ASSOCIATION BETWEEN PATIENT-LEVEL PREDICTORS AND GYNECOLOGIC CANCER AMONG WOMEN WITH AUTOIMMUNE DISEASES

by

Zahra Bahrani-Mostafavi

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Health Services Research

Charlotte

2019

Approved by:

_____
Dr. Larissa R. Brunner Huber

_____
Dr. Reuben Howden

_____
Dr. Yvette M. Huet

_____
Dr. Wei Sha

_____
Dr. Christine Richardson

ABSTRACT

ZAHRA BAHRANI-MOSTAFAVI. Association Between Patient-Level Predictors and Gynecologic Cancer Among Women with Autoimmune Diseases (Under the direction of Dr. LARISSA R. BRUNNER HUBER)

Cancer and autoimmunity are two major chronic diseases among women. Based on previous studies that cancer and autoimmune diseases (AD) are both the accumulative effect of genetics and environmental exposures, it was imperative to recognize the profile of comorbidities and how they relate to patients' characteristics. This dissertation research was conducted to investigate the association between patient-level predictors and gynecologic cancer (GYNC) among U.S. women with AD by examining 2 aims: 1) to evaluate the association between patient-level predictors and GYNC among AD patients and 2) to determine the various subpopulations of AD patients with increased likelihood of GYNC based on patients' characteristics and comorbid conditions, using Classification Tree Analysis. This study was a secondary data analysis of 2007-2013 Florida State Inpatient Data samples of the Healthcare Cost and Utilization Project (HCUP). The study population included women with any AD diagnosis. The key predictors –including age, race/ethnicity, insurance type, length of stay in hospital, and median income level, in addition to 31 comorbidities used by Elixhauser et al.–were chosen based on previous findings in literature on their association with cancer globally. In this study, it was found that older age had decreased odds with GYNC among women with AD (45-65 years old: OR = 0.90, 95% CI: 0.82-0.99; and > 65 years old: OR = 0.79, 95% CI: 0.70-0.88). Medicaid holders and self-pay patients, and patients with GYN related procedures such as hysterectomy had increased odds of GYNC among patients

with AD (Hysterectomy: OR = 41.38, 95% CI: 37.40-45.78). Comorbidities such as AIDS/HIV, coagulopathy, weight loss, fluid and electrolyte disorders, renal failure, and obesity were found to have strong associations with GYNC among women with AD. Using predictive analytics some comorbidities such as coagulopathy, rheumatoid arthritis, and chronic pulmonary disease along with hysterectomy showed to be strong predictors of GYNC among specific populations of ADs. The unique combinations of characteristics that described subgroups of patients at risk for GYNC can be used as a potential risk assessment for GYNC as well as for early detection and/or prevention tools. The strong correlation between potential predictors and GYNC may lead the women with AD to be recommended for yearly cancer screening for early detection, and better management of possible GYN cancer toward women health.

ACKNOWLEDGEMENTS

I would like to express my utmost appreciation to my advisor Dr. Larissa R. Brunner Huber for her mentorship throughout my research study and to every member of my dissertation committee: Dr. Wei Sha for her technical supervision on data analytics; Dr. Yvette M. Huet for her valuable direction in pursuing my doctoral study in HSR program; Dr. Reuben Howden for his mentorship on extra-curricular activities during my study; and Dr. Christine Richardson for her mentorship on the science of cancer biology. Special Thanks to Dr. James Studnicki for providing my research data.

My deepest gratitude goes to my husband Taghi for his endless love and support throughout many years of my study and his constant encouragement for me to attain my dream. Most significant appreciation goes to my children Ana, Behrooz, and Amanda for their ongoing love and to my dearest granddaughter Zelda who brought me the joyful time with her birth while doing my research study.

My greatest love and appreciation goes to my late parents who taught me the respect for knowledge and encouraged me to pursue higher education. I express my appreciation to all my siblings, nieces and nephews for their love, support, and guidance throughout my study years.

Special thanks to all my classmates, colleagues, and friends for their friendship and encouragement during my doctoral program. Lastly, I would like to acknowledge the UNC Charlotte Graduate School for supporting my doctoral study with Graduate Assistant Support Plan and the 2019 Summer Research Fellowship awards.

DEDICATION

I dedicate this dissertation to my daughter Anahita Zohreh, and to my granddaughter Zelda Afsaneh, the 2 youngest women of my life. May they pick up the "gauntlet" and pursue their aspiration toward the betterment of the world, the earth, and the mankind.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACS	American Cancer Society
AD(s)	Autoimmune Disease(s)
AER	Absolute Excess Risk
AHRQ	Agency for Healthcare Research and Quality
APS-1	Anti-phospholipid Syndrome, Primary
APS-2	Anti-phospholipid Syndrome, Secondary
AS	Ankylosing Spondylitis
AUC	Area Under Curve
BRCA1	Breast Cancer Antigen 1
BRCA2	Breast Cancer Antigen 2
BSF	Boot Strap Forest
BSO	Bilateral Salpingo-oophorectomy
CNS	Central Nervous System
C5a	Complement 5 component a
DT	Decision Tree
ECI	Elixhauser Comorbidity Index
EOC	Epithelial Ovarian Cancer
FPR	False Positive Rate
GYNC	Gynecologic Cancer
GYN	Gynecologic
HCUP	Healthcare Cost and Utilization Project
HIV	Human Immunodeficiency Syndrome

HL	Hodgkin's Lymphoma
HPV	Human Papilloma Virus
HRT	Hormone Replacement Therapy
ICD-9-CM	International Classification of Diseases, 9th Revision, Clinical Modification
LOS	Hospital Length of Stay
ITP	Autoimmune Thrombocytopenic Purpura
MG	Myasthenia Gravis
MS	Multiple Sclerosis
NCI	National Cancer Institute
NEC	Not Elsewhere Classified
NF-KB	Nuclear Factor-k B
NIH	National Institutes of Health
NO	Nitrogen Oxide
OAP	Overall Accurate Prediction
OC	Ovarian Cancer
RA	Rheumatoid Arthritis
ROC	Receiving Operation Curve
SES	Socioeconomic Status
SHRR	Standardized Hospitalization Rate Ratio
SID	State Inpatient Data
SLE	Systemic Lupus Erythematous
SSC	Systemic Sclerosis
STAT 3	Signal Transducers and Activators of Transcription 3

WG	Wagner's Granulomatous
WHO	World Health Organization
WPV	Wrongly Predicted Value

CHAPTER 1: INTRODUCTION

1.1 Background

Cancer and autoimmunity are two major chronic diseases among women.[1,2] Cancer is a major health problem in the U.S. as well as many other parts of the world.[1] Despite rapid advancement in cancer research during the past decades, cancer continues to be a worldwide killer.[3,4] At the present time, cancer is the second leading cause of death in the U.S. after heart disease.[5-8] However, according to projections made by the World Health Organization (WHO) in 2013, it is expected that cancer will surpass ischemic heart disease as the leading cause of death by the year 2030.[4] The American Cancer Society (ACS) has estimated that 1,762,450 new cases of cancer will occur in the U.S. in 2019, and that an estimated 606,880 individuals will die from the disease.[1] Nearly, 47% of these deaths will be in women, and almost 12% of these deaths will be due to gynecologic cancer (GYNC).[1]

GYNC refer to a group of cancers, in which the initiation of each cancer occurs in a specific woman's reproductive organ, and each is named after the place where it started. Each GYNC is unique, and involves different signs, symptoms, and risk factors.[9] There are five major types of GYNC, which include: cancers of cervix, ovaries, uterine, vagina, and vulva. Additional types of GYNC include fallopian tube cancer and primary peritoneal cancer.[9]

Autoimmune diseases (ADs) are referred to the group of disorders that occurs when the immune system attacks the host's own organs and tissues instead of protecting them.[10,11] Autoimmune related diseases affect 5-10% of men and women of all ages, races, ethnicities, and socioeconomic statuses in developed countries[12,13] with women generally having a 2.7 times greater risk of acquiring an AD than men.[11,14,15] The National Institutes of Health (NIH) considers autoimmunity to be a major women's health issue.[11]

ADs include but are not limited to diseases such as rheumatoid arthritis (RA), juvenile RA, systemic lupus erythematous (SLE), ankylosing spondylitis (AS), inflammatory bowel disease (Crohn's disease and ulcerative colitis), psoriasis, autoimmune thyroid disease (Hashimoto thyroiditis and Graves' disease), vitiligo, multiple sclerosis (MS), type 1 diabetes, and celiac disease,[11,15] Sjogren's syndrome, antiphospholipid syndrome-secondary (APS-2), primary biliary cirrhosis, autoimmune hepatitis, scleroderma, anti-phospholipid syndrome primary (APS-1), autoimmune thrombocytopenic purpura (ITP), and myasthenia gravis (MG).[14]

1.1.1 *Gynecologic Cancer Incidence, Risk Factors, and Treatment*

GYNC accounts for 6.2% of all cancers and is the fourth leading cause of cancer death for women in the U.S.[1,9,16] The ACS estimated an annual diagnosis of more than 109,000 GYNC cases, with over 33,000 deaths in 2019.[1] There are some known risk factors associated with each GYNC. For example, older age, obesity (specifically fatty abdomen), and hormone therapy (estrogen) are common risk factors for uterine cancer.[17,18]

With cervical, vaginal, and vulvar cancers, human papilloma virus (HPV) infection and smoking are the major risk factors.[9,19-21] Older age and family history are known risk factors for ovarian cancer.[18,22] However, the genetic mutation called BRCA1 or BRCA2, and endometriosis are common risk factors for both ovarian and fallopian tube cancers.[9,22] Race/ ethnicity is also a risk factor for GYNC. White women have a higher incidence of OC followed by Hispanic, Asian/Pacific Islander, Black, and American Indian/Alaska Native women.[23] However, Black women have a 31% more likelihood of dying from OC than White women.[9,23] Although endometrial and cervical cancer mortality is twice as much in Black women than White women,[23] ethnicity seems to not be a foremost factor for these cancers. Rather, socioeconomic status and health disparities are the factors associated with endometrial and cervical cancer.[23]

Treatment of GYNC is complex and dependent on the type of cancer and the stage at which the cancer has been diagnosed.[9] Possible treatment includes surgery, radiation therapy, hormonal therapy, and chemotherapy.[3] Surgeries for GYNC include different types of hysterectomy (removal of uterus) including partial or total for uterine, ovarian, cervical, and some advanced cases of vaginal cancers. For local vaginal and vulvar cancers, local excision, or vaginectomy and vulvectomy procedures are done, respectively.[9,20,21] The hysterectomy procedures are performed through abdominal surgery, or by laparoscopic procedure, based on gynecological oncologists' recommendations.

1.1.2 *Autoimmune Diseases Incidence, Risk Factors, and Treatment*

ADs vary widely from affecting a single organ such as pancreatic islets in Type I diabetes mellitus, or the central nervous system in multiple sclerosis (MS), to affecting

many organs as is the case in SLE.[13] In addition, ADs are the underlying cause of 100 chronic illnesses[11,13] in which a combination of two major factors — hereditary and environmental factors, particularly infection — are involved.[2,13,24-26]

So far, researchers have identified 80-100 different types of ADs and they think there are at least 40 additional diseases with an autoimmune etiology.[11] In addition, it has been estimated that ADs are responsible for more than $100 billion in direct health care costs annually,[2,11,27,28] which place autoimmunity among the diseases that are most costly and the hardest to diagnose and treat.[10,11] Whether autoimmunity is affecting a single tissue or multiple organs, the mechanism of pathogenicity is due to the disruption of balance within an immune system, where the ability to tolerate self-constituents must be maintained in a normal healthy individual.[10,11] Consequently, treatment often involves therapies that target activated immunity against self. The most common treatment for ADs is immunosuppressant agents, which tend to control the autoimmune activities and to maintain the self-tolerance. However, this kind of treatment is often accompanied by the pitfalls of immunosuppressants,[10] including problems such as susceptibility to multiple infections and cancer.

There are several risk factors involved with ADs. One is gender, with women at greater risk of having ADs than men because of women's enhanced immune systems.[11,14] This specifically is due to hormonal factors including estrogen and progesterone – the female sex hormones – along with prolactin hormone, which are immune stimulants. Another risk factor is age. ADs affect young to middle-aged women, more than older women and men of any age. This is clearly due to the female reproductive hormones levels being active and highest level during the reproductive age span.[11] With regards to

race/ethnicity, Black, American Indian, and Latinas are generally more susceptible to ADs than Caucasians.[11] Exposure to some environmental agents such as drugs, metals, or pollutants may also increase the risk of developing ADs.[11] Moreover, hereditary and infections are two major risk factors for the onset of autoimmunity.[26]

1.2 Significance

The immune system plays a significant role in the pathogenesis of autoimmune diseases and cancer.[26-28] Autoimmunity causes dysregulation of immune response through immune complex-induced activation of complement system (a system that enhances immunity) in patients with AD. This enhancement of the immune system occurs through complement fifth component (C5a) as an inflammatory mediator,[29] which causes chronic inflammation. This inflammatory chronic inflammation in turn acts as a promoter of tumorigenesis[25] through induction of mutations, genomic instability, early tumor promotion, and enhanced tumor angiogenesis.[29,30]

Because of the high prevalence of ADs among women, and the important role of chronic inflammation in initiation of both autoimmunity and tumorigenesis and cancer,[30,31] more elucidation on the association between GYNC and ADs is essential. There has been limited research performed on the association between GYNC and ADs. Previous studies have mostly addressed only one type of AD with one or a few types of female cancer.[11] For example, Parikh-Patel et al.[32] investigated cancer risk for multiple types of cancer among cohorts of both sexes in RA patients in California; however, only a few types of female cancer were studied.[32] In another study, Hemminki et al.[12] studied the effect of multiple ADs on risk and survival in five female cancers among hospitalized

Swedish patients.[12] However, there has not yet been any comprehensive association study done among the U.S. population to examine GYNC in relation to most known ADs.

Considering cancer and ADs are both the accumulative effect of genetics and environmental exposure,[24,25,27] and that 90-95% of all cancers are linked to environmental factors and living effects,[27] conducting a comprehensive study to investigate the association between patient-level predictors and several GYNC among AD patients in the U.S. population was essential. Along with the diverse nature of ADs and GYNC, and the multiple risk factors involved with both diseases, it was important to recognize the profile of comorbidities and how they related to patients' characteristics. The tie between patient-level predictors and GYNC may help to discover novel risk factors among AD patients with GYNC.

Limited research has been done on the association between cancer and comorbidities. Ogle et al.[33] studied the association of several cancers with multiple chronic diseases; however, none of the cancers included in the study were GYNC. Therefore, the current study was conducted to evaluate GYNC in relation to comorbidities and patient characteristics among women with AD. This present study shows that patient's characteristics and comorbidities can be used as predictors of GYNC, among patients with AD.

