can also be caused by genetic syndromes, such as:
- Neurofibromatosis, type 1, which is caused by mutations in the gene *NF1*. This syndrome leads to an increased risk of many tumors, including somatostatinomas.
- Multiple endocrine neoplasia, type I (MEN1), caused by mutations in the gene *MEN1*. This syndrome leads to an increased risk of

tumors of the parathyroid gland, the pituitary gland, and the islet cells of the pancreas [8].

- Diabetes: Pancreatic cancer is more common in people who have diabetes. The reason for this is not known. Most of the risk is found in people with type 2 diabetes. This type of diabetes most often starts in adulthood and is often related to being overweight or obese. It's not clear if people with type 1 (juvenile) diabetes have a higher risk. In some people, though, the cancer seems to have caused the diabetes (not the other way around). This can happen when cancer spreads through the pancreas and damages enough of the insulin-making cells to cause diabetes [8].
- Chronic pancreatitis: Chronic pancreatitis is a long-term inflammation of the pancreas. This condition is linked with an increased risk of pancreatic cancer (especially in smokers), but most people with pancreatitis never develop pancreatic cancer. A small number of cases of chronic pancreatitis are due to an inherited gene mutation. People with this inherited (familial) form of pancreatitis have a high lifetime risk for developing pancreatic cancer [8].
- Cirrhosis of the liver: Cirrhosis is a scarring of the liver. It develops in people with liver damage from things like hepatitis and alcohol use. People with cirrhosis seem to have an increased risk of pancreatic cancer [8].

- Stomach problems: Infection of the stomach with the ulcer-causing bacteria *Helicobacter pylori* (*H. pylori*) may increase the risk of getting pancreatic cancer. Some researchers believe that excess stomach acid might also increase the risk [8].

2.4.3 *Risk factors with unclear effects*

- Diet: Some studies linked pancreatic cancer and diets that include a lot of red meat, pork, and processed meat (such as sausage and bacon). Others have found that diets high in fruits and vegetables may help reduce the risk of pancreatic cancer. But not all studies have found such links, and the exact role of diet in relation to pancreatic cancer is still being studied.
- Physical inactivity: Some research has suggested that lack of physical activity might increase pancreatic cancer risk. But not all studies have found this.
- Coffee : Some older studies have suggested that drinking coffee might increase the risk of pancreatic cancer, but more recent studies have not confirmed this.
- Alcohol : Some studies have shown a link between heavy alcohol intake and pancreatic cancer. This link is still not certain, but heavy alcohol use can lead to conditions such as chronic pancreatitis and cirrhosis, which are known to increase pancreatic cancer risk [8].

2.5 Symptoms or Early Warning Signs of Pancreatic Cancer

The average general health article on pancreatic cancer states flatly that this type of cancer shows no early symptoms. Most people diagnosed with pancreatic cancer are already in the advanced stage of the disease by the time it's caught, and the typical prognosis is death within one years. Only 4 percent of pancreatic cancer patients live beyond five years. Although they are bunch of early warning signs or symptoms for pancreatic cancer, most of people might ignore these signs and not seek immediate medical attention. In other hand, even patients seek medical service immediately, healthcare providers might not diagnose the disease correctly.

Some signs and symptoms associated with pancreatic cancer are namely;

- *Diabetes, especially if it comes on suddenly***.** Recently, the Mayo Clinic published startling research showing that 40 percent of pancreatic cancer patients had been diagnosed with diabetes one to two years before discovering they had a pancreatic tumor. Researchers believe the diabetes is caused by tumors that simply haven't been detected yet. The problem is, diabetes is very common, and the majority of diabetes isn't pancreatic cancer, so doctors are trying to develop screening tools to tell the difference. Diabetes is a condition in which the body does not make or properly use a pancreatic hormone called insulin. Insulin helps the body use glucose (sugar) efficiently. Normally, insulin allows glucose to enter cells and be used for energy. In the case of diabetes, either the body does not produce enough insulin or the amount that is produced is not fully

effective. Research studies suggest that new-onset diabetes in people over the age of 50 may be an early symptom of pancreatic cancer. A sudden change in blood sugar levels in diabetics who previously had well-controlled diabetes may also be a sign of pancreatic cancer [68].

- *Yellowing of the eyes or skin (Jaundice).* Even a small pancreatic tumor can block the bile duct in the head of the pancreas, causing bile to build up. This causes jaundice. Cancers that start in the head of the pancreas are near the common bile duct. These cancers can press on the duct and cause jaundice while they are still fairly small, which may allow these tumors to be found at an early stage. But cancers that start in the body or tail of the pancreas don't press on the duct until they have spread through the pancreas. By this time, the cancer has often spread beyond the pancreas as well. When pancreatic cancer spreads, it often goes to the liver. This can also lead to jaundice [69].
- *Itchy skin, palms, and soles of feet.* A little-known side-effect of jaundice is itchy hands and feet. It's due to a skin reaction to the bilirubin, the yellowish brown liver chemical that causes jaundice [69].
- Lack of appetite. Changes in appetite are common with cancer and cancer treatment. People with poor appetite or appetite loss may eat less than usual, not feel hungry at all, or feel full after eating only a small amount. Ongoing appetite loss may lead to weight loss, not getting the nutrients from food that the body needs, and loss of muscle mass and strength, all of which are serious complication [70]. An Italian study found that six to

eight months before being diagnosed with pancreatic tumors, patients reported a sudden drop in their appetite and a tendency to feel full after eating very little [71].

- *Changes in taste.* In the same Italian study, some of the patients surveyed said they'd suddenly lost their taste for coffee, wine, and smoking. In fact, they said, they felt "disgust" for the smell and taste of coffee and alcohol [70].
- *Abdominal pain.* Pancreatic cancer sufferers remember this pain as a gnawing pain, rather than a sharp cramp or ache, and it radiates toward the back. A characteristic clue: the pain goes away when you lean forward [70].
- *An enlarged gall bladder.* The same blockage of the bile duct that causes jaundice can also cause an enlarged gallbladder, as the bile builds up behind the duct. The good news is that an enlarged gallbladder can be seen on imaging tests, and it may even be possible for a doctor to feel it during a physical exam [70].
- *Pale, floating, smelly stools.* If a pancreatic tumor prevents digestive enzymes from reaching the intestine, the result is an inability to digest fatty foods. So when end up with loose, smelly "floaters" as a result of the excess fat, however, in particular, can be an early clue and is too often overlooked [70].
- *Dark, tarry stools.* Bleeding in the upper intestines causes this symptom [70].

- *Sudden, unexplained weight loss.* Weight loss is not always, as many people mistakenly believe, a sign of advanced cancer that's spread to the liver. It can also happen because a lack of pancreatic enzymes is causing fat to pass through the body undigested [70].

2.6 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. Most patients stay in the hospital for one to two weeks following the Whipple surgery. Other patients are suffering from leukemia require up to a week or more of continuous chemotherapy [8].

2.7 Cost

Pancreatic cancer ranks 11th in incidence, and fifth in cancer deaths, with 29,000 affected annually. Accurate estimates of the cost of pancreatic cancer are unavailable; existing estimates are variable or not generalizable. This paper presents detailed cost estimates for pancreatic cancer by service, age, and gender. Total annual costs are $4.9 billion, (men: $3.0 billion, women: $1.9 billion). Total direct costs are $881 million, with 71% ($627.1 M/$881.3 M) for those over 65 years. Total hospital costs are 77% ($679.5 M/$881.3 M) of total direct costs. Total indirect costs are $4.0 billion, with 63% ($2,518.43 M/$4,018

M) for those 45 to 64 years. Mortality costs are $3.7 billion, 93% ($3,739 M/$4,018 M) of indirect costs. The surgical cost burden may be less than indicated previously, with most hospitalizations not including a major procedure, and average operating room costs accounting for only 9% ($1,045/$11,055) of hospital costs. Women have significantly less cancer-directed surgery than men [9].

2.8 Mortality

With improve of the medical equipment, accurate diagnosis and prevention and treatment of the major chronic disease causing the most death. In the United States, pancreatic cancer is the fourth leading cause of cancer-related death in both men and women. In 2013, more than 45,000 people in the United States will be diagnosed with pancreatic cancer and more than 38,000 will die of this disease. Because pancreatic cancer usually is diagnosed at an advanced stage, the survival rate is extremely low compared with those of many other cancer types. The pancreatic cancer mortality rate was 5.40/100,000 (males 5.88/100,000, females 4.89/100,000), ranking 6th among all cancers [10].

Table 2.5: 5 year survive rate of pancreatic cancer patient

Stage	***5-year observed survival***
Stage IA:	14%
Stage IB	12%
Stage IIA	7%
Stage IIB	5%
Stage III	3%
Stage IV	1%

In figure 11, it shows the survival rate difference between exocrine type cancer and endocrine type cancer. In the chart, exocrine type pancreatic cancer has much worse survival rate than endocrine type cancer.

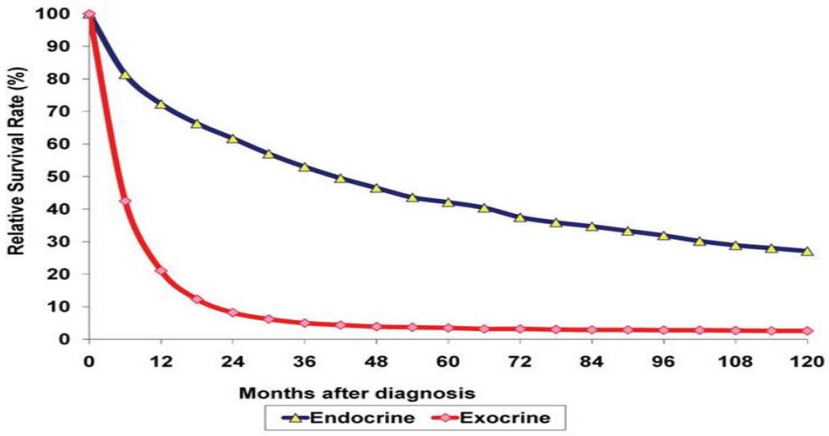


Figure 11. Cancer of the Pancreas: Relative Survival Rates (%) by Histologic Subtype [72]

2.9 Research Gap in Literature

It is well known that left time for living is essential to a cancer patient, and how to maximally extend cancer patients' life is one of the hottest research topics in nowadays. Although as detailed above much has been researched into the etiology and associations of factors with incidence and progression of pancreatic cancer, however, there is no study yet looking at a comprehensive view of the cofactors and comorbidities that might influence the incidence and the outcomes of pancreatic cancer.

Also it is not yet known how those cofactors and comorbidities will affect the outcomes of pancreatic cancer patient's lifespan 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 pancreatic cancer patients, and how demographic, hospital characteristics and comorbidities will affect lifespan of pancreatic cancer patients. 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 Three

METHODOLOGY

3.1 Research data files

The National (Nationwide) Inpatient Sample (NIS) is the largest all-payer inpatient care database in the United States, containing data on more than seven million hospital stays. Its large sample size is ideal for developing national and regional estimates and enables analyses of rare conditions, uncommon treatments, and special populations. The following links provide detailed documentation for the NIS [11].

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 2005-2009 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.

In this research study, NIS 2005-2009 data files were used as source data sets, and only pancreatic cancer patients' information were retrieved from 5 years NIS data files. The following 30 variables, table 3.1, will be involved in building logistic regression model. Those variables are associated with pancreatic cancer disease ad primary diagnosis, and dependent variable for this study is variable DIED.

Table 3.1: Data Variables Used for Analysis

Study Variables	Original Name in	Variable value notes
AGE_GROUP	AGE_GROUP	Age <=20, 1; Age>21 and <=40, 2; Age>41 and <=60, 3; Age >61 and <=80, 4, Age >80,5
FEMALE	FEMALE	Female =1 , Male=0
PAY1	PAY1	1=Medicare, 2=Medicaid, 3=Private insurance,4=Self-pay,5=No charge,6=Other, Categorical Variable
RACE	RACE	1 = White, 2 = Black, 3 = Hispanic, 4 = Asian/Pacific, 5 = Native Am., 6 = Other, Categorical Variable
ZIPInc_Qrtl	ZIPInc_Qrtl	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
Acquired immune deficiency syndrome	CM_AIDS	Diagnosed =1, Not diagnosed =0
Alcohol abuse	CM_ALCOHOL	Diagnosed =1, Not diagnosed =0
Deficiency anemias	CM_ANEMDEF	Diagnosed =1, Not diagnosed =0
Rheumatoid arthritis/collagen vascular	CM_ARTH	Diagnosed =1, Not diagnosed =0
Chronic blood loss anemia	CM_BLDLOSS	Diagnosed =1, Not diagnosed =0
Congestive heart failure	CM_CHF	Diagnosed =1, Not diagnosed =0
Chronic pulmonary disease	CM_CHRNLUNG	Diagnosed =1, Not diagnosed =0
Coagulopathy	CM_COAG	Diagnosed =1, Not diagnosed =0
Depression	CM_DEPRESS	Diagnosed =1, Not diagnosed =0
Diabetes, uncomplicated	CM_DM	Diagnosed =1, Not diagnosed =0
Diabetes with chronic complications	CM_DMCX	Diagnosed =1, Not diagnosed =0
Drug abuse	CM_DRUG	Diagnosed =1, Not diagnosed =0
Hypertension	CM_HTN_C	Diagnosed =1, Not diagnosed =0
Hypothyroidism	CM_HYPOTHY	Diagnosed =1, Not diagnosed =0
Liver disease	CM_LIVER	Diagnosed =1, Not diagnosed =0
Lymphoma	CM_LYMPH	Diagnosed =1, Not diagnosed =0
Fluid and electrolyte disorders	CM_LYTES	Diagnosed =1, Not diagnosed =0
Other neurological disorders	CM_NEURO	Diagnosed =1, Not diagnosed =0
Obesity	CM_OBESE	Diagnosed =1, Not diagnosed =0
Paralysis	CM_PARA	Diagnosed =1, Not diagnosed =0
Peripheral vascular disorders	CM_PERIVASC	Diagnosed =1, Not diagnosed =0
Psychoses	CM_PSYCH	Diagnosed =1, Not diagnosed =0
Pulmonary circulation disorders	CM_PULMCIRC	Diagnosed =1, Not diagnosed =0
Renal failure	CM_RENLFAIL	Diagnosed =1, Not diagnosed =0
Peptic ulcer disease excluding bleeding	CM_ULCER	Diagnosed =1, Not diagnosed =0
Valvular disease	CM_VALVE	Diagnosed =1, Not diagnosed =0
Weight loss	CM_WGHTLOS	Diagnosed =1, Not diagnosed =0

3.2 Goals and Objectives

The overall goal of the project is to identify how the comorbidities associated with pancreatic cancer diagnosis affect the life of pancreatic cancer patients across the United States. Specifically the objectives are to determine:

1) What clinical factors (such as number and types of comorbidities) influence the mortality of pancreatic cancer patients
2) Whether the inpatient mortality rate amongst pancreatic cancer patients differ with race, age, gender or socio –economic status
3) Whether certain types of comorbidities influence the life of pancreatic cancer patients which means certain comorbidities can either make patients live longer or live shorter.

