SMOKING IN THE UNITED STATES AIR FORCE: TRENDS, MOST PREVALENT DISEASES AND THEIR ASSOCIATION WITH COST

THESIS

Michail Gkoutouloudis, Captain, Hellenic Army

AFIT/GCA/ENV/11-S02

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/GCA/ENV/11-S02

SMOKING IN THE UNITED STATES AIR FORCE: TRENDS, MOST PREVALENT DISEASES AND THEIR ASSOCIATION WITH COST

THESIS

Presented to the Faculty

Department of Systems and Engineering Management

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Cost Analysis

Michail Gkoutouloudis

Captain, Hellenic Army

September 2011

AFIT/GCA/ENV/11-S02

SMOKING IN THE UNITED STATES AIR FORCE: TRENDS, MOST PREVALENT DISEASES AND THEIR ASSOCIATION WITH COST

Michail Gkoutouloudis

Captain, Hellenic Army

Approved:

//*signed*//	10 Aug 2011
Dirk P. Yamamoto, Lt Col, USAF (Chairman)	Date
//*signed*//	21 Jun 2011
Eric J. Unger, Lt Col, USAF (Member)	Date
//*signed*//	21 Jun 2011
Edward D. White, PhD, (Member)	Date

Abstract

This research focuses on the smoking rates among the Active Duty Air Force (ADAF) personnel and the association of smoking and cost of hospitalization because of diseases related to smoking. The analysis of the data taken from the Air Force Web HA questionnaire provides information about the relationship between the smoking rates of the ADAF personnel and specific socio-demographic characteristics. The analysis of a second dataset associated with the cost of hospitalization, provides a list with the most prevalent diseases related to smoking with the highest cost. Moreover, a Regression Analysis tries to explore potential predictors that could anticipate the cost of the most prevalent diseases related to smoking.

The Contingency Analysis showed that smoking in the U.S. Air Force is more prevalent among the enlisted, males, and the younger age groups. The Pivot Table Analysis demonstrated that ischemic heart disease and cerebrovascular disease present the highest cost. In addition, the enlisted personnel exhibit higher total cost compared to the officers, but the situation is reversed when referring to the average cost. Furthermore, while smoking is more prevalent among the younger age groups, the cost consequences of smoking are more intense in the older age groups. The Regression Analysis exhibited that the variables, related to socio-demographic characteristics, that explain better the cost of hospitalization are the age group of 45-60, the enlisted personnel, and all the pay ranks of the officers, while the diseases that affect more the cost of hospitalization are ischemic heart disease, cerebrovascular disease, malignant neoplasms of the urinary bladder, and other arterial diseases.

To my father, who taught me that hard work is always rewarded…

Acknowledgments

I would like to thank my committee, Lt Col Dirk P. Yamamoto, Lt Col Eric J. Unger, and Dr. Edward D. White for their support and guidance. But I would like to express my sincere appreciation and gratitude especially to my thesis advisor, Lt Col Dirk. P. Yamamoto, for his patience and persistence in requesting and getting the necessary data for this thesis. I am also indebted to him for the consistency he demonstrated, and the guidance and help he provided to me throughout the course of my effort in writing this thesis. Moreover, a special thanks to my stats professor, Dr. Edward D. White, for all the statistical guidance and insight.

I would also like to thank my sponsor, Dr. Brenda Moore, for the initial support and encouragement to start writing this thesis and enrich it with more information. She stood by me as a sponsor and as a friend.

Special thanks for her concern and love go to Annette R. Robb, Director of the International Student Office at AFIT.

Finally, I would like to express my deepest gratitude and love to my parents. Without their patience and support, I wouldn't have been able to achieve any of this.

Michail Gkoutouloudis

Table of Contents

Page

Page

Page

List of Figures

Page

Page

Page

List of Tables

Page

Page

SMOKING IN THE UNITED STATES AIR FORCE: TRENDS, MOST PREVALENT DISEASES AND THEIR ASSOCIATION WITH COST

I . Introduction

Background

Smoking is undoubtedly one of the most severe and serious social issues, and scientists and sociologists talk about it as a social phenomenon that affects various fields of human activity. Smoking is not merely a personal choice at the individual level, but affects society and thereby has become a public and social phenomenon. Smokers frequently face serious diseases that often lead to death such as lung cancer, pancreatic cancer, kidney cancer, emphysema, chronic bronchitis, coronary heart disease and cerebrovascular disease.

Smoking is a harmful habit not just for the smoker, but also for the other people surrounding the smoker. The detrimental effects of second-hand smoke include coughing, headaches, sore throat, eye irritation and dizziness. In addition, the dangerous effects that smoking can have on pregnant women and newborn babies cannot be ignored. Women who smoke are approximately 30% more likely to experience infertility than other women. Also, women who continue smoking while they are pregnant are twice as likely to have problems with their pregnancies in the third trimester (Diwan, 2010).

Even though anti-smoking campaigns have increased significantly in recent years and there are no more advertisements or billboards promoting smoking, the number of smokers increases every year and, accordingly, the number of deaths caused by the harmful effects of smoking (Mallin, 2002). “Smoking is the leading preventable cause of death in the U.S. More than 400,000 people die each year due to

smoking, with $167 billion spent in annual health-related economic losses" (Smith et al., 2008). The health consequences of smoking result in a substantial economic toll on people, employers, and society. Smoking results in cost effects that account for billions of dollars in annual medical care expenditures. The cost effects attributable to smoking include: cost of hospitalization, cost of physician visits, partial income loss due to disability and foregone future income due to premature death.

Problem Statement

Smoking among the active duty members of the U.S. military is one of the most alarming problems. Tobacco use by military personnel is an increasingly upsetting issue, because tobacco use can affect the alertness and readiness of troops during their deployment, and the general image and effectiveness of the military. Recent reports suggest that smoking has become more popular among those on active duty and especially those deployed in battlefields (Emanuel, 2010). At least one in three service members is a tobacco user of some sort, according to the Institute of Medicine (IOM) study (Emanuel, 2010). This number becomes higher when referring to those service members that are engaged in combat operations (Emanuel, 2010). Since smoking is clearly an issue that concerns the military, the government has implemented many measures in the past and continues to do so, in an effort to reduce smoking rates and eliminate tobacco use among its members. The Department of Defense (DoD), under the Health Promotion Policy Directive 1010.10 initiated in 1986, tried to improve and maintain the readiness and the quality of life of DoD personnel by replacing the Directive 6015.18, "Smoking in DoD Occupied Buildings and Facilities", and establishing a policy on smoking in DoD buildings and facilities (Arvey and Malone, 2008). Directive 1010.10 was more than a mere educational program and included restrictions concerning tobacco use. Directive 1010.10 also

included education and detailed information on the health effects and risks of smoking, aiming to prevent personnel from smoking and, in this way, enhancing their quality of life. Although Directive 1010.10 was extended in 1994 by Directive 1010.15 and implemented restrictions on indoor smoking, the tobacco control policy in DoD has largely remained unchanged, with smoking rates among active duty members of the U.S. military remaining high (Arvey and Malone, 2008).

The prevalent social problem of smoking in the U.S. military brings about a lot of consequences such as the aggravation of the DoD healthcare budget, the deterioration of military fitness levels and the mitigation of deployment readiness. Tobacco costs the Defense Department more than $1.6 billion a year in medical care and lost work days. The Pentagon laid out a plan in 1999 to reduce smoking rates by 5% a year by 2001, and could not achieve that goal (Riechman, 2009). Military tobacco users have been found to be more likely to have injuries during their training and have a higher probability of discharge within the first year of their service, compared to non smoking personnel (Klesges et al., 2001). In addition, tobacco users miss part of their training or miss duty days far more frequently than their non-smoking cohorts because of an illness related to smoking or aggravated because of smoking (Klesges et al., 2001). Those military personnel who smoke tend to be less productive and do not perform satisfactorily on physical tests relative to their non smoking colleagues (Conway and Cronan, 1988). A study that measured the factors affecting the performance of the physical fitness tests among the military population indicated that smoking was a more potent and firmer predictor of physical fitness than weight (Haddock et al., 2007).

Tobacco use includes the utilization of cigarettes, cigars, pipe tobacco and oral tobacco forms such as chew, snuff, dip and snooz. The main addictive substance in

tobacco is nicotine, which could be considered dangerous, as it is an addictive drug in any form. And like other addictive substances, it creates dependence and subsequently unpleasant withdrawal symptoms. Researchers have proven that the pharmacologic and behavioral characteristics that designate nicotine addiction are similar to the addiction that drugs such as heroin and cocaine provoke (American Heart Association, 2010). An addiction consists of the good feelings that result when an addictive substance is present and the bad feelings when it is not present, and nicotine addiction creates exactly the same symptoms, being one of the hardest addictions to break. Tobacco use and in the same sense nicotine use, create serious diseases and increases the risk of developing hardened arteries and heart attacks (American Heart Association, 2010).

Despite the vast research on the phenomenon of smoking and the heightened awareness of its detrimental effects upon health, society and government, and the noticeable publicity about litigation against tobacco companies, statistics indicate that the percent of adults who smoke in the United States increases every year, with a more pointed increase in smoking among persons 18 to 24 years of age (Mallin, 2002). After a 40-year decline, the U.S. smoking rate has fluctuated around 20% since 2005. Nearly 47 million adults make use of tobacco and the majority of them are male smokers and people living under the poverty level. (American Council for Drug Education, 2010).

Another severe problem associated with smoking and use of tobacco is the passive or environmental tobacco smoke (ETS), more commonly known as secondhand smoke. Demographics have shown that between 70% and 90% of non-smokers in the United States population are subject to secondhand smoke (University of Minnesota, 2010). It has been estimated that from the smoke emitted from one

cigarette, only 15% is inhaled by the smoker and the remaining 85% is released into the air for everyone to inhale. According to one study, secondhand smoke is the third leading preventable cause of death and disability in the United States after active smoking and alcohol use (University of Minnesota, 2010).

Tobacco use is one of the most significant health issues that the U.S. military faces today. In 2002, it was estimated that among military members, 33.8% were smokers, with the Marines holding the highest rate (38.7%) and the Air Force the lowest (27.0%) (Pyle et al., 2007). Smoking is responsible for a wide range of health problems, such as injuries, poor performance on fitness tests, and increased days of sickness (Pyle et al., 2007). In addition, tobacco use, apart from the harmful health effects that causes, is a serious financial burden for the U.S. The cost effects of smoking in the U.S. military result in high healthcare expenses, productivity loss, lost work days because of absenteeism and early discharge of active duty personnel--something that is more often observed in the Air Force. Air Force recruits who smoked, compared to non-smokers, were more likely to be discharged prematurely, burdening the DoD budget with an annual cost of $130 million, exceeding training expenditures (Klesges et al., 2001).

Additionally, there is concern due to the increased use of smokeless tobacco among military recruits and military members (Severson et al., 2009). The personnel of the U.S. military represent a remarkable percentage of the total population using smokeless tobacco. The use of smokeless tobacco is increasing among military personnel and its prevalence is found to be approximately twice, compared to the general population (Severson et al., 2009). Smokeless tobacco is mistakenly believed to be a safer alternative to smoking tobacco and that its use does not influence human health as much as smoking tobacco. However, smokeless tobacco has been found to

be more addictive than smoking tobacco and its users are more likely to become smokers than non smokeless tobacco users (Ebbert et al., 2006).

Few studies have investigated the reasons that cause smoking initiation among those who have never smoked, recividism among former smokers, or increased smoking frequency among current smokers. Some studies provide evidence that deployment of military personnel in battlefields is an important factor that affects both smoking initiation and relapse among non-smokers and former smokers, and additionally increases the tobacco consumption among current smokers (Poston et al., 2008). The deployment of active military personnel to active combat zones has increased over the last 20 years, since the U.S. participation in the Gulf War. It has been noted that there might be a relationship between the deployment and increased tobacco consumption among current smokers, initiation of smoking among never smokers or relapse among former smokers (Poston et al., 2008). The reasons most commonly quoted for smoking initiation or increased tobacco use during deployment are boredom, operational stress and anxiety. In addition, the lack or the limited availability of alternative activities such as gyms and movie theaters in an operational environment could increase tobacco consumption. Moreover, the misconception that the dangers that smoking causes are minimal, in comparison to the risks the deployed personnel face in the battlefield and the military environment of an operational theater may encourage tobacco use or increase the overall attitude of lenience toward smoking (Poston et al., 2008).

Military smoking is an increasingly important issue, because tobacco use negatively affects troop readiness and productivity, and in addition increases medical and training costs (Arvey and Malone, 2008). Given these effects, banning smoking within the military would be considered by many to be both militarily and fiscally

prudent. In 1985, the DoD conducted research on military smoking issues and found that tobacco use rates among military members were significantly higher than U.S. civilian rates and, additionally, it concluded that smoking affects readiness and estimated the cost effects of smoking related to healthcare (Arvey and Malone, 2008). On March 10, 1986, DoD announced an intense anti-smoking campaign through directive 1010.10 (Arvey and Malone, 2008). Directive 1010.10 was not just a mere educational program on quitting smoking, but went further than that, setting restrictions and specifying where individuals could smoke on military installations and when smoking would be permitted (Arvey and Malone, 2008). Directive 1010.10 also tried to educate and inform military members about the risks of tobacco use and tried to prevent personnel from initiating smoking, and to help personnel quit. Practitioners were educated, during the routine health examinations, to advise people about the risks related to smoking, the health benefits of abstinence and where they could get help to quit smoking (Arvey et al., 2008). Smoking prohibitions in indoor facilities were made more specific by Directive 1010.15, an extension of Directive 1010.10. Despite this extension, tobacco control policy has made small steps and has changed little since 1986 (Arvey et al., 2008).

In conjunction with the policy change, cessation assistance is offered to active duty military members. The program incorporates education techniques and nicotine replacement therapy, such as nicotine patches and nicotine gum, to assist in quitting the harmful smoking habit. The anti-tobacco policy tries to discourage individuals turning to alternative methods of tobacco use such a chewing or smokeless tobacco. This policy is amplified by the prescription and use of specific drugs that help kicking the habit of smoking, such as Chantix and Zyban (Commander, Submarine Forces Public Affairs, 2010).

Tobacco smoking deserves special consideration, since it affects the health, the quality and the readiness of the military personnel. In this way, tobacco smoking merits increased deliberation as an accessional benchmark of the quality of the military personnel, for various reasons. One of them is that the DoD suffers a serious financial burden from tobacco use. In 1998, DoD healthcare costs were estimated to have been inflated by $584 million annually, and in the same year, it was estimated that smoking created an additional cost of $346 million because of the annual cost of lost productivity (Larson et al., 2007). Moreover, smoking negatively affects the basic military training of recruits. Studies in the Navy found that smoking was one of the factors that predicted attrition in the first year of service and that 1,500 more recruits would graduate the after the 15-month period of training, if only non-smokers were recruited (Larson et al., 2007). The same findings are consistent with studies and researches on tobacco use in the Air Force. Smoking was one of the strongest predictors for discharge from training, compared to other predictors like demographics, education or even alcohol or drug use. In addition, estimates proved that recruits who smoke are related to an additional encumbrance of $18 million for the Air Force budget per year, because of excess training costs (Larson et al., 2007). It should be noted that smokers tend to have higher rates of absenteeism and are more often subject to injuries, compared to non-smokers. This fact has implications for organizational costs and productivity and consists of an additional predictor for the educational credentials and mental ability of the military personnel (Larson et al., 2007). In conclusion, smoking status could be considered as a predicative personnel quality benchmark.

Despite the fact that tobacco cessation measures and policies are a significant component of military health promotion programs, approximately one third of the

DoD personnel use tobacco, which is a percentage very close to the smoking rate among U.S. civilians and creates doubts about the physical and mental quality of the Army Forces (Larson et al., 2007). In addition, the U.S. military has always acted as a role model for society. Recent studies show that military members see themselves as role models for the rest of the society and in this way, a smoke-free and healthy military could be the benchmark of pride and consistency (Hoffman et al., 2008). Career military members and the military personnel stationed in supervisory roles should provide appropriate and healthy models. Moreover, they should render themselves responsible for the transmission and dissemination of an influential message in changing the conception and admittance of tobacco use by military members (Nelson and Pederson, 2008).

It is apparent that the issue of smoking has been the subject of ample research, therefore, considerable literature on the issue exists both in terms of the general population overall as well as the more specific issues of the United States and the U.S. military. Studies have shown that recently, the rates of smoking among the general population of the United Studies have decreased, while other studies have documented a high predominance of smoking among the military personnel, before and after their admission into the military (Nelson and Pederson, 2008). Smoking is more intense among deployed military members, because of stress, boredom, family separation and lack of other alternatives of entertainment. The military has adopted a subset of tobacco related objectives, which include the reduction of smoking and the elimination of the use of smokeless tobacco. The tobacco cessation programs are focusing on reducing the acute and alarming issue of smoking among the military members.

This study focuses on examining and analyzing the smoking rates among the active duty members of the Air Force of the United States. Smoking is a severe phenomenon for the Air Force today because it is negatively associated with readiness, fitness level and health quality of the personnel. Tobacco use in the Air Force is connected to premature death from diseases related to smoking, economic losses to society and a remarkable burden on the healthcare governmental budget. Huge healthcare expenditures and yearly lost productivity are the results of the high rates of smoking among the Active Duty of Air Force (ADAF). Moreover, this study tries to classify the most prevalent diseases related to smoking, according to their total cost, for which ADAF members have been conveyed to hospital. Smoking and high medical care costs are intimately connected, creating a huge burden on the healthcare budget of the DoD. In addition, this study makes an effort to detect any potential relationship between the cost of hospitalization of ADAF personnel because of smoking related diseases and various predictors related to socio-demographic characteristics of the population of ADAF.

The rates of smoking among the ADAF can be classified according to age, gender and rank. Tobacco use is more popular and widespread among the younger ADAF and especially among the enlisted ranks. Factors such as gender, age group and pay rank, affect the intensity of smoking, the health standard of the U.S. Air Force personnel and the magnitude of the relevant economic losses.

The first part of this study is based upon data extracted from the Air Force Web Health Assessment (AF Web HA) questionnaire, more specifically from the section of AF Web HA which refers to demographics and questions associated to smoking and tobacco use. Web HA is an online questionnaire completed by military members as part of annual medical assessments. The demographics give substantial

information about the profile of the interviewee which include gender, age and pay rank. The Tobacco-Use section of AF Web HA gives information about current smoking status of Air Force personnel and this section is used in this research for measuring the smoking trends among the ADAF members and their association with specific socio-demographic characteristics. The Tobacco-Use section of AF Web Ha questionnaire is given in Appendix A (AF Web HA, 2010).

The second and third part of this study is based on data obtained from the Air Force Medical Support Agency's Healthcare Informatics Division (AFMSA/SG6H). This dataset includes cost data of direct and network care, provided to ADAF personnel, because of smoking related diseases. The information extracted from this dataset is used for two purposes:

- For rating the most prevalent diseases related to smoking, according to their total cost, and providing additional information associated with the socio-demographic characteristics of the population

- For trying to detect any potential relationship between the cost of hospitalization and various variables affecting this cost

The purpose of this study is to analyze statistically the data obtained from the AF Web HA records, present the current smoking status of Air Force, and make a resource about who smokes more according to gender, age and pay rank. Additionally, this study focuses on the most prevalent diseases that are associated with smoking, and analyses them on a cost basis, in order to sort them out according to their cost and track any relationship between this cost and any characteristics referred to the ADAF members.

Research Objectives

Research Questions

1) How is smoking affected by the socio-demographic characteristics of the ADAF population?

    - How is smoking affected by pay rank?
    - How is smoking affected by gender?
    - How is smoking affected by age?

2) Which diseases cost most to the U.S. Air Force, according to their total cost of hospitalization?

3) How is the cost of hospitalization affected by gender, age, pay rank and each disease separately?

Research Focus

The initial area of research focuses on determining and measuring smoking rates in the U.S. Air Force, specifically the active duty members. The measurement of these rates is based on data, extracted from the AF Web HA questionnaire data, which gives important information about the tobacco use in Air Force, sorted by gender, age and pay rank. The secondary area of the research is exploring the hierarchy of the most prevalent diseases related to smoking, according to their cost and providing some information about the cost of these diseases, relating it to more specific characteristics associated with the population of ADAF. The third area of the research is based on the detection of any predictability of the cost by variables related to socio-demographic characteristics of the ADAF personnel.

Assumptions

The data sets used in this research demand the establishment of some assumptions, which allow better manipulation of them to make them useable for this analysis. Starting with the first data set of AF Web HA, rows with blank cells referring to age, gender and pay rank were assumed to be erroneous and were deleted. Moreover, there were some rows referring to the rank of Warrant Officer. Since this pay rank no longer exists in the Air Force, and the rows referring to this pay rank were very few, they were deleted. These actions were taken for a better manipulation of the data set and for the elicitation of undistorted results.

The second data set, including the cost of hospitalization of ADAF because of diseases related to smoking, was reformulated, as below:

- The columns Diagnosis 2 up to Diagnosis 9 (Secondary Diagnoses) were excluded from the data set. Only the Diagnosis 1 column, which includes the ICD-9 coding of the Primary Diagnosis, was kept for this research.

- The Primary Diagnosis, and subsequently the whole data set, was restricted to the ICD-9 codes which refer to the most prevalent smoking related diseases, according to Smoking Attributable Mortality, Morbidity and Economic Costs (SAMMEC, 2010). The most prevalent diseases related to smoking and their ICD-9 codes, according to SAMMEC, are given below in Table 1.

Table 1. Most Prevalent Diseases Related to Smoking (SAMMEC, 2010)

Disease Category	ICD 9 Codes
MALIGNANT NEOPLASMS	
Lip, Oral Cavity, Pharyx	140-149
Esophagus	150
Stomach	151
Pancreas	157
Laryx	161
Trachea, Lung, Bronchus	162
Cervix Uteri	180
Kidney and Renal Pelvis	189
Urinary Bladder	188
Acute Myeloid	205
CARDIOVASCULAR DISEASES	
Ischemic Heart Disease	410-414, 429.2
Other Heart Disease	390-398, 415-417, 420-429.1, 429.3-429.9
Cerebrovascular Disease	430-438
Atherosclerosis	440
Aortic Aneurysm	441
Other Arterial Disease	442-448
RESPIRATORY DISEASES	
Pneumonia, Influenza	480-487
Bronchitis, Emphysema	490-492
Chronic Airway Obstruction	496

- There were two rows with the index unisex (U) for gender. Those were deleted.

- There were three rows with the index Air Force (AF), ten rows with the indices Warrant 1, Warrant 2, Warrant 3 (W1, W2, W3) and 2 rows with the index XX for pay rank. Those were deleted.

- There were 2 rows with the index zero and nine for age. Those were deleted, also.

- The cost was expressed in ThenYear Dollars. The use of cost data, which incorporates time value of money associated with inflation, demands its conversion to Constant Year Dollars. The procedure and

method of this conversion is presented and described in a detailed way in Chapter III.

The above described assumptions were made for a better management of the data sets and for the exclusion of some erroneous inputs that would distort the analyses and the results of this study. Furthermore, the restriction of the field of the research to the most prevalent diseases related to smoking according to SAMMEC enables the researcher to focus on those diseases that provoke the majority of the health problems related to smoking and investigate their influence on cost. It is acknowledged here that ICD codes are judgment calls of the medical provider and can, in theory, be incorrect. However, for this research, it is assumed that these are accurate diagnoses. Also note that it is assumed for this research that smoking is the primary cause of the diagnoses.

Preview

The discussion will begin with a review of the existing literature on smoking worldwide, in the United States, the Department of Defense and the Air Force. In Chapter III, the methodology used in this study will be presented, explaining which methods and what kind of analyses were used in each case. In Chapter IV, a Contingency Analysis will be developed to determine the rates of smoking among ADAF. Subsequently, a Pivot Table Analysis will be developed and will be graphically presented the cost rating of the most prevalent diseases related to smoking. The next step will be the development and presentation of a Regression Analysis for the exploration of potential statistical relationship between cost and various variables regarding the population examined in this study. Finally, the

conclusions of the research will be discussed, along with the efficacy of the cessation policy and what additional measures could be taken in the framework of the promotion of quitting smoking.

II. Literature Review

Tobacco Use and its Health and Cost Effects Around the World

Numerous studies have been done all around the world focusing on smoking and the harmful effects on human health and, consequently, on society. The World Health Organization (WHO) has been dedicated to the fight against smoking for many years and has conducted many studies on the detrimental consequences of the tobacco use. Every year WHO organizes campaigns against smoking in numerous countries, trying to inform people of the adverse health effects of smoking while launching programs for the cessation of tobacco use. In 2008, WHO published the "WHO Report on the Global Tobacco Epidemic, 2008: the MPOWER Package." This report refers to the smoking problem as a devastating epidemic that threatens the lives of one billion men, women and children during the 21st century: "Prompt action is crucial. The tobacco epidemic already kills 5.4 million people per year from lung cancer, heart disease and other illnesses. Unchecked, that number will increase to more than 8 million a year by 2030" (WHO, 2008).

Tobacco use is spread throughout the world because of successful direct and indirect marketing, low prices, lack of awareness of its effects on health and the economy, and ineffective policies against smoking. While the tobacco epidemic might be destructive, it is preventable and it can be significantly decreased if prompt action is taken. The WHO has established the MPOWER, a set of six significant measures against smoking: 1) raise taxes on tobacco products, 2) ban of marketing, sponsorship and advertisements of tobacco products, 3) the protection of non-smokers and people that suffer from second-hand smoking, 4) better information and awareness about the harmful effects and dangers of smoking, 5) offer of help to those who want to try and

quit smoking and 6) effective monitoring of the tobacco use epidemic and of the application of cessation policies.

The WHO report emphasizes that there still are crucial issues to be resolved, in order that further steps can be taken towards the extinction of the smoking problem. Among these issues are: 1) the weak monitoring and the lack of data on tobacco related diseases and deaths, which would propel effective tobacco control, 2) the inadequate implementation of smoke-free laws (only 5% of the global population is protected by these laws according to the WHO report), 3) insufficient establishment of cessation programs, 4) the unawareness of the full extent of health risks smoking induces in the majority of smokers, 5) the economic power of tobacco industries and the ineffective enforcement of bans on tobacco advertising, promotion and sponsorship and 6) the relevant low prices of tobacco products and their low taxation.

WHO and the MPOWER set of policies focus on these issues, in order to fight against the tobacco epidemic. Moreover, this report emphasizes the power of the tobacco industries and their dynamic marketing of their products. According to the WHO, the tobacco industry as a whole is a disease vector and spreads its epidemic through direct and indirect promotion in every angle of the planet.

The developed countries are already experiencing the harmful health and economic effects of smoking and now on the list are low-income and poor countries without any tobacco control or effective policies against tobacco use. Poverty is one of the long-term net economic effects of smoking. The tobacco industry's objective is to attract more users and to convert them into addicted smokers and this addiction disproportionately hurts the poor. After striking the wealthy and developed countries, smoking strikes poor countries now, augmenting the gap between wealthy and poor countries, since a smoker in a poor country in order to purchase tobacco, deprives

himself and his family from basic necessities such as food, shelter, education, and healthcare. In addition, the tobacco industry targets women and adolescents, trying to expand its clientele and create more addicted users. The tobacco industry is well-funded and more politically powerful and its strength can be restricted only through severe unbiased political action.

Young people and adolescents are also targeted by sophisticated and misleading advertising campaigns and tobacco industries spend millions on advertisements, trying to create more smokers, presenting smoking as a kind of emancipation, glamour and independence (Mackay and Eriksen, 2002). The foreword of Dr. Gro Harlem Brundtland concludes stating the aim of the Tobacco Atlas and the Tobacco Free Initiative, which is the enhancement of the global awareness of tobacco consumption and its effects in every aspect of human life, and the construction of new and the strengthening of existing actions against the devastating phenomenon of smoking.

This literature review will focus on several trends, beginning with male smoking, where worldwide almost one billion males smoke. This includes 35 percent in developed countries and 50 percent in developing countries (Mackay and Eriksen, 2002). Moreover, the smoking rates among men have peaked but they are declining at a slow tempo (Mackay and Eriksen, 2002). Educated men tend to give up smoking more than uneducated men. This fact implies that smoking is transforming into a habit of the low-education and the low-financial status men (Mackay and Eriksen, 2002).

The current number of female tobacco users is estimated at 250 million worldwide. This rate is analyzed in more detail, consisting of 22 percent of female smokers in developed countries and 9 percent of female smokers in developing countries (Mackay and Eriksen, 2002). The tobacco industries, in an effort to gain

more female “clients” and expand their market share, promote special advertising campaigns using misleading icons of emancipation and allurement. In addition, they launch special tobacco products for women, the so called “feminized cigarettes”, trying to create more female smokers (Mackay and Eriksen, 2002).

The Atlas makes a reference to youth smoking, mentioning that the majority of smokers begin using tobacco before reaching adulthood. The factors that contribute to the rise of smoking rates among adolescents are the specialized tobacco industry advertising, the relatively low prices of tobacco products and easy access to them. Starting the harmful habit of smoking during adolescence, makes teenage smokers even more addicted to it, and expands the danger of contracting smoking related diseases, such as heart disease and lung cancer, in their 30s or 40s (Mackay and Eriksen, 2002).

It is very noticeable that, while the consumption presents an image of stabilization or even decreasing in some countries, worldwide, the number of people smoking increases, especially because of the expansion of the world’s population.

Those who smoke prefer mainly cigarettes. Ninety six percent of tobacco product sales are from cigarettes (Mackay and Eriksen, 2002). The Atlas gives a short list of the regions of the planet that consume the biggest share of cigarette production worldwide. Tobacco sales and consumption are greatest in: “Asia, Australia and the Far East (2,715 billion cigarettes), followed by the Americas (745 billion), Eastern Europe and Former Soviet Economies (631 billion) and Western Europe (606 billion)” (Mackay and Eriksen, 2002).