CHAPTER 2: LITERATURE REVIEW

2.1 Cancer and Immunity

According to the ACS estimation of 2019, cancer is the 2nd leading cause of death for males and females of all ages, and the first leading cause of death for women age 40-79.[1] Cancer is the result of cumulative effects of living and the interaction of environmental exposures and genetic disposition.[34] It is likely that most cancer risk factors act as either tumor initiators (mutagenic agents) or tumor promoters (non-mutagenic agents).[31] In a healthy state, the body's immune system has the role of cancer suppression and inhibits cancer progression. For instance, patients who are immune-compromised by an infection like HIV, are more susceptible to develop some cancers such as Kaposi's sarcoma, central nervous system lymphomas, lung cancer, Hodgkin's lymphoma, and invasive cervical cancer.[26,35] Conversely, evidence suggests that although the immune system may generate anticancer responses, it may also promote cancer.[26] In a state of cancer promotion, the immune mediators, such as cytokines, chemokines and free radicals, may trigger tissue damage leading to chronic inflammation,[36,37] and consequently increase the risk of cancer promotion.[26,30,38-40] In a recent study, Yoneda et al. explored the effect of (C5a-C5a) complement system — a major component of the innate immune system which works with the immune system to trigger an inflammatory response to fight against infection — activation in cancer promotion and autoimmune causation.[29]

Therefore, immune responses have roles in both the promotion and suppression of carcinogenesis.[26] Figure 2.1 illustrates that an immune response is necessary to protect the host against the development of cancer;[26] however, the activation of the immune

system may lead to loss of self-tolerance (the recognition of self from non-self)[36] and initiation of autoimmunity. Whenever the immune system recognizes self-constituents, ADs will result.[15] In turn, chronic inflammation associated with autoimmune diseases causes tumor development through induction of mutations, genomic instability, early tumor promotion, and enhanced tumor angiogenesis (i.e. the growth of new blood vessels needed for tumor progression; Figure 2.2).[30]

2.2 Cancer, Autoimmunity, and Inflammation

Autoimmunity causes dysregulation of immune response in patients with AD. The association between autoimmunity and certain cancers, in particular lymphoma is well established.[37, 41] Since the immune system plays a significant role in the pathogenesis of autoimmune diseases and cancer,[26] it is essential to first examine the underlying molecular mechanisms of inflammation and its relation to cancer and autoimmunity. It is also established that inflammation plays a critical role in increasing cancer risk, and tumorigenesis.[30] There is evidence that an inflammatory microenvironment is an essential component of all tumors, as the vast majority of malignancies (90-95%) are linked to lifestyle and environmental factors, while only 5-10% of cancers are caused by germ-line mutations. For example, approximately 35% of all cancers have been attributed to diet, 30% to tobacco smoking, 20% to obesity, 18% to infection, and 7% to radiation and pollutants.[3,4,27,30] Many of these environmental causes of cancer and risk factors are associated with some form of chronic inflammation.[27,30,37]

Chronic inflammation sets in if acute inflammation (the innate immunity of the body) is prolonged. There are multiple factors involved in inflammatory pathways that lead to tumorigenesis. Two such factors, nuclear factor-kB (NF-kB) and signal

transducers and activators of transcription 3 (STAT3), have major roles in cancer initiation and progression.[27] Similarly, environmental agents such as bacterial endotoxins, carcinogens like cigarette smoke and radiation, and hyperglycemia promote inflammation through activation of NF-kB or STAT3.[29] Thus, factors such as NF-kB or STAT3 can be considered as targets for drug therapy, and prevention of cancer and inflammatory diseases.[27,42] The activation of the aforementioned factors can be prevented by eliminating modifiable risk factors,[4] through promoting healthy diets, avoiding smoking, preventing environmental pollutants and radiation exposures, and controlling infection.

Another example of an inflammatory factor is Nitric Oxide (NO), which is produced by many cell types. It plays a vital role in host defense and immunity, including the modulation of inflammatory responses. Some studies indicate that excessive NO is produced during the active course of RA diseases and SLE,[43] two of the most common ADs involved with most cancers. Thomsen et al.[44] studied the NO synthase activity in human GYNC and demonstrated that high levels of NO synthase activity are present in malignant tissue from GYNC while the enzyme activity is below detectable levels in gynecological tissue from non-cancer patients. This finding suggests that high NO synthase activity is related to malignancy[44] and can be applied to GYNC and their associations with inflammations at the molecular level.

2.3 Development of Autoimmunity in Cancer Patients

Similarly, the mechanism of development of ADs in cancer patients has also been the subject of curiosity. Sigel et al.[45] have reported that the majority of survivors of childhood cancers had grown into adulthood with comorbidities associated with cancer, or the treatment of cancer.[45] Some of these adulthood comorbidities could include ADs.

Holmqvist et al.[46] investigated the development of several ADs — such as Sjogren's syndrome, autoimmune haemolytic anemia, Addison's disease, polyarthritis nodosa, chronic rheumatic heart disease, localized scleroderma, ITP, Hashimoto's thyroiditis, pernicious anemia, insulin-dependent diabetes mellitus, and sarcoidosis — in adult life after certain childhood cancers such as leukemia, HL, renal tumors, and CNS neoplasia in Scandinavia.[46] In this investigation, the authors used standardized hospitalization rate ratios (SHRRs), and absolute excess risks (AERs) to calculate the increase in the risk of having ADs among the survivors of childhood cancers. Holmqvist et al. observed an increased SHRR of 1.4 (95% CI=1.3-1.5) of all the ADs under study combined, which corresponded to an AER of 67 per 100,000 person-year.[46] One explanation for the increased risk of ADs is that the persistent immune abnormalities after treatment with chemotherapy predispose the person to the development of autoantibodies — a central entity to the pathogenesis of many ADs. Therefore, the cancer itself as well as the immunosuppressive treatment, and the increased number and types of infections during cancer treatment, could alter the immune system as a whole, leading to the production of autoantibodies and development of autoimmunity.[46]

2.4 Gynecologic Cancer

GYNC, as a group of reproductive cancers, are the deadliest female cancer after breast cancer, with an estimated 33,100 annual deaths from the disease in the U.S.[1] The most common GYNC in term of incidence are uterine corpus, ovaries, uterine cervix, vulva, and vagina and other genital cancers, respectively. Ovarian cancer is the deadliest of all, with an estimated 13,980 deaths followed by uterine corpus with over 12,000 deaths annually for all ages (Table 2.1).[1]

2.4.1 *Gynecologic Cancer Types, and Incidence Rate*

A. Uterine Corpus Cancer is cancer of the upper section of the uterus (womb), and is the most common GYNC among women with more than 60,000 cases in the U.S. annually.[9,17] According to Siegel et al., among all GYNC cancers the death rate has risen from 2012 through 2016 for uterine corpus by 2.9% for all ages, with the highest incidence in ages of 70 years or older.[1] There are different types of uterine cancer. Endometrial cancer is the most common type and is cancer of the endometrium, which is the lining of the uterus,[9] and includes multiple subtypes such as uterine papillary serous carcinoma,[47] and endometrioid carcinoma.[17,18] Uterine sarcoma is another type of uterine cancer. This type is rare and occurs in outer layer muscles (myometrium) or other tissues of the uterus.[9,17,47] Routine diagnostic tests for uterine cancers are endometrial biopsy or a transvaginal ultrasound among high-risk women, based on a gynecologist's recommendation.[9]

B. Ovarian Cancer (OC) starts in the ovaries, the female reproductive glands, and can metastasize to other parts of the body.[22] As indicated before, OC is the second most common GYNC after uterine cancer with more than 22,000 cases in the U.S. annually.[1,9,22] A woman has a 1 in 75 chance of getting ovarian cancer in her lifetime. OC is also the most lethal member of GYNC (Table 2.2), and ranks fifth in causes of cancer deaths among women in the U.S.[16,22] It is known as the "silent killer" due to the lack of symptoms during the early phase of tumorigenesis, which causes 75% of patients to be diagnosed at later stages, and consequently reduces the survival rate for these advanced stage patients to only 30–40%.[22,48] If diagnosed at an earlier stage, the 5-year survival rate increases to 90%.[22] For 2016, the ACS reported over 14,000 deaths for OC across all

age categories. This makes OC the 5th leading cause of cancer death in women after pancreatic cancer. Among GYNC, OC ranks 4th for women aged 40-59 years and 5th for women aged 60-79 years.[1,4] There are three major types of malignant ovarian tumors: epithelial tumors which start from the outer surface cells of ovaries are the most common type of ovarian tumors and account for 85-90% of all OCs. Germ cell tumors, which start from the cells that produce the eggs within ovaries, account for only 2% of OCs while stromal tumors that arise from structural tissue cells of ovaries where female sex hormones (i.e. estrogen and progesterone) are produced account for 1% of all OCs.[22] At the present time, the most common detection tests for OC are recto-vaginal exams, trans vaginal ultrasounds, and CA-125 blood tests used in combination based on gynecologists' recommendations.[9]

C. Cervical Cancer starts in the cervix, the lower narrow end of the uterus, and is the third most common GYNC in the U.S., with over 13,000 new cases of invasive cervical cancer annually.[9,49] Cervical cancer is the only GYNC that can be prevented by routine screening via Pap smears and is curable when detected and treated early.[9,49] Most cervical cancers are squamous cell carcinomas, meaning that they develop from cells in the exocervix (outer layer of the cervix). The other cervical cancers are adenocarcinomas, or cancers that develop from gland cell. The less common cervical cancers, called mixed carcinomas, have features of both squamous cell carcinomas and adenocarcinomas.[9,49] During 2016, cervical cancer continued to be the 2nd leading cause of cancer death after breast cancer, among female ages 20-39, with about 470 deaths annually in this age group.[1] This high incidence rate of cervical cancer emphasizes the need for HPV vaccination expansion among adolescents and young women.[50]

D. Vulvar Cancer starts at the vulva, the outer part of the female genital organs. Vulvar cancer is rare and accounts for only 4% of GYNC and 0.6% of all cancers in women in the U.S.[20]

There are several types of vulvar cancer: squamous cell carcinomas which starts in squamous cells, the main skin cells type; adenocarcinoma which begins in gland cells and includes 8% of all vulvar cancers; melanomas which typically develop from pigment-producing skin cells, but can start in the vulva and are relatively rare making up 6% of all vulvar cancers; sarcoma which is very rare and accounts for less than 2% of all vulvar cancers; and basal cell carcinoma which also occurs very rarely on the vulva.[20]

E. Vaginal Cancer starts at the vagina (birth canal). It is also a rare GYNC and only about 0.1% of women will develop vaginal cancer in their lifetime.[19,21] There are several types of vaginal cancer: adenocarcinoma, which accounts for 10%; squamous cell carcinoma, which accounts for 9%; melanoma, which accounts for 3%; and sarcoma which affects the deep wall of the vaginal canal and accounts for 3% of all vaginal cancers.[23] There are two subtypes of vaginal sarcoma. Rhabdomyosarcoma is the most common type of vaginal sarcoma and is mostly found in children while leiomyosarcoma is often found in adults older than 50 years old.[19,21] Other cancers that initially started in adjacent organs like the vulva or cervix, but spread to the vagina, are also known as vaginal cancers.[19,21]

F. Fallopian tube cancer starts in the fallopian tube (the tubes that connect ovaries to the uterus). It is also a rare cancer, with symptoms and treatment similar to OC, yet, a better prognosis than OC.[22,48]

2.4.2 *Gynecologic Cancer Treatment*

Treatment of GYNC is complex and dependent on the type and stage of cancer at the time of diagnosis.[9,16] Possible treatment includes surgery, radiation therapy, hormonal therapy, and chemotherapy.[9] Surgeries for GYNC include different types of hysterectomy (removal of uterus) including partial or total for uterine, ovarian, cervical, and some advanced cases of vaginal cancers. For local vaginal and vulvar cancers, local excision, or vaginectomy and vulvectomy procedures are done, respectively.[9,19,20] The hysterectomy procedures are performed through abdominal surgery, or by laparoscopic procedure,[9] based on gynecological oncologists' recommendations.

2.4.3 *Gynecologic Cancer Risk Factors*

There are some known risk factors associated with each GYNC. For example, older age, fatty abdomen, and hormone therapy (estrogen) are common risk factors for uterine cancer,[9,17,18] in addition to modifiable risk factors. These modifiable risk factors include but are not limited to obesity, smoking, alcohol consumption, infection, and diabetes.[4] With cervical, vaginal, and vulvar cancers, human papillomavirus (HPV) infection and smoking are the major risk factors.[19-21,49,50] Older age and family history of OC and breast cancer are known risk factors for ovarian cancer.[22] However, the genetic mutation called BRCA1 or BRCA2, and endometriosis are common risk factors for both ovarian and fallopian tube cancers.[9,22]

2.5 Gynecologic Cancer and Disparity

2.5.1 *Disparity by Socioeconomic Status*

Lower socioeconomic status (SES), whether at the individual or society level, is usually associated with lower health outcomes, and higher mortality.[1,51] Egen et al.

performed a study on the effect of poverty on health outcomes in the U.S., comparing the health of poorest counties with wealthiest counties, based on median household income. They demonstrated the true impact of socioeconomic disparity on U.S. health outcomes and mortality rate.[51] The prevalence of behaviors such as smoking and obesity that increase the incidence and mortality rate of cancers, are seen much higher among residents of the poorest areas compared with that of the wealthiest areas (smoking: 27.6% vs. 13.8% and obesity: 36.7% vs. 24.4%).[1,51] Also poverty is associated with lack of facilities for cancer screening, late diagnosis of cancer, and lack of optimal treatment.[51]

The high mortality rate among people with lower SES is normally associated with later stage diagnosis of cancer, and inadequate treatment, even with preventable cancers like cervical cancer. Egen et al. demonstrated that during 2012-2016, the mortality rate of cervical cancer was twice as high in residents of poorer counties in the U.S. than those residents of affluent counties.[1,51] Also, the mortality rate of uterine corpus was 15% higher in poorer counties than in affluent counties (RR=1.15, 95% CI: 1.11-1.19). However, for ovarian cancer, there was no association between affluence and the death rate (RR=1.00, 95% CI: 0.97-1.03),[1,51] perhaps due to the nature of late stage diagnosis of ovarian cancer.