3.3 Research Hypotheses of the project

- Are there statistically significant associations between certain types of comorbidities and mortality Pancreatic Cancer patients
- Are there statistically significant influence of comorbidities on life of Pancreatic Cancer patients

In table 3.2, it lists the 2 research hypotheses of this study, all independent variables are associated with pancreatic cancer disease, and pancreatic cancer is the primary diagnosis in NIS 2005-2009 data set. The description of independent variables lists in table 3.1.

Table 3.2: Study Hypotheses and Corresponding Statistical Tests

Research Question	Hypothesis	Independent Variables	Outcome Variable (Independent variable)	Inferential Analyses
Do Type of Comorbidities significantly affect DIED (In hospital mortality)?	Hypothesis 1	AGE_GROUP FEMALE PAY1 RACE ZIPInc_Qrtl CM_AIDS CM_ALCOHOL CM_ANEMDEF CM_ANEMDEF CM_ARTH CM_BLDLOSS CM_CHF CM_CHRNLUNG CM_COAG CM_DEPRESS CM_DM CM_DMCX CM_DRUG CM_HTN_C CM_HYPOTHY CM_LIVER CM_LYMPH CM_LYTES CM_NEURO CM_OBESE CM_PSYCH CM_PULMCIRC CM_RENLFAIL CM_ULCER CM_VALVE CM_WGHTLOSS	DIED	ROC Logistic Regression
Do certain types of comorbidities have influence on life of pancreatic cancer patients	Hypothesis 2	AGE_GROUP FEMALE PAY1 RACE ZIPInc_Qrtl CM_AIDS CM_ALCOHOL CM_ANEMDEF CM_ARTH CM_BLDLOSS CM_CHF CM_CHRNLUNG CM_COAG CM_DEPRESS CM_DM CM_DMCX CM_DRUG CM_HTN_C CM_HYPOTHY CM_LIVER CM_LYMPH CM_LYTES CM_NEURO CM_OBESE CM_PARA CM_PERIVASC CM_PSYCH CM_PULMCIRC CM_RENLFAIL CM_ULCER CM_VALVE CM_WGHTLOSS	DIED	Odds Ratios Logistic Regression

3.4 Research Design & Methods

In this project we plan to utilize the datasets obtained from the Nationwide Inpatient Sample (NIS) database towards our analyses of Pancreatic Cancer patients. The NIS is the largest all-payer inpatient care database in the United States containing data from 2005 to 2009. 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 2005 to 2009 and the statistical analysis software SAS 9.2 will be employed to extract the datasets and perform the analyses. The figure12 below illustrates the conceptual model employed in this research project.

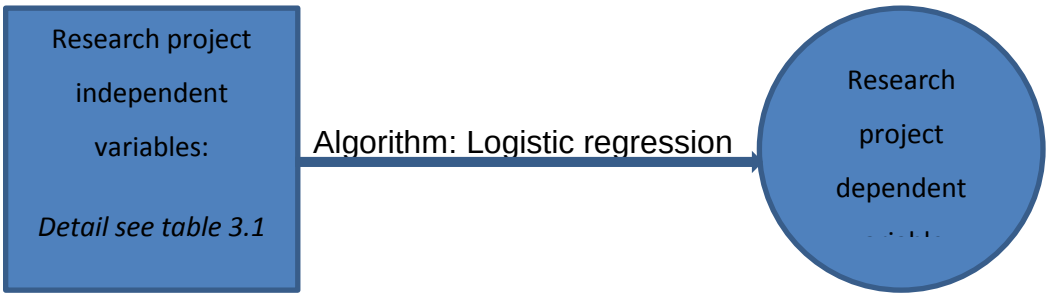


Figure 12: Conceptual model for identifying the factors determining Mortality

Essentially it shows the factors that are hypothesized to affect the research outcomes of the mortality for the pancreatic 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; demographics group delineates the race and age of the patient as also the type of insurance (Private, Medicare and the like). The outcomes of interest as identified in this proposal are the mortality. 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. Predictive logistic regression with ROC curve(Receiver Operating Characteristic) and Odd Ratios model will be setup and validated, it will be employed to determine the risks and ratios for the various comorbidities also with age groups, gender, race and socio-economic status influencing mortality of pancreatic cancer patients.

3.5 Statistical Methodology

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

Parametric methods such as:

- Logistic regression models - ROC curve(Receiver Operating Characteristic)
- Logistic regression models – Odd Ratio

3.5.1 *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 [14]. 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:

$$logit(p) = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \ldots + b_kX_k$$

where p is the probability of presence of the characteristic of interest. The logit transformation is defined as the logged odds:

$$odds = \frac{p}{1-p} = \frac{probability\ of\ presence\ of\ characteristic}{probability\ of\ absence\ of\ characteristic}$$

and

$$logit(p) = \ln\left(\frac{p}{1-p}\right)$$

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. β0 is the intercept in the regression equation, β1 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 [15].

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 [16]. The next chapter presents the results arising from conducting these statistical tests and the discussions on the results is also presented. Logistic regression

3.6 SAS data modeling used in research project

SAS (Statistical Analysis System; not to be confused with SAP) is a software suite developed by SAS Institute for advanced analytics, business intelligence, data management, and predictive analytics. SAS is a software suite that can mine, alter, manage and retrieve data from a variety of sources and perform statistical analysis on it [20]. SAS provides a graphical point-and-click user interface for non-technical users and more advanced options through the SAS programming language [20]. SAS programs have a DATA step, which retrieves and manipulates data, usually creating a SAS data set, and a PROC step, which analyzes the data [20].

3.6.1 *Logistic regression models - ROC curve (Receiver Operating Characteristic)*

A Receiver Operating Characteristic Curve (ROC) is a standard technique for summarizing classifier performance over a range of trade-offs between true positive (TP) and false positive (FP) error rates (Sweets, 1988). ROC curve is a plot of *sensitivity (the ability of the model to predict an event correctly)* versus *1-specificity* for the possible cut-off classification probability values π_0 [17].

A representation and interpretation of the area under a receiver operating characteristic (ROC) curve obtained by the “rating” method, or by mathematical predictions based on patient characteristics, is presented. It is shown that in such a setting the area represents the probability that a randomly chosen diseased subject is (correctly) rated or ranked with greater suspicion than a randomly chosen non-diseased subject. Moreover, this probability of a correct ranking is the same quantity that is estimated by the already well-studied nonparametric Wilcoxon statistic. These two relationships are exploited to (a) provide rapid closed-form expressions for the approximate magnitude of the sampling variability, *i.e.,* standard error that one uses to accompany the area under a smoothed ROC curve, *(b)* guide in determining the size of the sample required to provide a sufficiently reliable estimate of this area, and *(c)* determine how large sample sizes should be to ensure that one can statistically detect differences in the accuracy of diagnostic techniques [17].

Following is the SAS code used to obtain the ROC curve results.

```
ods graphics on;
proc logistic data=p0509 plots(only)=(roc(id=obs) effect);
```

```
    MODEL DIED(EVENT='1') =AGE_GROUP FEMALE PAY1 RACE
    ZIPInc_Qrtl CM_AIDS CM_ALCOHOL CM_ANEMDEF CM_ANEMDEF
    CM_ARTH CM_BLDLOSS CM_CHF CM_CHRNLUNG CM_COAG
    CM_DEPRESS CM_DM CM_DMCX CM_DRUG CM_HTN_C
    CM_HYPOTHY CM_LIVER CM_LYMPH CM_LYTES CM_NEURO
    CM_OBESE CM_PSYCH CM_PULMCIRC CM_RENLFAIL CM_ULCER
    CM_VALVE CM_WGHTLOSS/
    scale=none
    clparm=wald
    clodds=pl
    rsquare;
run;
ods graphics off;
```

3.6.2 *Logistic regression models – Odd Ratio*

The odds ratio (OR) is one of several statistics that have become increasingly important in clinical research and decision-making. It is particularly useful because as an effect-size statistic, it gives clear and direct information to clinicians about which treatment approach has the best odds of benefiting the patient. Significance statistics used for the OR include the Fisher's Exact Probability statistic, the Maximum-Likelihood Ratio Chi-Square and Pearson's Chi-Square [18].

The calculation of the odds ratio is quite simple. The formula is as follows:

$$\text{Odds ratio} = \frac{PG_1 / (1 - PG_1)}{PG_2 / (1 - PG_2)}$$

Where "PG_1" represents the odds of the event of interest for Group 1, and "PG_2" represents the odds of the event of interest for Group 2 [19].

Following is the SAS code used to obtain the odds ratio (OR) results.

```
ods output "Odds Ratios"=p0509_odds;
proc logistic data=p0509 ;
        MODEL DIED(EVENT='1') =AGE_GROUP FEMALE PAY1 RACE
        ZIPInc_Qrtl CM_AIDS CM_ALCOHOL CM_ANEMDEF CM_ARTH
        CM_BLDLOSS CM_CHF CM_CHRNLUNG CM_COAG CM_DEPRESS
        CM_DM CM_DMCX CM_DRUG CM_HTN_C CM_HYPOTHY CM_LIVER
        CM_LYMPH CM_LYTES CM_NEURO CM_OBESE CM_PARA
        CM_PERIVASC CM_PSYCH CM_PULMCIRC CM_RENLFAIL
        CM_ULCER CM_VALVE CM_WGHTLOSS
                / selection = stepwise;
run;
data work.p0509_odds2;
        set work.p0509_odds;
        format OddsRatioEst lowercl uppercl 5.3;
        lcl2=lowercl;
        ucl2=uppercl;
        OR='OR'; LCL='LCL'; UCL='UCL';
run;
ods html close;
data _null_;
pct=0.75/nobs;
call symputx("pct", pct);
call symputx("pct2", 2*pct);
set p0509_odds nobs=nobs;
run;
proc template;
        define Style styles.foreststyle;
        parent = styles.analysis;
        style GraphFonts  from GraphFonts /
        'GraphFootnoteFont' = ("<MTsans-serif-unicode>",10pt,italic);
```

```
        end;
        run;
        ods escapechar '^';
        ods listing sge=off image_dpi=100 style=styles.foreststyle;
        ods graphics / reset width=7in height=8in
        imagename="OddsRatioPlot_0509";
        title "Odda Ratios on Mortality by Study (Year 2005-2009)";
        title2 h=8pt 'Odds Ratio and 95% CL';
        footnote j=l "This graph is created using SAS^{unicode '00ae'x} 9.2
        SGPLOT Procedure";

proc sgplot data=work.p0509_odds2 noautolegend;
          scatter y=Effect x=OddsRatioEst /
                markerattrs=graphdata2(symbol=diamondfilled size=3);
          scatter y=Effect x=OddsRatioEst / xerrorupper=ucl2 xerrorlower=lcl2
                markerattrs=graphdata1(symbol=squarefilled size=3);
          scatter y=Effect x=or / markerchar=OddsRatioEst x2axis;
          scatter y=Effect x=lcl / markerchar=lowercl x2axis;
          scatter y=Effect x=ucl / markerchar=uppercl x2axis;
          refline 1 8  / axis=x;
          refline 0.1 1 / axis=x lineattrs=(pattern=shortdash) transparency=0.5;
          inset '              Live'  / position=bottomleft;
          inset 'Die'  / position=bottom;
          xaxis type=log offsetmin=0 offsetmax=0.35 min=0.1 max=8
               minor display=(nolabel) ;
          x2axis offsetmin=0.7 offsetmax=0.05 display=(noticks nolabel);
          yaxis display=(noticks nolabel) offsetmin=&pct2 offsetmax=&pct;
  run;
title;
```

3.7 Data Analyses

The data obtained from the NIS is year 2005-2009. These data were received in the form of a compressed American Standard Code for Information Interchange (ASCII) data. These data were uploaded into SAS 9.2 and converted into an active SAS data file using the corresponding syntax. The active data set was then saved as an Excel file and a SAS file in order to be used as needed in those systems. The variables analyzed were selected based on the information of already analyzed data as reported by ACS and SEER, and the review of literature. The variables are namely;

AGE_GROUP,FEMALE,PAY1,RACE,ZIPInc_Qrtl,acquired immune deficiency syndrome, alcohol abuse, deficiency anemias, rheumatoid arthritis/collagen vascular diseases, chronic blood loss anemia, congestive heart failure, chronic pulmonary disease, coagulopathy, depression, diabetes uncomplicated, diabetes with chronic complications, drug abuse, hypertension (combine uncomplicated and complicated), hypothyroidism, liver disease, lymphoma, fluid and electrolyte disorders, metastatic cancer, other neurological disorders, obesity, paralysis, peripheral vascular disorders, psychoses, pulmonary circulation disorders, renal failure, peptic ulcer disease excluding bleeding, valvular disease, weight loss for pancreatic cancer patients. Pancreatic cancer was identified using the Clinical Classifications Software (CCS) for the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM). After identification of pancreatic cancer disease code, namely; 1570, 1571, 1572, 1573, 1574, 1578 and 1579,

the data was reduced to include only relevant information pertaining to the interested variables and to minimize the processing time of the software output. Complex sampling of frequency distribution, descriptive statistics and measures of association were analyses performed. The output and graphs will be shown in the result chapter.

Chapter IV

RESULTS

4.1 Overview

This chapter will provide the findings and interpretation of the results. The results presented are graphical representations of the findings, images of data and tables. The data were derived from complex sampling of univariate statistics carried out to show the frequency distribution of race and gender of inpatients with pancreatic cancer for calendar year 2005-2009 of the NIS dataset, descriptive statistics for age of inpatients with pancreatic cancer and measures of association between pancreatic cancer and some health conditions. There are 2 core research project results 1) prediction of in hospital mortality of pancreatic cancer patient (ROC curve) 2) prediction of influence of commodities of pancreatic cancer patients life.