The Tobacco Atlas focuses on the cost of smoking to the economy and to the smoker. Commencing with the cost to the economy, the tobacco companies claim that smoking and subsequently the production of tobacco products benefits the economy

and if all the tobacco control measures were to go into action, then tax revenues would decrease dramatically (Mackay and Eriksen, 2002). Many people that work in the tobacco industry would be unemployed and the economy would be called to face a serious hardship. But the tobacco companies avoid mentioning the economic losses that economy suffers from smoking. Tobacco use creates great losses to governmental economies, to the employers and to the environment because of the healthcare expenses due to smoking related diseases, absenteeism, decreased productivity, loss of foreign exchange because of the import of tobacco products, accidents, and deforestation because of careless fires caused by smoking or loss of land that could be used to cultivate food instead of tobacco.

Regarding the cost to the smoker, the main cost is the money spent on tobacco and cigarettes, which diverts money away from buying food, clothing or shelter. Moreover, a smoker may experience the loss of income because of illness and the loss of family income because of the time taken by the family members to look after a smoker. Smokers often have to deal with higher healthcare or insurance expenses, facts that dramatically decrease their net income.

Education is the most substantial part of the process of tobacco control. All the anti-smoking measures or any taxation and legislative intervention would not be meaningful without the understanding of their effectiveness. The purpose of the anti-smoking education is to focus not only on the harmful effects of smoking, but also aims to teach people, especially young people, how they could refuse this harmful habit. The Tobacco Atlas cites the efforts of quitting smoking and which techniques can be successfully utilized to quit the use of tobacco. The most popular techniques are: “Social support, clinics, quitlines, internet sites, skills training, nicotine

replacement therapy (NRT), and other pharmaceutical treatments" (Mackay and Eriksen, 2002).

The last part of the Tobacco Atlas makes some prognostics about the future of the tobacco epidemic. The most prevalent prediction is that the tobacco epidemic is increasing and expanding, while shifting from developed countries to the developing ones. Moreover, it is predicted that more women will be smoking in the future (Mackay and Eriksen, 2002). The remiss legislative interventions and the lack of structured and scientific information about the harmful effects of smoking, and the role of the powerful tobacco industries in the developing countries reinforce the expansion of the epidemic. The Tobacco Atlas describes the future as "bleak" unless immediate and considerable action is taken now. Studies, research reports and the several anti-smoking policies have proven that smoking rates can be significantly decreased if every government and nation takes sustained and decisive measures against the epidemic. (Mackay and Eriksen, 2002)

Smoking and its Effects in the United States

Tobacco use in the United States, along with exposure to tobacco smoke, are two of the most preventable causes of premature deaths due to chronic diseases, negative financial effects to society, and an economic impairment of the country's healthcare system. It has been estimated that at least 30% of all cancer related deaths, almost 80% of the deaths associated with chronic obstructive pulmonary disease and early cardiovascular disease and deaths related to it, are primarily engendered by the harmful habit of smoking (Adhikari et al., 2008). In order to assess the extent of the economic loses and the magnitude of the burden on the healthcare system of the United States because of smoking, the same team of Adhikari et al., conducted a study, which was an analysis of SAM (Smoking-Attributable Mortality) and of YPLL

(Years of Potential Life Lost) because of smoking, based on data of the Centers for Disease Control's (CDC) SAMMEC (Attributable Mortality, Morbidity, and Economic Costs) system. The analysis focuses on the years 2000-2004 and indicates that during this period the use of cigarettes and the exposure to cigarette smoke was responsible for at least 443,000 premature deaths, approximately 5.1 million YPLL and $96.8 billion in productivity losses annually in the United States (Adhikari et al., 2008).

The same analysis uses the sex and the age of the smokers and people exposed to tobacco smoke as leading variables, and is focused on nineteen adult and four infant disease categories. According to this analysis, during the period of 2000-2004, the estimated annual averages of deaths provoked by smoking were 269,655 deaths among males and 173,940 deaths among females in the United States (Adhikari et al., 2008).

It is worth mentioning that, among the nineteen adult diseases, the most prevalent diseases attributable to smoking were lung cancer, ischemic heart disease and COPD (Chronic Obstructive Pulmonary Disease). Percentages of deaths among adults 35 years or older, indicate that 41% of smoking associated deaths were engendered by cancer, 32.7 % by cardiovascular diseases and 26.3% by respiratory diseases. Along with the adult deaths, it was estimated that 776 infants died annually due to smoking during pregnancy, and 49,400 cases of lung cancer and heart disease annually were related to second-hand smoking (Adhikari et al., 2008).

Citing the economic effects of smoking, the same analysis mentions that for the same period of 2000-2004, the average productivity loss assignable to smoking was $96.8 billion, where $64.2 billion was attributed to males and $32.6 billion to females (Adhikari et al., 2008). Even though the smoking rates have declined

significantly compared to 1960s when they had reached their peak, the number of deaths attributed to diseases related to smoking is almost the same, because population has increased. This increase of the population contributes to the increase of the absolute number of deaths, even though the rates of smoking attributable diseases have relatively decreased (Adhikari et al., 2008).

During the period of 2000-2004, the total economic burden of smoking was $193 billion per year, including healthcare expenditures (which had been calculated to be almost $96 billion) and productivity losses (approximate estimation was $97 billion). This burden is 325 times larger than $595 million, which was the total cost of investments in tobacco control and cessation programs in fiscal year 2007 (Adhikari et al., 2008). Tobacco control and cessation programs could expedite the decline in smoking rates and subsequently the reduction in expenditures related to productivity losses and healthcare expenditures related to smoking.

Morbidity and Mortality Weekly Report (MMWR) published an article in September 2010 regarding the national and state adult smoking prevalence, reporting that even though the prevalence of smoking has declined the past 30 years in the United States, it continues to be the leading cause of cardiovascular diseases, multiple cancers and pulmonary diseases. Combined, these diseases cause the death of approximately 443,000 people annually and encumber the governmental budget with $ 193 billion annually, including healthcare expenditures and productivity losses. Even though the smoking rates have decreased over the past 30 years, the phenomenon of smoking is still one of the most alarming and widespread in the country (Dube et al., 2010). The report is based on 2009 data from the National Health Interview Survey. According to this data set, in 2009, 20.6% (46.6 million) of the adults of the United States were current smokers. Of these 46.6 million smokers,

36.4 million (78.1%) were regular smokers smoking on daily basis, and 10.2 million (21.9%) smoked on some days. In addition, smoking was more prevalent among men (23.5%) than women (17.9%) (Dube et al., 2010). Referencing racial groups, smoking was less prevalent among Asians (12.0%) and Hispanics (14.5%), compared to non-Hispanic Blacks (21.3%) and non-Hispanic Whites (22.1%). Smoking was most prevalent among multiple races (29.5%) and American Indians/Alaska Natives (23.2%) (Dube et al., 2010). Counting the smoking prevalence according to regions, the Midwest stands for the highest prevalence (23.1%) followed by the South (21.8%), and the West with the lowest prevalence (16.4%) (Dube et al., 2010). Smoking prevalence varies when it is observed by education level. Smoking rates were higher in 2009 among adults with a General Educational Development certificate (GED) (49.1%), and they tended to decline as the education level increased, reaching their lowest value (5.6%) among those with a graduate level degree. It is remarkable that smoking prevalence was higher among people living below the federal poverty level (31.1%), compared to those living at or above this level (19.4%) (Dube et al., 2010). The MMWR article concludes with the importance of tobacco control and cessation programs, referring especially to the states with the lowest smoking prevalence (Utah and California) and how successful and effective their long-running tobacco control programs have been (Dube et al., 2010). The article emphasizes the importance of anti-smoking strategies, such as price increases on tobacco products, concise smoke-free policies, and well organized campaigns and their implementation combined with access to efficient treatments and services.

Centers for Disease Control and Prevention (CDC) website provides valuable and important information about the smoking trends in the U.S., which correlate with the information previously given in the article from MMWR. The page is called Fast

Facts, last updated in September 2010 and last reviewed in October 2010, and besides the smoking rates of the U.S. in 2009, provides additional financial information, concerning the money spent in the advertising and promotions by the tobacco industry, the amounts available for tobacco control programs and the cost of second-hand smoking. The percentages of adults in the United States that were current smokers in 2009, are given in Table 1 (CDC, 2010):

Table 2. Percentages of U.S. Adults Current Smokers in 2009 (CDC, 2010)

Category	**Percentage**
All U.S. Adults	20.6 %
American Indian/Alaska Native Adults	23.2 %
White Adults	22.1%
African American Adults	21.3%
Hispanic Adults	14.5%
Asian American Adults	12.0%

An adult is defined as a person 18 years or older and a current smoker is considered a person who has reported that he/she has smoked at least 100 cigarettes during their lifetime and at the time of interview declared that they smoked every day or some days. Each day, approximately 1,000 persons under the age of 18 years old begin the harmful habit of smoking while every day 1,800 adults of 18 years old or older, begin tobacco use on a daily basis (CDC, 2010). CDC states that smoking costs almost $193 billion annually, an amount that consists of $97 billion lost in productivity and $96 billion in healthcare. A remarkable piece of information is that second-hand smoking costs more than $10 billion annually, a cost which is composed

of healthcare expenditures, morbidity and mortality (CDC, 2010). The same web page provides additional information about the funds spent for tobacco control and cessation programs. According to this report, in 2008, $24.4 billion was available to states, funds concentrated by excise taxes and legal settlements, for tobacco control programs, but only a small percentage (<3%) was spent for this purpose (CDC, 2010). Moreover, enormous amounts of money were spent by the tobacco industry, in order to reach its promotion aims. In 2006, $12.5 billion was spent totally for advertising campaigns (CDC, 2010).

MMWR, in an older article, makes a distinction between smoking morbidity and smoking mortality. The article talks about the cigarette smoking attributable morbidity in the United States in 2000 and there is a reference that labels the difference between morbidity and mortality. Data related to mortality indicate the number of individuals that die each year because of a disease attributed to smoking, while morbidity data is associated with the prevalence of persons that bear a disease affiliated to smoking (MMWR, 2003). The article focuses on the diseases attributable to smoking morbidity and mentions that in the United States, in 2000, approximately 8.6 million people had serious diseases related to smoking and chronic bronchitis and emphysema, which are accountable for a percentage of 59% of all smoking attributable diseases (MMWR, 2003). More specifically the article mentions:

> In 2000, an estimated 8.6 million (95% CI=6.9-10.5 million) persons in the United States had an estimated12.7 million (95% CI=10.8-15.0 million) smoking-attributable conditions. For current smokers, chronic bronchitis was the most prevalent (49%) condition, followed by emphysema (24%). For former smokers, the three most prevalent conditions were chronic bronchitis (26%), emphysema (24%), and previous heart attack (24%). Lung cancer accounted for 1% of all cigarette smoking-attributable illnesses. (MMWR, 2003)

Tobacco Use in the United States Military and Air Force

There is not vast literature on the topic of tobacco use in the U.S. Air Force. The majority of the articles and the studies focus on the association between smoking and the enlisted ranks of the Air Force, the use of smokeless tobacco and smoking during deployment. A common theme of these studies is the necessity of the implementation of a more active control and cessation smoking policy.

A study on active duty members of the U.S. Air Force, published in 2000, provides costs of smoking for active duty personnel of the Air Force for the year 1997. The article mentions that almost 25% of male and 27% of female active duty personnel aged between 17 and 64 years were smokers in 1997 (MMWR, 2000). Moreover, the estimated costs of current smoking, according to a study conducted in 1997 for the ADAF members, reached approximately the amount of $107.2 million per year, which was composed of $20 million for medical care expenses and $87 million for lost workdays (MMWR, 2000). The $20 million of healthcare expenses represent 6% of the total budget of the Air Force delegated to medical care expenditures and the $87 million of lost workdays was comprised of $76 million for lost workdays among males and $11 million among females (MMWR, 2000). The DoD estimated that in 1995, $584 million was spent annually in the healthcare sector because of smoking attributable diseases and $346 million of lost productivity occurred.

A similar study, presenting and analyzing the costs of mortality and morbidity attributed to smoking within the DoD, was conducted by Helyer et al. and published in 1998, using data from the year 1995 and the methodology of the Centers for Disease Control and Prevention. The population was comprised of active duty members of DoD, their families, retirees, and their dependents aged under 69 years

old. The study mentions that in 1995, the prevalence of smoking among the active duty personnel of the DoD was 31.6%. 54.6% of active duty members were never smokers, while 13.8% were former smokers (Helyer et al., 1998). The study makes a distinction between direct and indirect costs. The direct costs include direct healthcare costs, productivity loses and premature deaths. The total direct healthcare costs counted for $584 million, the largest amount was attributed to hospitalization costs, 77%, while 18% was ascribed to physicians' fees (Helyer et al., 1998). Male smokers were responsible for the largest share of the direct healthcare costs, 74%, and the majority of them belonged to age group 35 to 64 years (Helyer et al., 1998). The study rates smoking related diseases according to their share of responsibility in provoking a premature death. The cardiovascular diseases were responsible for 45% of the premature deaths attributable to smoking, neoplasms and lung cancer accounted for 35% of deaths and respiratory diseases were found at the third place of this assortment, with the percentage of 19% (Halyer et al., 1998). The premature deaths associated with smoking accounted for 16% of the deaths in the population of the DoD, almost one in six deaths (Halyer et al., 1998). In 1995, active duty members were hospitalized for 9,239 days because of a smoking related disease, and the cost connected with those days was almost $1 million. The cost of smoke breaks totaled $345,199,197 (Halyer et al., 1998). Enlisted personnel accounted for the 32.6 % of the current smokers, while among the officers ranks, 9.5% of smokers was accounted to the pay ranks O1-O3 and 7.1% to the pay ranks O4-O10 (Halyer et al., 1998).

One of the major concerns of the DoD in recent years is the unhealthy lifestyle of the military population and its dependents, and its consequences, financial and social, on the DoD itself. Tobacco use, overweight and obesity, and high alcohol consumption (referred as "TOBESAHOL") are the principal unhealthy behaviors of

the active members of the U.S. military, which adversely affect the quality of their health level, because of the numerous diseases caused by TOBESAHOL, and create costs of billions of dollars because of medical care expenses, lost productivity and premature decay (The Lewin Group, 2010). The Office of the Assistant of Defense, using the "Military Health System (MHS) Cost of Disease Estimator (CoDE), and based on a User-defined scenario that includes the TRICARE Prime beneficiary Air Force population stationed in CONUS," conducted a report that estimates the rates and costs of TOBESAHOL among active duty members of the Air Force and their dependents, and the Air Force retirees aged under 65 years old and their dependants, for Fiscal Year 2008. Of the $774 million of DoD medical costs due to TOBESAHOL in 2008, $174 million were attributed to problems generated by smoking (The Lewin Group, 2010). The tobacco use in this study was defined as the use of cigarettes and smokeless tobacco. Referring to the smoking rates among the active duty personnel of the report, an approximate number of 392,000 TRICARE Prime adult enrollees, expressed in a percentage of 46% of total TRICARE PRIME adult enrollees, were current smokers, and men were more likely than women to be moderate to heavy smokers. Correspondingly, young adults were more susceptible to be current smokers than older adults (The Lewin Group, 2010).

Some years ago a survey was conducted, based on every trainee entering the USAF enlisted force from August 1995 to August 1996 in order to provide information on the factors affecting trainees that urge them to smoke. The sample of the survey consisted of 32,144 trainees entering the enlisted ranks of the USAF for the period August 1995 – August 1996 and the data were collected on the basis of four general domains: demographic data, the background of smoking, coefficients related to tobacco use, and other risk factors. The results showed that the trainees that were

married, those that came from families with high income, those with low education level and Euro-Americans were more susceptible to smoke (Haddock et al, 1998). The survey demonstrated that one of the most forceful predictors of the smoking status of the trainees, participating in the survey was their concept of the social attractiveness of smoking (Haddock et al, 1998).

In their research, Klesges et al. showed that 28.5% of the 29,044 recruits who entered the Basing Military Training (BMT) of the Air Force from August 1995 to August 1996 were smokers. Smokers were 1.8% more likely to be discharged from the BMT during the first year, compared to non-smokers. Among the Air Force recruits, of the 14% discharged, 19.4% were smokers and 11.8% non-smokers (Klesges et al., 2001). The associated excess training costs of discharged recruits reached the amount of $18 million per year for the Air Force and assuming that, the same ratio of recruits prematurely dismissed because of smoking was applied to other services of U.S. military, the total military annual excess costs of training would approximately account for $130 million (Klesges et al., 2001). In addition, the investigation mentions that smoking status, compared to the rest of the demographic predictive variables used for this study, was the best single predictor of the premature discharge of recruits from the BMT of the Air Force (Klesges et al., 2001).

Despite the anti-tobacco measures implemented in recent years, such as the free-of-charge tobacco treatments, the regulation of the prices of the tobacco products, and the designation of military buildings as smoke free; the smoking trends among the active duty military personnel in the United States remain high and present an increasing trend that is remarkably higher compared to civilians (Haddock et al., 2009). The study was conducted with the aid of 15 focus groups from four USAF installations and nine focus groups from two U.S. Army installations, and was

concentrated on junior enlisted personnel and on those who directly supervise them, aged from 18 to 24 years old (Haddock et al., 2009). The results demonstrated that the factors that encourage tobacco use among junior enlisted ranks were smoke breaks, the easy access to tobacco products in the military installations, the social attractiveness of smoking, anxiety and boredom, and the apprehension of gaining weight. On the other hand, the factors that discourage tobacco use were the severe smoking bans in all military installations and vehicles, the inconvenience of smoking in designated areas, and the influence of the supervisors (Haddock et al., 2009).

Another study, published in 2009, focuses on the reasons for tobacco use among soldiers of U.S. Army. Soldiers in the Army use tobacco in order to fight stress, relax, socialize and make friends. Moreover, the majority of the soldiers in this study believed that the use of tobacco could help them to face the psychological and physical anxieties derived from the requirements of training and deployment. Some of them used tobacco products because of issues related to boredom and sleep deprivation (Nelson et al., 2009). Some of them used Smokeless Tobacco (SLT) as a less harmful alternative to smoking, despite being well aware of the adverse relationship between SLT and oral health (Nelson et al., 2009). In conclusion, the study suggests that the Army regulations and smoking restrictions should be more severe regarding the use of SLT. The team of Haddock et al. conducted a study in 2001, using the entire population of the Air Force Basic Military Training recruits for the period August 1995-August 1996, focusing on the use of smokeless tobacco among this population. The conclusions of this survey revealed that SLT is a powerful predictor of smoking initiation and the users of SLT appeared to be more susceptible to risky behaviors, such as dangerous driving (driving while intoxicated, not using seat belts) and the usage of alcoholic beverages (Haddock et al., 2001). Those who

tended to make extensive use of SLT were caucasians, while minorities were less likely to use SLT. SLT is used more frequently by recruits with high-income household backgrounds, suggesting that low income may be a barrier to the use of SLT (Haddock et al., 2001). The main finding of the research was the ascertainment that SLT is a strong predictor for tobacco use initiation, and that anti-smoking and cessation regulations and measures should include strategies that ban the use of SLT among the ranks of the Air Force (Haddock et al., 2001).

Many U.S. military personnel report fighting stress with smoking. Stein et al. investigated the relationship between high levels of stress and tobacco use among active duty members of the U.S. military, using the survey of Health-Related Behaviors of the DoD administered during the period September 2002 – February 2003. The study demonstrated that individuals that smoke or use smokeless tobacco, reported combating higher levels of stress related to family and work issues, compared to former or never smokers (Stein et al., 2008). Also, 18.39% of the participants were experiencing stress related to deployment, 15.52% were facing problems with a coworker, 15.42% were having problems with a supervisor, while 7.82% were combating stress derived from relationships. Finally, 6.24% reported stress because of health problems (Stein et al., 2008). In all cases, tobacco users were more susceptible to other negative behaviors, such as drinking alcohol and careless driving compared to non smokers (Stein et al., 2008). The studysuggested that tobacco use as a method for coping with stress is not effective and smoking makes an individual less likely to use “positive coping strategies.”

Another study, similar to the above mentioned, cites the relationship between cigarette smoking and military deployment, based on analyses conducted during the period of March 2007 – April 2007. Smith et al. in their study mention that

deployment is a decisive factor for smoking initiation and particularly for smoking recidivism. Among individuals that had never smoked before, 2.3% began smoking after deployment, while among former smokers, the percentage of those who reported resumption of smoking after deployment was 39.4%. The total percentage of smoking increase after deployment was 57% (Smith et al., 2008).

A team of researchers attempted to evaluate the smoking status and the status of tobacco use cessation (TUC) policies implemented for active duty members of the DoD, using and analyzing data collected from a new Military Treatment Facility (MTF) TUC evaluation tool in 2007. The study reported that, in 1997, $20 million was spent for medical care expenses associated with smoking for active duty AF personnel, and their cost of lost productivity for the same year due to smoking reached $87 million (Fraser et al., 2009).

In 2004, the medical cost of smoking to the DoD accounted for $1.3 billion, while a more recent study mentions that the annual cost to the DoD of tobacco use, comprised of healthcare expenditures, lost productivity, and decreased readiness, amounts to $1.6 billion (Fraser et al., 2009). The study focuses more on the smoking trends among the active duty members of the DoD for 2007 and compares them to the corresponding civilian members. The resulting investigation showed that the percentage of current smokers among the active duty personnel of the DoD for 2007 was 19.1%, slightly lower than the percentage of current smokers of the general population of the country for the year 2006, which was 20.8% (Fraser et al., 2009). Valuable information can be extracted from this study, concerning the rates of lifetime smoking, current smokers, everyday smokers and someday smokers for the Air Force (AF) for the year 2007, which are compared to the corresponding rates of the total

Military Health System (MHS) and the CDC National Benchmark rates, according to a survey executed in 2006. These rates are given in Table 3.

Table 3. CY 2007 Prevalence Smoking Rates (Fraser et al., 2009)

Prevalence Smoking Rates Computed for:	AF	MHS	CDC National Benchmark
Lifetime Smoking	40.7%	46.7%	50.2%
Current Smokers	16.1%	19.1%	20.8%
Everyday Smokers	64.1%	66.3%	80.1%
Someday Smokers	32.9%	33.7%	19.9%

The percentage of users of smokeless tobacco for the Air Force in 2007 was 4.5%, while the prevalence of smokers and users of smokeless tobacco at the same time was 0.9% (Fraser et al., 2009). The study emphasizes the fact that despite the implementation of several tobacco controls and cessation programs, the percentages of smokers in the DoD still remain high, and suggests a series of recommendations. The tobacco control policy of the DoD should be updated, including more severe policies such as the pricing of tobacco products sold in military facilities. Moreover, there should be an "inter-departmental communication" among the several forces of the DoD, for an enhanced collection of data, concerning the efficacy of the several anti-tobacco policies and the medication used in them (Fraser et al., 2009). The Medical Treatment Facilities (MTFs) should apply a more scrutinized observance of the Nicotine Replacement Therapy (NRT) and non-NRT medications and of their results, and perform cost-benefit analyses which would provide beneficial information about the quit rates and the effectiveness of these therapies (Fraser et al., 2009).

The most recent study about smoking and its association with mental health disorders among active duty military members was released in February 2011. This study used data from the 2005 DoD Survey of Health Related Behaviors (HRB) Among Active Duty Military Personnel in order to extract information regarding the smoking trends among the four forces of the U.S. military, and the relationship between smoking and mental health. The survey was based on a population of 13,603 subjects and the majority of the population was aged between 21 and 34 years old. Sixty-six percent of the population consisted of White, non-Hispanic individuals, 44% had some college education, 44.6% of the respondents were not married and 49.2% of the respondents were married with their spouse at home. Moreover, the biggest part of the population comprised of enlisted subjects, 82.2%, and 56% of them that had been deployed in the past three years (Schroeder, 2011). The results, regarding the smoking trends among the four forces of the U.S. military, showed that the Army had the highest prevalence of smoking at 31.9%, while the Marine Corps accounted for the lowest percentage of smoking prevalence with 12.9% (Schroeder, 2011).

Regarding the association between mental health and smoking, the respondents that had received mental counseling in the past were 67% more likely to smoke. The study reports that the ranks of officers were less likely to smoke compared to the enlisted ranks. In addition, the survey makes a reference to the association of smoking and some behavioral characteristics, such as the usage of alcohol and the absence of physical exercise. The respondents who reported being "heavy drinkers" were over four times more likely to be smokers, while those who didn't perform a workout at least 3 times a week had an increased likelihood of smoking (Schroeder, 2011). The study underscores the medical and occupational morbidity for the active duty members of DoD caused by mental disorders and their

association to smoking. This morbidity could be controlled and subsequently decreased by providing increased support in the field of diagnosis and treatment of mental disorders and by launching a more drastic smoking cessation policy for the members of the DoD (Schroeder, 2011).

Another report, prepared by the Research Triangle Institute (RTI) reported that the prevalence of cigarette use in the DoD population in 2005 was 32.2%, while the Air Force had the lowest percentage of prevalence of cigarette use, 23.3%, compared to the Army, Navy and Marine Corps (Bray et al., 2006). Male smokers exceed female smokers with 33.5% versus 24.2% prevalence. Among racial groups, Whites, non–Hispanics are most likely to smoke (36%). Moreover, individuals with an education level of high school or less are more susceptible to cigarette smoking. Marital status affects smoking, too, as unmarried participants of the study represent the highest percentage of smokers, 38.1%. The lowest pay ranks of E1 to E3 had the highest prevalence of cigarette use, 45.9%, almost ten times larger than the corresponding prevalence for the pay ranks of O4 to O10 (Bray et al., 2006). For the Air Force, 14.5% of the respondents reported they started smoking after joining, while 39% of the current smokers among the active duty members confessed they started smoking after joining the Air Force (Bray et al., 2006). Thirty-three percent of the Air Force respondents reported that the availability of tobacco products in Air Force installations makes it easy for someone to smoke, and among the reasons that explain cigarette use, 24.7% of the Air Force participants of the study, which was the highest rate, reported cigarette use in order to relax and calm down (Bray et al., 2006).

A significant issue related to smoking is the productivity loss within the DoD. Smokers present a higher productivity loss compared to the rest of the population, and the most frequent types of productivity loss are “leaving work earlier, being late for

work by 30 minutes or more, and working below normal performance level" (Bray et al., 2006). Moreover, smokers are more susceptive to working accidents than nonsmokers.

The most recent data available for current smoking rates is for 2008 and notes that the current smoking prevalence among military personnel is 31%, a rate which remains mostly steady since 2002. The main contributing factors that keep this rate unchanged for the last year are stress, deployment, boredom, the easy access to tobacco products, and in some cases sleep deprivation. Among the four services of the U.S. military, the Marine Corps present the highest prevalence of smoking, 37%, and Air Force the lowest at 23%. Rates referring to the use of smokeless tobacco in the DoD show the Air Force to hold the lowest rate of 9% and the Marine Corps the highest rate, 22%. The majority of smokers in the DoD are male, single, White, enlisted, and between the ages of 18 and 20 years old, and usually of low education level. The goal of the DoD was to implement an anti-smoking policy that could decrease the smoking prevalence to below 12% by 2010, but this has not been managed. The Air Force has applied the most severe tobacco control and cessation measures of all the branches of the U.S. military, including the ban of smoking during Basic Military Training, restricting smoking to very specific areas, and the prohibition of smoking advertisements in Air Force publications (Legacy for Longer Healthier Lives, 2011).

III. Data and Methodology

Overview

The purpose of this chapter is to define and delineate the data sets and methodology used to answer the research questions formulated in Chapter I. This chapter will start with a discussion related to the data sets used for this research, how the data were aggregated and used and what kind of conclusions and information was extracted from each of them. The next step will be to discuss what kind of analysis was used in each data set in order to answer the research questions. This analysis follows the partition of the research questions, since this partition is compatible with the nature of the data sets and their use. Next, each analysis will be further analyzed and presented in a more meticulous way, since each analysis answers questions of different nature.

In the first data set we have a Contingency Analysis followed by a graphical presentation of Excel diagrams, while in the second data set we have a Multiple Regression analysis. The second case is more complicated since the dependent variable, which is the cost of hospitalization of active duty members of the United States Air Force, hospitalized because of diseases related to smoking, is explained by many independent variables associated with age, gender, pay rank, and the frequency of the appearance of the most prevalent diseases related to smoking which present the highest cost. The second data set is used, additionally, to elicit information related to the cost of the diseases and to compose a list of the ten most “expensive” diseases. The frequency of the appearance of the diseases, mentioned in this list, is used as one of the independent variables in the regression models, which are built in order to explain how the cost is affected by the age, gender, pay rank and the diseases themselves. Lastly, the regression models will be further discussed, focusing on which

variables are used and why, how the variables were formed, and which diagnostics tests were used, in order that the validity of the models is better explained.

Data Sources

Two data sets were used in this research in order to deduce the results needed to answer the research questions. The first data set is a data set based on the Air Force Web Health Assessment (AF Web HA) questionnaire, answered by active members of the Air Force, throughout the years 2005 – 2009. The Web HA questionnaire is a part of the annual Preventative Health Assessment (PHA) exam and it is mandatory for all Active Duty Air Force members. It is divided into 17 sections, covering demographics and all the health issues for the member. This research is focused on sections 1 and 8, which are Demographics and Tobacco Use. The data set was provided by the Healthcare Informatics Division, AF/SG6H in San Antonio Texas, in November 2010, and includes information for the period from 2005 to 2009. This data set is used for the Contingency Analysis and the formulation of the smoking rates among the AF active duty members according to age, gender and pay rank.