Siegel et al. in an assessment study for cancer control demonstrated the importance of SES in reducing the burden of cancer in U.S. They reported the ACS estimation of cancer epidemiologists that about 22% of all cancer death in 2018 would not occur if all American had the same SES as college graduates. Overall, according to their estimation cancer death rates could be averted by 34% among Americans aged 25-75 years, if SES disparities were eliminated.[45] The estimation of avertable deaths by

eliminating educational disparities for ages 25-74 years old for GYNC was:1% for ovarian cancer, 10% for endometrial cancer, and 55% for cervical cancer.[45]

2.5.2 *Disparity by Race/Ethnicity in Gynecologic Cancer*

Cancer incidence is different among people of different races/ethnicities. This difference is due to SES inequalities which impact the high-quality cancer prevention, early detection, and treatment as well as exposure to risk factors.[1] Overall, race/ethnicity is a risk factor for all GYNC. Sigel et al. compared the incidence and mortality rates of cervical cancer among four race categories in the U.S. for 2011-2016. They found the highest incidence rate per 100,000 population was among Hispanics (9.6), followed by Non-Hispanic Blacks (9.2), American Indians/Alaska Natives (9.2), Non-Hispanic Whites (7.1), and Asian/Pacific Islanders (6.0).[1] However, the mortality rates per 100,000 population of cervical cancer was highest among Non-Hispanic Blacks (3.6), followed by American Indian/Alaska Natives (2.8), Hispanics (2.6), Non-Hispanic Whites (2.1), and Asian /Pacific Islanders (1.7).[1] These mortality rates show the impact of poverty in the disparity of cervical cancer among different racial/ethnic groups.[1,45] Also, the endometrial cancer mortality rate is twice as high in Black women (7.3) compared to White women (3.9), despite the higher incidence rate of endometrial cancer in White (24.8) vs. Black (21.8) women.[23] Although, race/ethnicity is a risk factor for endometrial cancer, but does not seem to be the foremost factor for endometrial cancer death. Rather, SES and health disparities are the factors associated with mortality rate of this cancer.[23] Similarly, although White women have a higher incidence of OC than Black women,[23] Black women have a 31% higher mortality rate from OC than that of White women due to disparity.[23]

2.6 Economic Influence on Gynecologic Cancer

The level of economic development has a significant effect on mortality rates of GYNC.[4] Despite the high prevalence and mortality of GYNC among women worldwide, there has not been enough funding to facilitate investigation to ensure critical discoveries of GYNC, including discoveries related to causality, prevention, and treatments. Recently, Spencer et al. conducted a study on disparities in the allocation of research funding for GYNC. They investigated the distribution of National Cancer Institute (NCI) funding from 2007-2014 for 3 GYNC sites (ovary, cervix, and uterus) in comparison to 15 other cancer sites.[52] They used "cancer-specific lethality" scores, calculated by using the "site-specific mortality to incidence ratios multiplied by person-years of life lost per 100 new cases," to study the allocation of research funds by NCI.[52] In this study, the lethality score demonstrated the differences in mortality, incidence, and impact on person's living years for 18 different cancers. Among the 18 cancers under study, prostate cancer had the highest funding to lethality scores (Mean (SD)) of (1.812 (0.364)), followed by breast cancer (1.803 (0.105)), and melanoma (0.519 (0.075)). In comparison, GYNC ranked in the bottom half of NCI funding allocation. The funding to lethality scores for ovarian cancer was (0.097 (0.008)), cervical cancer was (0.087 (0.009)), and uterine cancer was (0.057 (0.006)). These scores translated into being ranked 10/18, 12/18, and 14/18, respectively. The impact of this underfunding of GYNC by NCI can be seen by low enrollment and smaller number of clinical trials available for patients with these cancers, and as a result fewer recommendations for treatment of GYNC.[52] This underfunding in GYNC studies makes more investigation in the field of GYNC essential for better management of these cancers.

2.7 Autoimmune Diseases

ADs are the collectives of clinically distinct disorders that have a misguided immune response in common.[13] ADs affect 5-10% of the world population,[12,13,53] with women generally having a 2.7 times greater risk of acquiring an AD than men.[11,14,53,54] Of the 50 million Americans coping with ADs, more than 75% are women.[14] ADs contribute disproportionately to morbidity and mortality among young to middle-aged women[28], and are amongst the top ten leading causes of all death among U.S. women age 65 and younger (Figure 2.3).[2] This high mortality rate among women with AD has been linked to cancer as well as other causes.[55]

Ramos-Casals et al. performed a review study, in which they used "big-data" methodology to do a geo-epidemiology investigation of ADs. They classified ADs into 2 groups, organ-specific ADs (AD damage is on a specific organ), and systemic ADs (AD affects a large number of organs and systems). In their study, they indicated that the imbalanced gender gap among patients with ADs is significantly dependent on the variation in ADs etiology. Additionally, the age of the patient at the time of onset of the disease, in particular with systemic ADs is also dependent on the type of ADs.[53] In their study, they demonstrated that with systemic ADs, 73.1% of patients were women, with a 3:1 female: male ratio (Table 2.2). The gender gap was most seen with Sjogren's syndrome (female: male ratio of 10:1) followed by systemic sclerosis (SSC), (female: male ratio of 5.2:1), SLE (female: male ratio of 5.1:1), and APS (female: male ratio of 5:1) (Table 2.2).[53]

2.7.1 *Autoimmune Diseases Incidence and Risk Factors*

Unlike cancer, ADs have not yet been recognized as a category of diseases. Rather, they are viewed as a single disease within multiple specialties.[11] This classification is due to the fact that ADs cross different medical specialties (e.g. RA is within rheumatology, Hashimoto's Thyroiditis is within endocrinology, MS is within neurology, and SLE is within dermatology).[11] Yet, ADs are the underlying cause of 100 chronic illnesses,[11,13,54] in which a combination of hereditary and environmental factors, and particularly infection are involved.[2,13,25,26] So far, over 100 different types of AD and AD like diseases (i.e. diseases with an autoimmune etiology) have been identified by researchers.[15] In addition, ADs are hard to diagnose and treat,[10] and as a result are very costly, with estimated indirect healthcare costs of more than $100 billion annually.[2,11,27,39]

The universal mechanism of pathogenicity among all autoimmune conditions is the disruption of balance within an immune system, where the ability to tolerate self-constituents must be maintained in a normal healthy individual.[10] Consequently, treatment often involves therapies that target activated immunity against self. The most common treatment for ADs is immunosuppressant agents, which tend to control the autoimmune activities and to maintain the self-tolerance. However, this kind of treatment is often accompanied by consequences,[10] including problems such as susceptibility to multiple infections and cancer.

The risk factors involved with ADs include gender, with women at greater risk of having ADs than men because of women's enhanced immune systems.[11,14] The female sex hormones — estrogen and progesterone — along with prolactin hormone, are immune stimulants and the major factors that make women more susceptible to ADs. In

contrast, androgen — a dominant male hormonal factor — is an immune suppressor.[14] Also, female gender seems to be a risk factor for poly-autoimmunity, the condition when more than one AD coexists in a single patient.[14,54] Another risk factor is age. ADs most commonly affect young to middle-aged women. This is clearly due to the female reproductive hormones levels being active and highest level during the reproductive age span.[11,56] MS and SLE are two examples in which younger women have an earlier onset than men,[14] while RA is more common among older individuals and affects both males and females.[11]

With regards to race/ethnicity, Black, American Indian, and Latinas are generally more susceptible to ADs than Caucasians.[11] Exposure to some environmental agents such as chemicals (aromatic amines and organophosphates), drugs (Procainamide, hydroxyzine, thiazides, calcium channel blockers, proton pump inhibitors, and interferon α),[56] metals (mercury, gold, and silver), or pollutants may also increase the risk of having ADs.[11] Moreover, hereditary and infections are two major risk factors for the onset of autoimmunity.[26]

2.8 Gynecologic Cancer and Autoimmunity

The association of some GYNC with autoimmunity at the molecular biology level in specific tissues have been the subject of studies in recent years. Yoneda et al. studied the C5a complement receptor (C5aR) mediated enhancement of malignancy in uterine cervical carcinoma stage I cells. They observed that C5aR invasion occurred more in deeper tissue. This result may explain the underlying reason for poor prognosis of cancer among AD patients, and in particular those with cervical cancer invasion.[29] In a study of the association between ADs and hepatobiliary cancer risk among U.S. adults, McGee et

al. concluded that organ-specific ADs may increase hepatobiliary cancer risk through local inflammation.[37] This finding suggests that any local inflammation in women's reproductive organs can increase the risk of GYNC. For example, endometriosis may increase the chance of endometrial cancer, or autoimmune oophoritis may increase the risk of ovarian cancer. The role of organ-specific autoimmunity in causing GYNC requires further investigation.

Despite the possible association between autoimmunity and GYNC, there has not been enough studies to investigate the details of the association of these two diseases. Because of the multiple risk factors involved with both GYNC and ADs, the need for a comprehensive investigation to study the association between comorbidities and patient characteristics with GYNC among AD patients is essential for the improvement of prevention, early detection, and advanced treatment of GYNC.

2.9 Patient-Level Predictors of Gynecologic Cancer Among Women with Autoimmune Diseases

2.9.1 *Patient Characteristics*

Patients characteristics such as patient demographics, and SES are non-prognostic predictors that can be used for the prediction of some diseases. Previous studies have used some of these characteristics to calculate the specific incidence cases in different cancers.[1,57] For example, Yost et al. used SES to calculate the breast cancer incidence in California for different ethnic groups.[57] However, no studies have used patient characteristics as predictors for GYNC among patient populations with ADs. The patient characteristics that may be important predictors of GYNC include patient's age, race/ethnicity, insurance type (pay source), SES (median income level), length of stay in hospital (LOS), and procedures such as hysterectomy, vaginectomy, and vulvectomy.

Hysterectomy is the removal of uterus, and is the second most common surgical procedure performed among reproductive aged women after Caesarean section.[58-60] The primary reason for hysterectomy among women 35-54 years is uterine fibroids. For older women, the most common reasons are uterine prolapse or GYNC.[59] According to the National Center for Health Statistics, an estimated 600,000 hysterectomies are performed annually in the U.S. with an annual cost of $5 billion, which makes hysterectomy a major public health concern.[58,61,62] Moreover, hysterectomy is costly, with mean total patient costs of $30,000-45,000 depending on hysterectomy type, operative time, and the length of stay in hospital.[62] Hysterectomy types include, abdominal, vaginal, or Laparoscopic surgery procedures. Each one of these procedures can be total or partial removal of the female reproductive organs and tissues, depending on the need of the patient at the time of surgery and the GYN oncologist's decision. Most hysterectomies are performed as inpatient surgery, however in past years, many hysterectomies are performed in outpatient facilities. Recent studies have demonstrated that SES and racial differences determine the utilization of inpatient vs. outpatient facilities for hysterectomy procedures (Bahrani et al. data not yet published). Although most hysterectomy procedures are done for treatment of uterine fibroid, but about 10% of all the hysterectomy procedures is used for treatment of GYNC, and about 11% of hysterectomies goes toward the treatment of endometriosis. Hysterectomy is also used for risk reducing and elective procedures such as bilateral salpingo-oophorectomy (BSO) when cancer genes BRCA1 and BRCA2 are present.[63,64] Endometriosis is considered a risk factor for ovarian cancer.[65-68] Pearce et al. in a large association study reported a 2-fold increase risk of endometrioid and low-grade serous subtype OC, and 3-fold increase risk of clear-cell OC, with endometriosis.[66] In

another meta-analysis study, Wang et al, conducted another association study between endometriosis and epithelial ovarian cancer (EOC), and hysterectomy.[67] Their results suggested 42% increased odds of association between endometriosis and EOC. [67] Therefore, based on finding evidence about high associations between endometriosis and some types of GYNC in literature, and knowing that hysterectomy is the procedure used for treatment of both endometriosis and GYNC, thus hysterectomy can be considered as a risk factor/predictor of GYNC.

2.9.2 *Comorbidities*

Comorbidities refer to the preexisting conditions of a patient, not including the primary disease under treatment.[33,69] Comorbidities in relation to cancer are the existence, severity, and the types of chronic conditions alongside the cancer.[69] The prevalence of comorbid conditions is usually higher among cancer patients of minority racial/ethnic groups, and those living with disparities.[69] Comorbidity has major role in the survival rate of cancers[70] and it should be considered as a major factor when managing cancer patents' treatment.[71]

Several chronic diseases such as chronic infections, immunity system diseases, and diabetes mellitus are associated with cancers. Infectious agents such as HPV in association with cervical cancer, HBV in association with liver cancer, or TB in patients with lung cancer are examples of infections associated with cancer.[37] Also, the association of immune system dysregulation such as RA with lymphoma and other hematological cancers, and autoimmune thyroiditis with thyroid cancer[29] are examples of immune system diseases as risk factors for cancer.[29,67] Comorbidities are extremely important in the clinical care of patient due to their great impact on all aspects of care,

e.g. prevention, screening, and prognosis.[33,69,72] Observing any association between comorbidities and GYNC can be valuable for improving care for GYNC patients with ADs.

2.10 Summary and Study Significance

Cancer continues to be a major cause of mortality worldwide and is the second most common cause of death after heart disease.[1,3] GYNC is the fourth leading cause of cancer death for women in the U.S. with an estimation of over 30,000 deaths from the disease.[1] GYNC refer to a group of cancers where initiation of each cancer occurs in a specific part of women's reproductive organs. Each GYNC is unique and involves different symptoms and risk factors. Like cancer, autoimmunity is a major chronic disease among women. ADs are among the top ten leading causes of death among U.S. women aged 65 and under, and the underlying cause of 100 chronic illnesses in which the combination of two major factors such as hereditary and environmental factors are involved.[2,11,27,39] Women have a 2.7 times higher risk of acquiring AD than men. In addition, ADs are responsible for over $100 billion in direct healthcare cost annually in the U.S.[2,11]

Because of the high prevalence of AD among women, and knowing that autoimmunity causes chronic inflammation, and chronic inflammation acts as a tumor promoter, more elucidation on the association between GYNC and AD is essential for the management of GYNC. Only a few studies have explored the possible association between autoimmunity and GYNC. Prior studies have mostly addressed only one type of AD with one or a few types of female cancer.[12] However, no comprehensive studies have been done among the U.S. population to evaluate associations between most ADs with all

types of GYNC. Considering cancer and AD are both the accumulative effect of genetics and environmental exposures,[24,25,27] and that 90-95% of all cancers are linked to lifestyle and environmental factors,[27] it is imperative to conduct a comprehensive study to investigate the association between patient-level predictors and several types of GYNC among AD patients in the U.S. population.

The significant impact of this research will be through the possible improvement of women's health in general by setting priorities for autoimmunity recognition as a serious disease, and cancer prevention and management through healthier living. Discovery of any novel risk factors associated with cancer and autoimmunity among women would have public health and public policy relevance. The strong correlation between AD and GYNC may lead women with AD to be recommended for yearly cancer screening for early detection, and better management of possible GYNC.

2.11 Objectives and Hypothesis

2.11.1 *Objectives*

The underlying goal of this study was to use data analytics to uncover novel GYNC risk factors. Therefore, it was imperative to investigate how patient-level predictors — the patient's characteristics and comorbid conditions — were related to GYNC among women with AD. The specific aims of this study were as follows:

1. To evaluate the association between patient-level predictors and GYNC among AD patients.
2. To determine the various subpopulations of AD patients with increased likelihood of GYNC based on patients' characteristics and comorbid conditions, using Classification Tree Analysis.