4.2 Expectation

A promising combination of one technique of state of the art classifier and a sample data of 36,510 patients' record, this study expect to help in identifying some the comorbidities of pancreatic cancer and use them as possible as prediction factors for identifying their influence on pancreatic cancer patient life. Those comorbidities are

AGE_GROUP,FEMALE,PAY1,RACE,ZIPInc_Qrtl,acquired immune deficiency syndrome, alcohol abuse, deficiency anemias, rheumatoid arthritis/collagen vascular diseases, chronic blood loss anemia, congestive heart failure, chronic pulmonary disease, coagulopathy, depression, diabetes uncomplicated, diabetes with chronic complications, drug abuse, hypertension (combine uncomplicated and complicated), hypothyroidism, liver disease, lymphoma, fluid and electrolyte disorders, metastatic cancer, other neurological disorders, obesity, paralysis, peripheral vascular disorders, psychoses, pulmonary circulation disorders, renal failure, peptic ulcer disease excluding bleeding, valvular disease, weight loss. Although, this study results will be cross verified by other 2 cancers: a) Breast cancer b) Stomach cancer.

4.3 Preliminary Results

4.3.1 Dataset for Study

The main data are derived from a larger dataset of The Nationwide Inpatient Sample (NIS) through the year 2005-2009. In NIS database set, we used 1) CORE File 2) Severity File.

We used following ICD-9 code, table 4.1, to retrieve pancreatic cancer patient with primary diagnosis only from NIS year 2005-2009 CORE file and Severity file using SAS 9.2 Enterprise Guide. Total 36,510 patients' records were obtained from query process.

Table 4.1: ICD-9 for pancreatic cancer

157.0	Malignant neoplasm of head of pancreas
157.1	Malignant neoplasm of body of pancreas
157.2	Malignant neoplasm of tail of pancreas
157.3	Malignant neoplasm of pancreatic duct
157.4	Malignant neoplasm of islets of langerhans
157.8	Malignant neoplasm of other specified sites of pancreas
157.9	Malignant neoplasm of pancreas, part unspecified

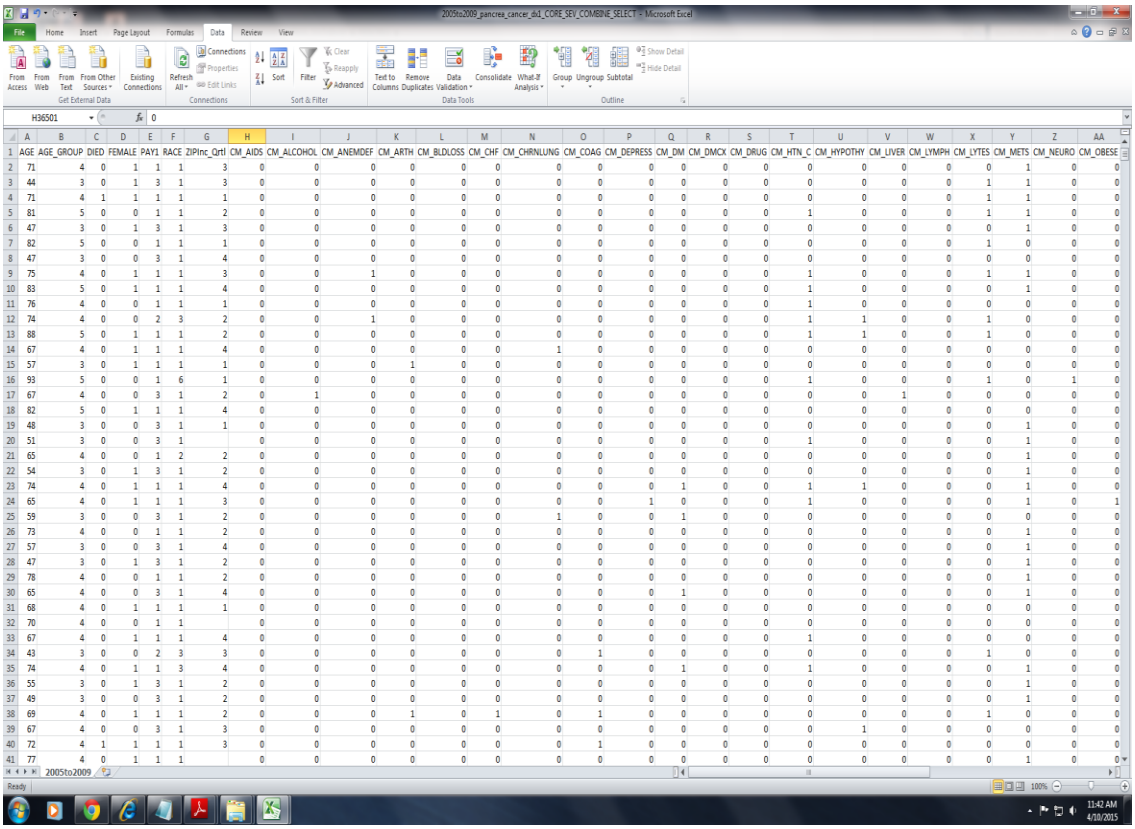


Figure 13: Screen shot of final study database

In above figure 13, it is a screen shot of the final study database we do research on. All comorbidities of pancreatic cancer has been covert to 0 and 1, see table 4 for the detail variables used in this research study.

In below table 4.2 (data from HCUPnet), it shows the admitting source of pancreatic cancer patients from 2000-2006. In the table, the number of

pancreatic cancer patients from emergency room has been increasing steadily every year, it can be interpreted as early diagnosis of pancreatic cancer is rare and symptoms of disease is not obvious. Usually patients will be diagnosed as pancreatic cancer somewhat late when they feel something wrong. In the statistic table 4.2, it indicates that 43% of pancreatic cancer patients were admitted from emergency room which means patients found something odd and rush to ER room for medical service. The sample data of admitted from other hospital and from long term care are too small, so there is no statistic value for those data.

Table 4.2: Pancreatic Cancer Patients' Admitting Source in NIS data file, Data from HCUP NIS Statistics [22]

	2000	2001	2002	2003	2004	2005	2006
Total number of discharges	30,583	30,656	31,046	30,979	32,217	33,603	33,420
Admitted from emergency department	10,479 34.27%	11,201 36.54%	12,549 40.42%	12,368 39.92%	13,615 42.26%	13,899 41.36%	14,363 42.98%
Admitted from other hospital	1,213 3.97%	1,155 3.77%	943 3.04%	1,246 4.02%	1,259 3.91%	1,373 4.08%	1,561 4.67%
Admitted from long term care	559 1.83%	529 1.72%	430 1.39%	410 1.32%	443 1.37%	323 0.96%	348 1.04%

4.3.2 Descriptive Statistics

The data below are existing analyzed data reported by the American cancer society and the surveillance, epidemiology, and end result program. Figure 11 shows the incidence rate of pancreatic cancer by age & sex and by age & race for calendar year 2005-2009. According to the data, the incidence rate per 100,000 persons is higher in males than females and higher in older individuals.

The second data shows a higher incidence amongst African Americans than Whites. This data was adjusted to the 2000 United States standard population.

4.3.2.1 A Review of Existing Data

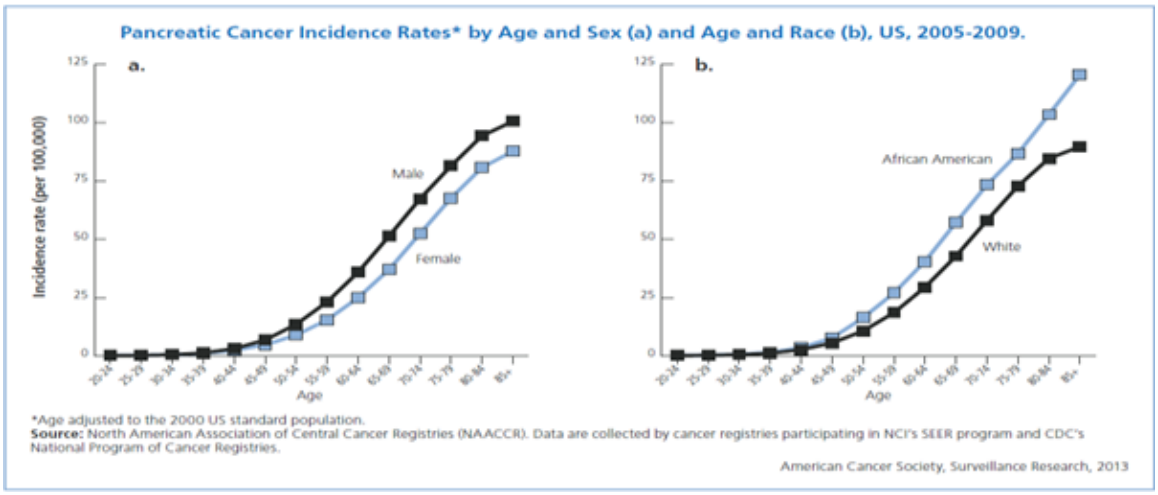


Figure 14: Pancreatic Cancer Incidence Rate. Data from American Cancer Society [21]

The data in figure14 shows the rate of incidence for pancreatic cancer is higher in males than females and this may be a correlation with the smoking habits of males and females. It has been known that smoking is more common in males than females; pancreatic cancer is about 30% more common in men than in women [7]. The data to the right shows that incidence of pancreatic cancer is higher in African Americans than Whites. Beside genetic reason between genders and races, this scenario may be related to the type of diet and their living style in the different groups, or certain disease like obesity might contribute to the results also. From the previous research study, obesity is higher in African Americans than Whites. The two graphs also indicate the rate of incidence of

pancreatic cancer increases with age, the likelihood of developing pancreatic cancer in the next 10 years is about four times higher at age 70 than at age 50 [23].

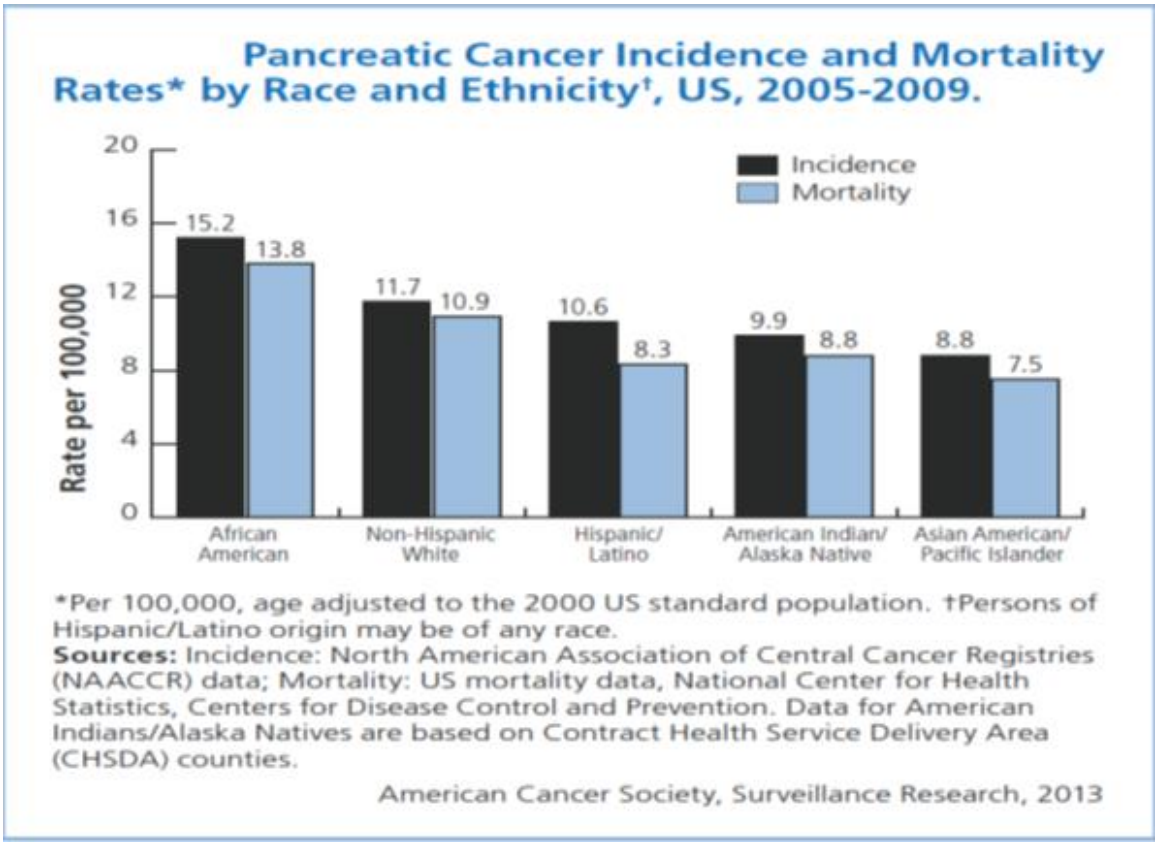


Figure 15: Pancreatic Cancer Incidence and Mortality. Data from American Cancer Society [21]

Figure 15 is another data showing the incidence and mortality of pancreatic cancer by race and ethnicity for calendar year 2005-2009. The rate per 100,000 people in the United States shows that African Americans have the highest incidence and mortality than any other race or ethnic group. In comparing the data from the American Cancer Society with those from SEER, a consistent

report was observed. From the figure 15, it shows that the mortality in pancreatic cancer is really high.