The second data set is a cost data set, presenting the cost of hospitalization (expressed in Then Year dollars) of the active duty members of Air Force due to a disease related to smoking, throughout the period 1999 – 2009. The cost data set was obtained from the Air Force (AF) Medical Support Agency's Healthcare Informatics Division (AFMSA/SG6H), located in San Antonio, Texas. The data came from direct care (on base, either inpatient or outpatient) and network care (off-base, provided by non-military medical providers). Any on-base cost data is determined using MEPRS (Medical Expense and Performance Reporting System) criteria. All off-base cost data represent what is charged to Tricare. For each individual of this data set, there is an ICD-9 code in column Diagnosis 1, which refers to the primary diagnosis for the

patient. There are 8 more columns named Diagnosis 2, Diagnosis 3 up to Diagnosis 9 and include additional information, based on ICD-9 coding, about secondary diagnoses. This research is based on the Diagnosis 1 (primary diagnosis) for the most prevalent diseases related to smoking, according to the assumptions presented in Chapter I.

Furthermore, the data set takes into consideration additional information such as age, gender, and pay grade of the patient and of course the current year, information that is composed of the independent variables of the regression models built with this data set. The range of cost fluctuates between $0.00 and $307,063.12, which created an initial problem in the distribution of the Y response, which in this case is the Cost. In Appendix B, the primary distribution of Cost is shown and it is apparent that many outliers exist that make the data set seem erroneous and indicate a poorly fitting regression line. Avoidance of these outliers led to the partitioning of the primary data set into 5 subsets. The result of this partition was the creation of the following 5 subsets presented in Table 4:

Table 4: Five Subsets and their Range of Cost

Five Subsets		
Number of Subset	**Description**	**Cost Range in $**
1	Low Cost	$0.00 - $600.00
2	Medium Cost	$600.01 – $1,800.00
3	High Cost	$1,800.01 – $11,000.00
4	Very High Cost	$11,000.01 - $30,000.00
5	Extremely High Cost	$30,000.01 – 307,064.00

The partition of the initial data set and the creation of 5 subsets lead to the structure of 5 regression models, one for each subset. The development of 5 regression models corresponding to different ranges of cost, verified evidence that for different levels of cost there are different variables affecting the Y response (cost).

In addition, the second data set of cost was used in a Pivot Table Analysis, for the extraction of additional information about the cost of hospitalization due to smoking and its correlation to several socio-demographic characteristics of the Air Force population and to the most prevalent diseases related to smoking. The procedure and the purpose of the Pivot Table Analysis are further analyzed later in this chapter.

Normalization of the Cost Data Set

The second data set, used for building the regression models, is a data set of the cost of the hospitalization expressed in Then Year Dollars, which are dollars that include the effects of inflation and/or reflect the price levels expected to prevail during that year (SCEA, 2011).

The comparison of cost over the course of many years demands the conversion of Then Year Dollars to Constant Year Dollars. Constant year Dollars are a method of comparing dollar amounts of several years, without the effects of inflation. In this way the dollar amounts are showed at the value they would have in a selected Base Year. The Constant Year Dollars method includes the division of Current Year Dollar estimates by appropriate price indices. This procedure is also known as deflating (SCEA, 2011). The conversion from Then Year to Constant Year Dollars is used to present the value of something over time, excluding the effects of inflation or deflation. This enables the researcher to compare cost over time and to normalize the data set and make it more eligible for regression analysis. The method used for

converting the Then Year Dollars to Constant Year Dollars was based on the Inflation Calculator of the Air Force. The Inflation Calculator of the Air Force is a tool, which allows the user to generate any desired set of inflation tables, for any base year, starting from the year 1949 and ending at the year 2060. This calculator enables the user to perform inflation conversions without using inflation tables, or, to generate the Inflations tables and use them in order to execute the appropriate conversions. In addition, the inflation tables generated by this calculator include all types of expenses of the Air Force and both the Raw and Weighted inflation indices.

The method used in this research for converting the Then Year to Constant Year Dollars was based on first generating the Weighted Inflation Indices, using the year 2009 as the Base Year. For this specific conversion (Then Year to Constant Year Dollars) it is appropriate to use the weighted indices. The cost data set of this research has a range of 11 years, from 1999 to 2009. The last year was used as the Base Year. The use of 2009 as Base Year converts all the Then Year Dollars to 2009 Constant Year Dollars. The category of expenses used in this research is Operations and Maintenance (3400), which incorporates the medical expenses for the members of Air Force. The Air Force Inflation Calculator was used in this research for generating the Weighted Inflation Indices for Operation and Maintenance (3400), for the period 1999 – 2009, using the year 2009 as the Base Year. These inflation indices are given in Table 5.

Table 5. USAF Weighted Inflation Indices Based on OSD (Office of the Secretary of Defense) Raw Inflation Rates - Base Year (FY) 2009

Fiscal Year	Operations & Maintenance (3400)
1999	0.831
2000	0.843
2001	0.855
2002	0.863
2003	0.876
2004	0.897
2005	0.929
2006	0.952
2007	0.976
2008	0.995
2009	1.007

After generating the inflation indices with the year 2009 as Base Year, the second step in the process of converting the Then Year Dollars to Constant Year Dollars is locating the weighted index that corresponds to the Then Year of the provided dollar amount. The third step is the division of the provided dollar amount by this weighted index. For example a dollar amount of $100 in Then Year Dollars of 1999 could be converted into Constant Year Dollars of 2009 by dividing the amount of $100 by the weighted index of 1999. Using Table 5, the amount $100 must be divided by 0.831. The division of $100 by 0.831 equals $120.34 ($100/0.831 = $120.34) and thus the 1999 Then Year Dollars is converted into 2009 Constant Year Dollars. This method of conversion was used for converting the whole cost data set from Then Year Dollars to 2009 Constant Year dollars.

Contingency Analysis

The first data set, based on the AF Web HA questionnaire, was used to detect if there is a relationship between smokers (dependent variable) and pay rank, gender, or age (independent variables). For this purpose, the Contingency Analysis of nominal

variables was used, determining if a relationship exists between two nominal variables. Other statistics such as t-tests, regressions and so on, apply to dependent variables that are continuous.

> The Contingency Analysis structures the data into a two-way table showing the groupings for each of two different variables. Once the contingency table has been constructed, it is easy to examine if the two variables are independent. The statistical test to use in this case is the chi-square test for independence. (Treloar, 2009)

The null hypothesis is that the two variables are independent. In case that the null hypothesis is rejected, when the chi-square value is large and the corresponding p-value is low, then a relationship between the two variables is identified.

JMP®, the statistical tool used in this research, when conducting the Contingency Analysis of two nominal variables, produces the Mosaic Plot, the Contingency Table and the Tests Report. The Tests Report gives the negative log-likelihood for categorical data, the Degrees of Freedom and the R-square (U) value. But the most important part of the Tests Report is the two Chi-square statistical tests of the hypothesis. “The Likelihood Ratio Chi-square test is computed as twice the negative log-likelihood for Model in the Tests table. The Pearson Chi-square is another Chi-square test of the hypothesis that the response rates are the same in each sample category.” (JMP, 2007) The Mosaic Plot is divided into small rectangles and each rectangle is proportional to a frequency count of interest, and in this research each rectangle shows the size of smokers and non smokers for each relative group (pay rank, gender and age), depending on the X-variable used each time. The Contingency Table appears as a simple two-way frequency table and for each factor level of the X-variable there is a row (like two rows for gender, one for males and one for females) and a column for each response level of the Y response (in this research there are two columns, one for the smokers and one for the non smokers). The

Contingency Table provides cell quantities, such as Count, Total%, Row%, Col%, and Expected which were used (in the Microsoft Excel® tool) in this research for a more graphic presentation of the smoking rates in Air Force and how these rates are affected by pay rank, gender, and age.

Pivot Table Analysis

The Pivot Tables option in Microsoft Excel® is one of the most powerful features of Microsoft Excel® and allows rapid, flexible, and dynamic analysis of a data set. The Pivot Tables feature is the most appropriate and quickest way of summarizing lengthy data sets into a compact format. Furthermore, it is a helpful tool which is used to find relationships within data that are hard to discover because of the amount and length of data, and to organize the data into an easier format to chart. In this research Pivot Tables was used with the aim to summarize the cost information of the second data set, after having deflated, and to reveal potential relationships between cost and a group of variables such as pay rank, age, gender, and diseases. The Pivot Table analysis was initially the leading tool for classifying the diseases according to their cost and answering the second research question of this study. In a second phase, the efficiency of the Pivot Tables was used in a very fruitful way and more valuable information was extracted for the compact organization of the initial data set. Apart from the table of the cost ranking of the diseases and the graphical depiction, the Pivot Table analysis enriched this research with tables and graphs exhibiting the total and average annual cost of hospitalization of the ADAF personnel and to the number of medical visits, sorted out by specific socio-demographic characteristics such as pay rank, gender and age, for the period from 1999 to 2009. Additionally, The Pivot Tables furnished this research with information which harmonizes with the subsequent regression analysis of the third research question.

Regression Analysis

Data Sets

The third research question of this study was answered using the Regression Analysis of multiple variables. The data set used for the regression analysis was the cost data, after having been deflated and converted into Constant Year Dollars with year 2009 as the base year. In this data set, Cost is the Y response. One basic step, before defining the x variables and creating dummy variables for the development of the regression modeling, was to analyze the Y response (cost) and see how it looked like in a Histogram plot. The Distribution option in JMP® produced a histogram of cost response and since cost was a continuous variable, the Distribution generated a histogram with a bar chart and an outlier box plot. The histogram demonstrated the existence of outliers, which are equal to extreme values, and indicated the division of the initial data set into subsets of different range of cost. The result of this division was the partition of the data set into five subsets and the construction of 5 regression models, one for each subset and its correspondent range of cost. The presentation of the Distribution of the initial data set and the five subsets is included in Appendix A.

Dependent Variable

The dependent variable, of the regression analysis developed in this research, is the cost of hospitalization of active duty members of Air Force, for the period 1999 – 2009, because of smoking related diseases. The initial data set did not present any uniformity because of the great range of cost, and for this reason it was subsequently partitioned into five subsets, using the range of cost as criteria. In this way, five regression models were built, each one for a different scale of cost, but the dependent variable for all the models remains the cost.

Independent Variables

The initial data set consisted of a small number of columns, which were the primary independent variables. Gender, age, pay rank, and primary diagnosis were the main columns of the data set, which were the original independent variables. Gender, pay rank and primary diagnosis were nominal variables while age was a continuous variable. The original independent variables were inserted into JMP and used for generating dummy variables, which recoded the independent variables and distinguished them into different treatment groups, taking the values 0 or 1 in order to indicate the absence or presence of some of their categorical effect. The following sections define each of the independent variables.

- Age. The original independent variable ‘Age” was divided into five dummy variables, each one corresponding to a different age group. There were five age dummy variables for the following age groups: Age 17-24, Age 25-34, Age 35-44, Age 45-60 and Age 61-87.
- Gender_1. The dummy variable Gender_1 was derived from the original categorical variable Gender, with the value 1 assigned to males and 0 to females.
- Enlisted. This dummy variable was created from the original categorical variable Pay_Rank and split all the pay ranks into two more general subcategories: enlisted and officers. All the enlisted ranks were given value 1 and the ranks of officers received value 0.
- CD, OCS. This dummy variable derived from the original categorical variable Pay_Rank and ascribed value 1 to the ranks of Cadet (CD) and Officer Candidate School (OCS) and value 0 to the rest of the ranks.

- E1, E2, E3, E4. Dummy variable derived from the original categorical variable Pay_Rank and ascribed value 1 to the ranks of Airman Basic (E1), Airman (E2), Airman First Class (E3), Corporal (E4), and value 0 to the rest of the ranks.
- E5, E6. Another dummy variable originated from the original Pay_Rank, which assigned value 1 to the ranks of Staff Sergeant (E5) and Technical Sergeant (E6) and value 0 to the rest of the ranks.
- E7, E8, E9. This dummy variable was originated from the original independent variable Pay_Rank and attributed value 1 to the ranks of Master Sergeant (E7), Senior Master Sergeant (E8), and Chief Master Sergeant (E9), and value 0 to the rest of the ranks.
- O1, O2, O3. This dummy variable was created from the variable Pay_Rank and ascribed value 1 to the ranks of Second Lieutenant (O1), First Lieutenant (O2), and Captain (O3), and value 0 to the rest of the ranks.
- O4, O5, O6. Another dummy variable derived from Pay_Rank, which assigned value 1 to the ranks of Major (O4), Lieutenant Colonel (O5), and Colonel (O6), and value 0 to the rest of the ranks.
- O7, O8, O9, O10. This dummy variable originated from Pay_Rank and attributed value 1 to the ranks of Brigadier General (O7), Major General (O8), Lieutenant General (O9) and General (O10), and value 0 to the rest of the ranks.
- Diseases. This continuous variable was a derivative of the original variable Primary Diagnosis, and it was used as a stepping stone for the creation of the dummy variables, associated with the cost and the frequency of the appearance of the diseases. The creation of this variable demanded the use of the table

with the classification of the diseases according to their cost. This assortment was part of the research done in the Pivot Table Analysis. This variable contains values from 1 to 11, according to the cost rating of diseases presented and described in the Pivot Table Analysis of Chapter IV.

- Dummy variables for Diseases, counted from 1 to 11. These dummy variables were created from the above described variable "Diseases", and they are associated with the cost and the frequency of appearance of the most prevalent diseases related to smoking with the highest cost (the eleven diseases that cost the most to the Air Force for the period 1999-2009, assorted by the Pivot table Analysis presented in Chapter IV). In each of the dummy variables, value 1 is attributed to the disease referred to the dummy variable, and value 0 to the rest of the diseases.

Summary of Data

The section of the regression analysis used the data set of the cost of hospitalization of active duty members of the United States Air Force due to smoking related diseases, for the period 1999 – 2009. The diseases used in this analysis were the ones registered in Primary Diagnosis and only the most prevalent diseases associated with smoking. The dependent variable was the cost, which was first converted into Constant Year Dollars, using the weighted inflation indices with base year 2009, of the Inflation Calculator of Air Force. The initial data set was partitioned into five data subsets because of the wide range of cost and the non-existence of uniformity. The independent variables used for building the five regression models, one for each data subset, were five dummy variables for age, gender, separation of the enlisted and officers' ranks; seven dummy variables for seven pay rank groups; and

eleven dummy variables related to the cost rating and to the frequency of the appearance of the diseases used in this research.

Building the Models

The regression analysis in this research was elaborated by applying Ordinary Least Squares (OLS) regression to the data subsets described previously. The OLS regression is the most common form of linear regression which maximizes the amount of explained variation in the dependent variable by minimizing the sum of squared distances between the observed responses in the dataset and the responses predicted by the linear approximation. The general form of OLS regression is given by the formula below:

$$y_i = \beta_o + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi} + \varepsilon_i$$

In this formula, y_i represents the dependent variable or the response, for a specific observation i. The independent variables, sometimes called control variables, are given by x_{1i}, x_{2i}, …, x_{pi} for p i observations, The coefficients of each of the p parameters are represented by β_1, β_2, …., β_p and the intercept or constant term in this equation is given by β_0. The ε_i represents the residual, which is the difference between the actual and the estimated function value.

Search for Predictive Variables

The procedure used for the selection of the most predictive regressors (x variables) was based on the t-statistics of the regressors and their p-values. When regressing the model with all the regressors, the t-ratio in the Parameter Estimates, gives the t-statistic test for a test of the null hypothesis that $\beta_i = 0$. If the null hypothesis is true for a regressor x_i, then this regressor has no effect on the regressant Y and can be deleted from the regression model. The column "Prob>|t|" in Parameter Estimates gives the p-values for a two-tailed test of the null hypothesis that $\beta_i = 0$.

Small p-values in this column sign that the corresponding regressor (x variable) does have an effect on the regressant Y and can be used in the regression model. In this research, for the OLS regression models, x variables were used whose t-statistic p-values were less than 0.05.

Model Diagnostics

The assessment of the appropriateness of a linear model, built through the regression analysis process, is based on some statistical indices given by the results of the regression and on some diagnostic tests. The indices consist of part of the results of the building process of the linear model, while the diagnostic tests are done by the researcher, in order to assess if the model fits the data well. The following sections define which indices were taken into consideration and what tests have been executed for the assessment of the goodness-of-fit of the models, built in this research.

R – Squared (R^2) and Adjusted R – Squared (Adj. R^2)

"By definition, R^2 is the fraction of the total squared error that is explained by the model. Thus values approaching one are desirable. But some data contain irreducible error, and no amount of modeling can improve on the limiting value of R^2" (Annis, 2008). R^2 is the relative measure of the predictability of a model and takes values between 0 and 1. The higher and closer to 1 the R^2 is, the better the model. The R^2 measures how well the linear model approximates the real data. Referring to the values the R^2 can take, an R^2 equal to 1 means that the regression line perfectly fits the data. The R^2 increases as more variables are added to the model. Here lies the drawback of the misleading use of the R^2: an increased number of variables included in the model would erroneously increase the value of R^2. For this reason, an alternative R^2 is used for the assessment of a model and this is the Adjusted R^2. The Adjusted R^2 is an alternative approach of R^2, but it penalizes the statistic when

additional variables are added to the model. The Adjusted R^2 is always less than or equal to R^2 and increases only when a new term inserted in the model improves it.

Influential Data Points – Cook's Distance

The outcome and accuracy of a least squares regression analysis could be distorted by the existence of one or more influential points. Influential points are data points with large effect on the slope of the regression line and on the estimated values and p-values of the independent variables. Including an influential point in the building procedure of a least squares regression model, could affect the accuracy of the model and distort the statistical significance of the regressors (independent variables). Cook's Distance is the diagnostic test used for detecting potential influential points. Any data point which, in the Overlay Plot of Cook's Distance, presents a value greater than 0.25, indicates that it might be a potential influential point and should be evaluated and eventually removed. Any points removed because of a large Cook's Distance value are mentioned in Chapter IV, and the correspondent model has been re-built without these points and the new results are given and compared to the previous ones.

Tests for Normality and Constant Variance

The diagnostic test used in this research for normality of model residuals is the Shapiro Wilk test. The Shapiro Wilk test demands the distribution of the studentized residuals, which is used for the test of normality. The test is based on the null hypothesis that the residuals are normally distributed, and thus the data are normally distributed. A p-value greater than 0.05 fails to reject the null hypothesis and in this case, the residuals are normally distributed and the data set is well modeled. On the contrary, a p-value less than 0.05 rejects the null hypothesis and in this case, the studentized residuals are not normally distributed and the data not well modeled.

The Breusch-Pagan test is used to test for homoscedasticity in a linear regression model. The key assumption in the Breusch-Pagan test is that the variance of the errors is constant across the observations. If the errors present constant variance, then they are called homoscedastic. For the assessment of this assumption the residuals are plotted and the null hypothesis is that the residuals exhibit constant variance. A p-value larger than 0.05 fails to reject the null hypothesis and in this case the errors exhibit constant variance and they are homoscedastic.

Multicollinearity

Multicollinearity occurs when two or more variable are collinear, meaning that they are linearly related and they measure substantially the same thing. In this case, the overall p-value of the model might be low but neither of the x variables makes a significant contribution to the model. The assessment method used for multicollinearity in this research is the Variation Inflation Factors (VIFs). High values of VIF scores, and particularly VIF scores larger than 5, mean that the fit of the model is affected by multicollinearity and variables with high VIF scores should be omitted and combined with other variables, for a better contribution to the model.

The Final Ordinary Least Squares Regression Models

, Five models were built, each one corresponding to a specific subset of the initial cost data set and to a particular range of cost. For every model presented in Chapter IV, the following properties are explained:

- The independent variables chosen every time and used in every model
- Information extracted from the Parameter Estimates, such as the estimated coefficients of each independent variable, the standard errors, the t-ratios, the p-

values, which must be less than 0.05 for each variable used in each model in order for the variable to be predictive and statistically significant, and the VIF scores.

- Information extracted from the Summary of Fit Section, such as the R^2 and the Adjusted R^2, which provides information about the goodness of fit of the model and how well the regression line fits the real data points.

- The overall p-value of the model, derived from the Analysis of Variance. A p-value less than 0.05 rejects the null hypothesis that the means are the same, and in this case, there is significant difference between the different variables.

- The results of the Shapiro Wilk and the Breusch-Pagan tests.

Summary

This chapter outlines the methodology used for analyzing the data sets used in this study and presents what tools were used in each data set and for the investigation of each research question. The first research question was answered through the analysis of the AF Web HA data base, using the Contingency Analysis and the visual presentation of the results with the aid of Microsoft Excel®. The investigation of the other two research questions has been conducted with the use of the data set referencing the cost of hospitalization of active duty members of U.S. Air Force. Specifically, the second research question was answered through the usage of Pivot Tables, which supplied the research with a list of the diseases with the highest cost and with additional valuable information about the total and average cost for each

group with specific socio-demographic characteristics. The third research question was investigated with the assistance of Ordinary Least Squares Regression Analysis. For this regression analysis the most prevalent diseases related to smoking were used, registered in the Primary Diagnosis, and the initial cost data set was partitioned in five subsets, each one corresponding to a particular range of cost.

The subsequent chapter, Chapter IV, presents and summarizes the results of each type of analysis used in this study, and answers each research question framed in Chapter I. Chapter V emphasizes and highlights the results obtained through the investigation of each research question.

IV. Results and Analysis

Overview

This chapter details the results of each of the methods discussed in Chapter III. First, the Contingency Analysis is presented with all the tables and figures extracted from this analysis. Information from the Mosaic Plots and from the Contingency Tables are used in Microsoft Excel®, in order to be presented in a more visual and descriptive way. Next, the Pivot Table Analysis presents a list with the most prevalent diseases related to smoking of the highest cost of hospitalization. Similarly, in this analysis, additional information related to total and average cost of groups sorted by age, gender and pay rank, is presented in graphs, executed with the aid of Microsoft Excel®. Finally, in the Regression Analysis section of this research, the models built for each range of cost and their correspondent results are presented. The results will be focused mainly on the predictive variables and the power of predictability of each model.

Contingency Analysis

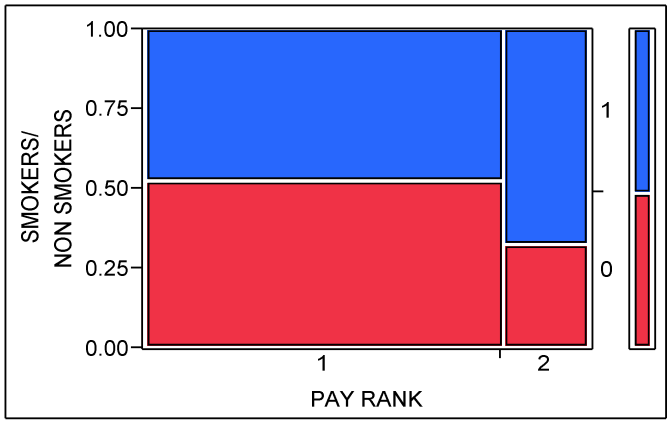
The Contingency Analysis report shows a Mosaic Plot, a Contingency Table and a Tests report. The Mosaic Plot is "a graphical representation of the two-way frequency table of Contingency Table" (JMP 2007). The Mosaic plot consists of rectangles. The area of each rectangle is proportional to the proportions of the Y variable in each level of the X variable (JMP 2007). The Contingency Table is a two-way frequency table, with a row for each factor level and a column for each response level (JMP, 2007). The Tests Report presents the results for two tests to determine whether the response level rates are the same across X levels (JMP, 2007). In this chapter, the results of three Contingency Analyses are presented (associated to the three socio-demographic variables of pay rank, gender and age) and emphasis is given

to the two Chi-Square tests, which actually are the drivers for the existence or relationship between smoking and pay rank, gender and age.

Smokers / Non-Smokers versus Pay Rank

The first Contingency Analysis refers to the relationship between Smoking and Pay Rank. The Mosaic Plot, the Contingency Table and the Tests Report for this Contingency Analysis are shown below in Figure 1.

Mosaic Plot



Contingency Table

PAY RANK by SMOKERS/NON SMOKERS

Count Total % Col % Row %	**SMOKERS**	**NON SMOKERS**	
ENLISTED	170521 42.46 87.30 52.41	154814 38.55 75.05 47.59	325335 81.01
OFFICERS	24798 6.17 12.70 32.52	51466 12.82 24.95 67.48	76264 18.99
	195319 48.64	206280 51.36	401599

Tests Report

N	**DF**	**-LogLike**	**RSquare (U)**
401599	1	4992.3428	0.0179

Test	**ChiSquare**	**Prob>ChiSq**
Likelihood Ratio	9984.686	0.0000*
Pearson	9791.725	0.0000*

Fisher's Exact Test	**Prob**	**Alternative Hypothesis**
Left	1.0000	Prob(SMOKERS/NON SMOKERS=1) is greater for PAY RANK=1 than 2
Right	0.0000*	Prob(SMOKERS/NON SMOKERS=1) is greater for PAY RANK=2 than 1
2-Tail	0.0000*	Prob(SMOKERS/NON SMOKERS=1) is different across PAY RANK

Figure 1: Contingency Analysis of Smokers/Non-Smokers versus Pay Rank

In the Mosaic Plot, Pay Rank=1 refers to Enlisted and Pay Rank=2 refers to Officers. Moreover, the red color represents the smokers and the blue the non smokers. It is visually obvious that the enlisted are more numerous than officers. In

addition, more than 50% of the enlisted are smokers, while the majority of the officers are non-smokers. See Table 6 and Figure 2.

Table 6: Comparison of Smokers and Non Smokers for each Pay Rank

	ENLISTED	OFFICERS
SMOKERS	52.41%	32.52%
NON SMOKERS	47.59%	67.48%

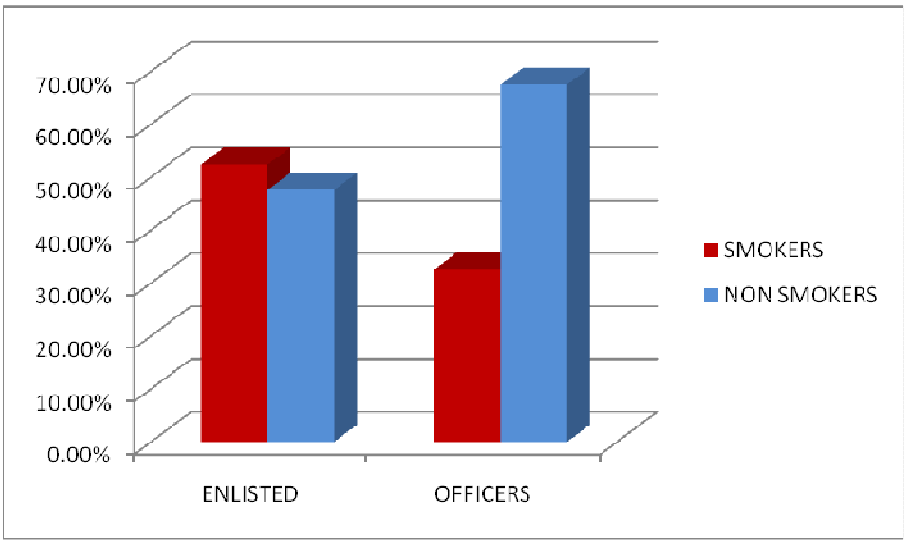


Figure 2: Comparison of Smokers and Non Smokers for each Pay Rank

From the table and graph above it can be deduced that the majority of enlisted members smoke, while the exact opposite phenomenon is observed among the officers' ranks. Only 32.52% of the officers' population consists of smokers, while the non smokers represent the high rate of 67.48%.

The Contingency Table of Figure 1 gives a couple of percentages for the smoking status of the Air Force sample of the data set used in the Contingency Analysis, and of each group separately. The smoking status of the sample is given in Table 7 and Figure 3.

Table 7: Smoking Status of the Air Force Population

	SMOKERS	NON-SMOKERS
AIR FORCE	48.64%	51.36%

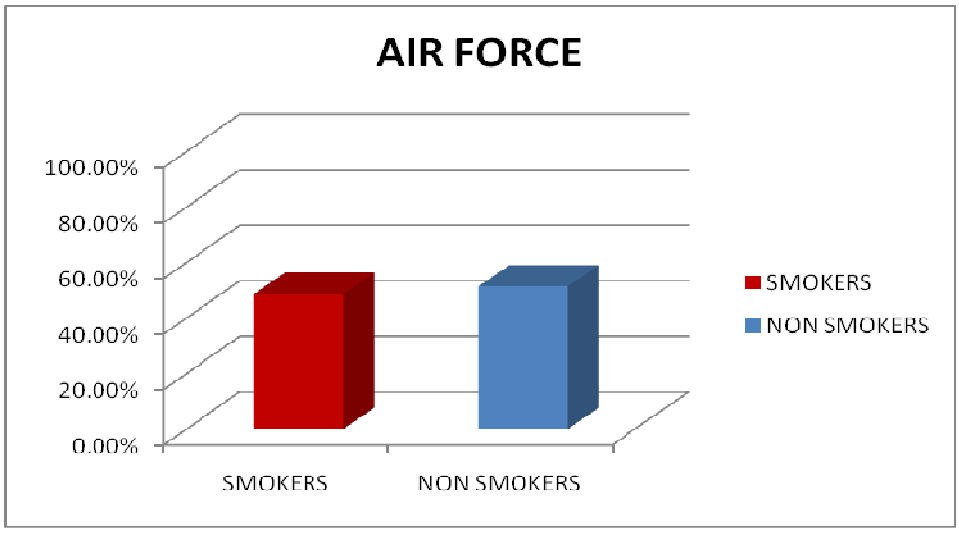


Figure 3: Smoking Status of the Air Force Population

The table and graph show that 48.64% of the ADAF population of the data set is smoking and 51.36% is not smoking. This leads to the conclusion that the Air Force population is divided into two large, almost equal, groups: smokers and non smokers. The percentage of smokers (48.64%) consists of 42.46% enlisted and 6.17% officers. The distribution of 48.64% is given by Table 8 and Figure 4.