Through investigating these questions, unique combinations of characteristics that describe subgroups of patients at risk for GYNC can be used as a potential risk assessment for GYNC as well as for early detection and/or prevention tools.

2.11.2 *Hypothesis*

For this current study, I hypothesized that:

1. There is an association between patient-level predictors and GYNC among women with AD.
2. A unique combinations of patient-level predictors can describe subgroups of AD patients at risk for GYNC among patients with AD.

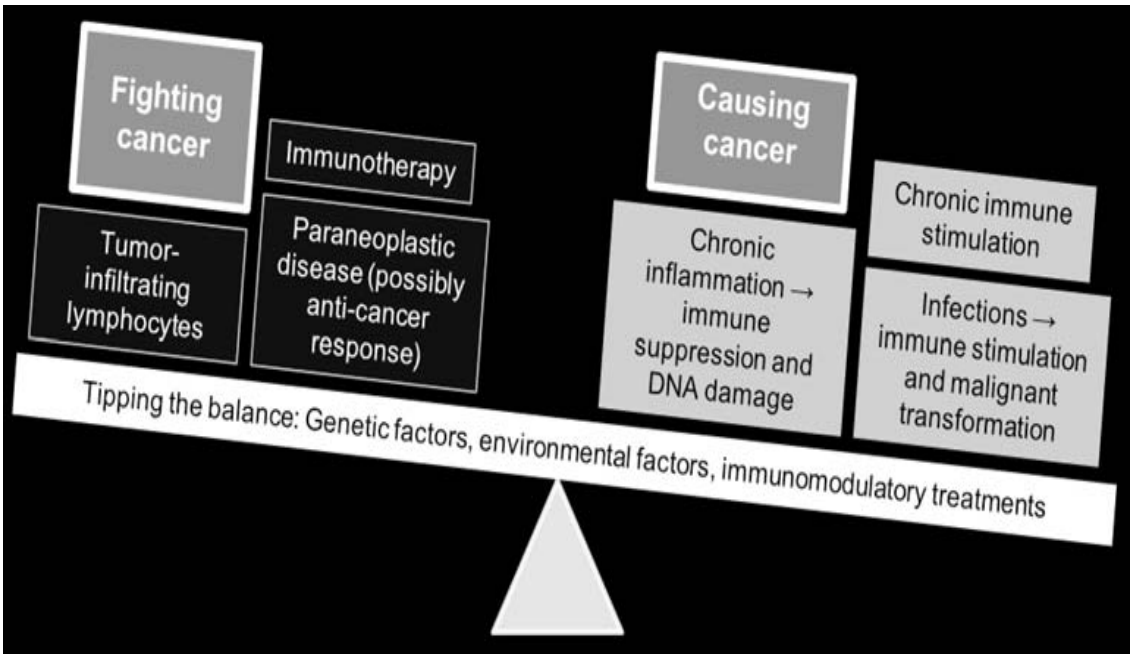

Figure 2.1: Immune System[26]

(*Reprinted from: "Multiple Associations Between a Broad Spectrum of Autoimmune Diseases, Chronic Inflammatory Diseases and Cancer." Franks, A. L. et al. Anticancer Research. 2012, 32: 1119-1136).*

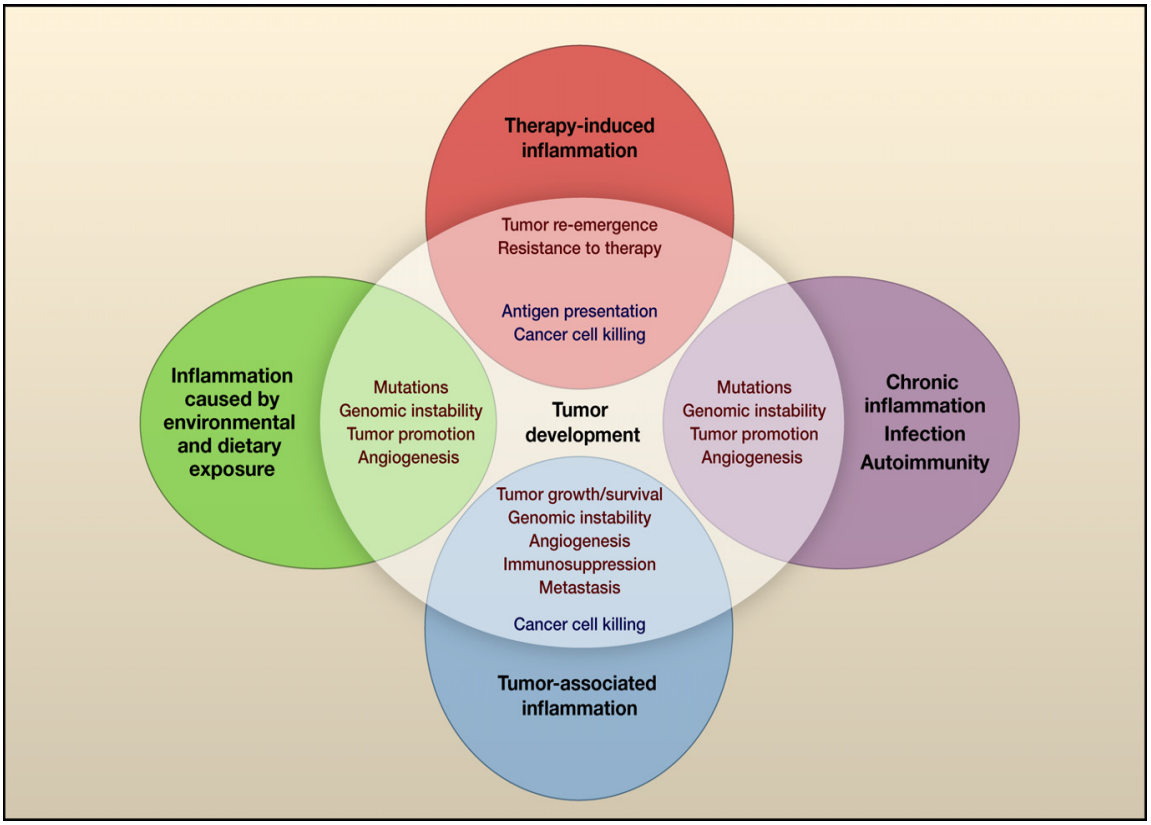

Figure 2.2: Chronic Inflammation and Autoimmunity in Tumorigenesis and Cancer[30]
(Reprinted from: "Immunity, Inflammation, and Cancer." Grivennikov SI, et al. Cell 140: 883-899, 2010).

Autoimmune diseases	Age, y						All ages[a]
	1–4	5–14	15–24	25–44	45–64	>65	
Group 1 (diseases with specific *ICD* categories)							
Autoimmune hemolytic anemia	0	2	1	2	11	77	93
Glomerulonephritis	3	6	5	44	88	745	893
Graves disease	0	0	2	4	3	15	24
Multiple sclerosis	0	0	3	254	620	514	1391
Myasthenia gravis	0	0	1	9	14	150	174
Myocarditis (I)	24	20	26	97	78	126	401
Polymyositis and dermatomyositis	0	0	1	13	46	112	172
Rheumatic fever and heart disease	1	1	17	177	582	2832	3613
Rheumatoid arthritis	0	0	1	14	183	1244	1442
Scleroderma	0	5	4	85	318	490	902
Systemic lupus erythematosus	0	9	62	338	353	356	1118
Subtotals for group 1	28[b]	43	123[b]	1037[b]	2296[b]	6661	10223
Group 2 (diseases without specific *ICD* categories)							
Addison disease	0	0	2	11	15	61	89
Chronic active hepatitis	1	1	1	12	51	135	201
Goodpasture syndrome	0	0	1	8	15	34	58
Idiopathic thrombocytopenia purpura	1	0	3	21	29	134	188
Type 1 diabetes mellitus	1	8	52	269[c]	0[c]	0[c]	330[c]
Myocarditis (II)	0	0	0	0	0	4	7
Pemphigus vulgaris	0	0	0	0	2	28	30
Pernicious anemia	0	0	0	0	2	68	70
Primary biliary cirrhosis	0	0	1	10	78	308	398
Relapsing polychondritis	0	0	0	0	3	24	27
Sjögren syndrome	0	0	0	1	16	43	60
Thyroiditis	0	0	2	2	2	0	6
Uveitis	0	0	0	0	0	0	0
Vitiligo	0	0	0	0	0	0	0
Subtotals for group 2	3	9	62	334	213	839	1464
Totals	31[b]	52[b]	185[b]	1371[b]	2509[b]	7500	11687
10th leading cause of death counts	23	45	75	800	1380	10135	11974

Note. ICD = International Classification of Diseases.
[a]All ages includes persons younger than 1 year and of unknown age.
[b]Subtotal or total count for autoimmune disease deaths exceeds the count for the 10th leading cause of death.
[c]Diabetes mellitus deaths were included only for persons younger than 35 years.

Figure 2.3: Counts of Deaths of Women with an AD as the Underlying Cause Compared with Official Counts for the 10th Leading Cause of Death, by Disease Category and Age: U.S. 1995[2]

(Reprinted from: "Autoimmune Diseases: A Leading Cause of Death among Young and Middle-Aged Women in the United States." Walsh, SJ, et al. American Journal of Public Health. 2000; 90:1463-1465).

Table 2.1 - Female:Male Ratios in Autoimmune Diseases[11,53]

(Adopted from American Autoimmune Related Diseases Association Site, and Autoimmunity Reviews 14(2015, 670-679)).

Type of Disease	Female: Male Ratio
Primary Sjogren's syndrome[1]	10:1
Systemic sclerosis[1]	5.2:1
Systemic lupus erythematosus[1]	5.1:1
Anti-phospholipid syndrome[1]	5:1
Polymyalgia rheumatic[1]	3:1
Giant cell artheritis[1]	2.4:1
Inflammatory myopathies[1]	2:1
Rheumatoid artheritis[2]	2.5:1
Antiphospholipid syndrome-primary[2]	2:1
Autoimmune thrombocytopenic purpura (ITP)[2]	2:1
Multiple sclerosis[2]	2:1
Myasthenia gravis[2]	2:1

[1] Adopted from reference #53; [2] adopted from reference # 11.

Table 2.2 - Estimated New GYNC Cases and Deaths for Females in United States, 2019[1]

(Adapted from: "Cancer Statistics, 2019," Siegel RL, et al. CA: Cancer Journal for Clinicians 2019, Jan-Feb; 69(1):7-34).

Cancer Type	Estimated new Case	Estimated Death
Uterine Cervix	13,170	4,250
Uterine Corpus	61,880	12,160
Ovary	22,530	13,980
Vulva	6,070	1,280
Vagina & Other Genital	5,350	1,430
Total	109,000	33,100

CHAPTER 3. METHODOLOGY

3.1 Conceptual Model

The current study is modeled after the theory of the bio-psychosocial model of Engel.[33,73] Engel's model is based on the belief that physical illness is caused by a complex interaction of biological, psychological, and sociocultural factors.[73] This model has been supplemented into the field of medicine as a new model since the introduction of this conceptual framework in 1997. Based on this model, cancer should not be treated as a single somatic disease. Rather the social, behavioral, and demographic characteristics of patients as well as other comorbid conditions should be considered in order to achieve a good prognosis, management, and/or prevention of the disease.[33,73] Examining patient characteristics such as age, race/ethnicity, socioeconomic status, and comorbid conditions in this current study may provide new insight into the prognosis and better management of GYNC among women with ADs (Figure 3.1).

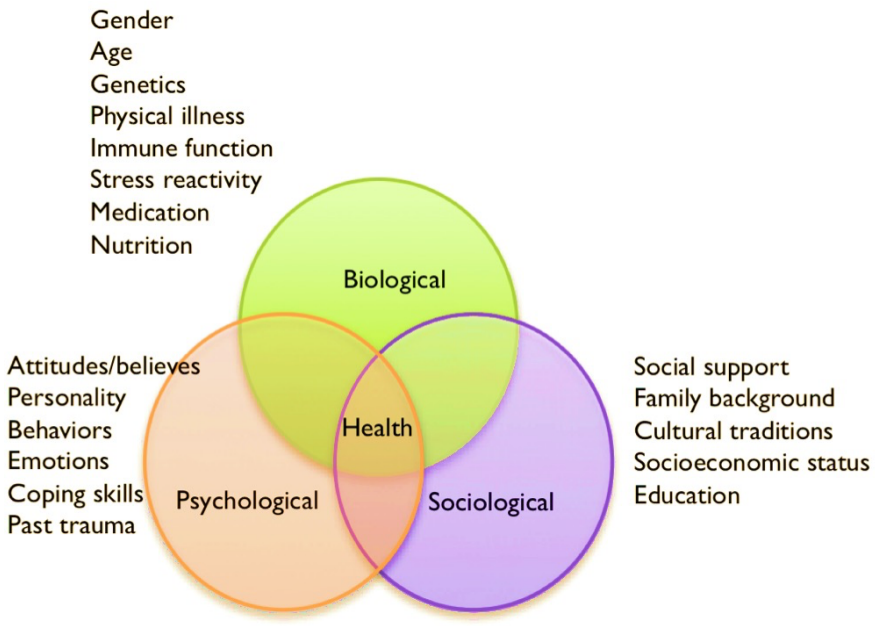

Figure 3.1: Conceptual Frameworks/Bio-Psychosocial Model
Adopted from: http://prespectivesclinic.com/health-psychology

3.2 Data Source

The current study was a cross-sectional, retrospective administrative data analysis. In this study, 2007-2013 Florida State Inpatient Data (SID) samples of the Healthcare Cost and Utilization Project (HCUP)[74] were used — to identify ADs, GYNC, comorbidities diagnosis, GYN related procedures, as well as patients' characteristics — to perform the data analysis.

HCUP databases are a family of administrative longitudinal databases comprising information on inpatient stays, outpatient surgeries, and emergency department visits in U.S. hospitals. It brings together the data collection efforts of state data organizations, hospital associations, private data organizations, and the federal government. Through HCUP data, a national information resource of encounter-level health care data is created. HCUP data are sponsored by the Agency for Healthcare Research and Quality (AHRQ).[74]

The data used for analysis of this study was a pooled dataset from SID samples of the years 2007-2013, containing patient discharge records from over 17 million patients seen at 200 acute hospitals in Florida. These records contained information on patient demographics, diagnoses, procedures, insurance pay sources, and hospital length of stay.[74-76]

3.3 Human Subject Protection

The HCUP data are publicly available to purchase. The Institutional Review Board of the University of North Carolina at Charlotte approved use of the HCUP data for this study.