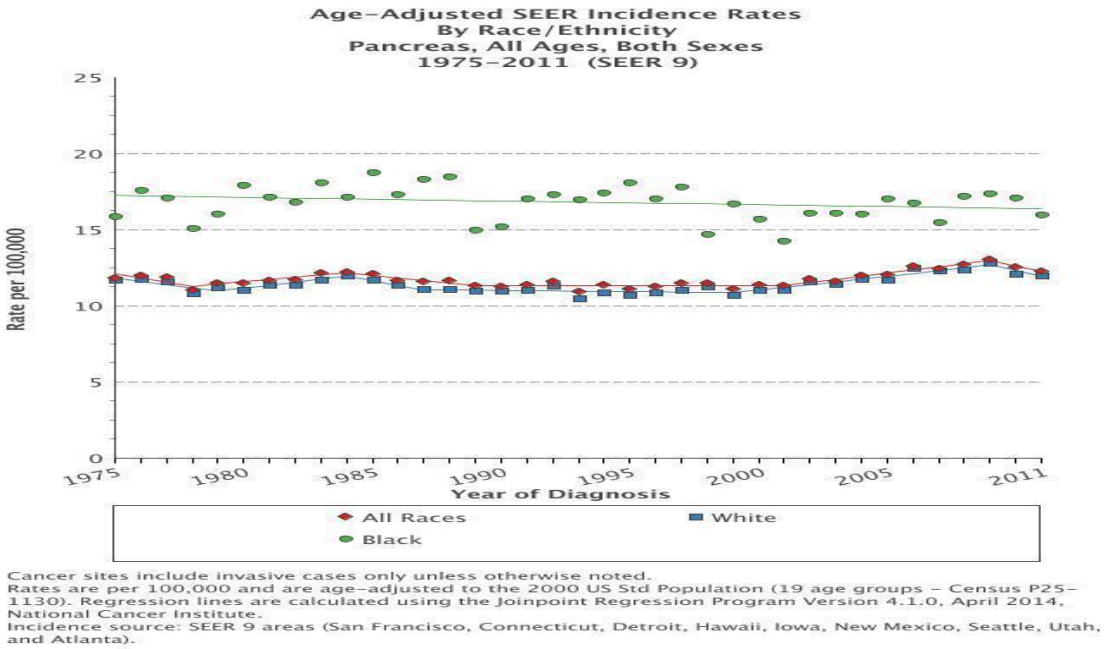


Figure 16: New Cases of Pancreatic Cancer by Race/Ethnicity. Data from SEER, NCI at NIH

Figure 16 is a data from SEER showing the percent of new cases of pancreatic cancer by age. From the plot, black group has significantly higher chance to develop in pancreatic cancer than other race groups. However, the trend of number of the black group in pancreatic cancer incident has a slightly going down while other race groups has a trend of number of incident going up.

4.3.2.2 Analyses of HCUP NIS Data

The results below are from the Nationwide Inpatient Sample (NIS) from 2005-2009, the data was retrieved base on ICD 9 code 1570-1579 from CORE and

SEVERITY data file (Pancreatic Cancer only). We use SAS 9.2 application to describe following variables in Table 4.3.

Table 4.3: Describe variables

Study Variables	Original Variable Name in the NIS Data Set	Variable value	Description
AGE_GROUP	AGE_GROUP	1,2,3,4,5	Age <=20, 1; Age>21 and <=40, 2; Age>41 and <=60, 3; Age >61 and <=80, 4, Age >80,5
FEMALE	FEMALE	1,0	Female =1 , Male=0
PAY1	PAY1	1,2,3,4,5,6	1=Medicare, 2=Medicaid, 3=Private insurance,4=Self-pay,5=No charge,6=Other, Categorical Variable
RACE	RACE	1,2,3,4,5,6	1 = White, 2 = Black, 3 = Hispanic, 4 = Asian/Pacific, 5 = Native Am., 6 = Other, Categorical Variable
ZIPInc_Qrtl	ZIPInc_Qrtl	1,2,3,4	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

In below table 4.4, we have mean values of all 5 study variables.

- In age group variable (AGE_GROUP), the mean values of pancreatic cancer patients is 3.89, which is located in age 60-80 range and is literally supported by both statistics from American Cancer Society, “The risk of developing pancreatic cancer increases as people age. Almost all patients are older than 45. About two-thirds are at least 65 years old. The average age at the time of diagnosis is 71.” [25] See figure 14 for the result from this study.
- In gender variable (FEMALE), the mean value is 0.509, male and female has a very close possibility to develop into pancreatic cancer. This result matches the statistic from American Cancer Society, “Men are about 30% more likely to develop pancreatic cancer than women. The difference in

pancreatic cancer risk was more pronounced in the past (when tobacco use was much more common among men than women), but the gap has closed in recent years."[25]

- In primary insurance payment variable (PAY1), the mean value 1.915 shows that most of pancreatic cancer patients are using Medicare and Medicare which means majority of pancreatic cancer patients either aged or low income, see figure 20.
- In race variable (RACE), the mean value 1.522 presents the most of pancreatic cancer patients are either white or black, see figure 21. According to American Cancer Society, African Americans are more likely to develop pancreatic cancer than whites. The reasons for this are not clear, but it may be due in part to having higher rates of other risk factors for pancreatic cancer, such as diabetes, smoking in men, and being overweight in women [23].
- In socio-economic variable (ZIPInc_Qrtl), the mean value is 2.5 which shows that income is not a significant factor to pancreatic cancer, see figure 22. However from the result in John Hopkins Medicine, "Low income is associated with an increased risk of pancreatic cancer for several reasons. First, individuals who earn lower incomes tend to smoke more and smoking doubles the risk of pancreatic cancer. "[23] The possible reason for this difference is that the variable ZIPInc_Qrtl categorical is setup too low which means 1, 2, 3 all under $45,000, that income should be ranged into low income.

Table 4.4: Summary Statistics on NIS 2005-2009 Pancreatic Cancer Patients Age group, Gender, Insurance, Race and Income status Results

The MEANS Procedure

Variable	Mean	Std Dev	Minimum	Maximum	N
AGE_GROUP	3.8908488	0.7166116	1.0000000	5.0000000	36509
FEMALE	0.5098673	0.4999095	0	1.0000000	36484
PAY1	1.9157063	1.1913380	1.0000000	6.0000000	36444
RACE	1.5221872	1.0968786	1.0000000	6.0000000	28530
ZIPInc_Qrtl	2.5075556	1.1192898	1.0000000	4.0000000	35669

Generated by the SAS System ('Local', X64_7PRO) on April 12, 2015 at 7:57:58 AM

In figure 17, the distribution curve on pancreatic cancer patients shows the

median age at diagnosis of pancreatic cancer was 69 years of age. This means that about half of all patients developed this disease when they are older than age 69.

Distribution analysis of: AGE

Basic Statistical Measures			
Location		Variability	
Mean	68.23139	Std Deviation	12.89842
Median	69.00000	Variance	166.36936
Mode	77.00000	Range	105.00000
		Interquartile Range	19.00000

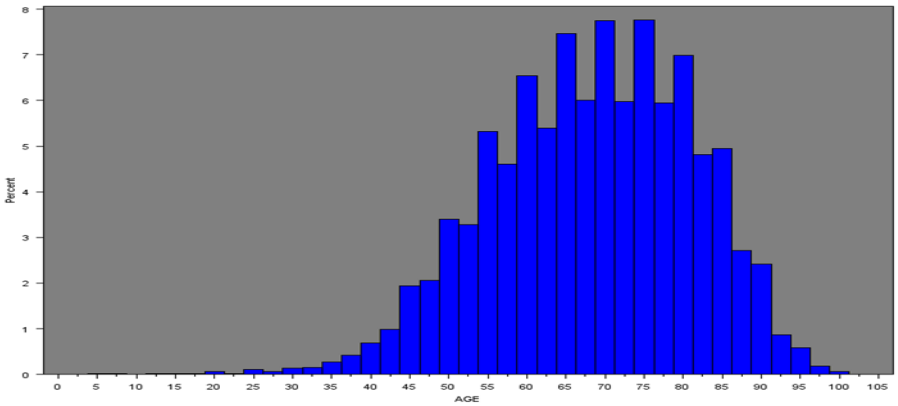
Basic Confidence Limits Assuming Normality			
Parameter	Estimate	95% Confidence Limits	
Mean	68.23139	68.09906	68.36372
Std Deviation	12.89842	12.80554	12.99268
Variance	166.36936	163.98176	168.80973

Tests for Location: Mu0=0			
Test		Statistic	p Value
Student's t	t	1010.648	Pr > |t| <.0001
Sign	M	18250	Pr >= |M| <.0001
Signed Rank	S	3.3307E8	Pr >= |S| <.0001

Missing Values			
Missing Value	Count	Percent Of All Obs	Percent Of Missing Obs
.	8	0.02	100.00

Distribution analysis of: AGE

The UNIVARIATE Procedure



Generated by the SAS System ('Local', X64_7PRO) on April 13, 2015 at 9:53:44 AM

Figure 17: Distribution analysis of Age

In figure 18, it shows pancreatic cancer incidence most often happens on

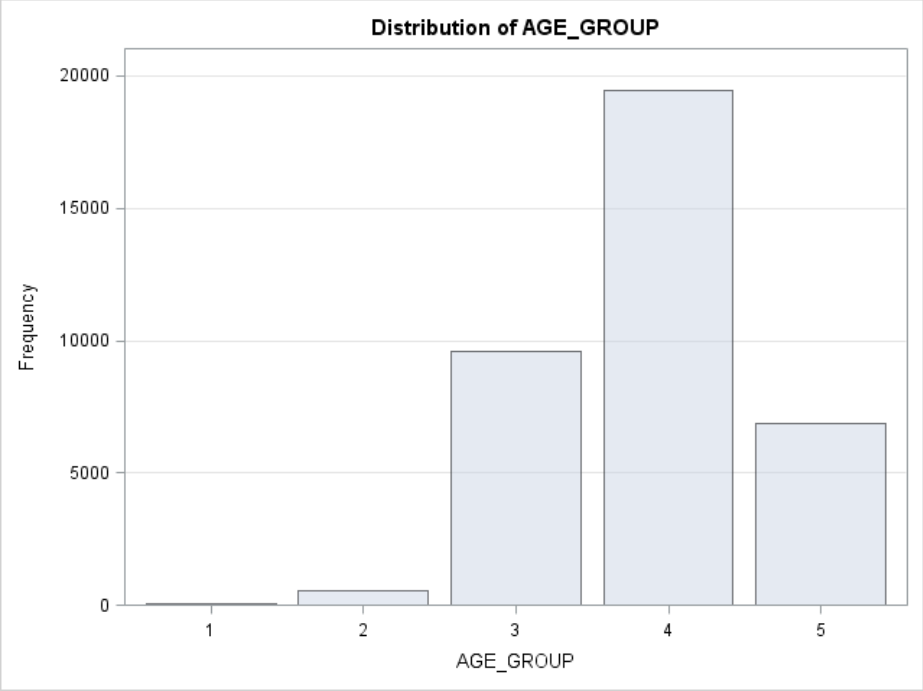
age group 4, which the age is between 60 and 80.

One-Way Frequencies on Age group

Results

The FREQ Procedure

AGE_GROUP	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	51	0.14	51	0.14
2	568	1.56	619	1.70
3	9574	26.22	10193	27.92
4	19438	53.24	29631	81.16
5	6878	18.84	36509	100.00



Generated by the SAS System ('Local', X64_7PRO) on April 13, 2015 at 9:51:02 AM

Figure 18: One-Way Frequencies on Age group

In figure 19, it shows the possibility of incidence to develop to pancreatic cancer is almost equal between male and female, this result is supported published data from American Cancer Society," Men are about 30% more likely to develop pancreatic cancer than women, but the gap has closed in recent years." [25]

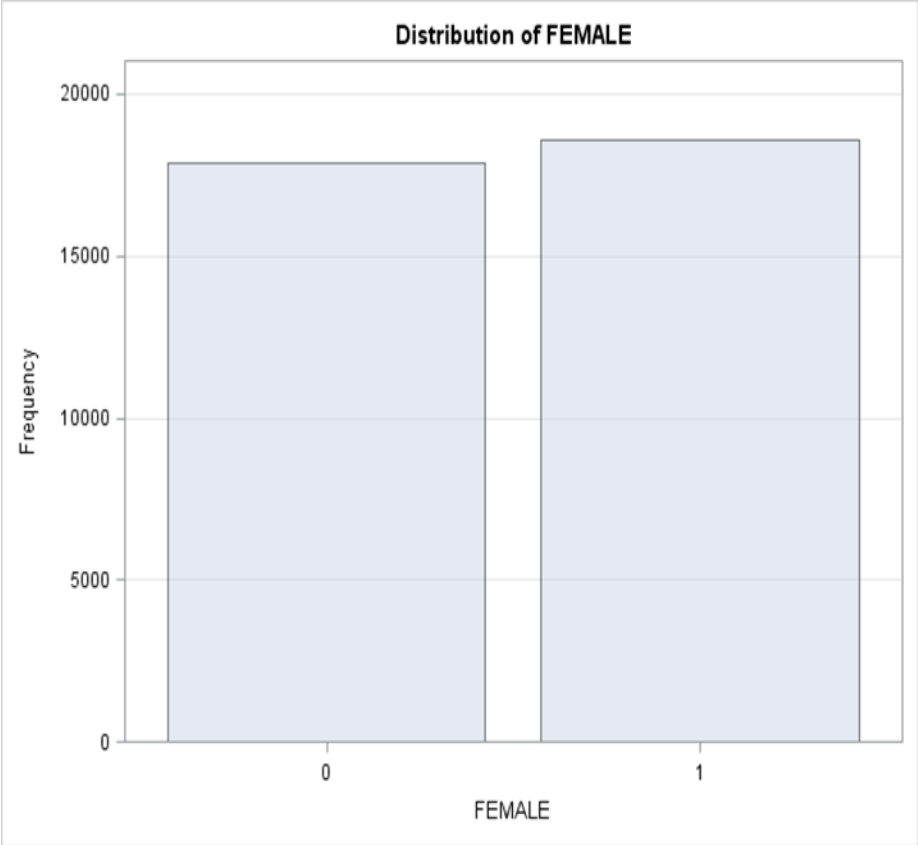
One-Way Frequencies on Gender (FEMALE=1, MALE=0)

Results

The FREQ Procedure

FEMALE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	17882	49.01	17882	49.01
1	18602	50.99	36484	100.00

Frequency Missing = 25



Generated by the SAS System ('Local', X64_7PRO) on April 13, 2015 at 9:44:27 AM

Figure 19: One-Way Frequencies on Gender

In figure 20, the result indicates that majority of pancreatic cancer patients were insured by either Medicare or Medicaid which means most of patients were either elderly or low income.