Table 8: Distribution of the Smokers' Population

	ENLISTED WHO SMOKE	OFFICERS WHO SMOKE	TOTAL
AIR FORCE	42.46%	6.17%	48.64%

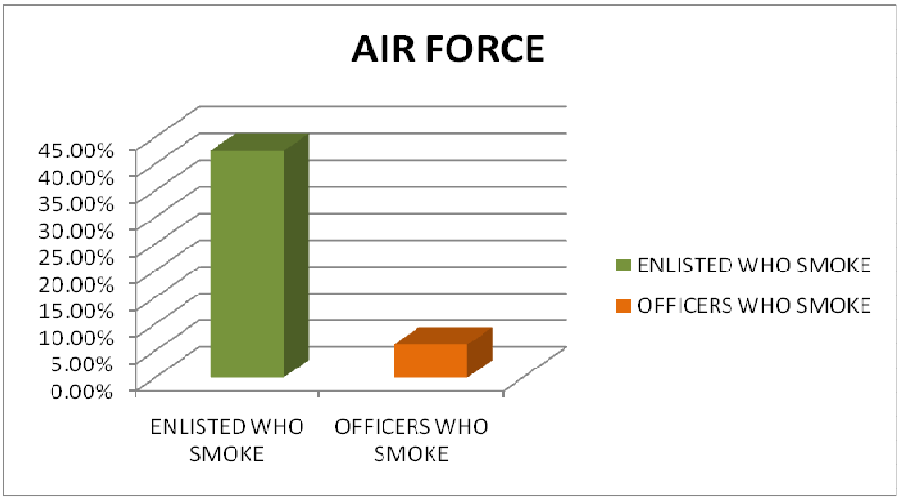


Figure 4: Distribution of the Smokers' Population

The same picture of the distribution of the smokers' population is given, but more detailed, if the smokers are considered a population of their own. The following graph (see Figure 5) presents this distribution. The graph shows that 87.30% of the smokers' population consists of enlisted and only 12.70% of officers. The percentages are indeed alarming and show that smoking is more prevalent among the enlisted.

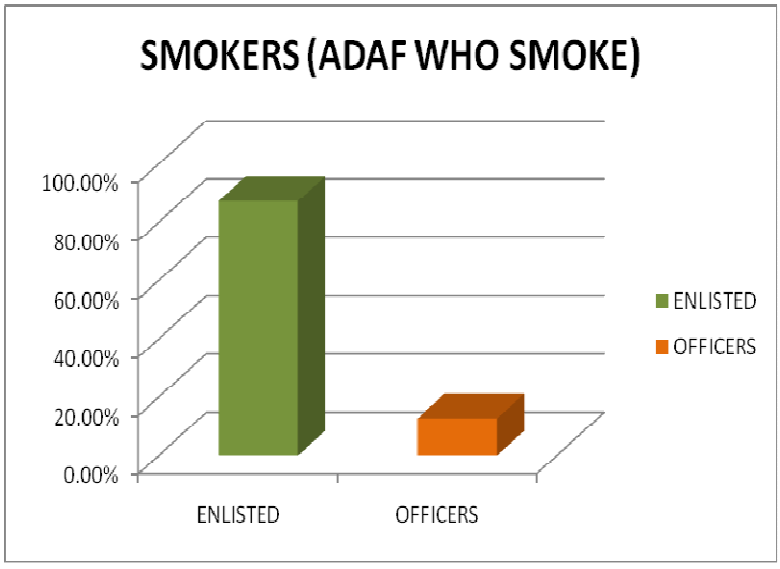


Figure 5: Distribution of the ADAF members that smoke

The Tests Report, which is the last part of the Contingency Analysis of Smokers / Non-Smokers versus Pay Rank, gives the results of Pearson and Likelihood Ratio tests. Both of these test whether the two variables, in this case smoking and pay rank, are independent or not. Both Pearson and Likelihood Ratio tests have the same assumptions, and the null hypothesis here is that the variables are independent, meaning that there is no relationship between them, and more specifically, smoking is not affected by pay rank. The Chi-Square test compares the observed cell frequencies with expected cell frequencies, and assumes a null hypothesis that the variables are independent (JMP, 2007). The expected values are calculated by multiplying the row total and column total, and then divide by the grand total. "The Chi-Square test is always valid if there are no empty cells (no cells with a cell frequency of 0), and if the expected cell frequency for all cells is five or greater" (JMP, 2007). Figure 6 gives the Contingency Table of Smokers and Non-Smokers for each pay rank with the observed and expected frequencies, and the results of the Tests Reports.

Contingency Table

PAY RANK By SMOKERS/NON SMOKERS

Count Expected	**SMOKERS**	**NON SMOKERS**	
ENLISTED	170521 158228	154814 167107	325335
OFFICERS	24798 37091.2	51466 39172.8	76264
	195319	206280	401599

Tests Report

N	**DF**	**-LogLike**	**RSquare (U)**
401599	1	4992.3428	0.0179

Test	**ChiSquare**	**Prob>ChiSq**
Likelihood Ratio	9984.686	0.0000*
Pearson	9791.725	0.0000*

Fisher's Exact Test	**Prob**	**Alternative Hypothesis**
Left	1.0000	Prob(SMOKERS/NON SMOKERS=1) is greater for PAY RANK=1 than 2
Right	0.0000*	Prob(SMOKERS/NON SMOKERS=1) is greater for PAY RANK=2 than 1
2-Tail	0.0000*	Prob(SMOKERS/NON SMOKERS=1) is different across PAY RANK

Figure 6: Contingency Table and Tests Report of Smokers/Non-Smokers versus Pay Rank

All the cells of the expected cell frequency of the Contingency Table of Figure 6 are greater than five, fact that indicates that the Chi-Square test is a valid test. Pearson test uses the observed and expected cell frequencies, while the Likelihood Ratio test uses a more complex formula (Schlotzhauer, 2007). The column Prob>ChiSq gives very low p-values for both tests. These very low p-values, which are less than the significance level of 0.01, give enough evidence to reject the null hypothesis of independence between smoking and pay rank, and indicate a relationship between the two variables.

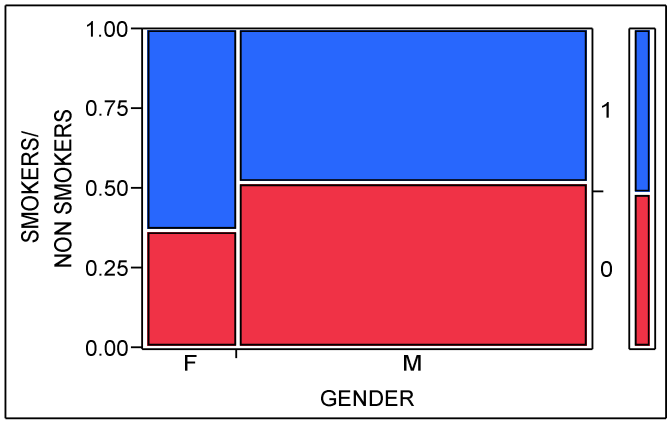
The Fisher's Exact Test is more suitable for small frequency tables and JMP® performs this test for 2x2 tables, but it cannot be executed for larger tables. JMP® presents the results for both one-sided test and two-sided test. The 2-tail p-value is the more suitable test and tests for independence between the two variables, and is

interpreted the same way as the Chi Square test. The p-value of 2-tail test is very low, less than the significance level, and in this case the null hypothesis is again rejected, meaning that a relationship between the two variables, smoking and pay rank, exists.

Smokers/ Non-Smokers versus Gender

In this Contingency Analysis, the existence of relationship between smoking and gender is examined. The Mosaic Plot, the Contingency Table, and the Tests Report of this Contingency Analysis are given below in Figure 7.

Mosaic Plot



Contingency Table

GENDER by SMOKERS/NON SMOKERS

Count Total % Col % Row % Expected	**SMOKERS**	**NON SMOKERS**	
FEMALES	31286 7.79 16.02 36.69 41467.5	53976 13.44 26.17 63.31 43794.5	85262 21.23
MALES	164033 40.84 83.98 51.85 153852	152304 37.92 73.83 48.15 162485	316337 78.77
	195319 48.64	206280 51.36	401599

Tests Report

N	**DF**	**-LogLike**	**RSquare (U)**
401599	1	3123.7184	0.0112

Test	**ChiSquare**	**Prob>ChiSq**
Likelihood Ratio	6247.437	0.0000*
Pearson	6178.605	0.0000*

Fisher's Exact Test	**Prob**	**Alternative Hypothesis**
Left	0.0000*	Prob(SMOKERS/NON SMOKERS=1) is greater for GENDER=F than M
Right	1.0000	Prob(SMOKERS/NON SMOKERS=1) is greater for GENDER=M than F
2-Tail	0.0000*	Prob(SMOKERS/NON SMOKERS=1) is different across GENDER

Figure 7: Contingency Analysis of Smokers/Non-Smokers versus Gender

In the Mosaic Plot, the red color represents the smokers and the blue color the non-smokers. From the Mosaic Plot, it is easily seen and understood that the majority of the ADAF population are males. Furthermore, the Mosaic Plot shows that the majority of females in the Air Force are not smokers, while 51.85% of the male population is smokers. The picture of smoking status between genders in the Air Force is better presented by Table 9 and Figure 8.

Table 9: Comparison of Smokers and Non Smokers for each Gender

	FEMALES	**MALES**
SMOKERS	36.69%	51.85%
NON SMOKERS	63.31%	48.15%

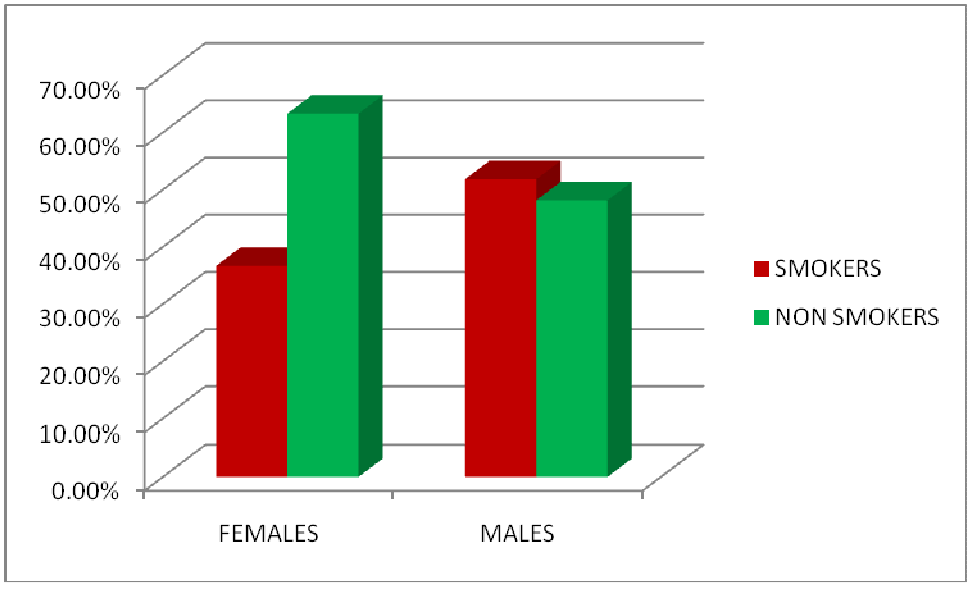


Figure 8: Comparison of Smokers and Non Smokers for each Gender

The data set used for the Contingency Analysis consists of 401,599 ADAF members and among them, 78.77% are males and 21.23% are females. These numbers are taken from the Contingency Table, where more detailed information is provided for this study and is shown below. Since the greatest part of the Air Force population is comprised of males, it comes naturally that the percentage of smokers among the male population will be by far higher than the female population. The difference of the percentages of smokers between the two genders is remarkably large and is shown in Table 10 and Figure 9.

Table 10: Distribution of Smokers' Population by Gender

	FEMALES WHO SMOKE	MALES WHO SMOKE
AIR FORCE	7.79%	40.84%

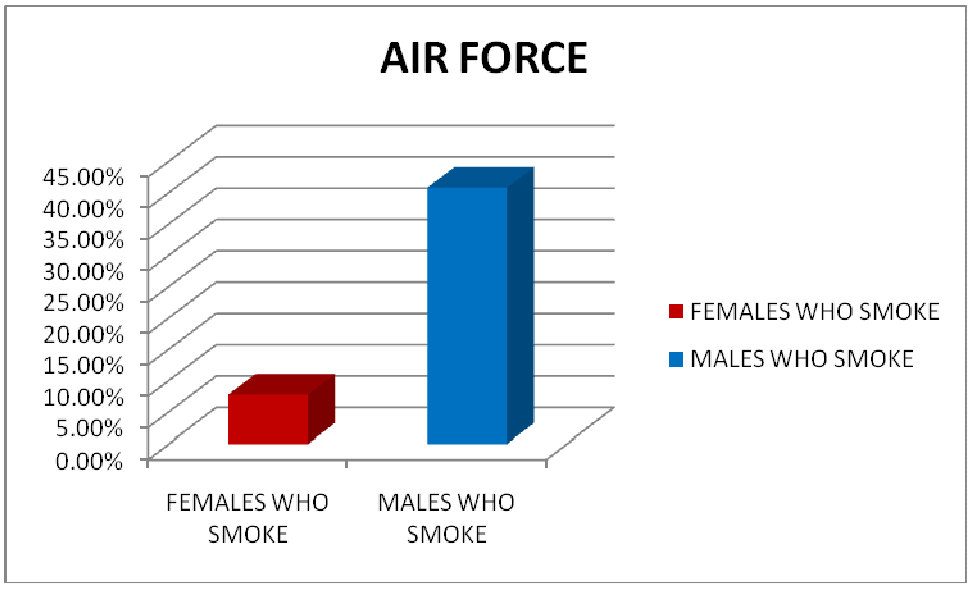


Figure 9: Distribution of Smokers' Population by Gender

The distribution of Smokers' Population references the distribution of this population out of the whole population of the Air Force of the data set used in the Contingency Analysis. This distribution shows that out of the total population of this data set, 7.79% consists of female smokers and 40.84% of male smokers. This is better shown if the smokers are considered a population of their own. The distribution of the population of the ADAF members that smoke is given below in Figure 10.

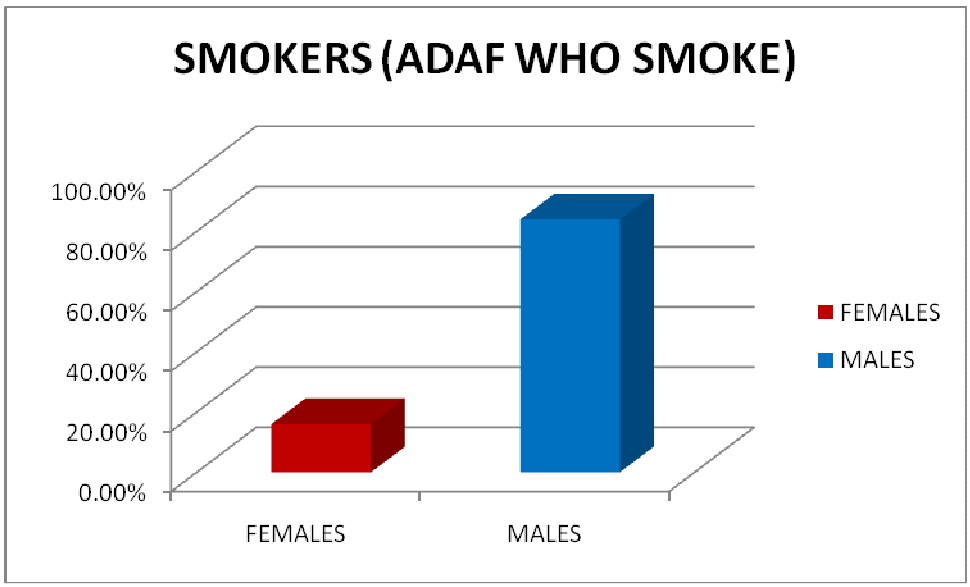


Figure 10: Distribution of the ADAF members that smoke by Gender

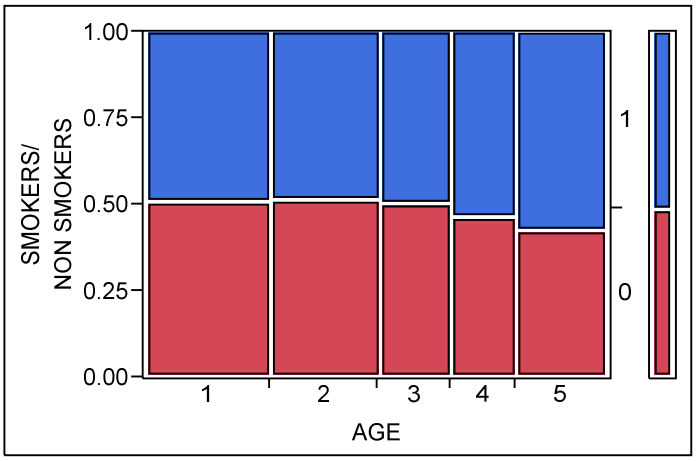
The graph in Figure 10 shows that 83.98% of the smoking population of the Air Force consists of males, and only 16.02% females. The gap between the two genders presented in this case is even larger, when analyzed according to only that part of the population of the Air Force which smokes.

The Contingency Table gives also the expected frequencies. In this Contingency Table, all the expected frequencies are larger than five, and this is an indication that the Chi-Square test, assuming the null hypothesis that the variables are

independent, is a valid test. The last part of the Contingency Analysis is the Tests Report, where the assumption of independence between the two variables is tested, which, in this case, the two variables are smoking and gender. The null hypothesis is that the two variables are independent and in this case is rejected, since the p-values of the Pearson and Likelihood Ratio are very low and are below the significance level of 0.01. This leads to the conclusion that the two variables are dependent and gender affects the smoking status of the ADAF population. This conclusion is further confirmed by the 2-tail p-value shown in Fisher's Exact Test Results. The 2-tail p-value is really low and is less than 0.10 of the significance level. The null hypothesis in this case is again rejected and the two variables are dependent, meaning that gender does affect the smoking status of ADAF members.

Smokers/ Non-Smokers versus Age

The last part of the Contingency Analysis includes the investigation of the existence of a relationship between smoking and age. This section of the Contingency Analysis examines if age is a variable that influences smoking, by testing which age range of ADAF personnel smokes the most. The Mosaic Plot, Contingency Table and Tests Report are given below in Figure 11.



Contingency Table

AGEG by SMOKERS/NON SMOKERS

Count Total % Col % Row % Expected	**SMOKERS**	**NON SMOKERS**	
Age 17-24	55993 13.94 28.67 50.81 53598.6	54212 13.50 26.28 49.19 56606.4	110205 27.44
Age 25-29	48510 12.08 24.84 51.46 45844.6	45752 11.39 22.18 48.54 48417.4	94262 23.47
Age 30-34	31530 7.85 16.14 50.00 30667	31525 7.85 15.28 50.00 32388	63055 15.70
Age 35-39	26451 6.59 13.54 46.54 27643.8	30388 7.57 14.73 53.46 29195.2	56839 14.15
Age 40+	32835 8.18 16.81 42.51 37565	44403 11.06 21.53 57.49 39673	77238 19.23
	195319 48.64	206280 51.36	401599

Tests Report

N	**DF**	**-LogLike**	**RSquare (U)**
401599	4	911.14729	0.0033

Test	**ChiSquare**	**Prob>ChiSq**
Likelihood Ratio	1822.295	0.0000*
Pearson	1816.918	0.0000*

1= age 17 – 24 , 2= age 25-29, 3= age 30-34, 4= age 35- 39, 5= age 40 +

Figure 11: Contingency Analysis of Smokers/Non-Smokers versus Age

The Mosaic Plot shows the distribution of ADAF personnel according to their age and their smoking status. The red color corresponds to that part of the personnel that smokes and AGEs 1-5 correspond to different age ranges. Number one represents the age range from 17 to 24 years old, number 2 the age range from 25 to 29, number three the age range from 30 to 34, number four the age range from 35 to 39 and number five the age range from 40 years old and up. From the Mosaic Plot it is seen that the largest part of ADAF personnel belongs to the age groups of 17 to 24 and 25 to 29, and more than half of the population of these age groups is smoking. A better presentation of the Mosaic Plot is given in Table 11 and Figure 12, where each age group is divided into smokers and non-smokers, and the distribution of the population of each age group is better displayed.

Table 11: Comparison of Smokers and Non Smokers for each Age Group

	Age 17-24	**Age 25-29**	**Age 30-34**	**Age 35-39**	**Age 40+**
SMOKERS	50.81%	51.46%	50%	46.54%	42.51%
NON SMOKERS	49.19%	48.54%	50%	53.46%	57.49%

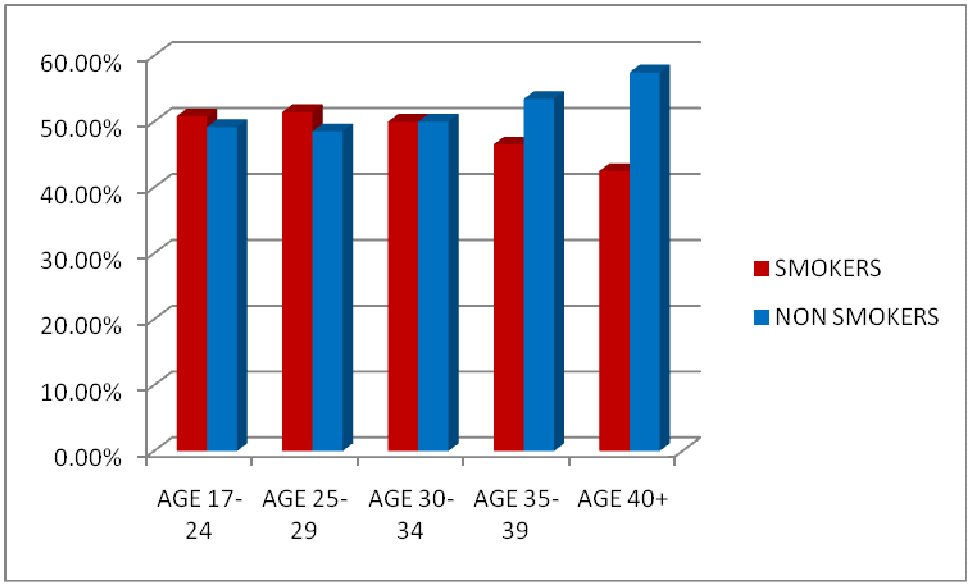


Figure 12: Comparison of Smokers and Non Smokers for each Age group

From Table 11 and Figure 12 it is deduced that more than half of the population of the first three age groups consists of smokers. Specifically, the age group of 30 to 34 years old is nearly evenly split between smokers and non-smokers. Younger age groups smoke more than middle age groups. The age groups of 35 to 39 and 40 years and up do not smoke that much, but even in these age groups, the percentages of smokers are not very low. The graph shows that almost half of the population of ADAF is smoking and this is alarming.

The Contingency Table provides valuable information which is related to which of the age groups smokes more and how the smoking population is distributed. It has been shown above, in the first part of the Contingency Analysis, that 48.64% of the Air Force population of the data set used in this part of this study is comprised of smokers. In Table 12 and Figure 13, the apportionment of 48.64% of smokers is displayed according to the five age groups.

Table 12: Apportionment of Smokers to five age groups

	ADAF aged 17-24 who smoke	**ADAF aged 25-29 who smoke**	**ADAF aged 30-34 who smoke**	**ADAF aged 35-39 who smoke**	**ADAF aged 40+ who smoke**	**Total**
Air Force	13.94%	12.08%	7.85%	6.59%	8.18%	48.64%

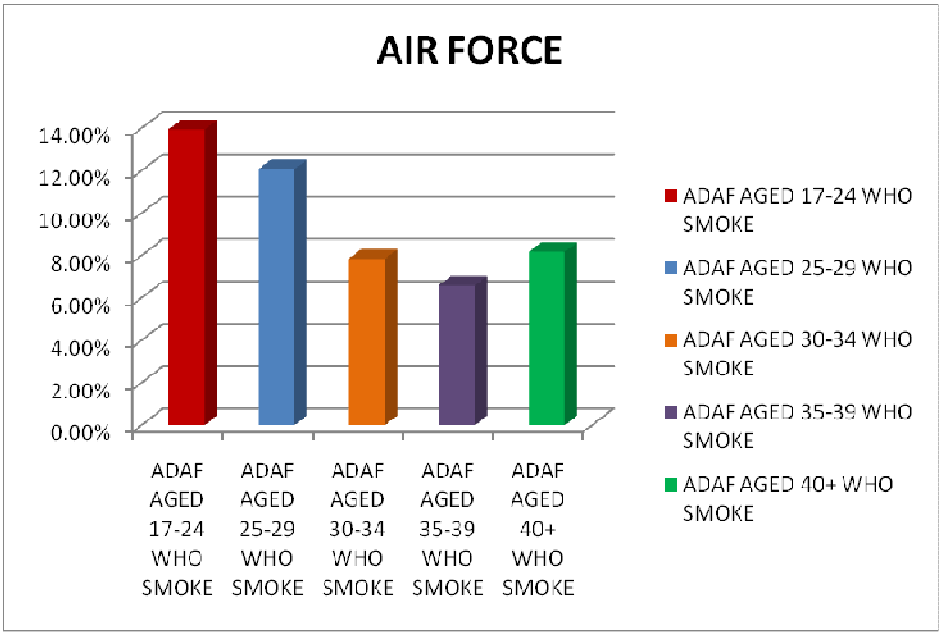


Figure 13: Apportionment of Smokers to five age groups

Figure 13 displays graphically the apportionment of smokers to five age groups and it is shown that the age group that smokes more than the others is the age group of 17 to 24 years old. On the other hand, the group that smokes the least is the one referencing the ages from 35 to 39 years old. In Table 12, where the percentages of smokers for each age group are presented, one can see that the percentages of smokers corresponding to the age groups of 17 to 24 and 25 to 29 are close to each other and together they constitute 26.02% of the smoking population out of the whole ADAF population of the data set. This is alarming for those who investigate and research the smoking issue in Air Force. More attention should be given to the young age ranges, where smoking is most prevalent.

The same picture of the smoking population is given, if smokers of the Air Force are considered a population of their own. The following Table 13 and Figure 14 show in a more detailed way the distribution of the ADAF personnel that smokes according to the five age groups.

Table 13: Distribution of the ADAF members that smoke to five age groups

	Age 17-24	Age 25-29	Age 30-34	Age 35-39	Age 40+
SMOKERS (ADAF WHO SMOKE)	28.67%	24.84%	16.14%	13.54%	16.81%

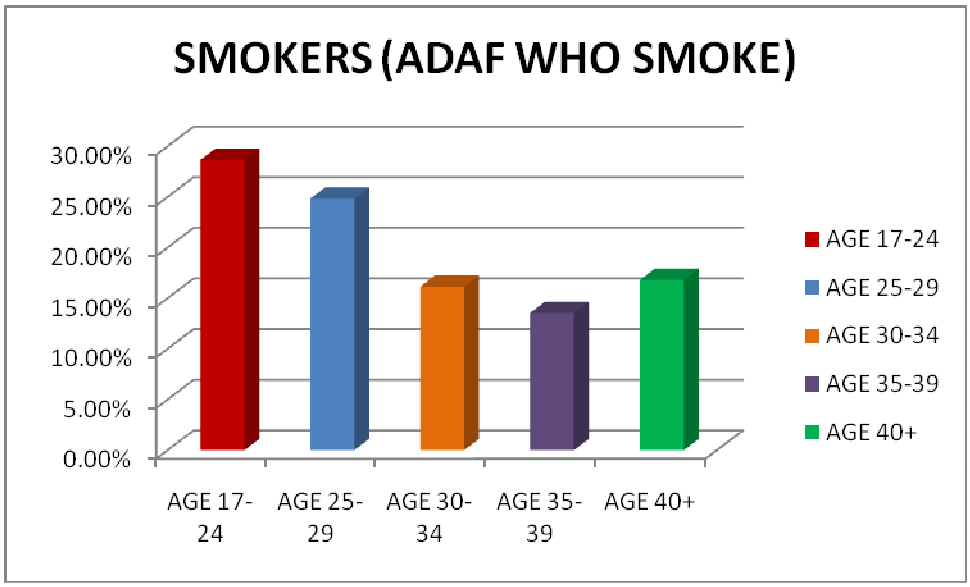


Figure 14: Distribution of the ADAF members that smoke to five age groups

Table 13 and Figure 14 give a more detailed picture of the distribution of the part of the ADAF personnel that smokes. The age group that smokes more than the others is the age group of 17 to 24 years old and the group that smokes the least is the group of 35 to 39 years old. If the percentages of smokers from the age groups 17 to 24 and 25 to 29 years old are added together, they constitute 53.51% of the smoking population of the Air Force. This means that more than half of the population of smokers in the Air Force consists of young people from 17 to 29 years old. The most productive part of the population of the Air Force is the group that smokes the most

and this is exceptionally alarming for the future health and quality of the USAF personnel.

In this Contingency Analysis, all the expected frequencies displayed in the Contingency Table, are larger than five, fact that points out that the Chi-Square test, testing the independence between the variables with the null hypothesis that the variables are independent, is a valid test. In the section of Tests Report, there is no Fisher's Exact Test Results, since in this case there is no a 2x2 Contingency Table. The Tests Report includes only the Pearson and Likelihood Ratio p –values. Both Pearson and Likelihood Ratio p-values are very low and less than the significance level of 0.01. There is enough evidence to reject the null hypothesis of independence between the two variables, where in this part of the Contingency Analysis, the two variables are smoking and age. Smoking and age are dependent and this means that age influences the smoking status of the ADAF personnel.

Pivot Table Analysis

The Pivot Table Analysis is based on another data set, which includes the cost of hospitalization of ADAF personnel due to diseases related to smoking. As mentioned in Chapter III, this data set was restricted to the most prevalent diseases related to smoking according to SAMMEC, and in this study only those diseases that had been registered in the Primary Diagnosis were used. The range of time of this data set covers the period from 1999 to 2009, and all the dollar values associated with the total cost of hospitalization because of diseases related to smoking, are expressed in Constant Year Dollars with base year as the year 2009. The Pivot Table Analysis begins with the presentation of a hierarchical list of the diseases with the highest cost, which is used later in the Regression Analysis for the creation of Dummy Variables. Furthermore, the same data set can be manipulated very easily with the aid of Pivot

Tables, a tool of Microsoft Excel®, tables and graphs are created, displaying additional information about the cost of hospitalization through the period 1999-2009 and the total and average cost of hospitalization for groups with specific socio-demographic characteristics related to age, gender, and pay rank.