3.4 Inclusion/Exclusion Criteria and the Study Population

In the current study, all the discharges with an International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM)[77] diagnosis code corresponding to ADs were selected (Table 3.1).[78,79] The ICD-9-CM codes for ADs were identified in the "Diagnostic Codes" columns (DX1-DX31) of the HCUP datasets. The HCUP Diagnosis Codes are the established codes representing the conditions that are related to the services that were provided during the hospitalization.[74]

The study population included women with one or more type(s) of AD (n=1,165,061). All male subjects and women without an AD diagnosis were excluded from the study. The major AD types commonly seen among women were categorized into nine AD categories based on the types of diseases and disorders,[53,78] and were identified using related ICD-9-CM codes indicated in Table 3.1.[78,79] These types included diseases and disorders such as arthritis-like diseases; dermatological diseases; digestive diseases; glandular diseases such as thyroid, adrenal, pancreas, reproductive, and salivary diseases; neurological disorders; muscular disorders; vascular and systemic diseases; sensory organs diseases, including auditory and ocular; and diseases affecting major organs such as the heart, kidney, liver, and lung (Table 3.1). The 9 AD categories were classified further into 5 groups, "All", A, B, C, and D, based on the body systems, and the tissues and organs affected by ADs (Table 3.1).[53]

3.5 Predictor/Independent Variables

In this study, patient-level factors such as comorbidities, and patient characteristics were used as independent variables. Of note, the terms "predictor

variables" and "predictive analysis" are being used in the statistical sense, as this was a cross-sectional study and causality could not be determined.

3.5.1 *Patient Characteristics*

Patient characteristics including age, race/ethnicity, LOS, pay source, and median household income were considered as predictor variables. In addition, types of surgical procedures performed — procedures such as hysterectomy, vaginectomy, and vulvectomy performed during years of 2007-2013 — were assessed as predictor variables. To screen for these characteristics, among the eligible population, categorical scores of (0) and (1) were used for the absence and presence of each condition in the data. The Patient Characteristics were as follow:

Age: The patient age in years was defined as the age of patient at admission, and was categorized as follow: <25 years old, 25-45 years old, 45-65 years old, and > 65 years old.

Race/ethnicity: The variable race was defined as patient race, and was categorized as: White, Black, Hispanic, and Others.

LOS: The LOS variable was defined as the hospital length of stay of the patient in days, and was used as an indicative of the severity of health conditions and higher cost. This variable was already calculated in the dataset, by subtracting admittance date of patient from dismissal date of hospitalization. LOS was categorized as: <1 day, 1-10 days, 11-20 days, 21-30 days, 31- 40 days, >40 days.

Pay Source: The pay source variable was defined as the principal payer code, and was categorized as: Medicare, Medicaid, Private Insurance, Self-Pay, and Other.

Socioeconomic Status: SES was defined based on patient income, and was characterized by the national quartile classification of the estimated median household income for the patient's ZIP Code (Zip INC_QRTL). The quartiles are categorized as category 1 to 4, indicating the poorest to wealthiest populations. These values have been obtained from ZIP Code-demographic data captured from Claritas Inc., a company that developed a set of geo-demographic segments for the U.S. These estimates have been updated annually; hence the values vary by year (Table 3.2).[80] The estimated median household income for the patients' ZIP Code was categorized as: Quartile 1(Q1), Quartile 2 (Q2), Quartile 3(Q3), and Quartile 4 (Q4).

Surgery Types: Surgery types were identified using ICD-9-CM codes[58] for abdominal hysterectomy, vaginal hysterectomy, subtotal hysterectomy, laparoscopic hysterectomy, radical hysterectomy, and other unspecified hysterectomies,[58] as well as vaginectomy, and vulvectomy procedures (Table 3.3).[9,81] The relative ICD-9-CM codes used to identify these surgeries were searched within the HCUP variables "Procedure Codes" (PR1-PR31). The Procedure codes are the codes representing all the significant procedures other than the principle procedures done for the patient.[74] The ICD-9-CM codes for the aforementioned procedures have been used in previous studies.[58]

3.5.2 *Comorbidities*

In the current study, comorbidity is referred to the presence of one or more medical conditions (preexisting) in addition to the GYNC. More comorbid conditions will indicate a higher risk of complications and death for patients. In addition, comorbidity is an indicator of hospital care utilization.[82] In this study, the comorbidity measures modeled after the Elixhauser Comorbidity Index (ECI)[82,83] were used to

measure risk adjusting and predicting outcomes. The ECI contained 31 different conditions, but for the purpose of this current study, they were grouped into 9 different categories (Table 3.4).[83] The comorbid conditions listed in Table 3.4 were used as predictor variables in this study, using ICD-9-CM codes[77,83] specific for these conditions, and were searched within the Diagnosis Codes, DX1-DX31 of the HCUP dataset.[74,83] To screen for these comorbid conditions among the eligible population, a categorical score of (0) or (1) was used for the absence or presence of each condition in the data.

3.6 Outcome/Dependent Variable

The outcome variable for this study was a diagnosis of any type of GYNC (Table 3.5). Specifically, the ICD-9-CM codes listed in Table 3.5 were used to create a dichotomized outcome variable indicating presence or absence of the outcome.[9,81] The ICD-9 codes used for GYNC have been used in previous studies.[58,81]

3.7 Data Analysis

3.7.1 *Data Subsets*

In this study, seven HCUP Florida SID datasets, collected from years 2007-2013 (Table 3.6) were inspected. The cleaned datasets were pooled (n=17,078,554) and the female population of the combined data was selected (n=9,590,428). Among the female population, subjects with one or more ADs were selected as the study population (n=1,165,061), using ICD-9-CM codes for AD types listed in Table 3.1. The relative ICD-9-CM codes were searched within the HCUP variables, the Diagnosis Codes, DX1-DX31.The ADs were further categorized based on their etiology, and were classified later into 5 groups, “All”, A, B, C, and D, based on the body systems, and the organs and tissues affected by ADs (Table 3.7).[53] The group “All” (n=1.165,061) included all of the

subjects with ADs among categories 1-9. The Group A (n=593,069), included patients within categories 1,2,7, and 8; Group B (n=238,568) included patients within categories 3, and 9; Group C (n=295,580) included patients within categories 5 and 6; and Group D (n=190,979) included patients within category 4. As it is shown in Table 3.1, there are distinct ICD-9-CM codes within each category. However, the ICD-9-CM code 279.49, although representing multiple disorders such as progesterone dermatitis, PANDAS, anti-synthetase syndrome, and autoimmune diseases not elsewhere classified (NEC), but is used in common among groups A-C. Consequently, the total frequencies of groups A-D exceed the frequency of group "All" (Table 3.7).

3.7.2 *Descriptive Analysis*

Descriptive analysis was conducted to describe the frequencies and percentages of all the predictor variables within levels of the outcome (AD subjects with GYNC vs those without GYNC), using SAS® version 9.4 (SAS Institue Inc., Cary, NC, 1989-2019). The descriptive analysis was restricted to the study population (the women with one or more type (s) of AD) – analysis was performed for each group of patients with AD (i.e. "All" and Groups A through D).

3.7.3 *Predictive Analysis, Logistic Regression*

The logistic regressions analyses of this current study were done using SAS® version 9.4. The regression model was used to measure the variability of all predictor variables in both levels of outcome. The method was done both in bivariate and multivariate models approach. Bivariate logistic regression was used to calculate unadjusted odds ratios (ORs) and 95% confidence intervals (CIs). This method was done to evaluate the relationship of each predictor variable with GYNC, for all five groups of

AD patients. For the multivariate model, a backward elimination stepwise logistic regression approach was used to create a reduced model. All predictor variables were entered, and only those variables that had a p-value of 0.20 or less were retained in the final adjusted model.[84] Similarly, this adjusted logistic regression approach was applied to all 5 groups of AD patients (i.e. "All" and groups A through D).

3.7. 4 *Predictive Analytics*

A. Bootstrap Forest

A classification tree analysis method was used to identify the effect of predictor variables on GYNC, to help in identifying the AD subpopulations with the highest likelihood of having GYNC,[76] and to investigate multilevel interactions of risk factors.[76,85] This predictive analytic aproach was conducted using the Bootstrap Forest (BSF) platform of JMP pro version 13 (JMP®, SAS Institue Inc., Cary, NC, 1989-2019) alongside SAS programing.[86-88] The BSF platform of JMP uses a decision tree classification in which data and variable subsets are used to reach an ideal model.[89] BSF is a method in which, many decision trees are created and response values are averaged for obtaining the final prediction.[90,91] In BSF methodology, a subset of data is chosen as training data to develop predictive models. The resulting models are used to separate a test sets in order to assess its performance. In this technique, samples are bootstrapped (sampling with replacement) from the training data, and a tree is fitted into the bootstrapped sample, where the variables are randomly selected for splitting at each node.[91,92] Therefore, each split in each tree is considerd as a random subset of independent variables. When maximum number of trees are reached, they get aggregated and the final model is built for a congregated conclusion.[91,92] Hence, the final prediction

for an observation in a BSF analysis is the average of predicted values for that observation over all the decision trees in the BSF analysis.[90]

To run the BSF analysis in this study, all the predictor variables were transformed into character type, due to the categorical nature of the variables, and the fact that the BSF analysis by JMP pro platform has special settings for categorical vs. continuous data, for an ideal predictive analysis. The modeling type of all the variables was changed to nominal or ordinal, depending on the nature of the variable (Table 3.8). In addition, each group of datasets, "All", A, B, C, and D, was randomly sampled in 1:1 and 1:5 ratios of GYNC positive vs. GYNC negative subjects, using SAS, to find the apropriate ratio for using in BSF analysis. This random sampling was done to obtain the best sample set for BSF analysis. For example, the study population data (group "All") were unbalaced with having far more GYNC negative (n=1,156,745) than GYNC positive (n=8,316) subjects, making the GYNC (+) to GYNC(–) ratio :139. These balanced samples were then called AD groups "All", "A", "B", "C", and "D" subsets. During BSF analysis, the samples subsequently were partitioned into 3 data-subsets: Training, Test, and Validation data- subsets by 80%, 10%, and 10% proportions, respectively (Figure 3.2).

The training and partitioning were done automatically by JMP pro 13 software, where the criteria are set to optimize quality measures such as Receiving Operation Curve (ROC), measuring the Area Under the Curve (AUC), R-square, and missclassification rate.[93] The validation is a process in which a portion of the dataset (data-subset validation) is used to estimate model parameters, while the training set (data-subset training) of dataset is used to assess the predictive ability of the model.[94] The

test portion of the dataset (data-subset test) is the set used for the final and independent assessment of the model's predictive ability.[94] The BSF menu for JMP pro is set for the criteria to be chosen for the predictive model (Figure 3.3).

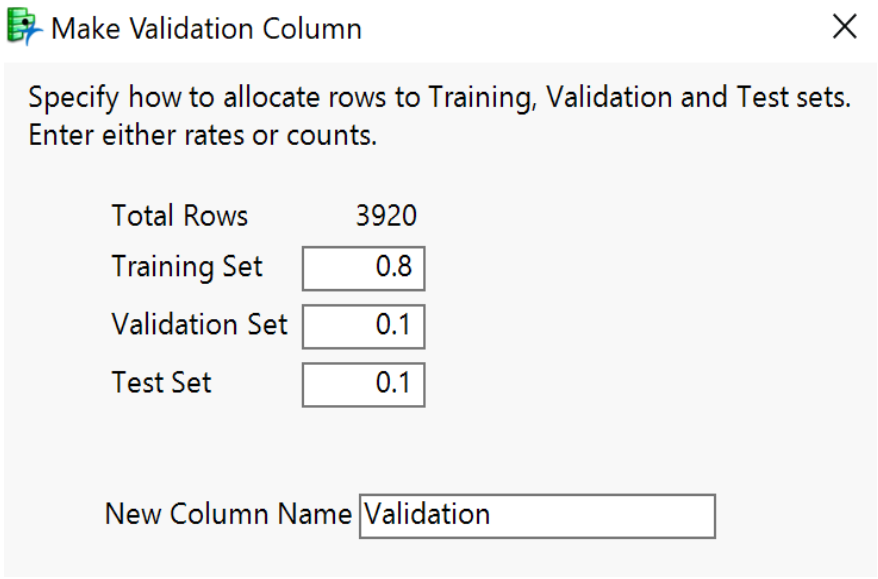

Figure 3.2: The BSF Validation Column Window for AD Group "D" Subset

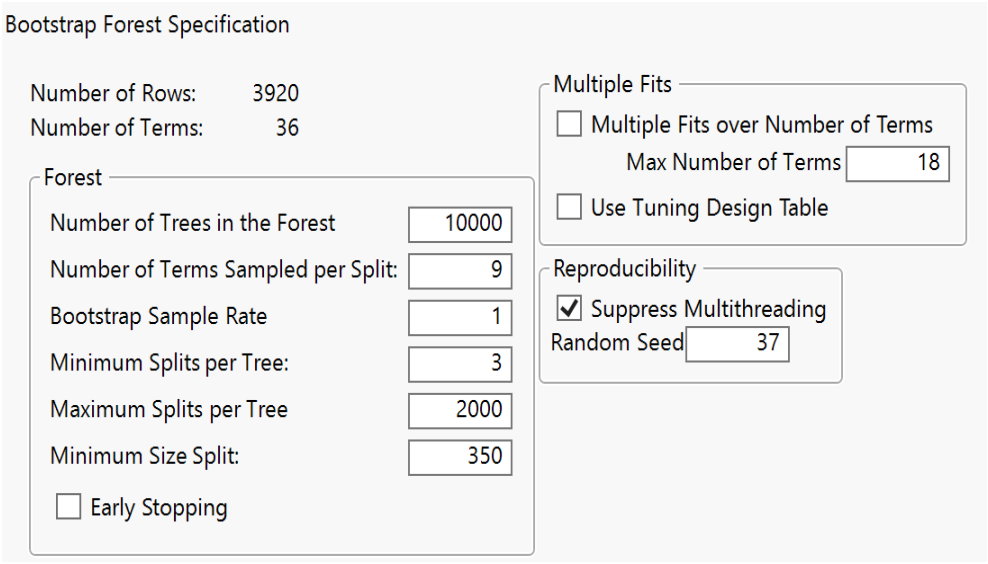

Figure 3.3: The BSF Menu for the Criteria to be Chosen for the Predictive Model for AD Group "D" Subset

The menu for this study stated the number of rows (dichotomous dependent variable, which is the subjects with and without GYNC) in each dataset group, and the number of terms (36 predictor variables used). Other criteria settings chosen in the menu are as follow:[89]

1. Number of trees in the forest, 10,000 trees to be grown in the model and then averaged togethe .
2. Number of terms sampled per split, 9 independent variables (columns) being considered as splitting candidates at each split.
3. Bootstrap sample rate, the proportion of observations to sample for growing each tree. At this point a new random sample is created for each tree. A default # of 1 was chosen.
4. Minimum splits per tree, the number 3 was chosen as the minimun number of splits for each tree.
5. Maximum splits per tree, the maximum number of splits for each tree, 2000 splits was chosen.
6. Minimum size split, the number 350 was chosen as the minimum number of observations necessary for a candidate split.
7. Early stopping, was not selected in this study for the purpose of continuation of partitionning until the specified number of trees (10,000) was reached.
8. Multiple fits over number of terms, 18 fits was selected for the purpose of creating a bootstrap forest for several numbers of terms sampled per split, with the maximum number of terms.