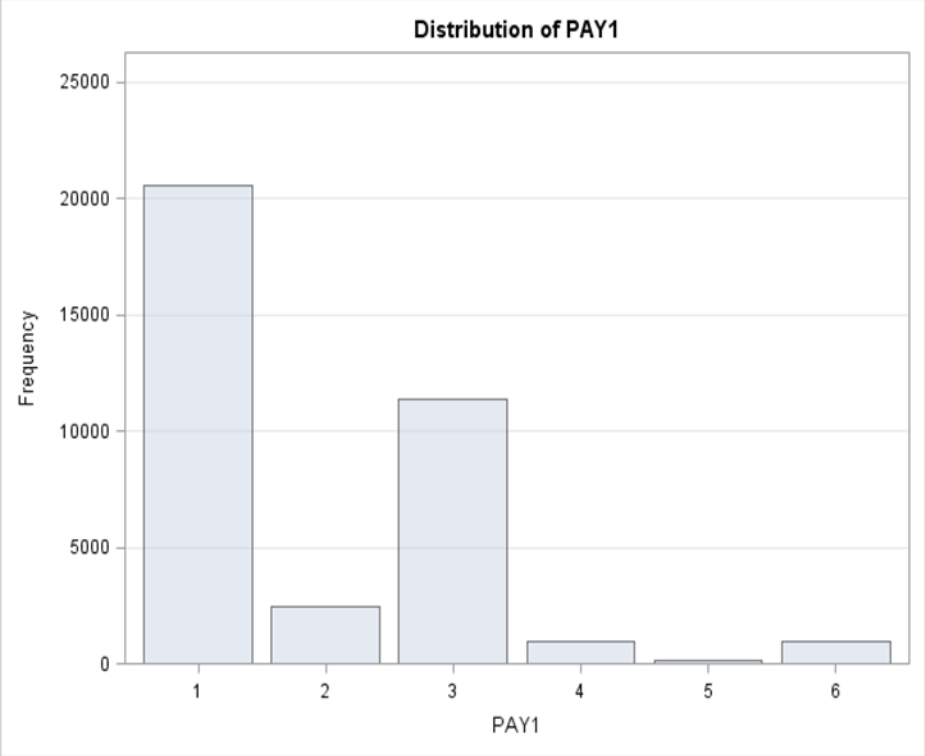
One-Way Frequencies on PAY1

Results

The FREQ Procedure

PAY1	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	20550	56.39	20550	56.39
2	2464	6.76	23014	63.15
3	11377	31.22	34391	94.37
4	992	2.72	35383	97.09
5	127	0.35	35510	97.44
6	934	2.56	36444	100.00

Frequency Missing = 65



Generated by the SAS System ('Local', X64_7PRO) on April 13, 2015 at 9:58:43 AM

Figure 20: One-Way Frequencies on PAY1

In figure 21, the result data shows that majority of pancreatic cancer patients were white and black.

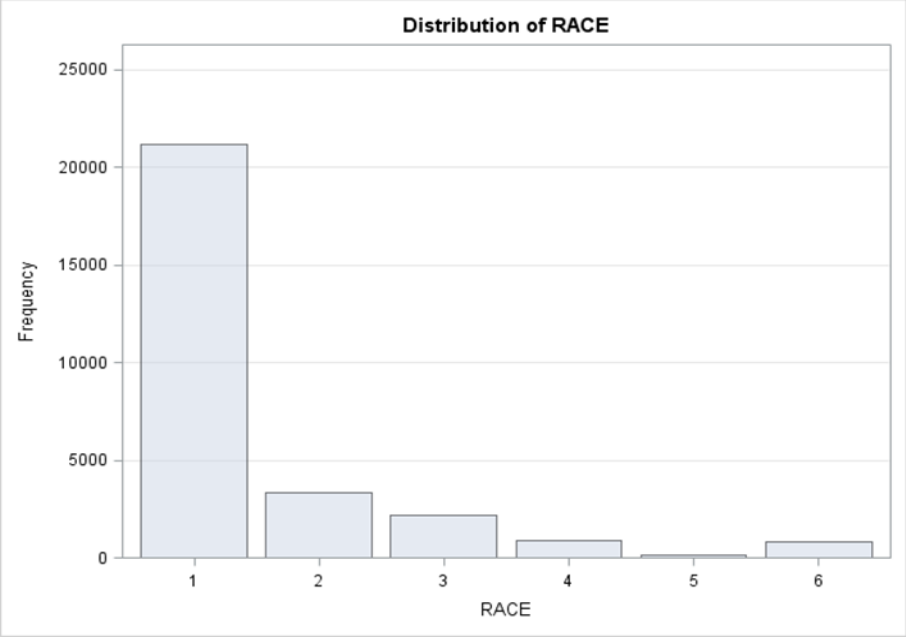
One-Way Frequencies on RACE

Results

The FREQ Procedure

RACE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	21182	74.24	21182	74.24
2	3331	11.68	24513	85.92
3	2203	7.72	26716	93.64
4	896	3.14	27612	96.78
5	117	0.41	27729	97.19
6	801	2.81	28530	100.00

Frequency Missing = 7979



Generated by the SAS System ('Local', X64_7PRO) on April 13, 2015 at 10:00:43 AM

Figure 21: One-Way Frequencies on RACE

In figure 22, it shows that there is no significant difference with income group, however, in published data shows that low income is one the factor in pancreatic cancer incidence. The reason why our study result has gap with the published one is category 1,2,3 all under $45,000 Which those categories should belong to low income level.

One-Way Frequencies on ZIPInc_Qrtl

Results

The FREQ Procedure

ZIPInc_Qrtl	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	8840	24.78	8840	24.78
2	8930	25.04	17770	49.82
3	8854	24.82	26624	74.64
4	9045	25.36	35669	100.00

Frequency Missing = 840

Generated by the SAS System ('Local', X64_7PRO) on April 13, 2015 at 10:03:18 AM

Figure 22: One-Way Frequencies on ZIPInc_Qrtl

4.4 Main Results

We have 2 main outputs in this research project,

1) Predictive the mortality rate amongst pancreatic cancer patients during their stay in hospital
2) Predictive the influence of comorbidities of pancreatic cancer (primary diagnosis) on patients' life - live longer or live short.

4.4.1 Predictive the mortality rate amongst pancreatic cancer patients during their stay in hospital

The figure 23 below is the ROC curve obtained from logistic regression model using SAS 9.2. The logistic regression model is to predict the mortality of in hospital stay pancreatic cancer patients, independent variables are in table4.

The area under the receiver operating characteristic (ROC) curve is frequently used as a measure for the effectiveness of diagnostic markers [24]. In output ROC curve, the area under the ROC curve (AUC) is 0.725 which is >0.5. It means when randomly pick one patient from the disease group (diagnosed as pancreatic cancer) and one from the no-disease group (not diagnosed as pancreatic cancer) and do the test on both. The patient with the more abnormal test result (Died) should be the one from the disease group (diagnosed as pancreatic cancer). The area under the curve is the percentage of randomly

drawn pairs for which this is true (that is, the test correctly classifies the two patients in the random pair).

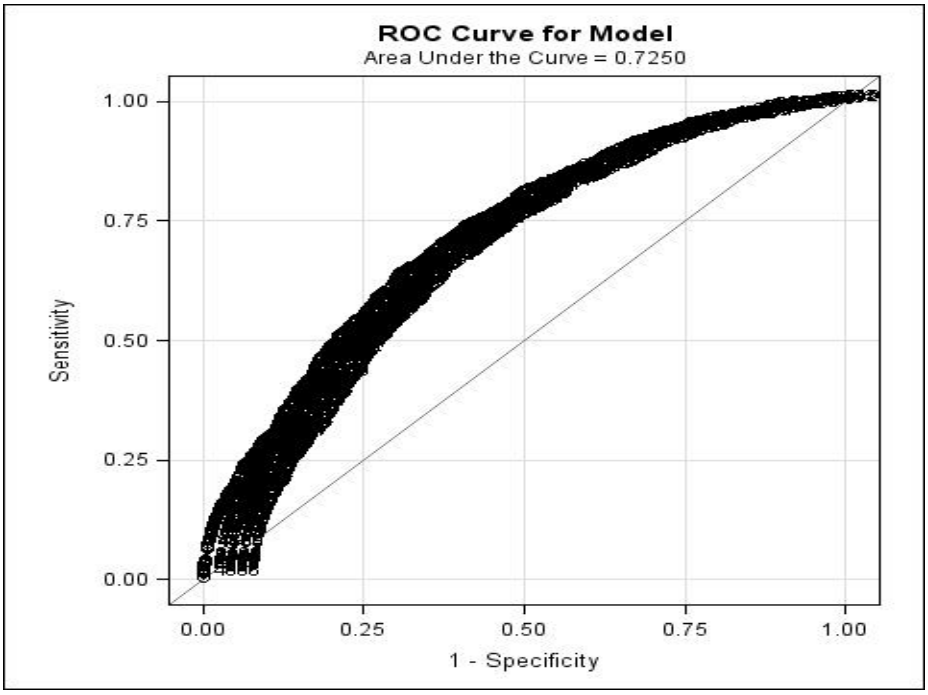


Figure 23: ROC curve model for Mortality Rate of pancreatic cancer patient during their stay in hospital. (Data from NIS 2005-2009)

4.4.2 Predictive the influence of comorbidities of pancreatic cancer (primary diagnosis) on patients' life

Every pancreatic cancer patient has certain comorbidities associated with, and how those comorbidities influence on pancreatic cancer patients' life? See table 3.1 for variables involved in this logistic regression model, the target prediction parameter is DIED. The results of odds ratio (OR) return from model show the predictive of certain variables (comorbidities) can affect the life (live longer or live shorter) of the pancreatic cancer patients.

We have 2 sets of results of odds ratio (OR) for each year (From year 2005 to 2009)

1) Odd ratio on all variables - includes both the 95% confidence internal (CL) overlap the null value (OR=1) and CL not overlap the null value.
2) Odds ratio on all variables with stepwise option –exclude statistical significance if it does overlap the null value (OR=1).

Table 4.5: Explanation of OR output in study

95% Low and upper Confidence interval (CL)	Statistical Significance	Description
LCL<1 & UCL>1	NO	The variable has no effect on pancreatic cancer patients' life
LCL<1 & UCL<1	YES	The variable will shorten the pancreatic cancer patients' life
LCL>1 & UCL>1	YES	The variable will extend the pancreatic cancer patients' life

In above table 4.5, it lists all 3 outputs of this research study, however, we are only interested in those variables with statistical significance as described below:

1) Both OR values of lower confidence internal(LCL) and upper confidence internal(UCL) less than 1 (life longer)
2) Both OR values of lower confidence internal(LCL) and upper confidence internal(UCL) more than 1 (life shorter)

If the LCL and UCL of a variable overlap the null value (OR=1), then it indicates the variable has no effect on pancreatic cancer patients' life. For example, the LCL and UCL values of OR in variable CM_ULCER are 0.038 and 2.319, which it is an overlap cross null OR null value 1, and then this variable does not report a measure's statistical significance.

In below figure 24 and figure 25, research study used logistic model running in SAS 9.2 to generate the odds ratio (OR) on comorbidities of pancreatic cancer patients against mortality variable (DIED). Study data files were retrieved from NIS 2005-2009.

1) Figure 24, Odds Ratio of all variables on Mortality
2) Figure 25, Odds Ratio with statistical significance variables only (Stepwise) on Mortality

Each set of result presents the OR values of the study variables with statistical significance or without statistical significance based on year 2005-2009. However, only those variables with statistical significance of CLC and UCL can be considered as good results.

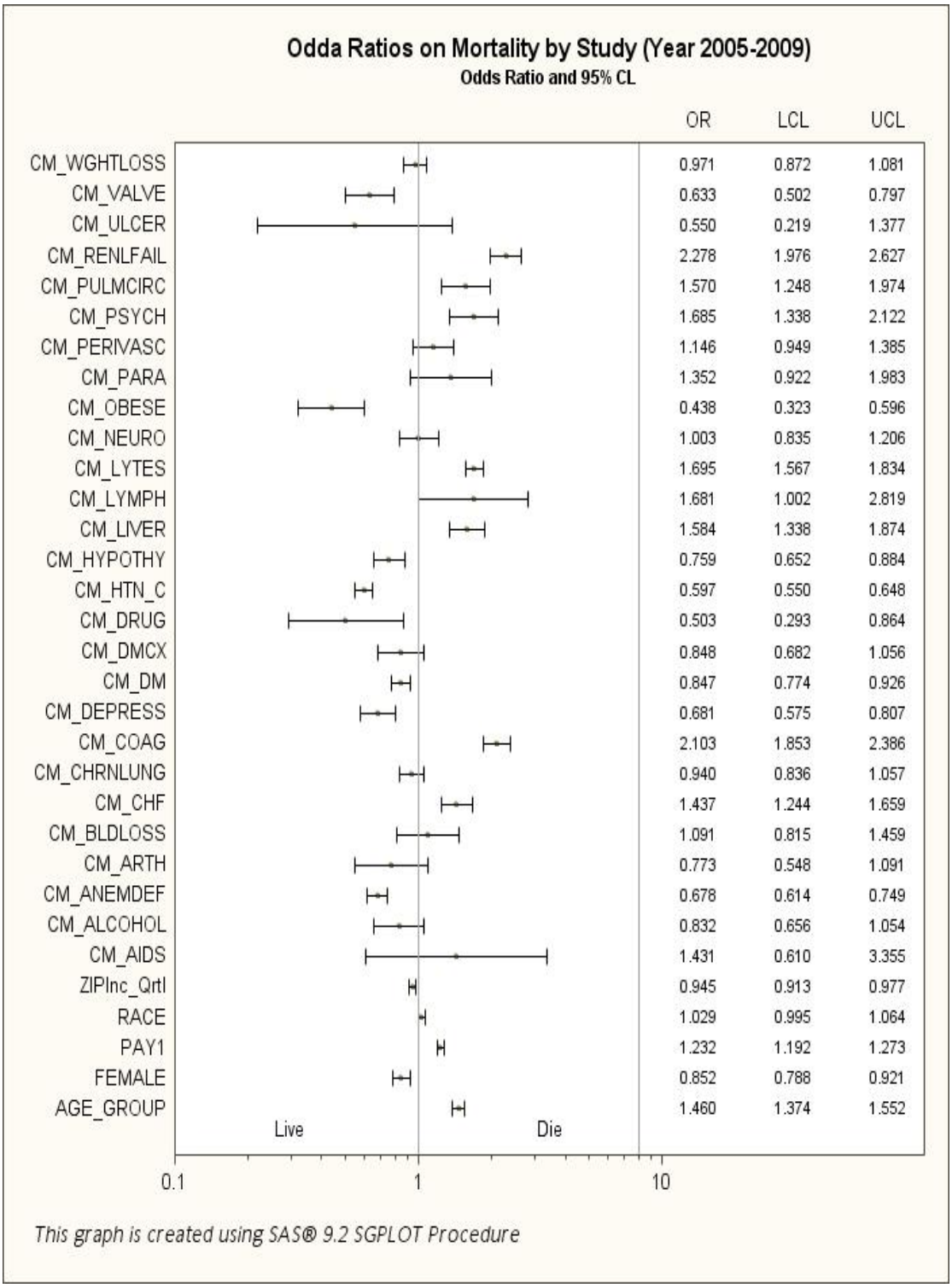


Figure 24: Odds Ratio of all variables on Mortality (Data from NIS Year 2005-2009)