Most Prevalent Diseases Related to Smoking and their Cost

As mentioned before, the data set used in the Pivot Table and Regression Analysis was narrowed to the most prevalent diseases related to smoking, according to a list of 18 diseases provided by SAMMEC. The total cost of hospitalization of those diseases was added throughout the years 1999-2009, and the result of this summation was the following list of the most prevalent diseases related to smoking with the highest cost, given in Table 14.

Table 14: Most Prevalent Diseases Related to Smoking with the Highest Cost

	MOST PREVALENT DISEASES RELATED TO SMOKING WITH THE HIGHEST COST	HIERARCHICAL RANK OF THE MOST PREVALENT DISEASES RELATED TO SMOKING ACCORDING TO THEIR COST
ISCHEMIC HEART DISEASE	$22,195,207.61	1
CEREBROVASCULAR DISEASE	$14,792,633.79	2
MALIGNANT NEOPLASMS OF TRACHEA, LUNG, BRONCHUS	$2,069,133.88	3
MALIGNANT NEOPLASMS OF LIP, ORAL, CAVITY, PHARYX	$1,863,827.99	4
OTHER HEART DISEASE	$1,548,527.68	5
MALIGNANT NEOPLASMS OF KIDNEY AND RENAL PELVIS	$1,445,977.67	6
MALIGNANT NEOPLASMS OF URINARY BLADDER	$1,038,325.76	7
BRONCHITIS, EMPHYSEMA	$968,831.96	8
MALIGNANT NEOPLASMS OF PANCREAS	$753,908.61	9
OTHER ARTERIAL DISEASE	$743,624.50	10
OTHERS	$1,773,335.43	11

From Table 14, it is deduced that the most “expensive” disease related to smoking is ischemic heart disease, with a cumulative cost of $22,195,207.61

throughout the years 1999-2009. The second place in the list is occupied by the cerebrovascular disease, with a cumulative cost of $14,792,633 for the period 1999-2009. It is seen that between these two diseases with the highest cost, there is a gap of approximately $12,000,000, which is a remarkably big gap. Ischemic heart disease is by far the disease with the highest cost because it is the most prevalent disease related to smoking compared to the rest of the diseases related to smoking. The rest of diseases present a cumulative cost of less than $2,000,000. This cost difference between ischemic heart disease and the rest of the diseases emphasized the importance that should be given to the prevention of this disease. Row 11 in the hierarchical ranking of the diseases includes the rest of the SAMMEC most prevalent diseases related to smoking and these diseases are presented in Table 15.

Table 15: Other Most Prevalent Diseases Related to Smoking

OTHER MOST PREVALENT RELATED TO SMOKING DISEASES
ATHEROSCLEROSIS
MALIGNANT NEOPLASMS OF ESOPHAGUS
MALINGNANT NEOPLASMS OF CERVIX UTERI
PNEUMONIA, INFLUENZA
CHRONIC AIRWAY OBSTRUCTION
AORTIC ANEURYSM
MALIGNANT NEOPLASMS OF STOMACH
MALIGNANT NEOPLASMS OF LARYNX

The classification of the most prevalent diseases related to smoking according to their cumulative cost for the period 1999-2009, with number one being the disease with the highest cost, is used in the Regression Analysis of this study for the creation of dummy variables. These dummy variables are used as regressors, trying to see which diseases affect the overall cost. The dummy variable of disease one refers to

ischemic heart disease, the dummy variable of disease two refers to cerebrovascular disease and so on, and the dummy variable of disease 11 refers to the list of the rest of the most prevalent diseases related to smoking. A graphical presentation of the diseases and their classification according to their cost is given by Figure 15.

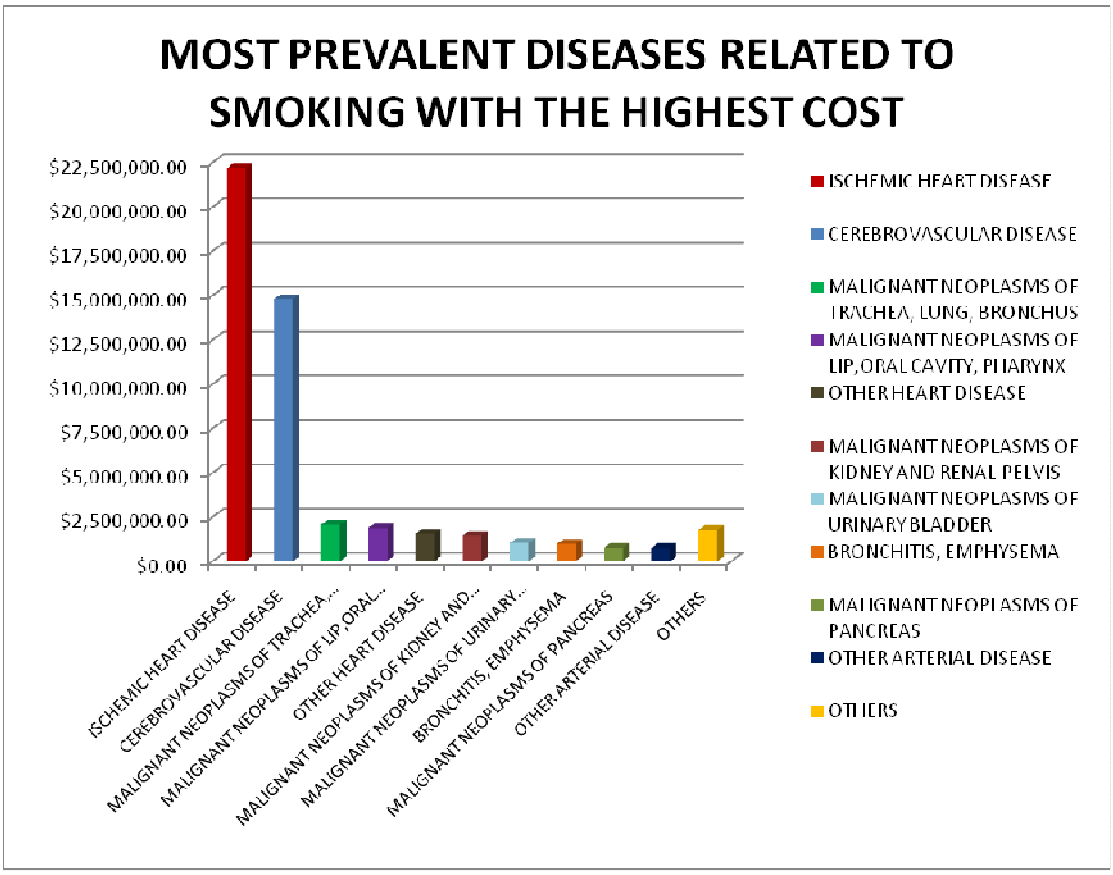


Figure 15: Most Prevalent Diseases Related to Smoking with the Highest Cost

With the aid of the graph, the cost gap between ischemic heart disease and the other diseases is visually presented. The total cost of all of the diseases, for the period that covers the years from 1999 to 2009, reaches the amount of $49,193,334. The distribution of this amount throughout the 11 year period from 1999 to 2009 is given by Table 16 and in a graphic by Figure 16.

Table 16: Total Annual Cost of the Most Prevalent Diseases Related to Smoking for the period 1999-2009

CY	TOTAL ANNUAL COST
1999	$905,758.27
2000	$2,986,421.97
2001	$3,932,715.58
2002	$3,900,035.24
2003	$2,221,221.96
2004	$5,257,462.32
2005	$5,496,775.31
2006	$5,631,356.73
2007	$6,787,907.41
2008	$7,106,103.24
2009	$4,967,576.85
GRAND TOTAL	$49,193,334.87

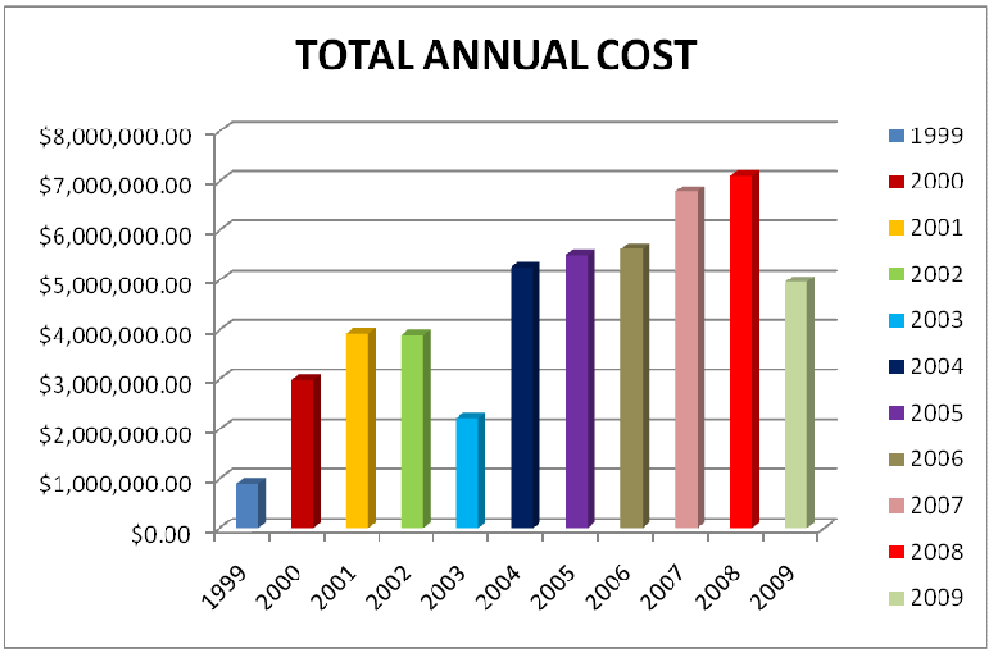


Figure 16: Total Annual Cost of the Most Prevalent Diseases related to smoking for the period 1999-2009

The graph in Figure 16 shows that the year with the highest total annual cost of hospitalization for ADAF members was 2008, where the total annual cost reached the amount of $7,106,103. The total annual cost and the grand total cost (the total cost

for the period 1999-2009) of hospitalization for males is remarkably higher than the corresponding one for females. Table 17 and Figure 17 give a better picture of the total annual cost for each gender.

Table 17: Total Annual Cost per Gender

TOTAL ANNUAL COST		
	FEMALES	**MALES**
1999	$125,110.65	$780,647.62
2000	$389,391.21	$2,597,030.76
2001	$435,860.60	$3,496,854.98
2002	$534,830.03	$3,365,205.21
2003	$207,179.58	$2,014,042.39
2004	$398,987.13	$4,858,475.18
2005	$579,500.25	$4,917,275.06
2006	$568,850.70	$5,062,506.02
2007	$830,986.43	$5,956,920.97
2008	$1,510,656.23	$5,595,447.01
2009	$598,674.02	$4,368,902.83
Grand Total	**$6,180,026.84**	**$43,013,308.03**

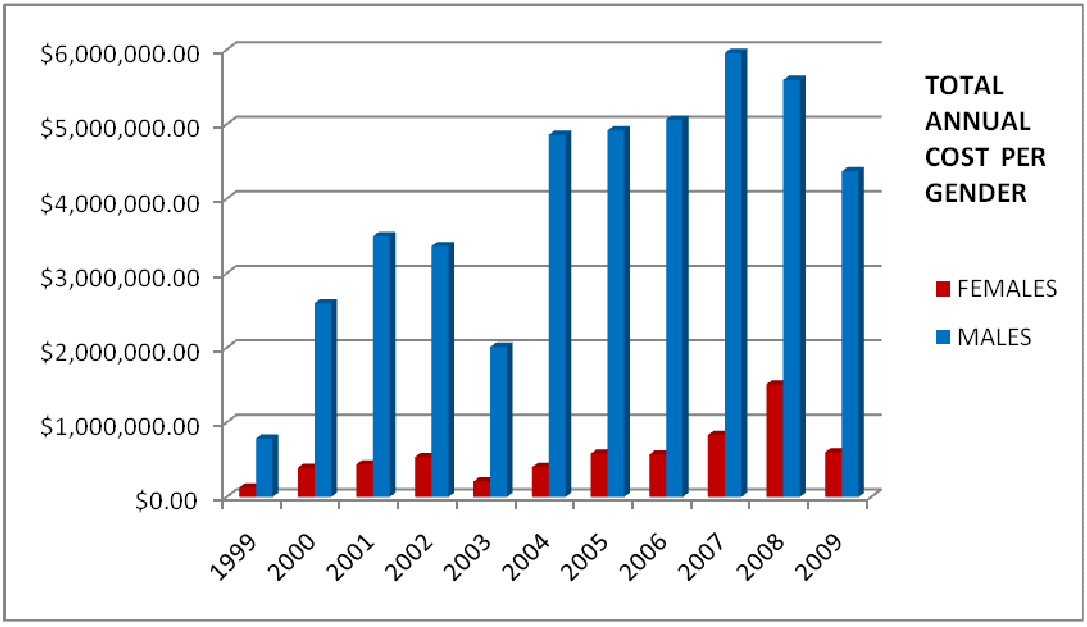


Figure 17: Total Annual Cost per Gender

Throughout the years 1999-2009, the cost of hospitalization for the male population of the Air Force is always higher than the corresponding cost for the female population. Hospital expenses for males reached their maximum in 007 ($5,956,920), while analogous expenses for females reached their maximum in 2008 ($1,510,656). In 2008, hospital expenses for female ADAF surpassed the limit of $1,000,000 for the first time.

The average cost of hospitalization for each gender is almost the same, and in one year, the average cost regarding females was higher than that one regarding males. There is no major difference between the grand average cost (the average cost for the period 1999-2009) for both genders. The following Table 18 and Figure 18 show the average annual cost per gender, while Table 19 and Figure 19 present the grand average cost per gender.

Table 18: Average Annual Cost per Gender

AVERAGE ANNUAL COST		
	FEMALES	**MALES**
1999	$1,097.46	$1,107.30
2000	$1,035.61	$1,057.42
2001	$945.47	$1,202.08
2002	$936.66	$996.21
2003	$3,092.23	$4,178.51
2004	$1,461.49	$2,809.99
2005	$2,138.38	$2,676.80
2006	$1,644.08	$2,387.97
2007	$1,486.56	$2,018.61
2008	$2,452.36	$1,925.48
2009	$1,153.51	$1,575.51

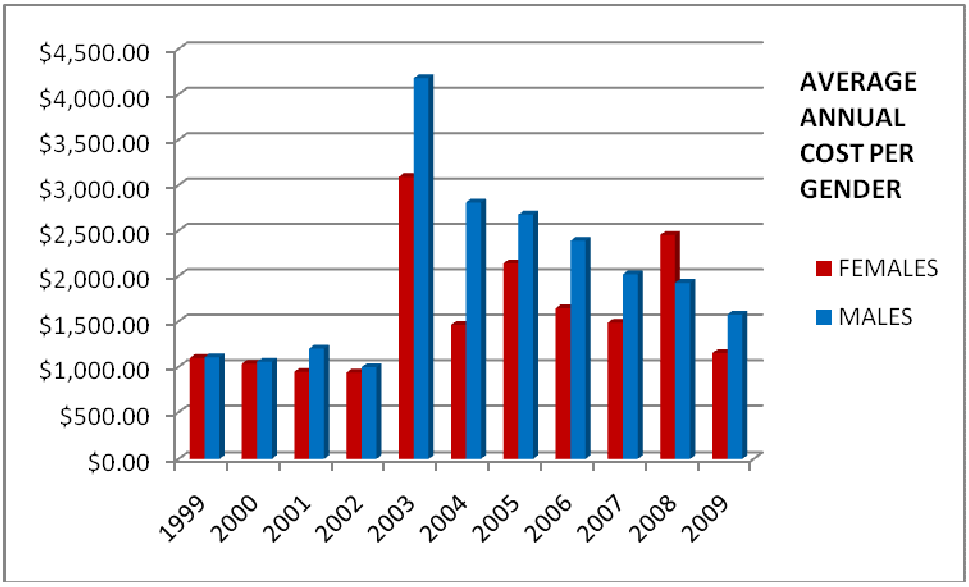


Figure 18: Average Annual Cost per Gender

Table: 19: Grand Average cost per Gender

	FEMALES	MALES
GRAND AVERAGE COST	$1,480.96	$1,774.04

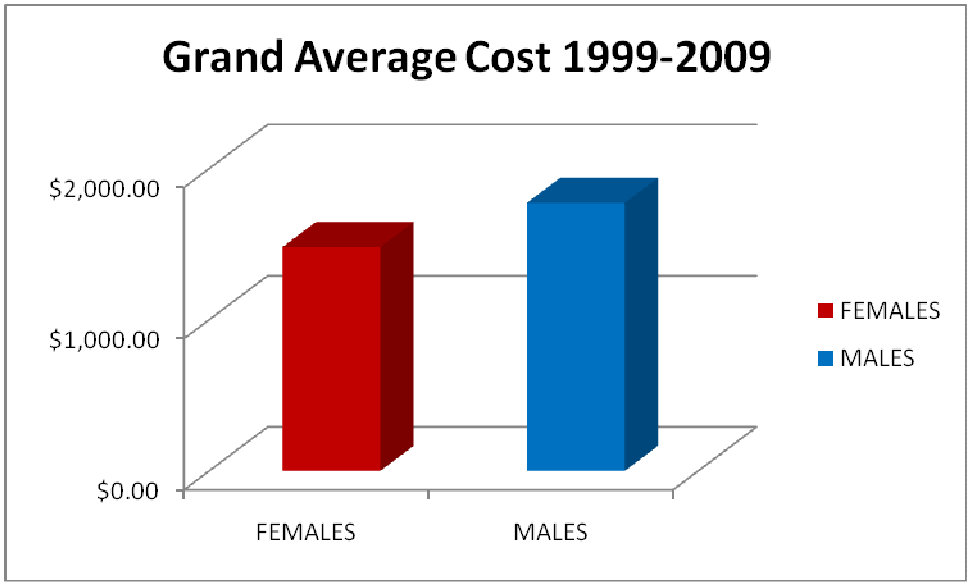


Figure 19: Grand Average cost per Gender

The above tables and figures present the average annual cost and the grand average cost per gender. In all cases, it is ascertained that the average annual costs for both genders are very close to each other. The average annual cost for males is always slightly higher than females, but in the year 2008, the female average annual cost surpassed the cost for males. The best way to measure and compare the average cost for each gender is the grand average cost per gender, which is the average cost for each gender for the whole period of 1999-2009. The graph in Figure 19 displays the bar chart of the grand average cost for each gender and is is shown that both genders do not differ that much concerning the average cost of hospitalization. It is understood that even though men are a larger portion of the military population, the average cost is almost the same for both men and women and the expenses of hospitalization do not differ considerably.

Another way to test if there is a significant difference between the average annual cost of each gender is the paired t-test. The paired t-test is a statistical test that compares the means of two groups of observations and tests to see if the average difference is significantly different from zero. The null hypothesis in this case is that there is no significant difference between the average annual cost of the two genders, and the alternative hypothesis is that there is significant difference between the average annual cost of the two genders. The paired t-test for the average annual cost of females and males is conducted with the significance level of $\alpha=0.05$. If the significance value of the two-tailed paired t-test is less than the significance level of $\alpha=0.05$, the null hypothesis is rejected and there is significant difference between the average annual cost of the two genders. The results of the paired t-test of the average annual cost of the two genders are shown below in Figure 20.

	FEMALES	*MALES*
Mean	1585.801	1994.170909
Variance	491376.5	952592.7179
Observations	11	11
Hypothesized Mean Difference	0	
df	18	
t Stat	-1.12712	
P(T<=t) one-tail	0.137246	
t Critical one-tail	1.734064	
P(T<=t) two-tail	**0.274492**	
t Critical two-tail	2.100922	

Figure 20: Two-Tailed Paired T-Test of the Average Annual Cost of Females and Males

It is seen in Figure 20 that the significance value of the two-tailed paired t-test of the average annual cost of the two genders equals to 0.274492, which is larger than the significance level of α=0.05. This indicates that the test fails to reject the null hypothesis and there is no significant difference between the average annual cost of females and males.

Age is another point of reference for the ADAF population. It is worth investigating the cost of hospitalization due to diseases related to smoking, with age being the point of reference for this cost. Total and average cost for each age group would be part of this investigation, and Tables 20 and 21 and Figures 21 and 22 provide valuable information for the grand total (the total cost during the period 1999-2009) and average cost of each age group of the Air Force population of the data set used for this part of the analysis.

Table 20: Grand Total Cost per Age Group

GRAND TOTAL COST PER AGE GROUP	
17-24	$3,330,039.14
25-32	$5,155,190.27
33-40	$14,090,162.82
41-48	$18,280,958.56
49-56	$7,059,219.94
57-64	$1,252,358.03
65-72	$22,817.90
73-80	$1,391.73
81-88	$1,196.47
GRAND TOTAL	**$49,193,334.87**

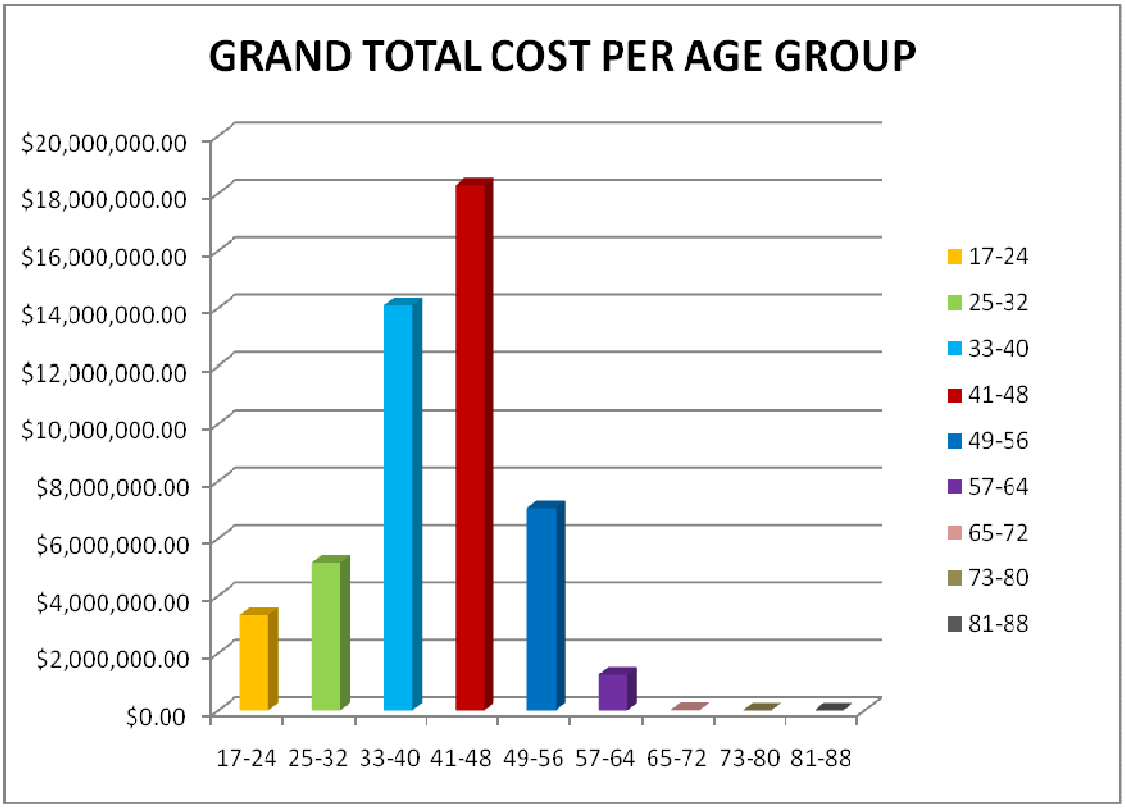


Figure 21: Grand Total Cost per Age Group

Table 21: Average Cost per Age Group

AVERAGE COST PER AGE GROUP	
17-24	$1,423.09
25-32	$1,524.30
33-40	$1,602.98
41-48	$1,854.81
49-56	$2,099.71
57-64	$1,920.79
65-72	$950.75
73-80	$173.97
81-88	$239.29
GRAND AVERAGE	**$1,731.00**

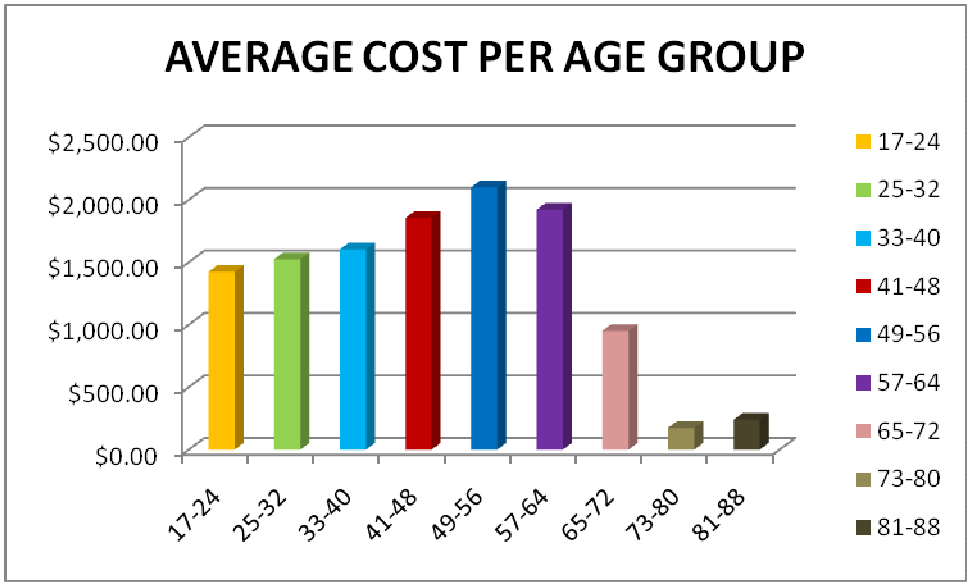


Figure 22: Average Cost per Age Group

Table 20 and Figure 21, presenting the grand total cost for each age group, show that the age group 41-48 is the group with the highest cost. The second highest grand total cost is for the age group 33-40 and the third is the age group 49-56. This classification of cost by the age groups indicates that ADAF personnel age33- 56 years that smoke, generate the highest cost of hospitalization. The picture of cost ranking by the age groups is slightly different when average cost is classified by the age groups. Table 21 and Figure 22 show that the age group with the highest average cost for the period 1999-2009 is the age group of 49-56 with an average cost of

$2,099.71, followed by the age group 57-64 with an average cost of $1,920.79, and age group 41-48 with an average cost of $1,854.81. It is inferred that the age groups of 41-48 and 49-56 are the groups with the highest grand total and average cost, and they are the age groups receive the most medical care due to diseases related to smoking. At this point it must be mentioned that the grand average cost, which is the general average cost of hospitalization for the whole population of the Air Force of the data set used in this part of the analysis is $1,731.

The above conclusions about the total and average cost of each age group can be visualized also by using thefrequency of visits to the hospital or to the doctor by the above mentioned age groups. Table 22 and Figure 23 show the frequency of visits, classified by age groups.

Table 22: Frequency of Visits per Age Group

FREQUENCY OF VISITS PER AGE GROUP	
17-24	2,340
25-32	3,382
33-40	8,790
41-48	9,856
49-56	3,362
57-64	652
65-72	24
73-80	8
81-88	5
GRAND TOTAL OF VISITS	28,419

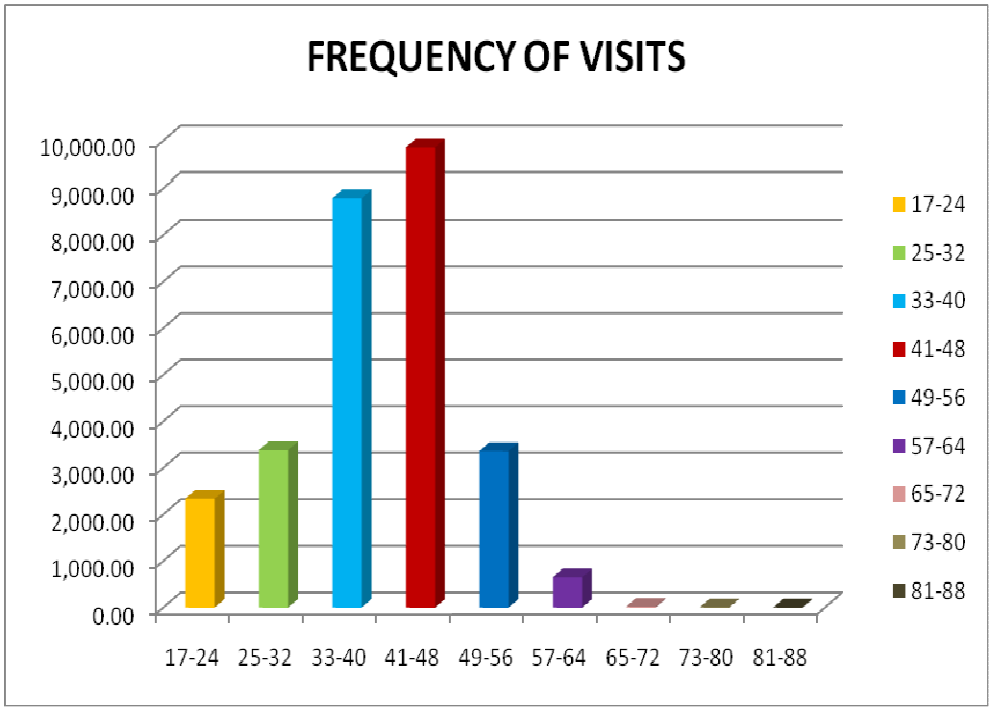


Figure 23: Frequency of Visits per Age Group

The age group with the highest number of visits is the age group of 41-48, with the age group of 33-40 following in second place. The same groups were among the ones with the highest grand total and average cost. The age groups of 25-32 and 49-56 have almost the same number of visits, and the age group of 49-56 was the one with the highest average cost. This classification of the frequency of visits, combined with the classification of grand total and average cost by the age groups, leads to the conclusion that smoking related diseases are most prevalent in the age range of 33 - 56.