The results of BSF analysis of all 5 subsets ("All", "A", "B", " C", and "D") were collected, and the predictive values were determined in terms of specificity, sensitivity, false positive rate, overall accurate prediction, and wrongly predicted values. These evaluating criteria were based on the values resulting from the data-subset Test, summarized in a table called Confusion Matrix in JMP pro 13, specified for categorical data (Figure 3.4).

Confusion Matrix

Training

Actual	Predicted Count	
GYN cancer	0	1
0	1330	234
1	506	1075

Validation

Actual	Predicted Count	
GYN cancer	0	1
0	175	27
1	64	146

Test

Actual	Predicted Count	
GYN cancer	0	1
0	159	35
1	54	115

Figure 3.4: Confusion Matrix for AD Group "D" Subset

The Confusion Matrix shows the test outcomes of true positive (TP), false positive (FP), false negative (FN), and true negative (TN). From these test outcomes, the predictive values listed above were calculated using the following formula:[93,95]

Sensitivity = TP/(TP + FN)

Specificity = TN/(TN + FP)

False Positive Rate = 1-Specificity = FP/(FP+TN)

Wrongly Predicted values = (FP + FN) / (TN + TP + FP + FN)

Overall Accurate Prediction = 1- Wrongly Predicted Value

The predicted values were calculated for all 5 AD groups "All", and "A-D" subsets. Also, the measurement of the relationship between predictor variables and the response is shown in a feature called Column Contribution, in which it shows which predictor in the model best predicts the response (Figure 3.5).

Column Contributions

Term	Number of Splits	G^2	Portion
Hysterectomy	4496	217.926817	0.7607
Age	4162	21.270253	0.0742
Fluid and Electrolyte Disorders	2669	9.59206516	0.0335
Diabetes Complicated	1153	9.35617803	0.0327
Diabetes Uncomplicated	2554	8.14417092	0.0284
Pay source	2798	4.95599688	0.0173
Obesity	1275	4.70133086	0.0164
Chronic Pulmonary Disease	1285	4.27895252	0.0149
Race	1758	2.29379043	0.0080
Depression	547	1.21980863	0.0043
Hypertension Uncomplicated	1633	1.19233983	0.0042
Household Income	1675	0.80443586	0.0028
Hypothyroidism	477	0.73171891	0.0026
Cardiac Arrhythmia	7	0.00137808	0.0000
Hospital Length of Stay	6	0.00103897	0.0000
Vaginectomy	0	0	0.0000
Vulvectomy	0	0	0.0000

Figure 3.5: The Column Contribution for AD Group "D" Subset

B.Decision Tree Analysis

Although the trees created by BSF were very infomative, they were very cumbersome due to a large number of trees (10,000 trees). To have a smaller comprehensive partition tree that showed the relationship of the GYNC binary response with the 36 predictors, the Decision Tree analysis technique from JMP® Pro Version 13 (SAS Institute Inc., Cary, NC, 1989-2019) was used. Decision Tree (DT) is a classification tree analysis method in which the partition algorithm searches all possible splits of the independent variables, and chooses the best splits to predict the optimal response (Figure 3.6).[96]

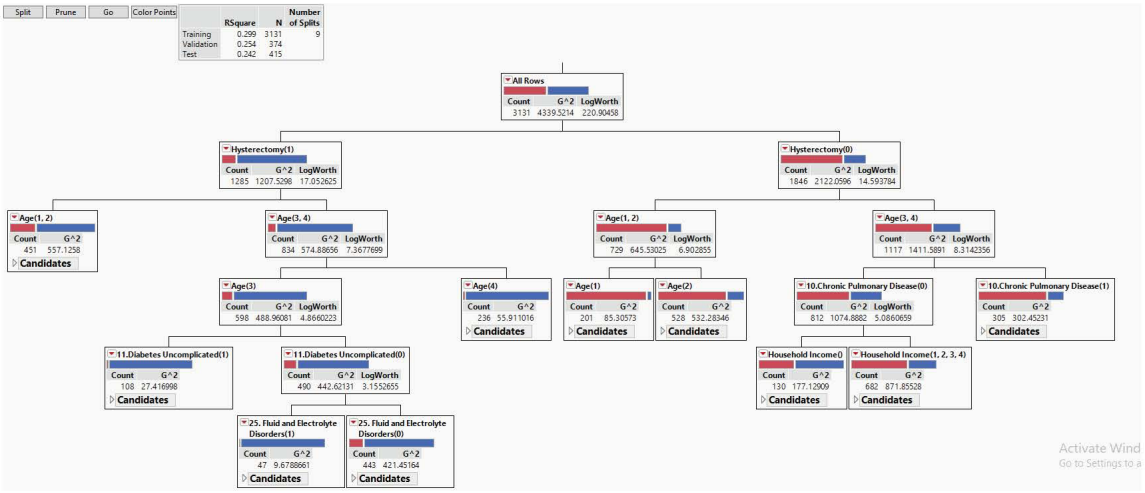

Figure 3.6: The Decision Tree for AD Group "D" Subset

In the DT analysis of this study, the categorical response was fitting the probabilities estimated for both levels of response, e.g. GYNC positive vs GYNC negative, thereby minimizing the residual log likelihood chi-square.[97] Also, the categorical factors got divided into 2 levels, where it considered all the possible groupings into 2 levels.[97] In DT analysis, generally the validation portion of the data table is used and selected randomly to create the predictive values in the form of "counts" for observations in each split, and the "logworth value." [96] The logworth statistics, used for node spliting are calculated as:

$-\log_{10}(\text{p-value})$.[97]

The adjusted p-value is calculated such that all the splitting events are taken into accounts.[97]

Another important value in DT analysis to report for categorical data is G^2, which is the likelihood ratio chi-square, and is twice the change in the entropy, where entropy is $\Sigma - \log(p)$, for each response, and p is the attributed probability of the response occurred.[95] The chosen G^2 is: G^2 test = G^2 parent - (G^2 left + G^2 right).[97,98]

Table 3.1 - List of Autoimmune Diseases and Related ICD-9-CM Codes[78,79]

Category	Sub-Category	AD Types	Diseases	ICD-9-CM Code
1		Arthritis like diseases	Psoriatic arthritis	696.0
			RA	714.0
			Juvenile RA	714.3, 714.30
			Palindromic rheumatism (PR)	719.31- 719.39
			Polymyalgia rheumatic	725
			Reactive arthritis (Reiter's disease	99.3
2		Dermato-logical diseases	Autoimmune progesterone dermatitis	279.49
			Dermatitis herpetiformis	694.0
			Pemphigus vulgaris	694.4
			Bullous pemphigoid	694.5
			Cicatricial pemphigoid	694.6, 694.61
			Linear IgG disease(LAD), Epidermolysis bullosa acquisita	694.8
			Bullous pemphigoid	694.9
			Discoid lupus erythematosus	695.4
			psoriasis	696.0, 696.1
			Lichen sclerosus, morphea	701.0
			Psoriasis	708.8
			Autoimmune urticaria	708.8
			Scleroderma	710.1
			Dermatomyositis	710.3
			Vitiligo	709.01
			Parry Romberg syndrome	349.89
			Herpes gestationis, pemphigoid gestationis (PG)	646.8
			Pyoderma gangrenosum	686.01
			Mucha-Habermann disease, Pityriasis lichenoides et varioliformis acuta	696.2
			Alopecia areata	704.01
			Lichen planus	697.0
			erythema nodosum	695.2
3		Inflam-matory bowel/ Digestive diseases	Crohn's disease	555.0
			Enteritis	555.9
			Ulcerative colitis	556.0-556.6 556.8,556.9
			Celiac disease	579.0
			Acalasia (esophagitis)	530.0

Continue on Next Page

4		Glandular diseases		
	4.1	Thyroid	Hashimoto thyroiditis	245.2
			Chronic thyroiditis, NOS	245.8
			Graves' disease	242.0
			Graves ophtholmopathy (thyroid eyes)	242.9
			Autoimmune thyroid disease,	245.8
	4.2	Adrenal	Addison's disease, autoimmune adrenal atrophy and insufficiency	255.4, 255.41
			Autoimmune Polyendocrine syndromes (APS type I, II, III), Schmidt syndrome	258.1
	4.3	Pancreas	Autoimmune pancreatitis (AIP)	557.1
			Diabetes Mellitus type I	250.01
			Juvenile diabetes	250.01
			Diabetes with Insulin dependency and other complications	250.03,25011,25013,25021,250.23,250.31, 250.33,250.41,250.43,250.51,250.53,250.61, 250.63,250.71,250.73,250.81,250.83,250.91,250.93
	4.4	Repro-ductive	Autoimmune oophoritis	614.1, 614.2
			Endometriosis	617.0
	4.5	Salivary	Sjogren's syndrome	710.2
5		Neuro-logical disorders	Multiple Sclerosis (MS)	340
			Paraneoplastic cerebral degeneration (PCD)	331.89,334.9
			Devic's disease	341.0
			Balo concentric sclerosis	341.1
			Transverse myelitis	341.2, 323.81
			Chronic Inflammatory demyelinating polyneuropathy (CIDP),	357.81
			Narcolepsy	347.0, 347.00, 347.01
			Guillain-Barre syndrome	357.0
			Lambert –Eaton syndrome	358.1, 358.30
			PANDAS	279.49
			Autoimmune encephalomyelitis(ADEM)	323.61, 323.81
			Stiff person syndrome (SPS)	333.91
			Restless leg syndrome(RLS)	333.94
			Reflex sympathetic dystrophy	337.20, 337.21, 337.22, 337.29

Continue on Next Page

6	Muscular disorders	Myasthenia Gravis	358.0,358.00, 358.01
		Juvenile myositis, dermatomyositis	710.3
		Polymyositis	710.4
		Inclusion body myositis (IBM)	359.71
		fibromyalgia, myositis	729.1
		neuromyotonia	333.90, 359.29
7	Vascular inflamma-tion, & Systemic disorders	Essential mixed cryoglobulinemia	273.2
		Autoimmune lymphoproliferative syndrome	279.41
		Pernicious anemia(PA)	281.0
		Anti-phospholipid syndrome primary (APS-1)	289.81
		Autoimmune hemolytic anemia, cold agglutinin disease	283.0
		Paroxysmal nocturnal hemoglobinuria (PNH)	283.2
		pure red cell aplasia (PRCA)	284.81
		Antiphospholipid syndrome	286.53
		immune thrombocytopenic purpura (ITP)	287.31
		Evans syndrome	287.32
		Anti-phospholipid syndrome secondary (APS-2)	287
		Henoch-Schonlein purpura(HSP)	287.0
		Autoimmune thrombocytopenic purpura (ITP)	287.31
		antiphospholipid syndrome (APS)	286.53
		Evens syndrome	287.32
		Autoimmune neutropenia	288.09
		microscopic polyangiitis (MPA), polyarthritis nodosa	446.0
		Wegener's granulomatosis, Churg-Strauss	446.4
		Giant cell arthritis, temporal arthritis	446.5
		Takayasu's arthritis	446.7
		Leukocytoclastic vasculitis	447.6
		essential mixed cryoglobulinemia	273.2
		autoimmune non-specific	279.4
		autoimmune neutropenia	288.09
		Sarcoidosis	135
		Bechet's disease	136.1
		Systemic Lupus Erythematosus (SLE)	710.0
		CREST syndrome (systemic sclerosis)	710.1

Continue on Next Page

			Sjogren's syndrome, Sicca syndrome	710.2
			Mixed Connective Tissue disorder (MCTD)	710.8
			Undifferentiated connective tissue disease (UCTD)	710.9
			thrombocytopenia	278.5
			Relapsing polychondritis	733.99
			Still's disease	714.2
			Chronic Lyme disease	z8881
			Parry Romberg syndrome	349.89
			Discoid Lupus Erythematosus (SLE)	695.4
			Amyloidosis	277.30
8		Sensory Organs Diseases		
	8.1	Auditory	Autoimmune inner ear disease (AIED)	388.8
			Meniere's disease	386.00
	8.2	Ocular	Uveitis	364.3, 364.00, 364.10
			Sympathetic ophthalmia (SO)	360.11
			Autoimmune retinopathy	362.89
			Pars planitis	363.21
			Mooren's ulcer	370.07
			Cogan's syndrome	370.52
			Discoid Lupus of eyelids	373.34
			Graves ophtholmopathy (thyroid eyes)	242.9
			Intermidiate uvitis (pars plantis)	364.3
			neuromyelitis optica (devics disease)	341.0
			uvitis	364.00
			uvitis chronic	364.10
			ocular cicatricial pemphigoid	694.61
			cirrhosis without alcohol	571.5
			Primary billiary cholangitis (PBC)	576.1
			Primary sclerosing cholangitis	576.1

Continue on Next Page

9		Major Organs Diseases		
	9.1	Cardiac	Rheumatic fever, acute rheumatic pericarditis	391.0
			Rheumatic fever, acute rheumatic endocarditis	391.1
			Autoimmune myocarditis	391.2
			Rheumatic fever, other	391.8
			Rheumatic fever, unspecific	391.9
			Rheumatic myocarditis	398.0
			Rheumatic myocarditis	398.9
			Rheumatic fever, heart disease unspecific	398.90
			Rheumatoid carditis, congestive heart failure	398.91
			Rheumatic fever, heart disease other	398.99
			Rheumatoid carditis	714.2
			Giant cell myocarditis	422.91
			Autoimmune Cardiomyopathy,	422.0
			Coxsakie Myocarditis	422.0
			Autoimmune myocarditis	429.0
	9.2	Kidney	IgA nephropathy	583.9
			Interstitial cystitis(IC)	595.1
			Good Pastures syndrome, anti-glomerular basement membrane nephritis	446.21
			Lupus nephritus	583.81
	9.3	Liver	Autoimmune hepatitis	571.42
			Chronic hepatitis	571.40
			Cirrhosis without alcohol	571.5
			Primary biliary cholangitis (PBC)	576.1
			Primary sclerosing cholangitis	576.1
	9.4	Lung	Lung purpura with glumeronephritus complex	446.21
			Anti-synthetase syndrome	279.49
			Lung Disease Sjogren	517.8

Table 3.2 - The Quartile Classification of the Estimated Median Household Income of Residents by the Patients ZIP Code[80]

Quartile Ranges of ZIPINC_QRTL by Year

YEAR	Quartile 1 $	Quartile 2 $	Quartile 3 $	Quartile 4 $
2007	1 - 38,999	39,000 - 47,999	48,000 - 62,999	63,000+
2008	1 - 38,999	39,000 - 48,999	49,000 - 63,999	64,000+
2009	1 - 39,999	40,000 - 49,999	50,000 - 65,999	66,000+
2010	1 - 40,999	41,000 - 50,999	51,000 - 66,999	67,000+
2011	1 - 38,999	39,000 - 47,999	48,000 - 63,999	64,000+
2012	1 - 38,999	39,000 - 47,999	48,000 - 62,999	63,000+
2013	1 - 37,999	38,000 - 47,999	48,000 - 63,999	64,000+

Adopted from www.hcupus.ahrq.gov/db/vars/siddistnote.jsp?