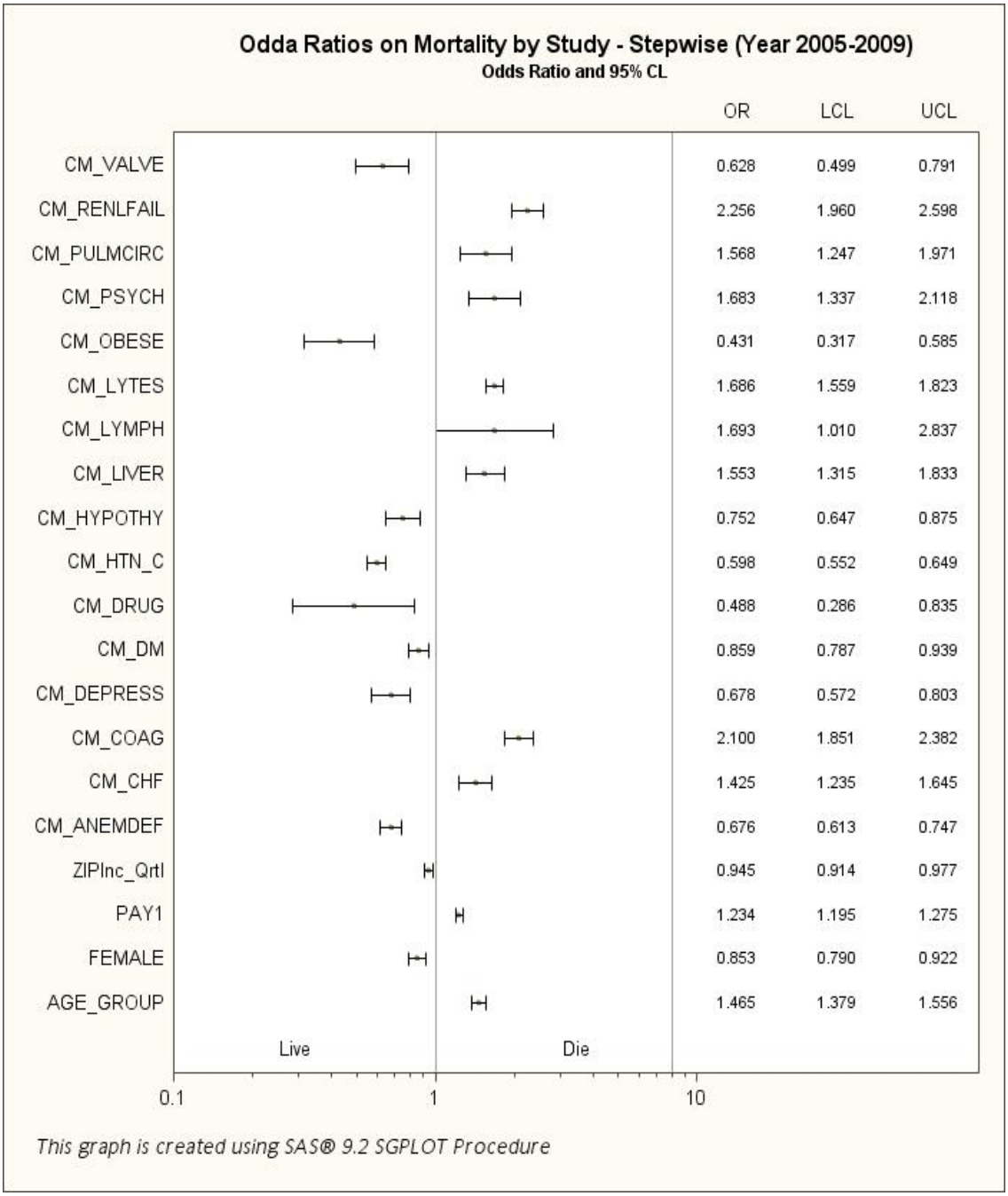


Figure 25: Odds Ratio with statistical significance variables only (Stepwise) on Mortality (Data from NIS Year 2005-2009)

In table 4.6, it lists all OR results return from logistic regression model. After filtering out the variables with LCL and UCL overlap the null value, those variables (in table 4.6, bold& Italic letters) with constant statistical significance

can be selected, which means those variables has influence on life of pancreatic cancer patients.

In table 4.7, this is the final list of variables can be considered as statistical significance which can affect pancreatic cancer patient life, either make patient live longer or make patients die sooner.

Table 4.6: OR results return from logistic regression model

Variable	**Influence on life** *+: live shorter* *-: live longer* *FALSE: no influence*	**Description**
AGE_GROUP	**+**	***AGE_GROUP***
FEMALE	**-**	***FEMALE***
PAY1	**+**	***PAY1***
RACE	FALSE	RACE
ZIPInc_Qrtl	**-**	***ZIPInc_Qrtl***
CM_AIDS	FALSE	acquired immune deficiency syndrome
CM_ALCOHOL	FALSE	alcohol abuse
CM_ANEMDEF	**-**	***deficiency anemias***
CM_ARTH	FALSE	rheumatoid arthritis/collagen vascular
CM_BLDLOSS	FALSE	chronic blood loss anemia
CM_CHF	**+**	***congestive heart failure***
CM_CHRNLUNG	FALSE	chronic pulmonary disease
CM_COAG	**+**	***coagulopathy***
CM_DEPRESS	**-**	***depression***
CM_DM	**-**	***diabetes, uncomplicated***
CM_DMCX	FALSE	diabetes with chronic complications
CM_DRUG	**-**	***drug abuse***
CM_HTN_C	**-**	***hypertension (combine uncomplicated***
CM_HYPOTHY	**-**	***hypothyroidism***
CM_LIVER	**+**	***liver disease***
CM_LYMPH	**+**	***lymphoma***
CM_LYTES	**+**	***fluid and electrolyte disorders***
CM_NEURO	FALSE	other neurological disorders
CM_OBESE	**-**	***obesity***
CM_PARA	FALSE	paralysis
CM_PERIVASC	FALSE	peripheral vascular disorders
CM_PSYCH	**+**	***psychoses***
CM_PULMCIRC	**+**	***pulmonary circulation disorders***
CM_RENLFAIL	**+**	***renal failure***
CM_ULCER	FALSE	peptic ulcer disease excluding bleeding
CM_VALVE	**-**	***valvular disease***
CM_WGHTLOSS	FALSE	weight loss

In below table 11, this is final list of variables can influence on life of pancreatic patients. The detail can be listed as follows:

1) If a patient is in high age group , he/she will die sooner than patients in lower age group
2) If a patient is female, she will live longer than male patients.
3) If a patient has no insurance, he/she will die sooner than patients have insurance
4) If a patient is in higher income group, he/she will live longer than patients in lower income group.
5) If a patient was diagnosed as deficiency anemias, he/she will live longer than patients have not.
6) If a patient was diagnosed as congestive heart failure, he/she will die sooner than patients have not.
7) If a patient was diagnosed as coagulopathy, he/she will die sooner than patients have not.
8) If a patient was diagnosed as depression, he/she will live longer than patients have not.
9) If a patient was diagnosed as diabetes, he/she will live longer than patients have not.
10) If a patient was diagnosed as drug abuse, he/she will live longer than patients have not.
11) If a patient was diagnosed as hypertension, he/she will live longer than patients have not.

12) If a patient was diagnosed as hypothyroidism, he/she will live longer than patients have not.

13) If a patient was diagnosed as liver disease, he/she will die sooner than patients have not.

14) If a patient was diagnosed as lymphoma, he/she will die sooner than patients have not.

15) If a patient was diagnosed as fluid and electrolyte disorders, he/she will die sooner than patients have not.

16) If a patient was diagnosed as obesity, he/she will live longer than patients have not.

17) If a patient was diagnosed as psychoses, he/she will die sooner than patients have not.

18) If a patient was diagnosed as pulmonary circulation disorders, he/she will die sooner than patients have not.

19) If a patient was diagnosed as renal failure, he/she will die sooner than patients have not.

20) If a patient was diagnosed as valvular disease, he/she will live longer than patients have not.

Table 4.7: Final list of variables can influence on life of pancreatic patients

Variable	Influence on life *+: live shorter* *-: live longer*	Description
AGE_GROUP	+	*AGE_GROUP*
FEMALE	-	*FEMALE*
PAY1	+	*PAY1*
ZIPInc_Qrtl	-	*ZIPInc_Qrtl*
CM_ANEMDEF	-	*deficiency anemias*
CM_CHF	+	*congestive heart failure*
CM_COAG	+	*coagulopathy*
CM_DEPRESS	-	*depression*
CM_DM	-	*diabetes, uncomplicated*
CM_DRUG	-	*drug abuse*
CM_HTN_C	-	*hypertension (combine uncomplicated and complicated)*
CM_HYPOTHY	-	*hypothyroidism*
CM_LIVER	+	*liver disease*
CM_LYMPH	+	*lymphoma*
CM_LYTES	+	*fluid and electrolyte disorders*
CM_OBESE	-	*obesity*
CM_PSYCH	+	*psychoses*
CM_PULMCIRC	+	*pulmonary circulation disorders*
CM_RENLFAIL	+	*renal failure*
CM_VALVE	-	*valvular disease*

4.5 Cross verification on main results with 3 different types of cancer

This study has provided a list of variables (see table 3.1) which has influence on pancreatic cancer patients life. However, those variables are based on pancreatic cancer with NIS data files from 2005-2009, and to find out if those variables (in table 4.8) has the same influence on other cancers, cross verification is necessary. In cross verification study, other 2 cancers have been selected as research target 1) Stomach cancer 2) Breast cancer.

In cross verification study, the same logistic regression model and the same amount of variables (See table 3.1) has been used to find their odds ratios and 95% CL. With 3 sets of output data: a) Pancreatic cancer b) Breast Cancer c) Stomach cancer, after filtering out those non-statistical significance variable (LCL<1 and UCL>1) and those variables not constant across all 3 cancers, for example, if a variable show + in pancreatic cancer and – in stomach cancer, then it will be excluded out.

In below figure 26, 27 and 28, those variables show their statistical significance (OR) on influencing 3 types of cancer patients' life constantly, either make patients live longer or live shorter.

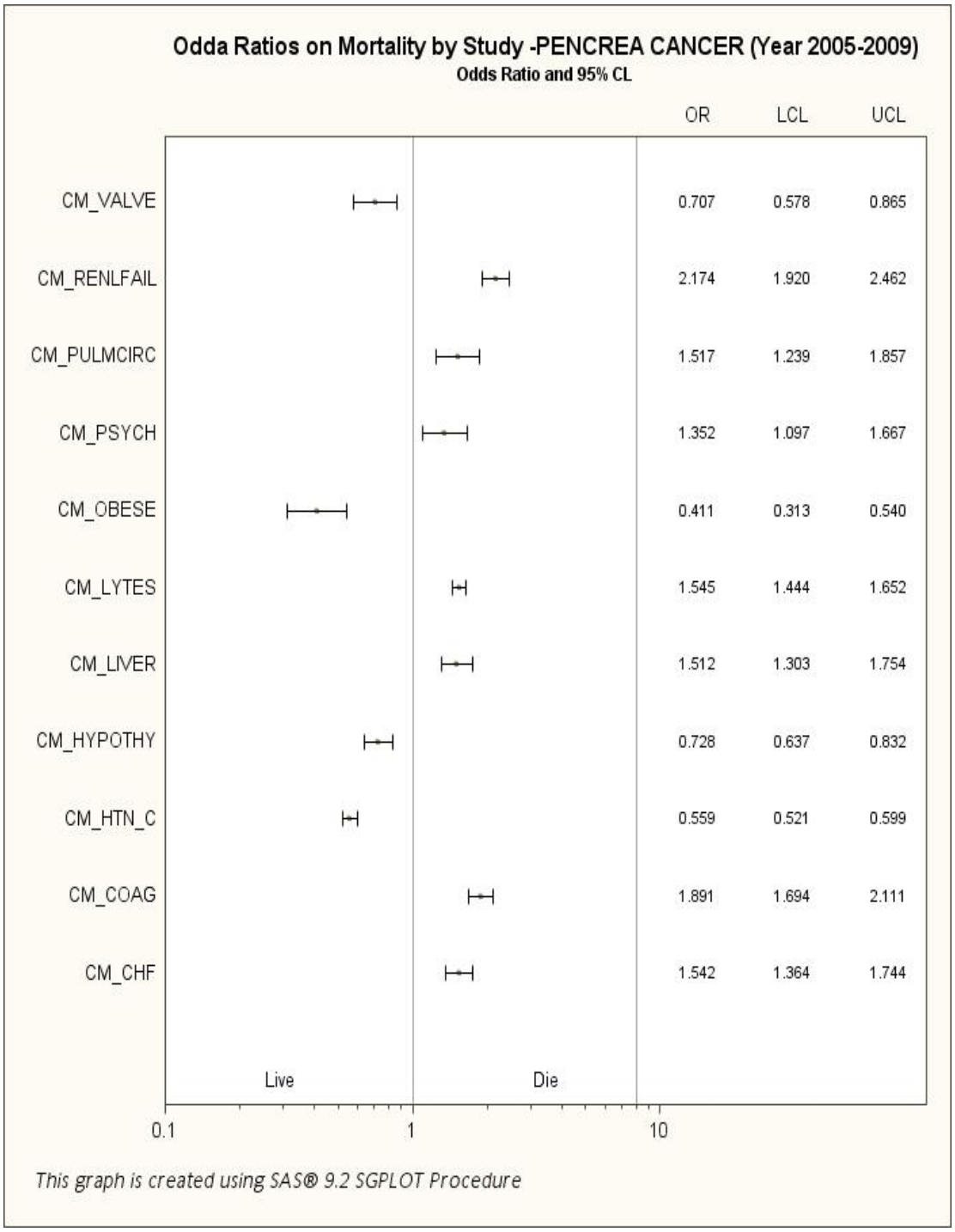


Figure 26: Variables can influence pancreatic cancer patients' life (Match with breast cancer and stomach cancer)