Table 23 and Figure 24 give a visual presentation of the classification of diseases related to smoking, according to the frequency of visits (the frequency that ADAF personnel visited a hospital or a doctor because of a smoking related disease).

Table 23: Classification of Smoking Related Diseases by the Frequency of Visits

	SMOKING RELATED DISEASES	FREQUENCY OF VISITS
1	ISCHEMIC HEART DISEASE	12482
2	CEREBROVASCULAR DISEASE	8133
3	BRONCHITIS, EMPHYSEMA	1977
4	OTHER ARTERIAL DISEASE	1005
5	ATHEROSCLEROSIS	927
6	MALIGNANT NEOPLASMS OF LIP,ORAL CAVITY, PHARYNX	766
7	OTHER HEART DISEASE	666
8	MALIGNANT NEOPLASMS OF TRACHEA, LUNG, BRONCHUS	639
9	MALIGNANT NEOPLASMS OF KIDNEY AND RENAL PELVIS	545
10	MALIGNANT NEOPLASMS OF URINARY BLADDER	527
11	OTHER DISEASES	752

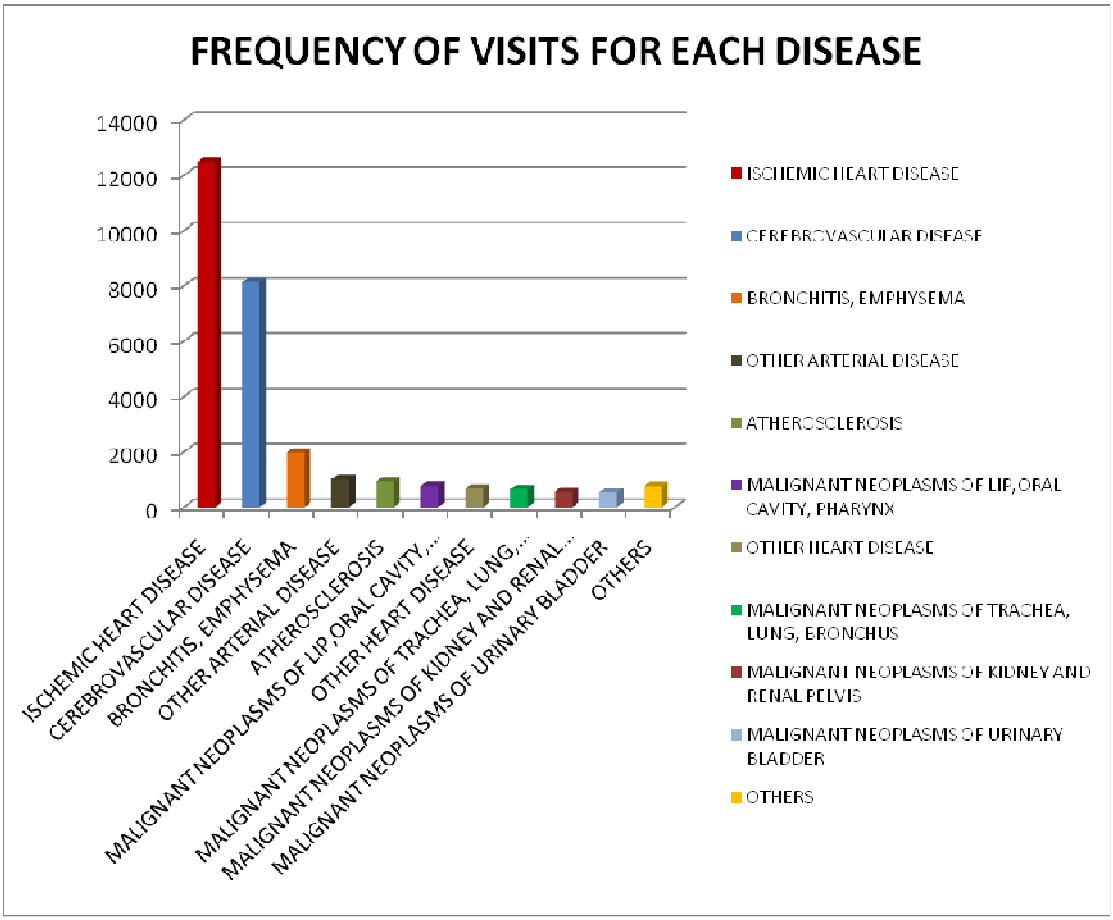


Figure 24: Classification of Smoking Related Diseases by the Frequency of Visits

This classification of diseases differs from the previous one showing the diseases with highest cost. Ischemic heart disease and cerebrovascular disease are the

two diseases with the highest cost and with the largest number of visits. The two assortments of diseases are not identical and this leads to the conclusion that a disease of high cost is not necessarily a disease with a large number of visits. That means that some diseases cost more than others and total cost is independent of the frequency of visits. In this case, the category Other Diseases includes the following: malignant neoplasms of cervix uteri, malignant neoplasms of esophagus, malignant neoplasms of pancreas, pneumonia- influenza, chronic airway obstruction, aortic aneurysm, malignant neoplasms of stomach and malignant neoplasms of larynx.

The last part of the Pivot Table Analysis includes a concentrated presentation of the cost of hospitalization of ADAF personnel during the period 1999-2009. Table 24 presents the grand total and average cost for the pay ranks of enlisted and officers and the grand total cost for the two genders separately.

Table 24: Concentrating Table of Cost for each Pay Rank and Gender

	ENLISTED	OFFICERS	GRAND TOTAL
GRAND TOTAL COST	$33,571,520.94	$15,621,813.92	$49,193,334.87
FREQUENCY OF VISITS	21,103	7,316	28,419
GRAND AVERAGE COST	$1,590.84	$2,135.29	$1,731.00
MALES	$29,440,395.99	$13,572,912.04	$43,013,308.03
FEMALES	$4,131,124.95	$2,048,901.88	$6,180,026.84
TOTAL	$33,571,520.94	$15,621,813.92	$49,193,334.87

From the above table, it is worthwhile noticing the grand total cost, frequency of visits, and the grand average cost for enlisted and officers. Here, the word ‘grand’ refers to the whole period of 1999-2009. Enlisted personnel are the majority and it comes naturally that their grand total cost and their frequency of visits are much higher than officers. But when it comes to the grand average cost, the grand average cost of officers is surprisingly higher than for enlisted. This means that the cost of hospitalization for an officer is higher than for an enlisted and since cost depends on

the kind of the disease and not on the pay rank, officers might be hospitalized for more “expensive” diseases, meaning that high-cost diseases are more prevalent among officers than enlisted. Figures 25 and 26 give a visual presentation of the grand total and the average cost of enlisted and officers.

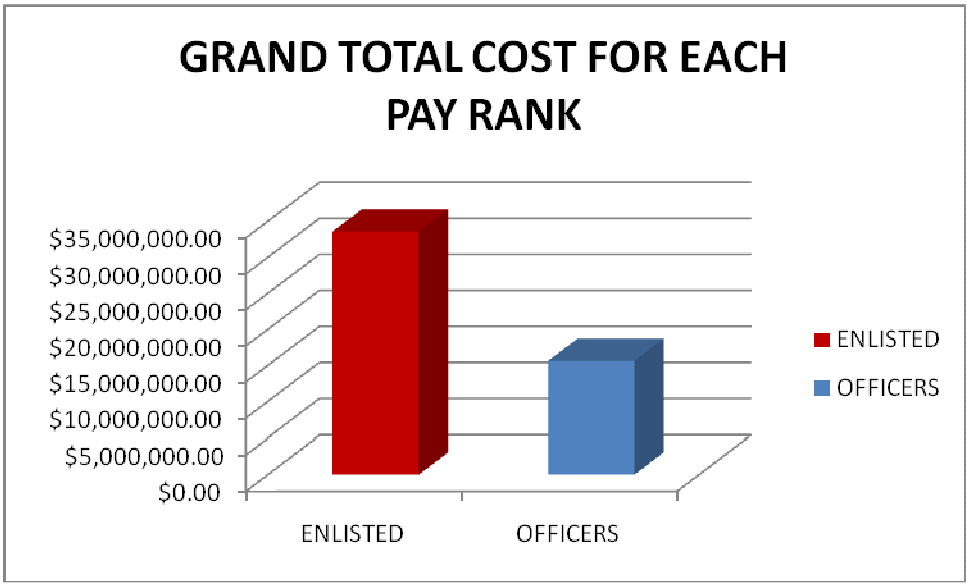


Figure 25: Grand Total Cost for each Pay Rank

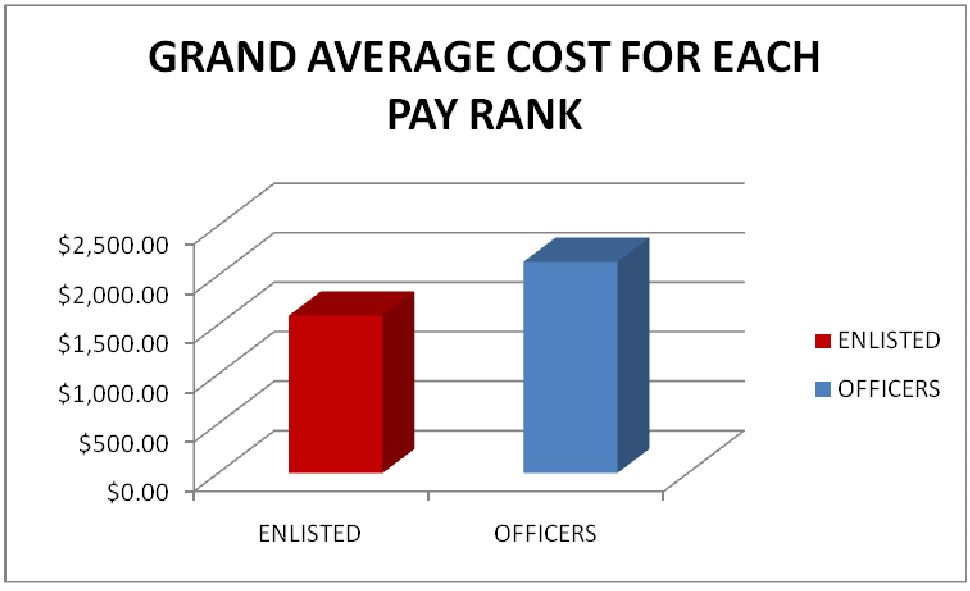


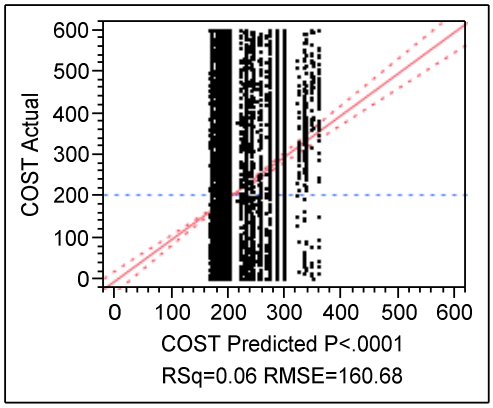
Figure 26: Grand Average Cost for each Pay Rank

Regression Analysis

Data Set 1 – Cost Range $0.00 - $600.00 – Model 1

As mentioned in Chapter III, the initial cost data set used for the Regression Analysis was divided into five subsets with different cost range. The first model, built in this part of the Regression Analysis, covers the cost range of $0.00 to $600.00. The Actual by Predicted Plot, the Summary of Fit, the Analysis of Variance (ANOVA) and the Parameter Estimates Report of this model are given in Figure 27.

Actual by Predicted Plot



Summary of Fit

RSquare	0.05878
RSquare Adj	0.058329
Root Mean Square Error	160.6838
Mean of Response	207.9587
Observations (or Sum Wgts)	18819

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	9	30328082	3369787	130.5144
Error	18809	485634728	25819	**Prob > F**
C. Total	18818	515962809		<.0001*

Parameter Estimates Report

Term	Estimate	Std Error	t-ratio	Prob>|t|	VIF
Intercept	299.06179	3.444336	86.83	0.0000*	.
AGE 45-60	-12.00913	2.838979	-4.23	<.0001*	1.1378766
ENLISTED	-102.9175	3.440053	-29.92	<.0001*	1.9634411
O1,O2,O3	-107.1333	6.055061	-17.69	<.0001*	1.3153296
DUMMY FOR DISEASE 1	-12.3346	2.50661	-4.92	<.0001*	1.1141692
DUMMY FOR DISEASE 4	62.903319	8.005478	7.86	<.0001*	1.0239846
DUMMY FOR DISEASE 7	47.597783	9.477384	5.02	<.0001*	1.0169059
DUMMY FOR DISEASE 10	37.920519	5.832891	6.50	<.0001*	1.0430684
O4,O5,O6	-94.54707	4.402874	-21.47	<.0001*	1.8204513
CD,OCS	36.904421	18.39044	2.01	0.0448*	1.0304815

Figure 27: Actual by Predicted Plot, Summary of Fit, Analysis of Variance and Parameter Estimates for Cost range $0.00-$600.00

In this first model, the most predictive variables for cost, with the lowest p-values of the t-statistic test, appear to be the following ones: AGE 45-60, ENLISTED, O1-O2-O3, O4-O5-O6, CD-OCS and the dummy variables for diseases 1, 4, 7 and 10.

All the variables were explained in Chapter III except for the dummy variables of the diseases. The most predictive diseases in this model are the following ones: disease 1 refers to Ischemic Heart Disease, disease 4 refers to Malignant Neoplasms of Lip, Oral, Cavity and Pharyx, disease 7 refers to Malignant Neoplasms of Urinary Bladder and disease 10 corresponds to Other Arterial Disease. The OLS regression model for the subset of $0.00 - $600.00 is given by the formula below:

Cost = 299.06-12.01(AGE 45-60)-102.92*(ENLISTED)-107.13*(O1,O2,O3)-12.33*(Ischemic Heart Disease)+62.90*(Malignant Neoplasms of Lip, Oral, Cavity, Pharyx)+47.60*(Malignant Neoplasms of Urinary Bladder)+37.92*(Other Arterial Disease)-94.55*(O4,O5,O6)+36.90*(CD, OCS)*

All the Variation Inflation Factors (VIF scores) of this model, which measure the redundancy among the explanatory variables, are below 5, meaning that no multicollinearity occurs. The p-value of F -statistic, which measures the overall model statistical significance, is shown in the Analysis of Variance in the Prob>F column. In this model, the p-value of F-statistic, for a 95% confidence level, is lower than 0.05 and indicates a statistically significant model. Both R^2 and Adjusted R^2 are presented in the Summary of Fit, and are values that measure the model performance. In this model both values are almost the same and equal to 0.058, indicating that this model explains approximately 5.8% of the variation in the dependent variable, which in this case is the cost of hospitalization. The particular low R^2 value and subsequently the low predictability of the model, indicate that the model does not provide a good fit of variables with the data and there is lot of variability not explained by the model.

The Cook's Distance Overlay plot in Figure 28 shows that there are no influential points.

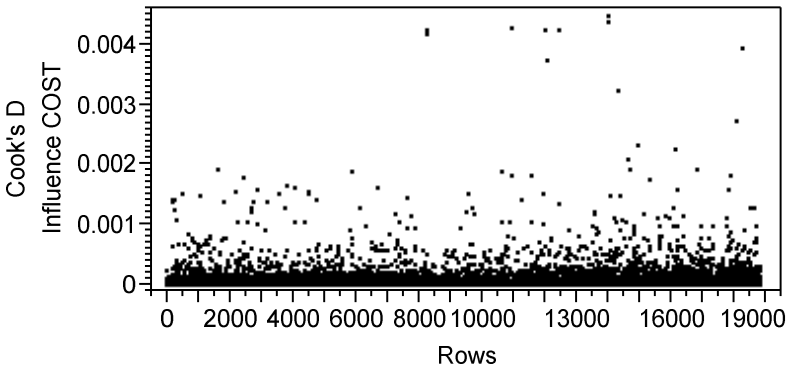
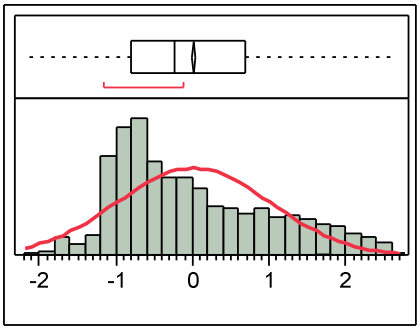


Figure 28: Overlay Plot of Cook's Distance for Model 1

The diagnostic test for Normality is the Shapiro-Wilk test, which demands a distribution of the studentized residuals. In the case of Shapiro-Wilk test, a p-value larger than 0.05 fails to reject the null hypothesis that the residuals are normally distributed. In this model the Shapiro Wilk test and the distribution of studentized residuals are given in Figure 29. The p-value is 0.01, lower than 0.05, and that means that the null hypothesis that the residuals are normally distributed is rejected, and the model does not pass the test for normality.



Fitted Normal
Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	-2.563e-7	-0.014289	0.0142884
Dispersion	σ	1.0000346	0.9900331	1.0102416

Goodness-of-Fit Test
KSL Test

D		Prob>D
0.107328	<	0.0100*

Figure 29: Shapiro-Wilk Test for Normality for Model 1

The test for constant variance is the Breusch-Pagan test, and tests if the variance of the errors is constant across the observations. In the Breusch-Pagan test, the null hypothesis is that the residuals exhibit constant variance and a p-value larger than 0.05 fails to reject the null hypothesis. This model fails to pass the Breusch-Pagan, since the p-value is very low, lower than 0.05 and thus the null hypothesis is rejected. The results of the Breusch-Pagan test were calculated with the aid of Microsoft Excel and are given in Table 25.

Table 25: Breusch-Pagan test for Model 1

	Input			
SSR	6.222E+10		Test Stat	46.71536
SSE	485634728		**P-value**	4.43E-07
N	18819			
df (reg)	9			

Data Set 2 – Cost Range $600.01 - $1,800.00 – Model 2

In Model 1 there was a lot of variability not explained by the data and for this reason the model demonstrated low predictability, even though the predictive variables had low t-statistic p-values, and the overall p-value of the F-statistic of the model was low, too. The same occurs with Model 2, which regards the cost range of $600.01 - $1800.00. The Summary of Fit, the Analysis of Variance, and the Parameter Estimates report of Model 2 are given in Figure 30.

Summary of Fit

RSquare	0.027174
RSquare Adj	0.025323
Root Mean Square Error	293.2935
Mean of Response	994.4549
Observations (or Sum Wgts)	5791

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	11	13886220	1262384	14.6753
Error	5779	497115830	86021	**Prob > F**
C. Total	5790	511002049		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t-ratio	Prob>|t|	VIF
Intercept	924.52264	18.2785	50.58	0.0000*	.
ENLISTED	102.28975	12.37734	8.26	<.0001*	2.393883
O1,O2,O3	107.45701	19.77797	5.43	<.0001*	1.3301553
O4,O5,O6	69.116268	14.08404	4.91	<.0001*	1.9318035
DUMMY FOR DISEASE 1	-39.79508	13.49212	-2.95	0.0032*	3.0627137
DUMMY FOR DISEASE 10	-115.4091	29.61618	-3.90	<.0001*	1.1884551
DUMMY FOR DISEASE 2	-41.32374	14.14094	-2.92	0.0035*	2.8695385
DUMMY FOR DISEASE 8	-153.7974	25.99338	-5.92	<.0001*	1.2590722
E5,E6	22.610802	9.804203	2.31	0.0211*	1.2787366
DUMMY FOR DISEASE 11	-74.36178	20.83754	-3.57	0.0004*	1.4674738
GENDER_1	27.247629	11.44658	2.38	0.0173*	1.0622896
O7,O8,O9,O10	163.22885	62.16223	2.63	0.0087*	1.0290769

Figure 30: Summary of Fit, Analysis of Variance and Parameter Estimates for Cost range $600.01-$1800.00

This model appears to have even lower predictability than Model 1. The R^2 value equals to 0.0272, and this means that the model explains approximately 2.72% of the variation in the dependent variable. The overall p-value of the F-statistic is lower than 0.05 and thus the null hypothesis that the explanatory variables in the model are not effective is rejected. The predictive variables of this model with low t-statistic p-values are: ENLISTED, O1-O2-O3, O4-O5-O6, O7-O8-O9-O10, E5-E6, GENDER_1 and the Dummy Variables for Diseases 1, 2, 8, 10 and 11. Disease 1 refers to Ischemic Heart Disease, disease 2 refers to Cerebrovascular disease, disease 8 refers to Bronchitis, Emphysema, disease 10 refers to Other Arterial disease and disease 11 refers to the rest of the most prevalent diseases related to smoking, which

are presented in a detailed way in Table 15. The OLS regression model for the cost range of $600.01 - $1800.00 is given by the formula below:

Cost=924.52+102.29(ENLISTED)+107.46*(O1,O2,O3)+69.12*(O4,O5,O6) + 163.23*(O7,O8,O9,O10)+22.61*(E5,E6)+27.25*(GENDER_1)-39.80*(Ischemic Heart Disease)-41.32*(Celebrovascular Disease)-153.80*(Bronchitis, Emphysema)-115.41*(Other Arterial disease)-74.36*(Other Diseases)*

The VIF scores of all the variables are below 5 and the model does not have any influential points, as shown in Cook's Distance Overlay Plot in Appendix C. In addition, the model does not pass the Shapiro-Wilk test for normality and the Breusch-Pagan test for constant variance.

Data Set 3 – Cost Range $1,800.01 - $11,000.00 – Model 3

The third model concerns the data set of cost range $1800.01-$11,000.00 and the Summary of Fit, the Analysis of Variance, and the Parameter Estimates report are given below in Figure 31.

Summary of Fit

RSquare	0.008385
RSquare Adj	0.007316
Root Mean Square Error	2443.441
Mean of Response	4578.227
Observations (or Sum Wgts)	2785

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	140406398	46802133	7.8390
Error	2781	1.6604e+10	5970403.4	**Prob > F**
C. Total	2784	1.6744e+10		<.0001*

Parameter Estimates Report

Term	Estimate	Std Error	t-ratio	Prob>|t|	VIF
Intercept	4835.6026	83.45806	57.94	<.0001*	.
AGE 25-34	-300.8502	134.2686	-2.24	0.0251*	1.0018253
ENLISTED	-281.2203	98.99104	-2.84	0.0045*	1.0066894
DUMMY FOR DISEASE 7	-951.7804	281.3771	-3.38	0.0007*	1.0053827

Figure 31: Summary of Fit, Analysis of Variance and Parameter Estimates for Cost range $1,800.01-$11,000.00

This model appears to have the lowest R-Squared and Adjusted R-Squared values, which means that the predictability of the model is very low. The R-Square equals to 0.008385, indicating that this model explains approximately 0.8385% of the variation in the dependent variable. This occurs because there is a lot of variability in the data, not explained by the model. The p-value of the F-statistic, which determines the overall statistical significance of the model, is very low. The p-values of the t-statistic of each explanatory variable are lower than 0.05 and the VIF scores of all of them are lower than 5. The most predictive variables for this cost range are: AGE 25-34, ENLISTED and the Dummy Variable for disease 7, which is the Malignant Neoplasms of Urinary Bladder. The equation of this model is given by the following formula:

Cost=4835.60-300.85(AGE 25-34)-281.22*(ENLISTED)-951.78*(Malignant Neoplasms of Urinary Bladder)*

The model does not appear to have any influential points, according to Cook's Distance Overlay plot, graphed in Appendix C. This model, like the previous ones, does not pass the tests for Normality and Constant Variance.

Data Set 4 – Cost Range $11,000.01 - $30,000.00 – Model 4

Model 4 regards the cost range of $11,000.01-$30,000.00 and the results of the Summary of Fit, the Analysis of Variance, and the Parameter Estimates Report, are presented below in Figure 32.

Summary of Fit

RSquare	0.013927
RSquare Adj	0.011433
Root Mean Square Error	4978.371
Mean of Response	17518.16
Observations (or Sum Wgts)	794

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	276876512	138438256	5.5858
Error	791	1.9604e+10	24784176	**Prob > F**
C. Total	793	1.9881e+10		0.0039*

Parameter Estimates

Term	Estimate	Std Error	t-ratio	Prob>|t|	VIF
Intercept	17958.049	259.8688	69.10	<.0001*	.
O7,O8,O9,O10	6105.6255	2885.988	2.12	0.0347*	1.0043652
AGE 35-44	-866.9508	354.9437	-2.44	0.0148*	1.0043652

Figure 32: Summary of Fit, Analysis of Variance and Parameter Estimates for Cost range $11,000.01-$30,000.00

The R^2 value of this model equals 0.013927, indicating that the predictability of the model equals 1.3927%. The F-statistic appears to have a p-value lower than

0.05 and the cost in this model is predicted by two variables, which are O7-O8-O9-O10 and AGE 35-44. The VIF scores of the regressors are below 5. The model is given by the following equation:

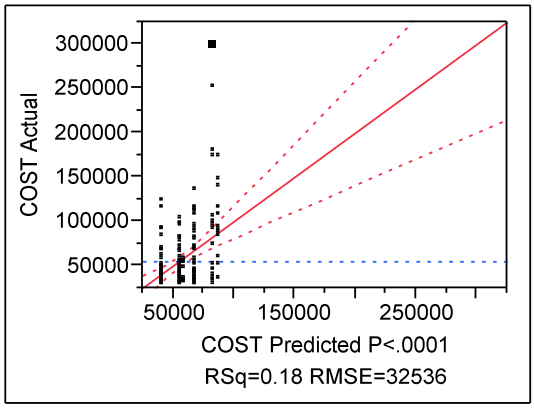
Cost=17958.05+6105.63(O7-O8-O9-O10)-866.95*(AGE 35-44)*

The Cook's Distance Overlay plot, presented in Appendix B, does not graph any influential points, but the model does not pass the tests for Normality and Constant Variance.

Data Set 5 – Cost Range $30,000.01 - $307,064.00 – Model 5

The last model concerns the extremely high cost range of $30,000.01-$307,064.00 and its Actual by Predicted plot, Summary of Fit, Analysis of Variance, and Parameter Estimates Report are given in Figure 33.

Actual by Predicted Plot



Summary of Fit

RSquare	0.175222
RSquare Adj	0.164273
Root Mean Square Error	32535.54
Mean of Response	55917.78
Observations (or Sum Wgts)	230

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	5.0825e+10	1.694e+10	16.0044
Error	226	2.3923e+11	1.0586e+9	**Prob > F**
C. Total	229	2.9006e+11		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t-ratio	Prob>|t|	VIF
Intercept	39587.422	3381.55	11.71	<.0001*	.
E1,E2,E3,E4	18783.413	7431.762	2.53	0.0122*	1.0800354
AGE 45-60	14670.658	4695.764	3.12	0.0020*	1.0930497
DUMMY FOR DISEASE 2	27735.602	4669.303	5.94	<.0001*	1.0549758

Figure 33: Actual by Predicted Plot, Summary of Fit, Analysis of Variance and Parameter Estimates for Cost range $30,000.01-$307,064.00

Compared to the previous four models, this model is the one, with the highest R^2 and Adjusted R^2 values. The R^2 value equals 0.175222, which means the model explains 17.5222% of the variation in the dependent variable. The p-value of the F-statistic is very low and lower than 0.05 and all the predictive variables have p-values of the t-statistic lower than 0.05 and VIF scores below 5. The regressors of this model are: E1-E2-E3-E4, AGE 45-60 and the Dummy Variable for disease 2, which is Cerebrovascular Disease. The equation of the model is the following:

Cost=39587.42+18783.41(E1,E2,E3,E4)+14670.66*(AGE 45-60)+27735.60*(Celebrovascular Disease)*

The Actual by Predicted Plot (See Figure 33) shows two potential influential points and the same can be seen from the Cook's Distance Overlay Plot, given below in Figure 34.

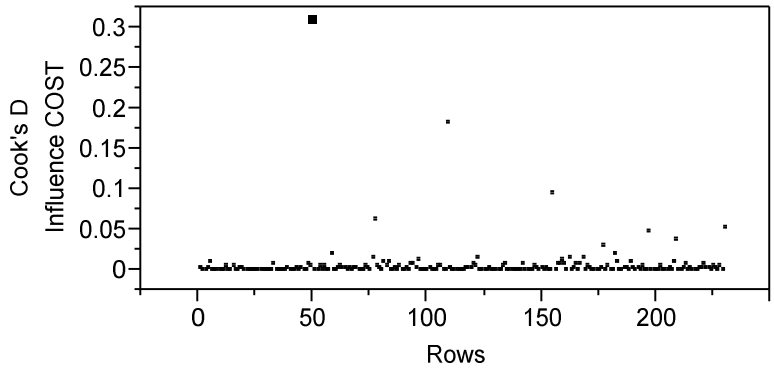
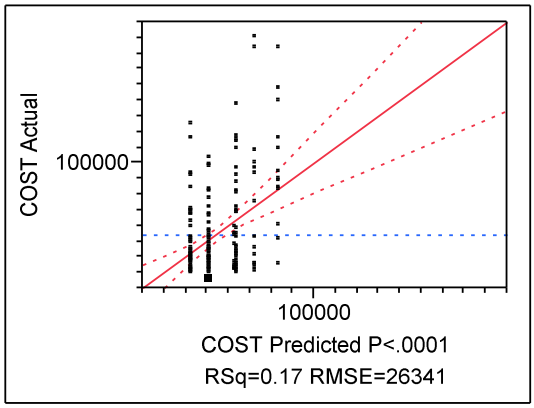


Figure 34: Cook's Distance Overlay Plot for Model 5

These two influential points correspond to the two extreme cost values of $252,301 and $307,063, which lie far enough from the rest of the values of this cost range. Subsequently, these two values must be removed and the model must be re-assessed and re-examined for its validity and predictability, excluding these two influential points. The Actual by Predicted Plot, the Summary of Fit, the Analysis of Variance, and the Parameter Estimates Report of the new re-assessed model are given in Figure 35.