Table 3.3 - The Surgery Types and Related ICD-9-CM Codes[58,81]

Procedure Type	Procedure Description	ICD-9-CM Code
Abdominal Hysterectomy	Hysterectomy, total or partial via abdominal surgery	68.4, 68.9, 68.49, 68.39, 68.69
Vaginal Hysterectomy	Hysterectomy, total or partial via vaginal procedure	68.5, 68.51, 68.59, 69.51
Laparoscopic Hysterectomy	Hysterectomy, total or partial, abdominal or vaginal via laparoscopic procedure	68.41, 68.31, 68.61, 68.51, 54.21, 68.71
Ovaries, Abdominal	Removal of ovaries, alone(oophorectomy) or as a part of hysterectomy via abdominal surgery	65.22, 65.52, 65.61, 65.51, 65.73, 65.49, 65.09, 65.39,
Ovaries, Laparoscopic	Removal of ovaries, alone or as a part of hysterectomy, abdominal or vaginal, via laparoscopy procedure	65.53, 65.63, 65.64, 65.23, 65.24, 65.25, 65.41, 65.01, 65.53, 65.31
None Specific	Non-specified hysterectomies	68.6, 68.7, 68.9, 68.3
Vaginectomy		70.32,70.33.70.61,70.63,70.64,70.69, 70.91,70.92,70.93
Vulvectomy		71.01,71.09,71.11,71.19, 71.61

Table 3.4 - The Comorbid Conditions Categories with ICD-9-CM Codes[77,83]

Disease Categories	ELX_#	Diseases	ICD-9 Code
1. Cardio-vascular Disorders	1	Congestive heart failure	398.91, 402.11, 402.91, 404.11, 404.13, 404.91, 404.93,428.0-428.9
	2	Cardiac arrhythmias	426.10, 426.11, 426.13, 426.2-426.53, 426.6-426.89, 427.0, 427.2, 427.31, 427.60,427.9, 785.0, V45.0, V53.3
	3	Valvular disease	093.20-093.24, 394.0-397.1, 424.0-424.91, 746.3-746.6, V42.2, V43.3
	4	Pulmonary circulation disorders	416.0-416.9, 417.9
	5	Peripheral vascular disorders	440.0-440.9, 441.2, 441.4, 441.7, 441.9, 443.1-443.9, 447.1,557.1,557.9, V43.4
	6	Hypertension	401.1, 401.9
	7	(combined): uncomplicated Hypertension, complicated	402.10, 402.90, 404.10, 404.90, 405.11, 405.19, 405.91, 405.99
2. Neurological Disorders	8	Paralysis	342.0-342.12, 342.9-344.9
	9	Other neurological disorders	331.9, 332.0, 333.4,333.5,334.0-335.9,340, 341.1-341.9,345.00-345.11, 345.40-345.51, 345.80-345.91, 348.1, 348.3, 780.3, 784.3
3. Chronic Diseases	10	Chronic pulmonary disease	490-492.8, 493.00-493.91, 494, 495.0-505, 506.4
	11	Diabetes, uncomplicated	250.00-250.33
	12	Diabetes, complicated	250.40-250.73, 250.90-250.93
	13	Hypothyroidism	243-244.2, 244.8, 244.9
4. Organ Failure/Diseases	14	Renal failure	403.11, 403.91, 404.12, 404.92, 585, 586, V42.0, V45.1, V56.0, V56.8
	15	Liver disease	070.32, 070.33, 070.54, 456.0, 456.1, 456.20, 456.21 571.0, 571.2, 571.3, 571.40-571.49, 571.5, 571.6, 571.8, 571.9,572.3,572.8, V42.7
	16	Peptic ulcer disease excluding bleeding	531.70, 531.90, 532.70, 532.90, 533.70, 533.90,534.70,534.90, V12.71
5. Immuno-Deficiency/ Infection	17	AIDS/HIV	042-044.9
6. Cancer	18	Lymphoma	200.00-202.38, 202.50-203.01,203.8-203.81, 238.6, 273.3, V10.71, V10.72, V10.79
7. Blood Disorders	22	Coagulopathy	2860-2869, 287.1, 287.3-287.5
	25	Fluid and electrolyte disorders	276.0-276.9
	26	Blood loss anemia	280.0
	27	Deficiency anemia	280.1-281.9, 285.9

Continue on Next Page

8. Mental/ Behavioral Disorders	28	Alcohol abuse	291.1, 291.2, 291.5, 291.8, 291.9, 303.90-303.93,305.00-305.03, V113
	29	Drug abuse	292.0, 292.82-292.89,292.9,304.00-304.93, 305.20-305.93
	30	Psychoses	295.00-298.9, 299.10-299.11
	31	Depression	300.4, 301.12, 309.0, 309.1, 311
9. Weight Related Disorders	23	Obesity	278.0
	24	Weight loss	260-263.9

Source: Comorbidity Measures for Use with Administrative Data. Anne Elixhauser, Claudia Steiner, D. Robert Harris and Rosanna M. Coffey Source: Medical Care, Vol. 36, No. 1 (1998), pp. 8-27.

Table 3.5 - The Gynecologic Cancer ICD-9-CM Codes[9,81]

Disease	ICD-9 Code
Malignant Neoplasm of:	
Uterus, Part Unspecified	179
Endocervix	180.0
Exocervix	180.1
Other Specified Sites of Cervix	180.8
Cervix Uteri, Unspecified Site	180.9
Placenta	181
Corpus Uteri, Except Isthmus	182.0
Isthmus	182.1
Other Specified Sites of Body of Uterus	182.8
Ovary	183.0
Fallopian Tube	183.2
Broad Ligament of Uterus	183.3
Parametrium	183.4
Round Ligament of Uterus	183.5
Other Specified Sites of Uterine Adnexa	183.8
Uterine Adnexa, Unspecified Site	183.9
Vagina	184.0
Labia Majora	184.1
Labia Minora	184.2
Clitoris	184.3
Vulva, Unspecified Site	184.4
Other Specified Sites of Female Genital Organs	184.8
Female Genital Organ, Site Unspecified	184.9
Carcinoma in Situ of:	
Cervix Uteri	233.1
Other and Unspecified Parts of Uterus	233.2
Unspecified Female Genital, Not Elsewhere Classified	233.3
Unspecified Female Genital Organ	233.30
Vagina	233.31
Vulva	233.32
Other Female Genital Organ	233.39
Neoplasm of Uncertain Behavior of:	
Uterus	236.0
Placenta	236.1
Ovary	236.2
Other and Unspecified Female Genital Organs	236.3

Adapted from: "Procedures to Treat Benign Uterine Fibroids in Hospital Inpatient and Hospital-Based Ambulatory Surgery Settings, 2013" AHRQ #200. 1-16. January 2016.

Table 3.6 - The List of Data Sources used to Examine the Research Question

Dataset	Year of Data Collection	Type of Data	State	Number of Observations (# of Patients)	Number of Variables in Data
1	2007	State Inpatient Data (SID)_ Core	Florida	2,563,370	282
2	2008	State Inpatient Data (SID)_ Core	Florida	2,571,753	283
3	2009	State Inpatient Data (SID)_ Core	Florida	1,303,082	289
4	2010	State Inpatient Data (SID)_ Core	Florida	2,640,092	294
5	2011	State Inpatient Data (SID)_ Core	Florida	2,656,249	297
6	2012	State Inpatient Data (SID)_ Core	Florida	2,670,520	296
7	2013	State Inpatient Data (SID)_ Core	Florida	2,673,488	295
Combined 2007-2013	2007-2013	State Inpatient Data (SID)_ Core	Florida	17,078,554	275

Table 3.7 - Autoimmune Diseases Groups "All", A, B, C, D and the Related Organs and Systems Affected

Group	Category*	AD Type:Systems And Organs Affected By Autoimmune Diseases	Number of Patients
All	1-9	All diseases including in Groups A, B, C, and D together	1,165,061
A	1, 2, 7, 8	Arthritis like diseases, Dermatological, Vascular and Systemic Disorders, and Sensory organs diseases (Auditory, Ocular)	593,069
B	3,9	Digestive system, and Major organs diseases, including:, Heart, Kidney, Liver, and Lung	238,568
C	5,6	Neurological and Muscular Disorders	295,580
D	4	Glandular Diseases including: Thyroid, Adrenal, Pancreatic, Reproductive, and Salivary	190,979

*See Table 3.1 for the list of "Category."

Table 3.8 - Categorical and Modeling Type of Predictor Variables

Variable	Categorical Type	Modeling Type
Congestive heart failure	Character	Nominal
Cardiac arrhythmia	Character	Nominal
Valvular disease	Character	Nominal
Pulmonary circulation disorders	Character	Nominal
Peripheral vascular disorders	Character	Nominal
Hypertension uncomplicated	Character	Nominal
Hypertension complicated	Character	Nominal
Paralysis	Character	Nominal
Other neurological disorders	Character	Nominal
Chronic pulmonary disease	Character	Nominal
Diabetes uncomplicated	Character	Nominal
Diabetes complicated	Character	Nominal
Hypothyroidism	Character	Nominal
Renal failure	Character	Nominal
Liver disease	Character	Nominal
Peptic ulcer disease excluding bleeding	Character	Nominal
AIDS/HIV	Character	Nominal
Lymphoma	Character	Nominal
Coagulopathy	Character	Nominal
Obesity	Character	Nominal
Weight loss	Character	Nominal
Fluid and electrolyte disorders	Character	Nominal
Blood loss anemia	Character	Nominal
Deficiency anemia	Character	Nominal
Alcohol abuse	Character	Nominal
Drug abuse	Character	Nominal
Psychoses	Character	Nominal
Depression	Character	Nominal
Median income level	Character	Ordinal
Age	Character	Ordinal
Race	Character	Nominal
Pay source	Character	Nominal
Length of stay	Character	Ordinal
Hysterectomy	Character	Nominal
Vaginectomy	Character	Nominal
Vulvectomy	Character	Nominal

CHAPTER 4: RESULTS

4.1 Descriptive Analysis Results

The pooled HCUP SID dataset used in this current study contained data on 17,078,554 patients who were discharged during the years of 2007-2013 from several Florida hospitals. Among this dataset, 9,590,428 (56.16%) patients were included because they were female. Amongst this female population, 1,165,061 (12.15%) women with one or more type(s) of AD were selected as the study population (group "All"). The study population were further classified into different groups (A, B, C, D) based on the systems, and the organs and tissues affected by autoimmunity (Figure 4.0, Table 4.1). As a result, group "All" included all the subjects with any AD types included in the study. Group A included patients with autoimmune arthritis like diseases, dermatological, vascular and systemic disorders, and sensory organs (auditory, ocular) diseases; group B included patients with autoimmunity affected digestive system, and major organs such as heart, kidney, liver, and lung; group C included autoimmune neurological and muscular disorders; and group D included autoimmune glandular diseases such as thyroid, adrenal, pancreatic, reproductive, and salivary gland diseases (Table 3.7).

4.1.1 *Patient Characteristics*

1. Patient characteristics of subjects among the AD Group "All": Approximately 1,165,061 women with one or more type(s) of AD were selected as the study population for the group "All". Among this population, there were 8,316 (0.71%) women who had one or more type(s) of GYNC, and 1,156,745 (99.29%) women who did not have any GYNC.

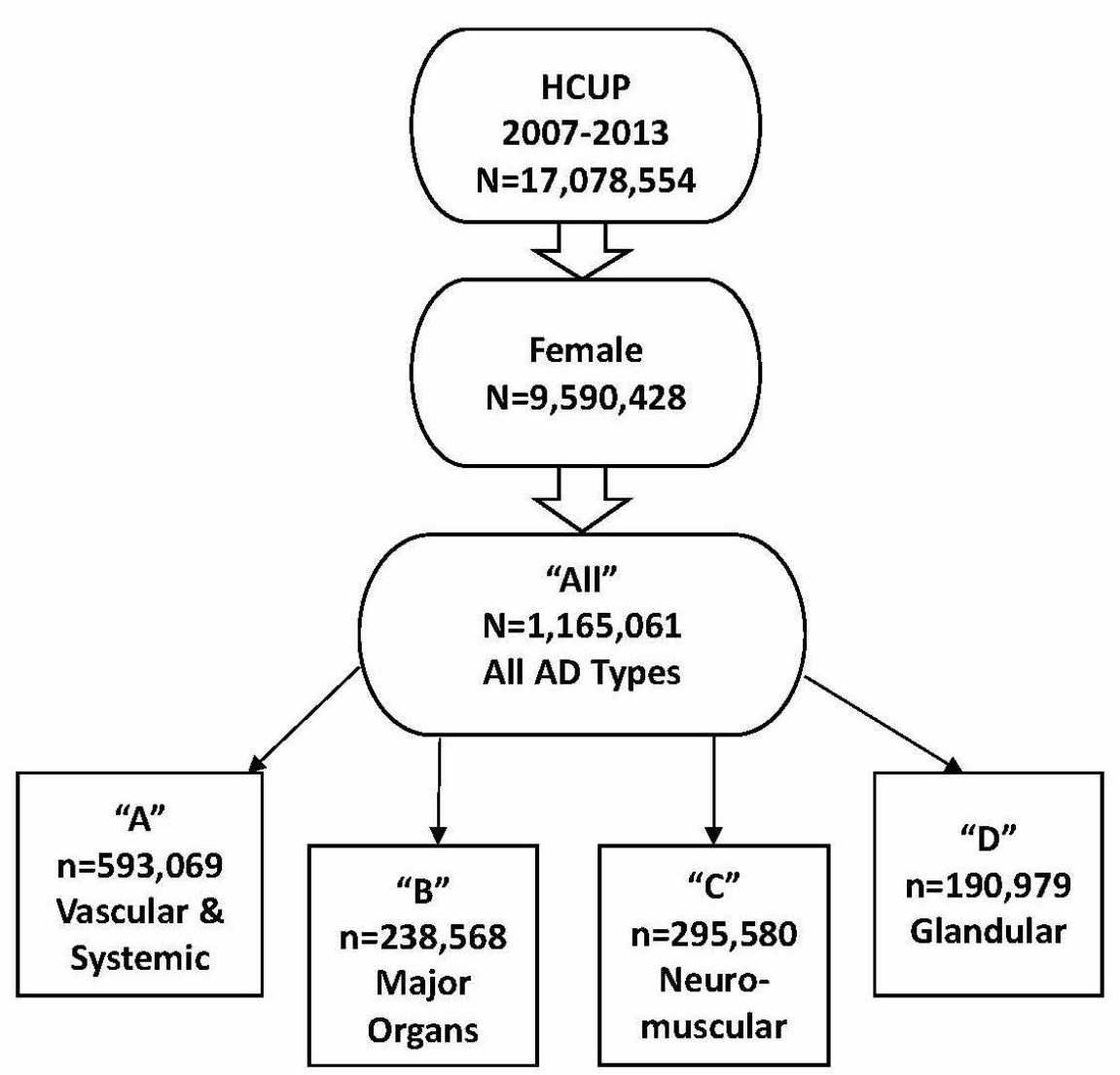

Figure 4.0: The Study Populations Data Pathway

Of the patients with GYNC in group "All", most were between the ages of 45-65 years old (43.80%) and White (65.50%). Nearly 64% belonged to the two lowest income levels, and nearly half had Medicare (50.79%). The majority of these women stayed in the hospital for 1-10 days (80.98%). In addition, among the patients with GYNC, over 31% had one or more types of Hysterectomies, 4.26% had Vaginectomy, and 2.53% had Vulvectomy (Table 4.2).