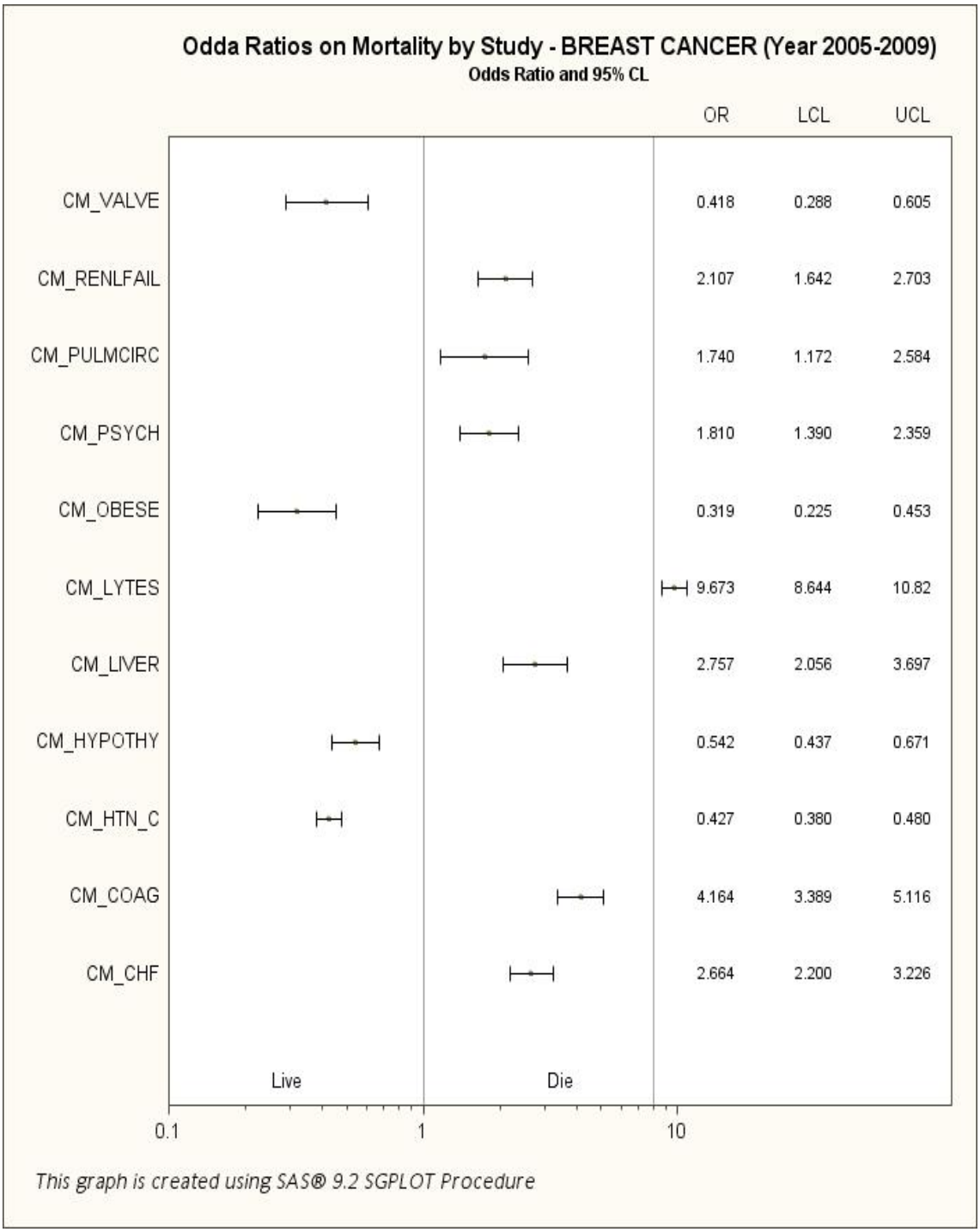


Figure 27: Variables can influence breast cancer patients' life (Match with pancreatic cancer and stomach cancer)

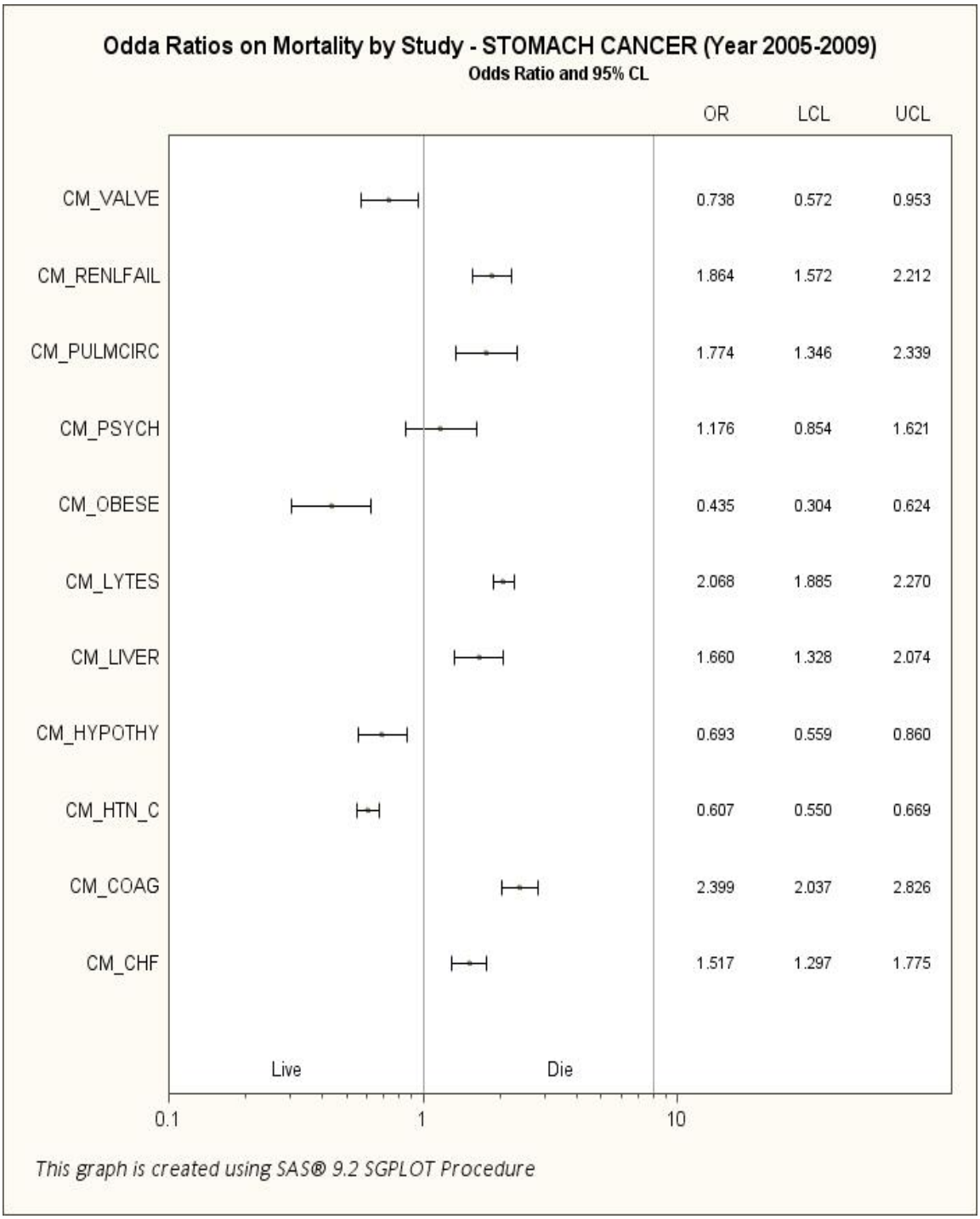


Figure 28: Variables can influence stomach cancer patients' life (Match with breast cancer and pancreatic cancer)

In table 4.8, this is list of 11 variables can influence on life of all 3 types of cancer patients, which is pancreatic cancer, breast cancer and stomach cancer. All 11 variables in table 4.8 have constant effect on patients' life, the detail is as follows:

1) If a patient was diagnosed as congestive heart failure, he/she will die sooner than patients have not.
2) If a patient was diagnosed as coagulopathy, he/she will die sooner than patients have not.
3) If a patient was diagnosed as hypertension, he/she will live longer than patients have not.
4) If a patient was diagnosed as hypothyroidism, he/she will live longer than patients have not.
5) If a patient was diagnosed as liver disease, he/she will die sooner than patients have not.
6) If a patient was diagnosed as fluid and electrolyte disorders, he/she will die sooner than patients have not.
7) If a patient was diagnosed as obesity, he/she will live longer than patients have not.
8) If a patient was diagnosed as psychoses, he/she will die sooner than patients have not.
9) If a patient was diagnosed as pulmonary circulation disorders, he/she will die sooner than patients have not.
10) If a patient was diagnosed as renal failure, he/she will die sooner than patients have not.

11) If a patient was diagnosed as valvular disease, he/she will live longer than patients have not.

Table 4.8: List of variables with constant influence on 3 types of cancer patients' life

Breast Cancer Variable	Influence on life	Pancreatic Cancer Variable	Influence on life	Stomach Cancer Variable	Influence on life	Description
CM_CHF	+	CM_CHF	+	CM_CHF	+	congestive heart failure
CM_COAG	+	CM_COAG	+	CM_COAG	+	coagulopathy
CM_HTN_C	-	CM_HTN_C	-	CM_HTN_C	-	hypertension (combine uncomplicated and complicated)
CM_HYPOTHY	-	CM_HYPOTHY	-	CM_HYPOTHY	-	hypothyroidism
CM_LIVER	+	CM_LIVER	+	CM_LIVER	+	liver disease
CM_LYTES	+	CM_LYTES	+	CM_LYTES	+	fluid and electrolyte disorders
CM_OBESE	-	CM_OBESE	-	CM_OBESE	-	obesity
CM_PSYCH	+	CM_PSYCH	+	CM_PSYCH	+	psychoses
CM_PULMCIRC	+	CM_PULMCIRC	+	CM_PULMCIRC	+	pulmonary circulation disorders
CM_RENLFAIL	+	CM_RENLFAIL	+	CM_RENLFAIL	+	renal failure
CM_VALVE	-	CM_VALVE	-	CM_VALVE	-	valvular disease

+: live shorter
-: live longer

Chapter V

DISCUSSION AND CONCLUSION

5.1 Summary

Each year approximately 45,000 people in the United States (and double this number in Europe) are now diagnosed with pancreatic cancer (adenocarcinoma). The prognosis is such that most of these people will have passed by the end of the first year. In the U.S., pancreatic cancer is 9th or 10th most commonly diagnosed cancer (depending on gender), but the fourth leading cause of cancer death in men and women [25]. Pancreatic adenocarcinoma is the fourth leading cause of cancer death and has an extremely poor prognosis: The 5-year survival probability is less than 5% for all stages. The only chance for cure or longer survival is surgical resection; however, only 10% to 20% of patients have resectable disease [26].

The lack of early detection of pancreatic cancer is mainly because of the little to no knowledge about the biology and etiology of pancreatic cancer, which in turn has not presented a systematic clinical approach in diagnosing patients with pancreatic cancer at the very early stage of the disease when they have a higher chance of survival. In hence, when a person was diagnosed as pancreatic cancer patient, it is life changing moment to a patient. How to maximally extend pancreatic cancer patients life has become a topic priority task for both patients and healthcare providers. The goal of the research study was to find out if certain

comorbidities associated with pancreatic cancer diagnosis can influence pancreatic cancer patients' life, which means can either make patients' life longer or shorter. In the research study, the data files were from NIS 2005-2009, research tool is SAS 9.2 and the algorithm used to recreate model is logistic regression.

Although, literatures to support this research study is not available, the finding from the study has been cross verified from 3 different types of cancer 2) pancreatic cancer b) breast cancer c)stomach cancer. The results from 3 types of cancer are highly matching and their odds ratio LCL and UCL values are all statistical significance.

5.2 Discussion and future research

This study has provided new angle of view to find certain hidden relationships between comorbidities and pancreatic cancer patients' life. The finding from this study shows that 11 comorbidities affect pancreatic cancer patients' life, they are:

1) Congestive heart failure

2) Coagulopathy

3) Hypertension (combine uncomplicated and complicated)

4) Hypothyroidism

5) Liver disease

6) Fluid and electrolyte disorders

7) Obesity

8) Psychoses

9) Pulmonary circulation disorders

10) Renal failure

11) Valvular disease.

The reason why above 11 comorbidities influencing pancreatic cancer patients' life can be categories into 3 parts:

1) Comorbidity itself
2) Medicines
3) Procedures

To each individual pancreatic cancer patient, the treatment plan is different from one to another depends on each patient's cancer stage, patient's condition and his/her comorbidities. From the finding, those comorbidities have significant strong relationship with pancreatic cancer patients' life, they can make patients' either longer or shorter (See table 13). Because the finding has been cross verified by 3 types of cancer, and the results are identical and constant, so the finding can be considered as a reliable and stable.

However, in this study, no clue can lead to a conclusion that this study finding is from comorbidities itself or the medications used on comorbidities or certain procedures have been practiced on patients or any combination treatment effects. All those possibilities need to be verified from further research, and the hidden pictures behind scene might be mined out using combination of mathematic algorithms and medical science.

5.3 Conclusion

From statistic data, more than 85% of pancreatic cancer diagnoses are made during the late stage of the disease, mortality rate of pancreatic cancer is extremely high. Only around 6% of people with pancreatic cancer survive more than 5 years after diagnosis. The main reason is because of the lack of reliable tools to diagnose the disease early enough so surgeons can remove the cancer before it spreads.

As a pancreatic cancer patient, any possible chances of extending his/her life are essential and countable.

This study aimed at finding the relationship between comorbidities and their influence on pancreatic cancer patients' life, and from the study, 11 comorbidities are confirmed that they can affect patients' life significantly. So far no literatures can be found on such kind of study, this study can be considered as a new research field in the future.