Actual by Predicted Plot



Summary of Fit

RSquare	0.171248
RSquare Adj	0.160148
Root Mean Square Error	26341.24
Mean of Response	53954.94
Observations (or Sum Wgts)	228

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	3.2116e+10	1.071e+10	15.4286
Error	224	1.5542e+11	693861153	**Prob > F**
C. Total	227	1.8754e+11		<.0001*

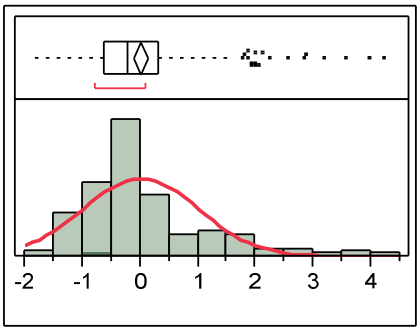
Parameter Estimates

Term	Estimate	Std Error	t-ratio	Prob>|t|	VIF
Intercept	42114.882	2747.569	15.33	<.0001*	.
E1,E2,E3,E4	20085.721	6018.051	3.34	0.0010*	1.0794089
DUMMY FOR DISEASE 2	20959.856	3831.182	5.47	<.0001*	1.0646598
AGE 45-60	8424.9607	3844.763	2.19	0.0295*	1.0998795

Figure 35: Actual by Predicted Plot, Summary of Fit, Analysis of Variance and Parameter Estimates, excluding the Influential Points

The new model, excluding the influential points, has not changed that much. Its predictability, according to the R^2 value, is 17.1248% and the p-values of the F and the t-statistic remain lower than 0.05. The VIF scores of the regressors are below 5. The new model does not pass the Shapiro-Wilk test of Normality. The results of this test are presented in Figure 36.

Studentized Residuals COST



Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	-8.119e-5	-0.131208	0.1310453
Dispersion	σ	1.0048196	0.9202783	1.1065984

-2log(Likelihood) = 648.228417495289

Goodness-of-Fit Test

Shapiro-Wilk W Test

W	Prob<W
0.887470	<.0001*

Figure 36: Shapiro-Wilk Test for Normality for the re-assessed model

The model does not pass the Breusch-Pagan test either. The results of the Breusch-Pagan test are given by Table 26.

Table 26: Breusch-Pagan test for the re-assessed model

	Input			
SSR	5.02E+19		Test Stat	54.00829
SSE	1.55E+11		P-value	1.12E-11
N	228			
df (reg)	3			

The re-assessed model, excluding the two influential points, explains almost the same variation of the dependent variable as model 5 does. Even though this last model yields the highest predictability, compared to the other four, it still cannot be

used to predict and explain the variation of cost, since there is a lot of variability in the data left unexplained. Notwithstanding, all the models developed above, provide valuable information about which factors could affect the cost of hospitalization.

Summary

This chapter presented the results of the Contingency Analysis, the Pivot Table Analysis, and the Regression Analysis. The Contingency Analysis proved the existence of relationship between smoking and pay rank, gender and age. Furthermore, it enriched this study with graphs and informative percentages of the smoking status of each group of smokers, among the ADAF personnel. The Pivot Table Analysis focused on the cost of hospitalization of the ADAF members because of smoking related diseases, and provided meticulous information about the total and average cost of hospitalization for each year and for several groups with different socio-demographic characteristics. Finally, the same data set of cost was used, in order to explore the statistical relationship between cost and several variables related to socio-demographic characteristics of the ADAF population and to the most prevalent diseases related to smoking.

V. Conclusions

Overview

This chapter uses the results of Chapter IV to answer the research questions initially proposed in Chapter I. After the assessment of the questions, an appraisal of the strengths and limitations of this study will be presented. This chapter closes with possible follow-up suggestions for further analysis in future studies.

Findings

In Chapter I, three research questions were defined, with research question number one was partitioned in three sub questions. After accomplishing a literature review and defining the methodology used in this study, two data sets were used with different tools, in order to analyze and answer the research questions, which are the object of this study research. The answers of all the research questions were based on the results of the previous chapter.

Research Question 1: How is smoking affected by the socio-demographic characteristics of the ADAF population?

- How smoking is affected by pay rank?
- How smoking is affected by gender?
- How smoking is affected by age?

This research question was answered after the analysis of the Web HA data set, with the aid of the Contingency Analysis tool. The Mosaic Plots, the Contingency Tables, and the Tests Reports were the products of the Contingency Analysis and demonstrated visually and statistically the existence of a relationship between smoking and pay rank, gender, and age. The Tests Reports of all three

Contingency Analyses developed in Chapter IV each reported very low Likelihood Ratio and Pearson p-values, which indicate a relationship between smoking and the socio-demographic characteristics of pay rank, gender, and age. Furthermore, the percentages of the contingency tables were used in Microsoft Excel® for a visual portrayal of the smoking status of Air Force, based on pay rank, gender and age.

More specifically, regarding the subquestion 'How is smoking affected by pay rank', the analysis showed that the majority of smokers among the ADAF personnel consists of enlisted personnel. First, the Contingency Analysis demonstrated that 48.64% of the whole Air Force population smokes, which is a remarkably high percentage. This percentage is the result of the Contingency Analysis done with the usage of the AF Web HA data set. The percentage of 48.64% regards the active duty personnel that have used any kind of tobacco products in their entire life. That means that 48.64% of the active duty personnel have smoked at least 100 cigarettes or used any other type of tobacco product at least 20 times in their entire life. Second, according to the previously stated information that the majority of smokers are enlisted, 48.64% of smokers, analyzed further, is comprised of 42.46% of enlisted and 6.17% of officers. Lastly, 52.41% of the enlisted population is smokers, which suggests negative consequences for the quality and readiness of this population.

There are more males in the Air Force than females and this fact helps answer the subquestion "How is smoking affected by gender". The percentage of 48.64% of smokers, if partitioned further under the criteria of gender, is comprised of 40.84% male and 7.79% female smokers. Additionally, if smokers are considered a population of their own, this population consists of 83.98% men and 16.02% women, and this fact underscores the prevalence of smoking among men in the Air Force.

The final subquestion of the first research question examines the relationship between smoking and age. It is shown in this research that almost every age group is nearly equally divided into two groups: smokers and non-smokers. In some age groups smokers are the majority, such as in the age groups 17-24 and 25-29. If the percentage 48.64% is broken up into age groups, the 13.94% belongs to the age group 17-24, 12.08% to the age group 25-29 (among these two age groups, smoking is most prevalent), 7.85% to the age group 30-34, 6.59% to the age group 36-39, and 8.18% to the group over 40.

All the above results which correlate smoking with pay rank, gender and age, should be used to better target smoking cessation programs and policies. The answer to the first research question verifies that smoking is more prevalent among the enlisted, males, and the young age groups. Half the enlisted population smokes. 83% of the smoking population is males. The age groups of 17-24 and 25-29 are the groups that smoke more.

Research Question 2: Which diseases cost more to the U.S. Air Force, according to their total cost of hospitalization?

The second research question was answered through the analysis of a different set of data than that of the first question. For the second research question, the data set of the cost of hospitalization of ADAF personnel because of smoking related diseases was used and analyzed with the assistance of Microsoft Excel, and particularly with the Pivot Tables tool. The most important product of this analysis was the list with the most prevalent diseases related to smoking with the highest cost. This list is given below in Table 27.

Table 27: List of the Most Prevalent Diseases Related to Smoking with the Highest Cost

ISCHEMIC HEART DISEASE	**1**
CEREBROVASCULAR DISEASE	**2**
MALIGNANT NEOPLASMS OF TRACHEA, LUNG, BRONCHUS	**3**
MALIGNANT NEOPLASMS OF LIP, ORAL, CAVITY, PHARYX	**4**
OTHER HEART DISEASE	**5**
MALIGNANT NEOPLASMS OF KIDNEY AND RENAL PELVIS	**6**
MALIGNANT NEOPLASMS OF URINARY BLADDER	**7**
BRONCHITIS, EMPHYSEMA	**8**
MALIGNANT NEOPLASMS OF PANCREAS	**9**
OTHER ARTERIAL DISEASE	**10**
OTHERS	**11**

The above table helps answer the second research question, but the manipulation of the data, in the process of answering the second research question, produced numerous results and outcomes. The Pivot Tables showed that the grand total cost of hospitalization for the period 1999-2009 was $49,193,334, where $43,013,308 concerned the male population of the Air Force and $6,180,026 the female population of the Air Force. Furthermore, the grand average cost (the definition "grand" refers to the period 1999-2009) for males was $1,774 and for females $1,480, indicating that the gap of average cost for both genders is not large and cost might not be affected by gender. Of importance is that the age groups with the highest grand total cost were the groups of 33-40 and 41-48 years. In addition, the groups with the highest grand average cost were the age groups 41-48, 49-56, and 57-64 years old. Supplementary information to the statistical analysis, regarding the age groups, is that groups 33-40 and 41-48 years old were the groups with the highest grand total number of visits to hospital or doctor. All this information, combined with the previous Contingency Analysis of the correlation between smoking and age, provides significant evidence that smoking is most prevalent among the young ages of 17 to 29, while the cost consequences of smoking are apparent in the older age groups of 33 to 48. A preventive anti-smoking policy, mostly focused on the younger ages

when people start smoking, could reduce the number of future smokers and save a considerable part of the Air Force budget spent on medical expenses related to smoking.

The outcome of the analysis of the frequency of visits to the hospital or to the doctor, due to a disease related to smoking, generated another list of diseases. In this list, the diseases were sorted by their frequency of appearance in the Primary Diagnosis column. Ischemic heart disease and cerebrovascular disease were the top two diseases, while being the diseases with the highest cost. This validates that these two diseases are the most prevalent diseases related to smoking. Finally, it is worthwhile to mention that the largest part of the grand total medical expenses of the Air Force related to smoking was due to the enlisted population. During the period 1999-2009, $33,571,520 was spent for the hospitalization of enlisted personnel and $15,621,813 was spent for the hospitalization of officers. The pattern of the grand average cost for each pay rank is reversed, with the grand average cost of enlisted was $1,590 and of the officers $2,135. Additionally, the general grand average cost of hospitalization for the whole Air Force population was $1,731.

The answer of the second research question revealed that the most prevalent diseases related to smoking are ischemic heart disease and cerebrovascular disease. Moreover, the largest portion of the medical expenses related to hospitalization corresponds to the enlisted population and the age range 33 to 48. Furthermore, smoking is most prevalent among men. All these conclusions could compose the main targets of a future, more effective anti-smoking campaign and of a beneficiary research for the shrinkage of the medical expenses of the ADAF personnel, related to smoking.

Research Question 3: How is the cost of hospitalization affected by gender, age, pay rank and each disease separately?

The third research question of this study was answered through the analysis of the same cost data set used in research question two. The tool used for the analysis of the data set and for answering the third question was the Regression Analysis, with the aid of JMP. Five models were developed, for five subsets of different cost range, in order to detect a relationship between cost and gender, age, pay rank, and the most prevalent diseases related to smoking with the highest cost. All five models generated low R^2 and low Adjusted R^2 values, meaning that the whole predictability of the models was of minimal importance. There was a lot of variability in the data set, not explained by the models, and for this reason the models did not provide a good fit for the variables with the data. The overall p-value of the F-statistic for all five models was very low, allowing the rejection of the null hypothesis that the explanatory variables are not effective and indicating that the explanatory variables are statistically related to the dependent variable. The same result occurred with the p-values of the t-statistic of each variable used in the models. All the variables resulted in low p-values and VIF scores below 5. The variables used more than one time in the five models developed in the Regression Analysis, are: AGE 45-60, ENLISTED, O1-O2-O3, O4-O5-O6, O7-O8-O9-O10, Ischemic Heart Disease, Cerebrovascular Disease, Malignant Neoplasms of Urinary Bladder, and Other Arterial Disease. These variables have greater effect and better explain the cost of hospitalization. The age that affects cost the most is 45- 60 years old. All the pay grades of officers explain and affect the cost more than the pay grades of enlisted. This might be due to the higher average cost of officers, shown in the Pivot Table analysis. The diseases that affect cost more are the diseases numbered 1, 2, 7 and 10 of the list (See Table 27) of the most prevalent

diseases related to smoking with the highest cost. These diseases correspond to Ischemic Heart Disease, Cerebrovascular Disease, Malignant Neoplasms of Urinary Bladder and Other Arterial Disease. The Regression Analysis proved that the two diseases with the highest cost and with the highest frequency of visits, Ischemic Heart Disease and the Cerebrovascular Disease, are the diseases that affect cost most, and can be used as explanatory variables of cost.

Strengths and Limitations

According to the research questions defined in Chapter I, this study tried to detect the existence of a relationship between smoking and several socio-demographic characteristics associated with the Air Force population, to present the status of smoking among the ADAF personnel, to investigate the factors that affect the cost of hospitalization due to smoking related diseases, and to examine which variables could be the most explanatory ones for the prediction of this cost. The various methods used for the investigation of the research questions and the results returned from the analysis, showed the strengths and the limitations of this study.

One of the strengths of this study is the fact that the Contingency Analysis proved the existence of a strong relationship between smoking and pay rank, gender and age. Moreover, this kind of analysis enriched the study with information and graphs about the smoking status of the ADAF personnel. A second strength of this research is for the findings regarding the assortment of diseases by their total cost and their frequency of visits during the period 1999-2009. This classification was used in the Regression Analysis for the creation of dummy variables, associated with the diseases, which were later used for the development of the OLS linear models. Furthermore, the Pivot Table Analysis enhanced the informative status of this study

about the cost of smoking, generating percentages and graphs associated with the total and average cost of each socio-demographic group of the U.S. Air Force. This type of analysis showed that the grand total cost of enlisted was much higher than the grand total cost of officers, but the grand average cost of officers was higher than the grand average cost of enlisted. This piece of information was confirmed later in the Regression Analysis, where the pay grades of officers were among the explanatory factors of the cost.

The study is limited as the OLS linear models, developed in the Regression Analysis, do not guarantee predictability, and likely cannot be used for future research. The variability, spread in the data, did not permit a good fit of the variables with the data. Nevertheless, the five models developed gave a number of variables that could be used in the future as explanatory variables of cost, in different data sets with lower variability. Even though the predictability of the models is very low and the models do not explain at a satisfactory level the variation in the cost, the same models demonstrated that the variables used as explanatory variables, include Ischemic Heart Disease, Cerebrovascular Disease, the age group of 45-60, and the pay grades of officers, which are variables shown to affect the cost in the Pivot Table Analysis as well. Underneath the Regression Analysis of this study there is strength, limited by the low predictability of the models.

Follow-Up Suggestions for Further Research

Opportunities for further research include the investigation of the average cost of officers. The Pivot Table Analysis revealed that the grand total cost of enlisted is remarkably higher, compared to the cost of the officers. On the other hand, the grand average cost of officers seems to be noticeably higher compared to the cost of

enlisted. Moreover, in the Regression Analysis it was shown that, among the explanatory variables of the cost, there was a variable referencing enlisted and variables associated with all the pay grades of officers. This fact proves that enlisted affect the total cost, but officers affect the average cost. The further research here lies in investigating the factors that influence the average cost of officers and render it higher compared to the one of enlisted.

Other research efforts should be directed at comparing the results of this study with analogous studies, elaborated in the other armed forces of the U.S. military. Identifying differences in the explanatory factors of cost and in the classification of the most prevalent diseases related to smoking with the rest of the armed forces, could grant a better and more scrupulous portrait of the smoking status of the Air Force.

Summary

Smoking is a social phenomenon and nowadays is characterized as an epidemic. It likely affects people of every race and social status. Smoking has become an alarming issue for the U.S. Air Force since, as demonstrated in this study, almost half the population of the ADAF personnel smokes. This study examined the association of smoking status and cost, provoked by smoking, with several socio-demographic characteristics of the Air Force population. These results could be used in the future, for a more effective and focused on specific groups, smoking cessation campaign and policy, for eliminating the smoking phenomenon and improving the quality of health, productivity, and readiness of the U.S. Air Force personnel.

Appendix A. AF Web HA Data Dictionary – Tobacco-Use section

SECTION 8: TOBACCO USE

Q8_1a **Type:** Numeric **Qcode:** T1

Section Number: 8-Tobacco Use **Question:** In your entire life, have you? (Check all that apply) Smoked at least one hundred cigarettes?

Description: Derived from question: "In your entire life, have you? (Check all that apply)" Smoked at least one hundred cigarettes?

Value:	**Description:**
1	Checked - Smoked at least one hundred cigarettes?

Q8_1b **Type:** Numeric **Qcode:** T1

Section Number: 8-Tobacco Use **Question:** In your entire life, have you? (Check all that apply) Smoked a pipe at least 20 times?

Description: Derived from question: "In your entire life, have you? (Check all that apply)" Smoked a pipe at least 20 times?

Value:	**Description:**
1	Checked - Smoked a pipe at least 20 times?

Q8_1c **Type:** Numeric **Qcode:** T1

Section Number: 8-Tobacco Use **Question:** In your entire life, have you? (Check all that apply) Smoked a cigar at least 20 times?

Description: Derived from question: "In your entire life, have you? (Check all that apply)" Smoked a cigar at least 20 times?

Value:	**Description:**
1	Checked - Smoked a cigar at least 20 times?

Q8_1d **Type:** Numeric **Qcode:** T1

Section Number: 8-Tobacco Use **Question:** In your entire life, have you? (Check all that apply) Used chewing tobacco or snuff at least 20 times?

Description: Derived from question: "In your entire life, have you? (Check all that apply)" Used chewing tobacco or snuff at least 20 times?

Value:	**Description:**
1	Checked - Used chewing tobacco or snuff at least 20 times?

Q8_1e **Type:** Numeric **Qcode:** T1

Section Number: 8-Tobacco Use **Question:** In your entire life, have you? (Check all that apply) I only use tobacco products occasionally

Description: Derived from question: "In your entire life, have you? (Check all that apply)" I only use tobacco products occasionally

Value:	**Description:**
1	Checked - I only use tobacco products occasionally

Q8_1f **Type:** Numeric **Qcode:** T1

Section Number: 8-Tobacco Use **Question:** In your entire life, have you? (Check all that apply) I have never used tobacco products

Description: Derived from question: "In your entire life, have you? (Check all that apply)" I have never used tobacco products

Value:	**Description:**
1	Checked - I have never used tobacco products

Q8_2a **Type:** Numeric **Qcode:** T2

Section Number: 8-Tobacco Use **Question:** Do you currently use any of the following tobacco products? (Check all that apply) Cigarettes

Description: Derived from question: "Do you currently use any of the following tobacco products? (Check all that apply)" Cigarettes. (**Asked if Q8_1f NE 1**).

Value:	Description:
1	Checked - Cigarettes

Q8_2b **Type:** Numeric **Qcode:** T2

Section Number: 8-Tobacco Use **Question:** Do you currently use any of the following tobacco products? (Check all that apply) Pipe

Description: Derived from question: "Do you currently use any of the following tobacco products? (Check all that apply)" Pipe. (**Asked if Q8_1f NE 1**).

Value:	Description:
1	Checked - Pipe

Q8_2c **Type:** Numeric **Qcode:** T2

Section Number: 8-Tobacco Use **Question:** Do you currently use any of the following tobacco products? (Check all that apply) Cigars

Description: Derived from question: "Do you currently use any of the following tobacco products? (Check all that apply)" Cigars. (**Asked if Q8_1f NE 1**).

Value:	Description:
1	Checked - Cigars

Q8_2d **Type:** Numeric **Qcode:** T2

Section Number: 8-Tobacco Use **Question:** Do you currently use any of the following tobacco products? (Check all that apply) Chewing tobacco or snuff

Description: Derived from question: "Do you currently use any of the following tobacco products? (Check all that apply)" Chewing tobacco or snuff. (**Asked if Q8_1f NE 1**).

Value:	Description:
1	Checked - Chewing tobacco or snuff

Q8_2e **Type:** Numeric **Qcode:** T2

Section Number: 8-Tobacco Use

Question: Do you currently use any of the following tobacco products? (Check all that apply) None of the above

Description: Derived from question: "Do you currently use any of the following tobacco products? (Check all that apply)" None of the above. (**Asked if Q8_1f NE 1**).

Value:	**Description:**
1	Checked - None of the above

Q8_3 **Type:** Numeric **Qcode:** T3a

Section Number: 8-Tobacco Use

Question: Do you now smoke cigarettes?

Description: Response to question: "Do you now smoke cigarettes?" (**Asked if Q8_2a = 1**).

Value:	**Description:**
1	Smoke cigarettes every day
2	Smoke cigarettes on some days

Q8_4 **Type:** Numeric **Qcode:** T7

Section Number: 8-Tobacco Use

Question: About how long ago was it that you started smoking cigarettes?

Description: Response to question: "About how long ago was it that you started smoking cigarettes?" (**Asked if Q8_3 = 1**).

Value:	**Description:**
1	Less than 1 month ago
2	1 month but less than 3 months ago
3	3 months but less than 6 months ago
4	6 months but less than 12 months ago
5	1 year but less than 5 years ago
6	More than 5 years ago
9	Don't know / Not sure

Q8_5 **Type:** Numeric **Qcode:** T8

Section Number: 8-Tobacco Use

Question: All together, for how many years have you been a regular smoker, not including the years that you had quit?

Description: Response to question: “All together, for how many years have you been a regular smoker, not including the years that you had quit?” (**Asked if Q8_4 = 5 or 6**).

Value:	**Description:**
1	1-2 years
2	3-5 years
3	6-10 years
4	11-15 years
5	16-20 years
6	More than 20 years
9	Don't know / Not sure

Q8_6 **Type:** Numeric **Qcode:** T9

Section Number: 8-Tobacco Use **Question:** On the average, how many cigarettes do you now smoke per day?

Description: Response to question: “On the average, how many cigarettes do you now smoke per day?” (**Asked if Q8_3 = 1**).

Value:	**Description:**
1	Less than 1 cigarette a day
2	1-10 cigarettes a day (half a pack)
3	11-20 cigarettes a day (1 pack)
4	21-30 cigarettes a day (1 and a half packs)
5	31-40 cigarettes a day (2 packs)
6	More than 40 cigarettes a day (more than 2 packs)
9	Don't know / Not sure

Q8_7 **Type:** Numeric **Qcode:** T10

Section Number: 8-Tobacco Use **Question:** Which best describes your intentions regarding quitting smoking?

Description: Response to question: “Which best describes your intentions regarding quitting smoking?” (**Asked if Q8_3 = 1**).

Value:	**Description:**
1	I intend to quit in the next 30 days and have tried for at least 24 hours in the past year
2	I intend to quit in the next 30 days
3	I intend to quit in the next 6 months
4	I do not intend to quit in the next 6 months

Q8_8 **Type:** Numeric **Qcode:** T11

Section Number: 8-Tobacco Use **Question:** About how long ago was it that you started smoking cigarettes?

Description: Response to question: "About how long ago was it that you started smoking cigarettes?" (**Asked if Q8_3 = 2**).

Value:	**Description:**
1	Less than 1 month ago
2	1 month but less than 3 months ago
3	3 months but less than 6 months ago
4	6 months but less than 12 months ago
5	1 year but less than 5 years ago
6	More than 5 years ago
9	Don't know / Not sure

Q8_9 **Type:** Numeric **Qcode:** T12

Section Number: 8-Tobacco Use **Question:** All together, for how many years have you smoked cigarettes, not including the years that you have quit?

Description: Response to question: "All together, for how many years have you smoked cigarettes, not including the years that you have quit?" (**Asked if Q8_8 = 5 or 6**).

Value:	**Description:**
1	1-2 years
2	3-5 years
3	6-10 years
4	11-15 years
5	16-20 years
6	More than 20 years
9	Don't know / Not sure

Q8_10 **Type:** Numeric **Qcode:** T13

Section Number: 8-Tobacco Use **Question:** On the average, when you smoked during the past 30 days, about how many cigarettes did you smoke each day?

Description: Response to question: "On the average, when you smoked during the past 30 days, about how many cigarettes did you smoke each day?" (**Asked if Q8_3 = 2**).

Value:	**Description:**

1	Less than 1 cigarette a day
2	1-10 cigarettes a day (half a pack)
3	11-20 cigarettes a day (1 pack)
4	21-30 cigarettes a day (1 and a half packs)
5	31-40 cigarettes a day (2 packs)
6	More than 40 cigarettes a day (more than 2 packs)
9	Don't know / Not sure

Q8_11 **Type:** Numeric **Qcode:** T14

Section Number: 8-Tobacco Use **Question:** On how many of the past 30 days did you smoke cigarettes?

Description: Response to question: "On how many of the past 30 days did you smoke cigarettes?" (**Asked if Q8_3 = 2**).

Value:	**Description:**
1	1-5 days
2	6-10 days
3	11-15 days
4	16-20 days
5	21-25 days
6	26-30 days
9	Don't know / Not sure

Q8_12 **Type:** Numeric **Qcode:** T15

Section Number: 8-Tobacco Use **Question:** Which best describes your intentions regarding quitting smoking?

Description: Response to question: "Which best describes your intentions regarding quitting smoking?" (**Asked if Q8_3 = 2**).

Value:	**Description:**
1	I intend to quit in the next 30 days and have tried for at least 24 hours in the past year
2	I intend to quit in the next 30 days
3	I intend to quit in the next 6 months
4	I do not intend to quit in the next 6 months

Q8_13 **Type:** Numeric **Qcode:** T3b

Section Number: 8-Tobacco Use **Question:** Do you now smoke a pipe?

Description: Response to question: "Do you now smoke a pipe?" (**Asked if Q8_2b = 1**).

Value:	Description:
1	Smoke a pipe every day
2	Smoke a pipe on some days

Q8_14 **Type:** Numeric **Qcode:** T18AF

Section Number: 8-Tobacco Use **Question:** Which best describes your intentions regarding quitting smoking a pipe?

Description: Response to question: "Which best describes your intentions regarding quitting smoking a pipe?" (**Asked if Q8_2b = 1**).

Value:	Description:
1	I intend to quit in the next 30 days and have tried for at least 24 hours in the past year
2	I intend to quit in the next 30 days
3	I intend to quit in the next 6 months
4	I do not intend to quit in the next 6 months

Q8_15 **Type:** Numeric **Qcode:** T3c

Section Number: 8-Tobacco Use **Question:** Do you now smoke cigars?

Description: Response to question: "Do you now smoke cigars?" (**Asked if Q8_2c = 1**).

Value:	Description:
1	Smoke cigars every day
2	Smoke cigars on some days

Q8_16 **Type:** Numeric **Qcode:** T19af

Section Number: 8-Tobacco Use **Question:** Which best describes your intentions regarding quitting smoking cigars?

Description: Response to question: "Which best describes your intentions regarding quitting smoking cigars?" (**Asked if Q8_2c = 1**).

Value:	Description:
1	I intend to quit in the next 30 days and have tried for at least 24 hours in the past year
2	I intend to quit in the next 30 days
3	I intend to quit in the next 6 months

4	I do not intend to quit in the next 6 months

Q8_17 **Type:** Numeric **Qcode:** T3d

Section Number: 8-Tobacco Use **Question:** Do you now use chewing tobacco or snuff?

Description: Response to question: "Do you now use chewing tobacco or snuff?" (**Asked if Q8_2d = 1**).

Value:	**Description:**
1	Use chewing tobacco or snuff every day
2	Use chewing tobacco or snuff on some days

Q8_18 **Type:** Numeric **Qcode:** T16

Section Number: 8-Tobacco Use **Question:** About how long ago was it that you started using chewing tobacco or snuff regularly?

Description: Response to question: "About how long ago was it that you started using chewing tobacco or snuff regularly?" (**Asked if Q8_2d = 1**).

Value:	**Description:**
1	Less than 1 month ago
2	1 month but less than 3 months ago
3	3 months but less than 6 months ago
4	6 months but less than 12 months ago
5	1 year but less than 5 years ago
6	More than 5 years ago
9	Don't know / Not sure

Q8_19 **Type:** Numeric **Qcode:** T17

Section Number: 8-Tobacco Use **Question:** On the average, when you smoked during the past 30 days, about how many cigarettes did you smoke each day?

Description: Response to question: "On the average, when you smoked during the past 30 days, about how many cigarettes did you smoke each day?" (**Asked if Q8_2d = 1**).

Value:	**Description:**
1	Less than 1 time a day

2	1-2 times a day
3	3-5 times a day
4	6-10 times a day
5	11-20 times a day
6	More than 20 times a day
9	Don't know / Not sure

Q8_20 **Type:** Numeric **Qcode:** T17aAF

Section Number: 8-Tobacco Use **Question:** Which best describes your intentions regarding quitting smokeless tobacco (chewing tobacco or snuff)?

Description: Response to question: "Which best describes your intentions regarding quitting smokeless tobacco (chewing tobacco or snuff)?" (**Asked if Q8_2d = 1**).

Value:	**Description:**
1	I intend to quit in the next 30 days and have tried for at least 24 hours in the past year
2	I intend to quit in the next 30 days
3	I intend to quit in the next 6 months
4	I do not intend to quit in the next 6 months

Q8_21 **Type:** Numeric **Qcode:** T4

Section Number: 8-Tobacco Use **Question:** About how long has it been since you last smoked cigarettes?

Description: Response to question: "About how long has it been since you last smoked cigarettes?" (**Asked if Q8_1a = 1 and Q8_2e=1**).