Reference

1. What is cancer. *Natl Cancer Institiute Natl Institutes Heal*. 2015. http://www.cancer.gov/cancertopics/what-is-cancer. Accessed Aril 12, 2015

2. Nugent FS, Stuart KE, Jay W. el al. Pancreatic Cancer. MedicineNet . 2015. http://www.medicinenet.com/pancreatic_cancer/article.htm. Accessed Aril 12, 2015

3. Francis W. Nugent, Keith E. Stuart, Jay W. Mark. Pancreatic Cancer, Page 3. MedicineNet . 2015. http://www.medicinenet.com/pancreatic_cancer/page3.htm#what_is_pancreatic_cancer. Accessed Aril 05, 2015

4. Cancer Statistics; 2014. Surveillance, Epidemiology, and End Results Program (SEER) http://seer.cancer.gov/statfacts/html/pancreas.html . Accessed April 12, 2015

5. Pancreatic Cancer. American Cancer Society. 2014 http://www.cancer.gov/cancertopic/pdq/treatmetn/pancreatic. Accessed Aril 05, 2015

6. Types of Pancreatic Tumors. John Hopkins Medicine 2012 http://pathology.jhu.edu/pc/BasicTypes2.php?area=ba. Accessed Aril 05, 2015

7. What are the risk factors for pancreatic cancer? American Cancer Society 2015. http://www.cancer.org/cancer/pancreaticcancer/detailedguide/pancreatic-cancer-risk-factors. Accessed Aril 09, 2015

8. Wilson LS, Lightwood JM. J Surg Oncol. 1999 Jul;71(3):171-81 Pancreatic cancer: total costs and utilization of health services. http://www.ncbi.nlm.nih.gov/pubmed/10404134

9. He Y, Zheng R, Li D, et al. 2015 Feb;27(1):29-37. doi: 10.3978/j.issn.1000-9604.2015.02.05. Pancreatic cancer incidence and mortality patterns in China, 2011. http://www.ncbi.nlm.nih.gov/pubmed/25717223

10. Rosenbaum BP, Kshettry VR, et al. 2015 Feb;129(2):173-81. doi: 10.1016/j.puhe.2014.11.011. Epub 2015 Feb 13. Diagnoses associated with the greatest years of potential life lost for in-hospital deaths in the United States, 1988-2010. http://www.ncbi.nlm.nih.gov.newproxy.downstate.edu/pubmed/25682904

11. Pancreatic cancer incidences statistics. Cancer Research UK 2014. http://www.cancerresearchuk.org/cancer-info/cancerstats/types/pancreas/incidence/uk-pancreatic-cancer-incidence-statistics#histology. Accessed March 21, 2015

12. What are the key statistics about pancreatic cancer? American Cancer Society 2014 http://www.cancer.org/cancer/pancreaticcancer/detailedguide/pancreatic-cancer-key-statistics. Accessed March 29, 2015

13. 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.

14. 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.

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

16. James A. Hanley, Ph.D. Barbara J. McNeil, M.D., Ph.D. "The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve" http://pubs.rsna.org/doi/pdf/10.1148/radiology.143.1.7063747

17. Receiver Operating Characteristic Curve (ROC). Eberly College of Science, Penn State University. STAT504. https://onlinecourses.science.psu.edu/stat504/node/163. Accessed April 3, 2015

18. Mary L. McHugh. The odds ratio: calculation, usage, and interpretation. Biochemia Medica 2009;19(2):120-6. http://dx.doi.org/10.11613/BM.2009.011

19. Encyclopedia of Research Design Encyclopedia of research design". 2010. doi:10.4135/9781412961288. ISBN 9781412961271

20. SAS Institute Inc. and World Programming Limited (England and Wales High Court (Chancery Division) July 23, 2010).

21. Special section: Pancreatic Cancer. http://www.cancer.org/acs/groups/content/@research/documents/document/acspc-038828.pdf. Accessed March 27, 2015

22. HCUP National (Nationwide) Inpatient Sample (NIS), Agency for Healthcare Research and Quality (AHRQ)

23. Pancreatic Cancer and African Americans. Johns Hopkins Medicine. 2014 http://pathology.jhu.edu/pc/nfptr/AA.php. Accessed March 18, 2015

24. Faraggi D, Reiser B. Estimation of the area under the ROC curve. Stat Med 2002;21:3093-106.

25. Pancreatic cancer prognosis. http://pancreatica.org/faq/pancreatic-cancer-prognosis. Accessed April 21, 2015

26. Páez D, Labonte MJ, et al. Pancreatic cancer: medical management (novel chemotherapeutics) Gastroenterol Clin North Am. 2012 Mar;41(1):189-209. doi: 10.1016/j.gtc.2011.12.004. Epub 2012 Jan 15.

27. What are risk factors for pancreatic cancer? Johns Hopkins Medicine 2014 http://pathology.jhu.edu/pancreas/BasicRisk.php?area=ba. Accessed April 28, 2015

28. Pancreatic Cancer Treatment. *Natl Cancer Institiute Natl Institutes Heal*. 2013. http://www.cancer.gov/cancertopics/pdq/treatment/pancreatic/HealthProfessional. Accessed January 11, 2013.

29. Wolfgang CL, Herman JM, et al. Recent progress in pancreatic cancer. *CA Cancer J Clin*. 2013;63(5):318-348. doi:10.3322/caac.21190.

30. Louzoun Y, Xue C, et al. A mathematical model for pancreatic cancer growth and treatments. *J Theor Biol*. 2014;351:74-82. doi:10.1016/j.jtbi.2014.02.028.

31. Lennon AM, Wolfgang CL, et al. The early detection of pancreatic cancer: what will it take to diagnose and treat curable pancreatic neoplasia? *Cancer Res*earch 2014;74(13):3381-3389. doi:10.1158/0008-5472.can-14-0734.

32. Lee ES, Lee JM. Imaging diagnosis of pancreatic cancer: a state-of-the-art review. *World J Gastroenterol*. 2014;20(24):7864-7877. doi:10.3748/wjg.v20.i24.7864.

33. Yang F, Jin C, et al. Emerging inorganic nanomaterials for pancreatic cancer diagnosis and treatment. *Cancer Treat Rev*. 2012;38(6):566-579. doi:10.1016/j.ctrv.2012.02.003.

34. 7. Poruk KE, Firpo MA, et al. Screening for pancreatic cancer. *Adv Surg*. 2014;48:115-136. doi:10.1016/j.yasu.2014.05.004.

35. Chang MC, Wong JM, et al. Screening and early detection of pancreatic cancer in high risk population. World J Gastroenterol. 2014;20(9):2358-2364. doi:10.3748/wjg.v20.i9.2358.

36. Poruk KE, Firpo MA, Adler DG, et al. Screening for pancreatic cancer: why, how, and who? Ann Surg. 2013;257(1):17-26. doi:10.1097/SLA.0b013e31825ffbfb.

37. Liu J, Xu D, Wang Q, et al. LPS induced miR-181a promotes pancreatic cancer cell migration via targeting PTEN and MAP2K4. Dig Dis Sci. 2014;59(7):1452-1460. doi:10.1007/s10620-014-3049-y.

38. Bardeesy NRA, DePinho R a. Pancreatic cancer biology and genetics. Nat Rev Cancer. 2002;2(12):897-909. doi:10.1038/nrc949.

39. Chang DK, Merrett ND, et al. Improving outcomes for operable pancreatic cancer: is access to safer surgery the problem? J Gastroenterol Hepatol. 2008;23(7 Pt 1):1036-1045. doi:10.1111/j.1440-1746.2008.05471.x.

40. Sarkar FH, Banerjee S, et al. Pancreatic cancer: pathogenesis, prevention and treatment. Toxicol Appl Pharmacol. 2007;224(3):326-336. doi:10.1016/j.taap.2006.11.007.

41. Oliveira-Cunha M, Siriwardena AK, et al. Molecular diagnosis in pancreatic cancer. Diagnostic Histopathol. 2008;14(5):214-222. doi:10.1016/j.mpdhp.2008.03.004.

42. Poruk KE, Firpo MA, Adler DG, et al. Screening for pancreatic cancer: why, how, and who? Ann Surg. 2013;257(1):17-26. doi:10.1097/SLA.0b013e31825ffbfb.

43. Sarris EG, Syrigos KN, et al. Pancreatic cancer: updates on translational research and future applications. Jop. 2013;14(2):145-148. doi:10.6092/1590-8577/1466.

44. Khawja SN, Mohammed S, et al. Pancreatic cancer disparities in african americans. Pancreas. 2015 May;44(4):522-7. doi: 10.1097/MPA.0000000000000323.

45. Wood LD, Hruban RH. Genomic landscapes of pancreatic neoplasia. J Pathol Transl Med. 2015 Jan;49(1):13-22. doi: 10.4132/jptm.2014.12.26. Epub 2015 Jan 15.

46. Okada K, Yamaue H. Nihon Rinsho. Current treatment and prognosis of pancreatic cancer in Japan and Western countries. Marm2015; 73 Suppl 3:177-80. Japanese. PMID: 25857010

47. Chan A, Diamandis EP, et al. Strategies for discovering novel pancreatic cancer biomarkers. J Proteomics. 2013;81:126-134. doi:10.1016/j.jprot.2012.09.025.

48. Bach P, Möhring C, et al. Retroperitoneal extravasation as the primary symptom of a pancreatic carcinoma. Urologe A. 2007 Nov;46(11):1548-50.

49. PANCREATIC CANCER FACTS 2014.; 2014. www.pancan.org.

50. Surgery for pancreatic cancer. American Cancer Society 2014. http://www.cancer.org/cancer/pancreaticcancer/detailedguide/pancreatic-cancer-treating-surgery. Accessed April 7, 2015

51. Wilson J. Report: Pancreatic Cancer Second Most Deadly by 2030.; 2014. http://www.cnn.com/2014/05/19/health/pancreatic-liver-cancer-deaths/.

52. THE ALARMING RISE OF PANCREATIC CANCER DEATHS IN THE UNITED STATES. 2012. https://www.pancan.org/wp-content/uploads/2013/01/incidence_report_2012.pdf

53. Chan A, Diamandis EP, et al. Strategies for discovering novel pancreatic cancer biomarkers. J Proteomics. 2013;81:126-134. doi:10.1016

54. Ansari D, Rosendahl a, et al. Systematic review of immunohistochemical biomarkers to identify prognostic subgroups of patients with pancreatic cancer. Br J Surg. 2011;98(8):1041-1055. doi:10.1002/bjs.7574.

55. Zhang P, Zou M, et al. Development of serum parameters panels for the early detection of pancreatic cancer. Int J Cancer. 2014;134(11):2646-2655. doi:10.1002/ijc.28584.

56. Michaud DS. Epidemiology of pancreatic cancer. Minerva Chir. 2004 Apr;59(2):99-111. PMID:15238885

57. Albert B. Lowenfe, Patrick Maisonneuve. Epidemiology and Prevention of Pancreatic Cancer. Japan Journal Clinical Oncology 2004;34(5)238–244

58. Pancreatic Cancer Action Network and American Association for Cancer Research Invite Applications for 2015 Research Grants. Cancer Res a J AACR. 2014. http://www.aacr.org/Newsroom/Pages/News-Release-Detail.aspx?ItemID=587#.VK7eL3u8o2I. Accessed March 1, 2015.

59. Prognosis of Pancreatic Cancer. Hirshberg Foundation for Pancreatic Cancer Research 2014. http://www.pancreatic.org/site/c.htJYJ8MPIwE/b.891917/k.5123/Prognosis_of_Pancreatic_Cancer.htm. Accessed April 23, 2015

60. Y. H. Shaib, J. A. Davila, et al. The Epidemiology of Pancreatic Cancer in the United States: Changes Below the Surface. Aliment Pharmacol Ther. 2006;24(1):87-94.

61. Vickie L. Shavers, Ph.D., Linda C. Harlan, Ph.D., et al, M.P.H. Racial/Ethnic Patterns of Care for Pancreatic Cancer. J Palliat Med. 2009 Jul; 12(7): 623–630. doi: 10.1089/jpm.2009.0036

62. Pancreatic Cancer Facts. Source for statistics: American Cancer Society: Cancer Facts & Figures 2014 and NCI Annual Plan & Budget Proposal For Fiscal Year 2012. http://www.pancreatic.org/site/c.htJYJ8MPIwE/b.5050503/k.40C9/Pancreatic_Cancer_Facts.htm. Accessed April 18, 2015

63. Garrido-Laguna I, Hidalgo M. Pancreatic cancer: from state-of-the-art treatments to promising novel therapies. Nat Rev Clin Oncol. 2015 Mar 31. doi: 10.1038/nrclinonc.2015.53. PMID:25824606

64. Pancreatic Cancer Report. Am Cancer Soc. 2013. http://www.cancer.org/cancer/pancreaticcancer/detailedguide/pancreatic-cancer-what-is-pancreatic-cancer. Accessed February 11, 2013.

65. Pancreatic Exocrine Tumors. Pancreatic Cancer Action Network. https://www.pancan.org/section-facing-pancreatic-cancer/learn-about-pan-cancer/types-of-pancreatic-cancer/exocrine/. Accessed April 25, 2015

66. Pancreatic Neuroendocrine Tumors (PNETs). Pancreatic Cancer Action Network. https://www.pancan.org/section-facing-pancreatic-cancer/learn-about-pan-cancer/types-of-pancreatic-cancer/endocrine-pancreatic-neuroendocrine-tumors/. Accessed April 25, 2015

67. Škrha P, Hořínek A, et al. [miRNA-192, miRNA-21 and miRNA-200: new pancreatic cancer markers in diabetic patients?]. Vnitr Lek. 2015 Spring;61(4):351-4. PMID: 25894267

68. Diabetes. Pancreatic Cancer Action Network. https://www.pancan.org/section-facing-pancreatic-cancer/learn-about-pan-cancer/symptoms/symptoms-diabetes-and-pancreatic-cancer/. Accessed April 29, 2015.

69. Signs and symptoms of pancreatic cancer. American Cancer Society. http://www.cancer.org/cancer/pancreaticcancer/detailedguide/pancreatic-cancer-signs-and-symptoms. Accessed April 28, 2015.

70. Raffaele Pezzilli. Screening Tests for Pancreatic Cancer: Searching for the Early Symptoms or the Population at Risk. JOP. J Pancreas (Online) 2004; 5(4):240-242. http://www.joplink.net/prev/200407/200407_news.pdf.

71. Melanie Haiken. 10 Early Warning Signs of Pancreatic Cancer. https://www.caring.com/articles/10-early-warning-signs-of-pancreatic-cancer. Accessed April 6, 2015.

72. Pancreatic Neuroendocrine Tumors (Islet Cell Tumors) Treatment (PDQ®). Nation Cancer Institute at NIH. http://www.cancer.gov/cancertopics/pdq/treatment/isletcell/HealthProfessional/page1. Accessed April 27, 2015.