Value:	**Description:**
1	Less than 1 month ago
2	1 month but less than 3 months ago
3	3 months but less than 6 months ago
4	6 months ago or more
9	Don't know / Not sure

Q8_22 **Type:** Numeric **Qcode:** T5

Section Number: 8-Tobacco Use **Question:** During the years that you smoked, about how many cigarettes per day did you smoke?

Description: Response to question: "During the years that you smoked, about how many cigarettes per day did you smoke?" (**Asked if Q8_1a = 1 and Q8_2e=1**).

Value:	**Description:**
1	Less than 1 cigarette a day
2	1-10 cigarettes a day (half a pack)
3	11-20 cigarettes a day (1 pack)
4	21-30 cigarettes a day (1 and a half packs)
5	31-40 cigarettes a day (2 packs)
6	More than 40 cigarettes a day (more than 2 packs)
9	Don't know / Not sure

Q8_23 **Type:** Numeric **Qcode:** T6

Section Number: 8-Tobacco Use **Question:** All together, for how many years did you smoke cigarettes, not including the years that you had quit?

Description: Response to question: "All together, for how many years did you smoke cigarettes, not including the years that you had quit?" (**Asked if Q8_1a = 1 and Q8_2e=1**).

Value:	**Description:**
1	Less than 1 year
2	1-2 years
3	3-5 years
4	6-10 years
5	11-15 years
6	16-20 years
7	More than 20 years
9	Don't know / Not sure

Q8_24 **Type:** Numeric **Qcode:** T24AF

ADDED 10/31/2008

Section Number: 8-Tobacco Use **Question:** During the past 12 months, have you stopped the use of any tobacco products for one day or longer because you were trying to quit?

Description: Response to question: "During the past 12 months, have you stopped the use of any tobacco products for one day or longer because you were trying to quit?" (**Asked if Q8_1f ne 1**).

Value:	**Description:**

1	Yes
2	No

Q8_25 **Type:** Numeric **Qcode:** T25AF

ADDED 10/31/2008

Section Number: 8-Tobacco Use

Question: When you last stopped the use of tobacco products for one day or longer, did you do anything to assist you with quitting?

Description: Response to question: "When you last stopped the use of tobacco products for one day or longer, did you do anything to assist you with quitting?" (**Asked if Q8_24 = 1**).

Value:	**Description:**
1	Yes
2	No

Q8_26a **Type:** Numeric **Qcode:** T26AF

ADDED 10/31/2008

Section Number: 8-Tobacco Use

Question: When you last stopped the use of tobacco products, what did you do to assist you with quitting? (Check all that apply) Consulted with (non-HAWC) Doctor, Nurse, or other health professional

Description: Derived from response to question: "When you last stopped the use of tobacco products, what did you do to assist you with quitting? (Check all that apply) Consulted with (non-HAWC) Doctor, Nurse, or other health professional" (**Asked if Q8_25 = 1**).

Value:	**Description:**
1	Checked - Consulted with (non-HAWC) Doctor, Nurse, or other health professional

Q8_26b **Type:** Numeric **Qcode:** T26AF

ADDED 10/31/2008

Section Number: 8-Tobacco Use

Question: When you last stopped the use of tobacco products, what did you do to assist you with quitting? (Check all that apply) Medication such as nicotine gum, patch, nasal spray, inhaler, lozenge, or prescription medication

Description: Derived from response to question: "When you last stopped the use of tobacco products, what did you do to assist you with quitting? (Check all that apply) Medication such as nicotine gum, patch, nasal spray, inhaler, lozenge, or prescription medication" (**Asked if Q8_25 = 1**).

Value:	**Description:**
1	Checked - Medication such as nicotine gum, patch, nasal spray, inhaler, lozenge, or prescription medication

Q8_26c **Type:** Numeric **Qcode:** T26AF

ADDED 10/31/2008

Section Number: 8-Tobacco Use

Question: When you last stopped the use of tobacco products, what did you do to assist you with quitting? (Check all that apply) Class

Description: Derived from response to question: "When you last stopped the use of tobacco products, what did you do to assist you with quitting? (Check all that apply) Class" (**Asked if Q8_25 = 1**).

Value:	**Description:**
1	Checked - Class

Q8_26d **Type:** Numeric **Qcode:** T26AF

ADDED 10/31/2008

Section Number: 8-Tobacco Use

Question: When you last stopped the use of tobacco products, what did you do to assist you with quitting? (Check all that apply) Call-line

Description: Derived from response to question: "When you last stopped the use of tobacco products, what did you do to assist you with quitting? (Check all that apply) Call-line" (**Asked if Q8_25 = 1**).

Value:	Description:
1	Checked – Call-line

Q8_26e **Type:** Numeric **Qcode:** T26AF

ADDED 10/31/2008

Section Number: 8-Tobacco Use **Question:** When you last stopped the use of tobacco products, what did you do to assist you with quitting? (Check all that apply) One-on-one session with HAWC staff

Description: Derived from response to question: "When you last stopped the use of tobacco products, what did you do to assist you with quitting? (Check all that apply) One-on-one session with HAWC staff" (**Asked if Q8_25 = 1**).

Value:	Description:
1	Checked - One-on-one session with HAWC staff

Q8_26f **Type:** Numeric **Qcode:** T26AF

ADDED 10/31/2008

Section Number: 8-Tobacco Use **Question:** When you last stopped the use of tobacco products, what did you do to assist you with quitting? (Check all that apply) Web-based support program

Description: Derived from response to question: "When you last stopped the use of tobacco products, what did you do to assist you with quitting? (Check all that apply) Web-based support program" (**Asked if Q8_25 = 1**).

Value:	Description:
1	Checked - One Web-based support program

Q8_26g **Type:** Numeric **Qcode:** T26AF

ADDED 10/31/2008

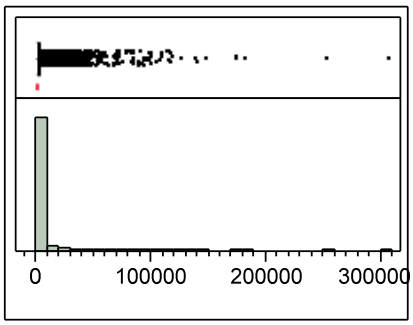
Section Number: 8-Tobacco Use **Question:** When you last stopped the use of tobacco products, what did you do to assist you with quitting? (Check all that apply) None of the above

Description: Derived from response to question: “When you last stopped the use of tobacco products, what did you do to assist you with quitting? (Check all that apply) None of the above” (**Asked if Q8_25 = 1**).

Value:	**Description:**
1	Checked - None of the above

Appendix B. Distributions of the initial data set and of the five subsets

Histogram



Quantiles

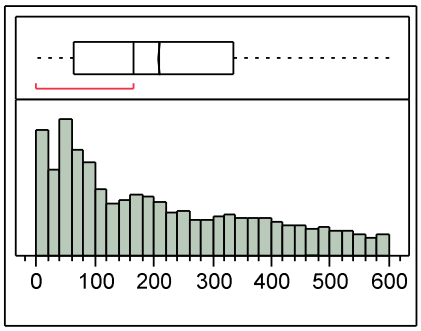
100.0%	maximum	307063
99.5%		38547
97.5%		15031
90.0%		2952
75.0%	quartile	879
50.0%	median	338
25.0%	quartile	103
10.0%		42
2.5%		9.3111
0.5%		0
0.0%	minimum	0

Moments

Mean	1731.0016
Std Dev	6685.8571
Std Err Mean	39.659999
Upper 95% Mean	1808.7371
Lower 95% Mean	1653.2661
N	28419

Figure 37: Distribution of the Initial data set

Histogram



Quantiles

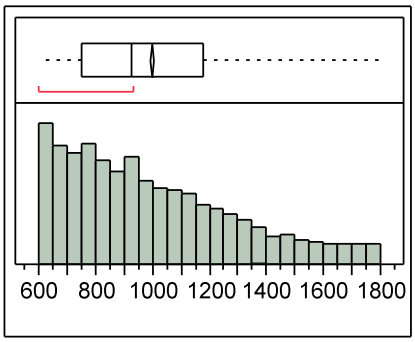
100.0%	maximum	599.99
99.5%		593.95
97.5%		559.09
90.0%		467.69
75.0%	quartile	334.98
50.0%	median	166.99
25.0%	quartile	64.67
10.0%		26.64
2.5%		0.00
0.5%		0.00
0.0%	minimum	0.00

Moments

Mean	207.9587
Std Dev	165.58556
Std Err Mean	1.207047
Upper 95% Mean	210.32462
Lower 95% Mean	205.59278
N	18819

Figure 38: Distribution of Cost for the range $0 - $600.00

Histogram



Quantiles

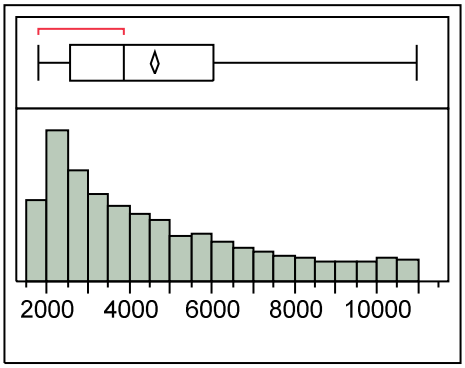
100.0%	maximum	1799.3
99.5%		1781.5
97.5%		1700.6
90.0%		1440.6
75.0%	quartile	1180.0
50.0%	median	928.9
25.0%	quartile	754.0
10.0%		654.2
2.5%		612.4
0.5%		602.7
0.0%	minimum	600.1

Moments

Mean	994.45493
Std Dev	297.07906
Std Err Mean	3.9038692
Upper 95% Mean	1002.108
Lower 95% Mean	986.80188
N	5791

Figure 39: Distribution of Cost for the range $600.01 - $1,800.00

Histogram



Quantiles

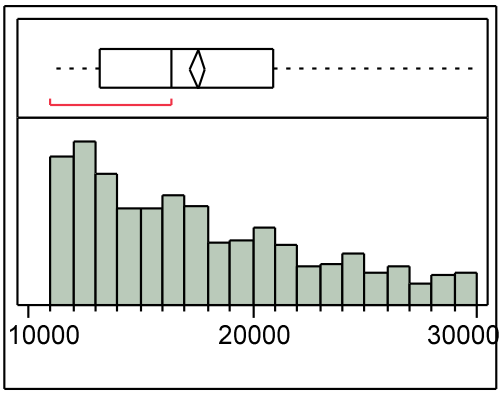
100.0%	maximum	10991.9
99.5%		10879.4
97.5%		10456.9
90.0%		8548.81
75.0%	quartile	6037.71
50.0%	median	3869.56
25.0%	quartile	2551.11
10.0%		2051.51
2.5%		1853.76
0.5%		1813.55
0.0%	minimum	1801.03

Moments

Mean	4578.2267
Std Dev	2452.428
Std Err Mean	46.471177
Upper 95% Mean	4669.3481
Lower 95% Mean	4487.1052
N	2785

Figure 40: Distribution of cost for the range $1,800.01 – 11,000.00

Histogram



Quantiles

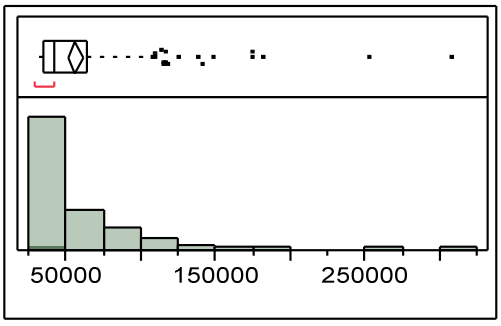
100.0%	maximum	29956
99.5%		29738
97.5%		28932
90.0%		25144
75.0%	quartile	20939
50.0%	median	16392
25.0%	quartile	13201
10.0%		11851
2.5%		11194
0.5%		11028
0.0%	minimum	11009

Moments

	17518.162
Mean	5007.077
Std Dev	177.69451
Std Err Mean	17866.969
Upper 95% Mean	17169.354
Lower 95% Mean	794
N	

Figure 41: Distribution of cost for the range $ 11,000.01 – $30,000.00

Histogram



Quantiles

100.0%	maximum	307063
99.5%		298575
97.5%		154285
90.0%		97739
75.0%	quartile	63980
50.0%	median	41818
25.0%	quartile	35372
10.0%		31727
2.5%		30155
0.5%		30027
0.0%	minimum	30022

Moments

Mean	55917.783
Std Dev	35589.829
Std Err Mean	2346.7238
Upper 95% Mean	60541.714
Lower 95% Mean	51293.852
N	230

Figure 42: Distribution of cost for the range $ 30,000.01 – $307,100.00

Appendix C. Cook's Distance Plots for Models 2, 3, and 4

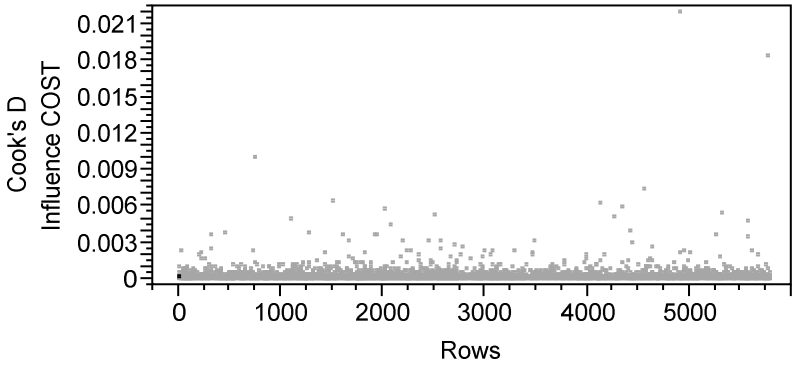


Figure 43: Cook's Distance Overlay Plot for Model 2

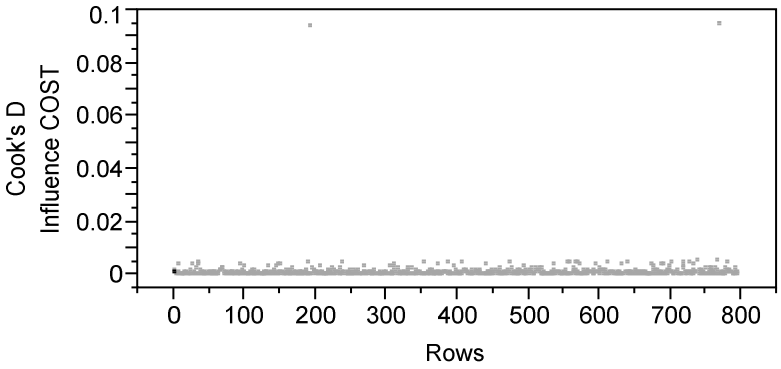


Figure 44: Cook's Distance Overlay Plot for Model 3

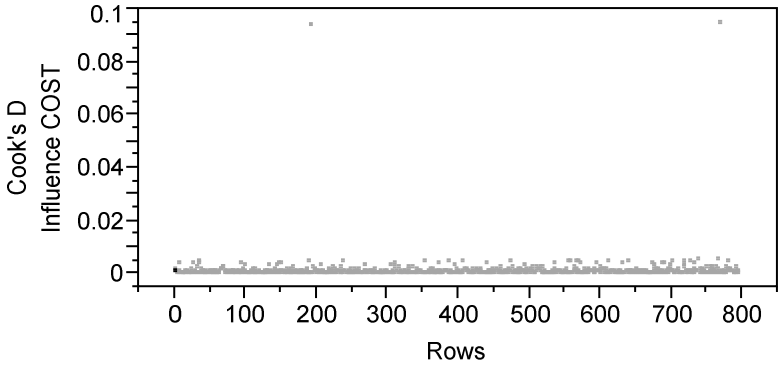


Figure 45: Cook's Distance Overlay Plot for Model 4

Bibliography

1. Adhikari, B., J. Kahende, A. Malarcher, T. Pechacek, and V. Tong. "Smoking-Attributable Mortality, Years of Potential Life Lost, and Productivity Losses- United States, 2000-2004", *MMWR Morbidity& Mortality Weekly Report*, 57: 1226-1228 (November 2008).

2. AF Web HA Data Dictionary, AFCHIPS4.AF_WEBHA.dbo.WEBHA_Analyst, Healthcare Informatics Division, AF/SG6H , San Antonio, Texas, 2 March 2010

3. American Council for Drug Education's. "Basic facts about Drugs: Tobacco," 10 December 2010. http://www.acde.org/common/Tobacco.htm.

4. American Heart Association. "Nicotine Addiction," 1 November 2010. http://www.americanheart.org/presenter.jhtml?identifier=4753

5. Annis, Charles. "R-squared." 8 April 2011. http://www.statisticalengineering.com/r-squared.htm.

6. Arvey, S.R. and Malone R.E. "Advance and retreat: tobacco control policy in the U.S. military," Mil Med, 173: 985-991 (October 2008).

7. Bray, Robert M., Laurel L. Hourani, Kristine L. Rae Olmsted, Michael Witt, Janice M. Brown, Michael R. Pemberton, Mary Ellen Marsden, Bernadette Marriott, Scott Scheffler, Russ Vandermaas-Peeler, BeLinda Weimer, Sara Calvin, Michael Bradshaw, Kelly Close, and Douglas Hayden. "Department of Defense Survey of Health Related Behaviors Among Active Duty Military Personnel - A Component of the Defense Lifestyle Assessment Program (DLAP)." Report to the Assistant Secretary of Defense (Health Affairs) under Cooperative Agreement No. DAMD 17-00-2-0057. December 2006.

8. "CDC chief says U.S. smoking rate of 20% is a paradox" ,*USA Today* (9/7/2010). 02 November 2010 http://www.usatoday.com/yourlife/health/medical/2010-09-07-smoking-rates_N.htm

9. CDC. "Fast Facts," 17 February 2011 http://www.cdc.gov/tobacco/data_statistics/fact_sheets/fast_facts

10. Commander, Submarine Forces Public Affairs. “Smoking To Be Extinguished On Submarines.” 16 Nov 2010. http://www. Navy.mil

11. Conway, T.L. and T.A. Cronan. “Smoking and Physical Fitness Among Navy Shipboard Personnel,” *Military Medicine*, 153:589-594 (1988).

12. Diwan, Piyush. “Smoking Still a Problem in the United States”. 30 November 2010. http://www.topnews.in

13. Dube, S.R., A. McClave, C. James, R. Caraballo, R. Kaufmann, and T. Pechacek. “Vital Signs: Current Cigarette Smoking among Adults Aged ≥ 18 Years- United States, 2009,” *MMWR Morbidity& Mortality Weekly Report*, 59: 1135-1140 (September 2010).

14. Ebbert, Jon O.,C. Keith Haddock, Mark Vander Weg, Robert C. Klesges, Walker .S.C. Poston, and Margaret DeBon. “Predictors of smokeless initiation in a young adult military cohort,” *American Journal of Health Behavior*, 30: 103-122 (Jan-Feb 2006).

15. Emanuel, Jeff. “Old enough to fight, but not old enough to light up.” 1 December 2010. http://www.cbsnews.com/stories

16. Fraser, James D., Richard G. Best, Joseph H. Kelly, Valerie Forman-Hoffman, Cecilia Cho, and Lanna J. Forrest. “Evaluation of Tobacco Use Cessation Programs in the Military Health System.” Report to TRICARE Management Activity Office of the Chief Medical Officer. February 2009.

17. Haddock, C. Keith, Jennifer E. Taylor, Kevin M. Hoffman, Walker S.C. Poston, Alan Peterson, Harry A. Lando, and Suzanne Shelton. “Factors which influence tobacco use among junior enlisted personnel in the United States Army and Air Force: a formative research study,” *American Journal of Health Promotion*, 23:241-246 (March – April 2009).

18. Haddock, C. Keith, Mark Vander Weg, Margaret DeBon, Robert C. Klegges, G. Wayne Talcott, Harry Lando, and Alan Peterson. “Evidence that smokeless tobacco use is a gateway for smoking initiation in young adult males,” *Preventive Medicine*, 32: 262-267 (March 2001).

19. Haddock, C. Keith, Robert C. Klesges, Gerald W. Talcott, Harry Lando, and Risa J. Stein. “Smoking prevalence and risk factors for smoking in a population of United States Air Force basic trainees,” *Tobacco Control*, 7:232-235 (September 1998).

20. Haddock, C.K., S.A. Pyle, W.S. Poston, R.M. Bray, and R.J. Stein. “Smoking and Body Weight as Markers of Fitness for Duty among U.S. military Personnel.” *Military Medicine*, 172:527-532 (May 2007).

21. Larson Gerald E., Stephanie Booth-Kewley, and Margaret A.K. Ryan. “Tobacco Smoking as an Index of Military Personnel Quality,” *Military Psychology*, 19: 273-287 (2007).

22. Hoffman, K.M., C.K. Haddock, W.S. Poston, J.E. Taylor, H.A. Lando, and S. Shelton. “A formative examination of messages which discourage tobacco use among junior enlisted members of the United States Military,” *Nicotine Tob Res*, 10: 653-661 (April 2008).

23. JMP. “Statistics and Graphics Guide, release 7.” 2007. p. 180-183. 2 April 2011. http://www.jmp.com/support/downloads/pdf/jmp_stat_graph_guide.pdf

24. Kaiserman, Murray J. *The Cost of Smoking in Canada, 1991*. Ontario: Health Canada, 1996.

25. Klesges, Robert C., C. Keith Haddock, Cyril F. Chang, G. Wayne Talcott, and Harry A. Lando. “The association of smoking and the cost of military training,” *Tobacco Control*, 10: 43-47 (March 2001).

26. Larson Gerald E., Stephanie Booth-Kewley, and Margaret A.K. Ryan. “Tobacco Smoking as an Index of Military Personnel Quality,” *Military Psychology*, 19: 273-287 (2007).

27. Legacy for Longer Healthier Lives. “Tobacco Use in the Military,” 9 April 2011 https://www.legacyforhealth.org/PDFPublications/Military_FactSheet.pdf

28. Mackay, Judith and Michael Eriksen. “The Tobacco Atlas,” Switzerland: World Health Organization, 2002.

29. Mallin, Robert. “Smoking Cessation: Integration of behavioral and drug therapies,” *Am Fam Physician*, 65: 1107-1114 (March 2002).

30. Miller, V.P., C. Ernst, and F. Collin. “Smoking-attributable medical care costs in the USA.” *Soc Sci Med*, 48:375-91 (1999).

31. MMWR. “Cigarette Smoking-Attributable Morbidity – United States, 2000”, *MMWR Morbidity& Mortality Weekly Report*, 52: 842-844 (September 2003).

32. MMWR. “Costs of Smoking among Active Duty U.S. Air Force Personnel—United States, 1997,” *MMWR Morbidity& Mortality Weekly Report*, 49: 441-445 (May 2000).

33. MMWR. “Tobacco use—United States, 1900-1999,” *MMWR Morbidity& Mortality Weekly Report*, 48:986-93 (5 November 1999).

34. Nelson, Jenenne, Linda L. Pederson, and Judene Lewis. “Tobacco use in the Army: Illuminating patterns, practices, and options for treatment,” Military Medicine, 174: 162-169 (February 2009).

35. Nelson, J.P. and L.L. Pederson. “Military Tobacco Use: A synthesis of the literature on prevalence, factors related to use, and cessation interventions,” *Nicotine and Tobacco Research*, 10: 755-790 (May 2008).

36. Poston, W.S.C., J.E. Taylor, K.M. Hoffman, A.L. Peterson, H.A. Lando, S. Shelton, and C.K. Haddock. “Smoking and deployment: Perspective of junior- enlisted U.S. Air Force and U.S. Army personnel and their supervisors,” *Military Medicine*, 173: 441-447 (May 2008).

37. Pyle, Sara A, C. Keith Haddock, Walker S. Carlos Poston, Robert M. Bray, and Jason Williams. “Tobacco use and perceived financial strain among junior enlisted in the U.S. military in 2002,” *Preventive Medicine*, 45: 460-463 (May 2007).

38. Riechman, Deb “Smoking in the military: An old habit dies hard,” *The Associated Press*, 17: 14-24 (September 2009).

39. Sarah R. Arvey and Ruth E. Malone. “Advance and Retreat: Tobacco Control Policy in the U.S. military,” *Military Medicine*, 173(10): 985-991 (October 2008)

40. SCEA: The Society of cost Estimating and Analysis. “SCEA Glossary,” 31 March 2011. http://www.sceaonline.org/prof_dev/glossary-c.cfm

41. Schlotzhauer, Sandra D. *Elementary Statistics Using JMP*. Cary, NC: SAS Institute Inc., 2007.

42. Schroeder, Erich W. The Association Between Mental Health and Cigarette Smoking in Active Duty Military Members. Case Number 88ABW-2011-0488. Air force Research Laboratory, 711th Human Performance Wing, Scholl of Aerospace Medicine, Graduate medical Education, Brooks City-Base TX, February 2011. (AFRL-SA-BR-SR-2011-0001).

43. Severson, Hebert H., A.L. Peterson, Judy A. Andrews, Judith S. Gordon, Jeffrey A. Cigrang, Brian G. Danaher, Christine M. Hunter , and Maureen Barckley. “Smokeless tobacco cessation in military personnel: a randomized controlled trial,” *Nicotine Tob Res*, 11: 730-738 (June 2009).

44. Skrivanek, Smita. “The use of dummy variables in regression analysis.” 5 April 2011. http://www.moresteam.com/whitepapers/dummy-variables.pdf

45. Smith, Besa, Margaret A.K. Ryan, Deborah L. Wingard, Thomas L. Patterson, Donald J. Slymen, and Caroline A. Macera. “Cigarette Smoking and Military Deployment: A Prospective Evaluation,” *American Journal of Preventive Medicine*, 35: 539-546 (December 2008).

46. Smoking Attributable Mortality, Morbidity and Economic Costs (SAMMEC). “ICD 9/10 Codes for Smoking-Attributable Mortality Fractions,” 5 October 2010. http://apps.nccd.cdc.gov/sammec/saf_reports.asp

47. Stein, Risa J., Sara A. Pyle, C. Keith Haddock, W.S. Carlos Poston, Robert Bray, and Jason Williams, “ Reported stress and its relationship to tobacco use among U.S. military Personnel,” *Military Medicine*, 173: 271-277 (March 2008).

48. The Free Dictionary. “Medical Dictionary,” 15 November 2010. http://medical-dictionary.thefreedictionary.com

49. The Lewin Group. *Cost of Overweight and Obesity, High Alcohol Consumption, and Tobacco Use within the TRICARE Prime Population, Summary Report for CONUS, FY 2008, Air Force*. Contract HHSP23320045017XI. Office of the Assistant Secretary of Defense, Health Affairs (OASD (HA)), TRICARE Management Activity, March 2010.

50. Treloar, Andrew. "Statistical Analysis." 1 April 2011. http://andrew.treloar.net/research/theses/phd/thesis-152.shtml

51. University of Minnesota, Division of Periodontology. "Tobacco use cessation program," 02 November 2010. http://www1.umn.edu/perio/tobacco/secondhandsmoke.html

52. World Health Organization. *WHO Report on the Global Tobacco Epidemic, 2008: The MPOWER Package*. NLM classification: WM 290. Geneva, World Health Organization, 2008

Vita

Captain Michail Gkoutouloudis was born in Sydney, Australia. He is an Officer in the Hellenic Army. He graduated from the Lyceum in Skoutari - Serron, Greece and enrolled in the Hellenic Officers Military Academy in 1995. He graduated with a Bachelor in Economics from the School of Law and Economics in Aristotle's University of Thessalonica, Greece in 1999, and was assigned to several posts of the Economic Services of the Hellenic Army, some of them to islands of the Eastern Aegean Sea and Cyprus.

In August of 2009, Captain Michail Gkoutouloudis entered the Cost Analysis Master's Program at the Air Force Institute of Technology's School of Engineering and Management. Upon graduation, he will be assigned to the Hellenic Army General Staff.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From – To)*
09-07-2011	Master's Thesis	Sep 2009 – Sep 2011

4. TITLE AND SUBTITLE

Smoking in the United States Air Force: Trends, Most Prevalent Diseases and their Association with Cost

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Gkoutouloudis, Michail, Captain, Hellenic Army (HA)

5d. PROJECT NUMBER
n/a

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)
Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT/GCA/ENV/11-S02

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Health and Wellness Center (HAWC)
Dr. Brenda Moore
Health Education / Tobacco Cessation
88 AMDS SGPZ Phone: 937-904-9362
2690 C St, Bldg 571, Wright-Patterson AFB OH 45433 e-mail:
brenda.moore.ctr@wpafb.af.mil

10. SPONSOR/MONITOR'S ACRONYM(S)
HAWC/88 AMDS SGPZ

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This research focuses on the smoking rates among the Active Duty Air Force (ADAF) personnel and the association of smoking and cost of hospitalization because of diseases related to smoking. Three types of analyses were used in this research. The Contingency Analysis was based on the data taken from the Air Force Web HA questionnaire. The Pivot Table Analysis and the Regression Analysis were based on a second data set associated with the cost of hospitalization.

The Contingency Analysis showed that smoking in the U.S. Air Force is more prevalent among the enlisted, males, and the younger age groups. The Pivot Table Analysis demonstrated that ischemic heart disease and cerebrovascular disease present the highest cost. Moreover, enlisted exhibit higher total cost compared to officers, but when referring to the average cost, the situation is reversed. The Regression Analysis exhibited that the variables, related to socio-demographic characteristics, that explain better the cost of hospitalization are the age group of 45-60, the enlisted personnel, and all the pay ranks of the officers, while the diseases that affect more the cost of hospitalization are ischemic heart disease, cerebrovascular disease, malignant neoplasms of the urinary bladder, and other arterial diseases.

15. SUBJECT TERMS
Smoking, Most Prevalent Diseases, Cost, Cerebrovascular Disease, Ischemic Heart Disease.

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Dirk P. Yamamoto, Lt Col, USAF
U	U	U	UU	160	**19b. TELEPHONE NUMBER** *(Include area code)* 937-255-3636x4511

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18

Prescribed by ANSI Std. Z39-18