Among women without GYNC, most were > 65 years old (42.83%) and White (65.93%). Nearly 66% belonged to the two lowest income levels, and 57.6% had Medicare. The majority of women without GYNC also stayed in the hospital for 1-10 days (85.3%). In addition, among the patients with no GYNC, 2.97% had one or more types of Hysterectomies, 0.33% had Vaginectomy, and only 0.11% had Vulvectomy (Table 4.2).

2. Patient characteristics of subjects among the AD groups A-D: Among women in groups A-B with GYNC, most were 65 years or over (49.1% and 50.8%, respectively) while among women in groups C-D with GYNC, approximately half were between 45-65 years old (53.1% and 47.6%, respectively). Also, for groups A-D, most women with GYNC were White (range: 58.3-78.2%), and belonged to the two lowest income levels (range: 30.3-33.2%). The majority of women with GYNC in groups A-C had Medicare (range: 56.1-60.2%), whereas those in group D had private insurance (41.78%). Also, the majority of women with GYNC in groups A-D stayed 1-10 days in the hospital (range: 76.1-89.4%). In addition, among the patients with GYNC, patients in groups A-D had one or more types of Hysterectomies (range:18.4-67.7%), Vaginectomies (range:1.2-14.1%), or Vulvectomies (range: 2.39-3.61%), with GYNC patients in group D having the highest number of Hysterectomy and Vaginectomy (Table 4.2.1).

In contrast, among patients without GYNC in groups A-B most were 65 years or over (range: 43.05-50.92%), while in group C most were between 45-65 years old (44.14%), and in group D most were between 25-45 years old (36.16%). Also, for groups A-D, most women without GYNC were White (range: 59.4-76.6%) and belonged to the two lowest income levels (range: 31.56-35.43%). The majority of women without GYNC

in groups A-D had Medicare (range: 35.59-63.28%) and also stayed in the hospital 1-10 days (range: 82.42-88.02%). Among the patients without GYNC in groups A-C, the numbers of all the 3 aforementioned procedures were low (0.07-0.89%). However, among patients without GYNC in group D, nearly 15% of women had one or more types of Hysterectomies (Table 4.2.1).

4.1.2 Patient Comorbidities

1. Patient comorbidities of the subjects among AD group "All": The frequencies and percentages for the comorbid conditions among patients with and without GYNC for group "All" are demonstrated in Table 4.3. The highest comorbidities among patients with GYNC for the group "All" were Fluid and Electrolyte Disorders (37.48%), Coagulopathy (32.65%), and uncomplicated Diabetes (19.80%). The highest comorbidities among patients without GYNC for the group "All" were uncomplicated Hypertension (41.82%), Fluid and Electrolyte Disorders (33.86%), and Chronic Pulmonary Disease (29.18%) (Table 4.3).

2. Patient comorbidities of the subjects among AD group A-D: The comorbid conditions for groups A-D are shown in Table 4.3.1. The highest comorbidities among patients with GYNC in all 4 AD groups A-D, were the same conditions as for the group "All" as mentioned before. Notably, in group A, uncomplicated Hypertension had a high prevalence (43.05%). Other notable comorbidities were Liver disease in group B (45.36%), Depression in group C (27.99%) and Obesity in group D (22.31%). The highest comorbidities among patients without GYNC were Cardiac arrhythmia in group A (25.37%); Complicated diabetes (22.28%) in group B; Depression (33.60%) with group C; and uncomplicated Diabetes (37.87%) with group D (Table 4.3.1).

4.2 Predictive Analysis - Logistics Regression Analysis

4.2.1 *Unadjusted Association between Patient-Level Predictors and GYNC among Patients with AD*

A. Unadjusted Results for Patient Characteristics

Tables 4.4, and 4.4.1 display the unadjusted odds ratios (ORs) and 95% confidence intervals (CIs) of the associations between select patient characteristics and GYNC for AD group "All", and the groups A-D respectively.

1. Group "All" patients with AD: For the group "All", compared to patients 25-45 years old, women 45-65 years old had 77% and women >65 years had 33% increased odds of having GYNC (OR = 1.77, 95% CI: 1.66-1.90 and OR = 1.33, 95% CI: 1.24-1.42, respectively). While Black women had decreased odds of having GYNC (OR = 0.90, 95% CI: 0.84-0.96), Hispanics and Other ethnicities had 10% increased odds of having GYNC (OR = 1.10, 95% CI: 1.03-1.18 and OR = 1.10, 95% CI: 1.02-1.20 respectively) when compared to White women. Patients with Private insurance had 44% increased odds of having GYNC (OR = 1.44, 95% CI: 1.37-1.52) as compared to patients with Medicare. Compared to the highest level of median income (Q4), all 3 lower income levels had decreased odds of having GYNC (Q1: OR = 0.81,95% CI: 0.75-0.88; Q2: OR = 0.81, 95% CI: 0.74-0.87; Q3: OR = 0.88, 95% CI: 0.81-0.95; Table 4.4). For hospital LOS, patients with a LOS of 1-10 days had decreased odds of having GYNC as compared to patients with a LOS of 21-30 days (OR = 0.50, 95% CI: 0.38-0.62). In addition, among patients with different GYN related procedures, all had increased odds of having GYNC, with Vulvectomy having the OR with the highest magnitude (OR = 24.56, 95% CI: 21.18-28.48) (Table 4.4).

2. Group A-D patients with AD: Similar to group "All", older women in groups A-D had statistically significant increased odds of having GYNC (Table 4.4.1). Also, like the group "All", compared to White women in the groups A-D, Black women had decreased odds of having GYNC (Table 4.4.1). Similar to group "All", compared to the Medicare beneficiaries, self-paid patients had decreased odds of having GYNC in groups A-B (OR = 0.78, 95% CI: 0.63-0.95, and OR = 0.58, 95% CI: 0.38-0.88). However, patients in group C-D had no association with GYNC when compared to Medicare beneficiaries (OR = 0.99, 95% CI: 0.71-1.38, and OR = 1.08, 95% CI: 0.88-1.32) (Table 4.4.1). For LOS in hospital, similar to patients in group "All", a dose-response effect was observed for patients in groups A-C. Specifically, as LOS increased, the odds of GYNC increased. However, for patients in group D, those with a LOS of 1-10 days had 26% increased odds of having GYNC (OR = 1.26, 95% CI: 0.87-1.81). But, this finding was not statistically significant. Moreover, in groups A-D patients with different GYN related procedures all had increased odds of having GYNC. This association was strongest in group B where women who had a Hysterectomy procedure had nearly 51 times the odds of GYNC (OR = 50.92, 95% CI: 43.57-59.50; Table 4.4.1).

B. Unadjusted Results for Patient Comorbidities

Tables 4.5, and 4.5.1 report the unadjusted ORs and 95% CIs for the associations between comorbidities and GYNC for AD group "All", and the groups A-D respectively. 1. Group "All" patients with AD: Comorbidities such as Weight Loss (OR = 2.02, 95% CI: 1.89-2.16), Coagulopathy (OR = 1.90, 95% CI: 1.80-1.97), Anemia due to blood loss (OR = 1.69, 95% CI: 1.50-1.90), AIDS/HIV (OR = 1.30, 95% CI: 1.02-1.70), Fluid and Electrolyte Disorders (OR = 1.17, 95% CI: 1.12-1.22), and Obesity (OR = 1.11, 95% CI:

1.04-1.17), were all associated with statistically significant increased odds of having GYNC (Table 4.5).

2. Group A-D patients with AD: Similar to group "All", in groups A-D comorbidities such as Weight loss and Blood loss anemia were associated with statistically significant increased odds of GYNC. In groups A-C, Coagulopathy and Fluid and Electrolyte Disorders were also associated with statistically significant increased odds for GYNC, with Coagulopathy having the greatest association (OR = 2.17, 95% CI: 2.04-2.30) for group A. Similarly, Obesity was associated with statistically significant increased odds of GYNC for groups B and D, with group D having the highest odds (OR = 2.13, 95% CI: 1.92-2.40) (Table 4.5.1).

4.2.2 *Adjusted Association between Patient-Level Predictors and GYNC among Patients with AD*

A. Adjusted Results for Patient Characteristics

Tables 4.6, and 4.6.1 display the adjusted ORs and 95% CIs for the associations between patient characteristics and GYNC for AD group "All", and the groups A-D respectively.

1. Group "All" patients with AD: In the adjusted model for the associations between patient characteristics and GYNC in group "All", the variables age, pay source, and GYN related procedures were retained. Compared to women age 25-45 years old, women <25 years old continued to have statistically significant decreased odds of GYNC although the magnitude of the association remained attenuated after adjustment (OR = 0.23, 95% CI: 0.16-0.33). However, after adjustment, older age was associated with decreased odds of GYNC (45-65 years old: OR = 0.90, 95% CI: 0.82-0.99; and >65 years old: OR = 0.79, 95% CI: 0.70-0.88). After adjustment, the magnitude of the association between

Medicaid holders and GYNC increased and retained its statistical significance (OR = 1.36, 95% CI: 1.23-1.51). Similarly, after adjustment, the magnitude of the association between self-pay patients and GYNC increased and retained its statistical significance (OR = 1.39, 95% CI: 1.15-1.67). However, after adjustment, Private insurance was associated with decreased odds of GYNC (OR = 0.90, 95% CI: 0.82-0.98). After adjustment, the magnitude of the association between GYN related procedures and GYNC remained similar to unadjusted findings and all retained their statistical significance, with Hysterectomy having the highest magnitude (OR = 41.38, 95% CI: 37.40-45.78, Table 4.6).

2. Group A-D patients with AD: When the study population was considered as groups A-D, only the procedure variable was retained in the model. After adjustment, the magnitude of the association between Hysterectomy and Vulvectomy with GYNC remained similar and both retained their statistical significance for groups A-D (Hysterectomy: A, OR = 44.61, 95% CI: 37.6-53.0; B, OR = 44.6, 95% CI: 33.5-59.3; C, OR = 67.87, 95% CI: 51.6-89.3; D, OR = 35.43, 95% CI: 29.3-42.80. Vulvectomy: A, OR = 39.00, 95% CI: 27.0-56.3; B, OR = 89.8, 95% CI: 45.6-177.0; C, OR = 34.37, 95% CI: 19.1-61.9; D, OR = 4.00, 95% CI: 2.56-6.19). Similar to unadjusted findings, the magnitude of the association between Vaginectomy and GYNC remained similar and retained its statistical significance for group D (OR = 2.21, 95% CI: 1.73-2.82). The magnitude of the association between Vaginectomy and GYNC for group A remained similar to the unadjusted findings, but did not retain statistical significance after adjustment (OR = 1.60, 95% CI: 0.94-2.69). Unlike the unadjusted findings, groups B-C did not have any association with GYNC after adjustment (Table 4.6.1).

B. Adjusted Results for Patient Comorbidities

Tables 4.7, and 4.7.1 display the adjusted ORs and 95% CIs for the associations between patient comorbidities and GYNC for AD group "All", and the groups A-D respectively.

1. Group "All" patients with AD: Similarly, in adjusted analysis for group "All", the magnitude of association between comorbidities such as AIDS/HIV (OR = 1.97, 95% CI: 1.42-2.74), Coagulopathy (OR = 1.26, 95% CI:1.18-1.34), Obesity (OR = 1.49, 95% CI; 1.36-1.62), Weight Loss (OR = 1.12,95% CI: 1.02-1.21), and Fluid and Electrolyte Disorders (OR = 1.11, 95% CI: 1.04-1.18) with GYNC remained similar, and all retained their statistical significance. However, with adjustment, Blood Loss Anemia, did not retain its statistical significance. Similar to unadjusted findings, Renal Failure retained its statistical significance (OR = 1.48, 95% CI: 1.26-1.72) (Table 4.7).

2. Group A-D patients with AD: After adjustment, comorbidities such as: Other Neurological Disorders and Chronic Pulmonary Diseases remained similar to unadjusted findings, and retained their statistical significance among group A-D. Also, the magnitude of the association between comorbidities such as AIDS/HIV, Weight Loss, Coagulopathy, Fluid and Electrolyte Disorders, and Alcohol Abuse with GYNC remained similar to unadjusted findings, and all retained their statistical significance after adjustment in groups A, C, and D. The magnitude of the association between Obesity and GYNC increased after adjustment and remained statistically significant among groups A-D (group A: OR = 1.47, 95% CI: 1.31-1.66; group B: OR = 1.61, 95% CI: 1.32-1.97; group C: OR = 1.22, 95% CI: 1.00-1.50; group D: OR = 1.51, 95% CI: 1.25-1.82 respectively, Table 4.7.1).

4.3 Predictive Analytics - Classification Tree Analysis

4.3.1 *Boot Strap Forest Results, using all Predictor Variables*

The BSF analyses were conducted to identify the effect of 36 predictor variables on GYNC, and to investigate the multilevel interactions between the risk factors. BSF analysis was applied to all 5 groups of AD patients: "All", A, B, C, and D.

4.3.1.1 BSF results for all patient-level predictors and GYNC among patients with AD

1. BSF analysis for group "All" patients with AD: For group "All", a random sample of 1:1 ratio of GYNC(+):GYNC(-) subjects were selected among the study population using SAS, to create the data subset for "All", called group "All" subset. The resulted sample included 16,600 AD patients, including 8,300 subjects with GYNC, and 8,300 subjects without GYNC. All 36 categorical transformed predictor variables were used in the BSF platform of JMP pro 13 to conduct the BSF analyses. As directed by JMP 13 pro menus, a 10,000-decision tree model was created as shown in the menu window for the group "All" subset below (Figure 4.1).

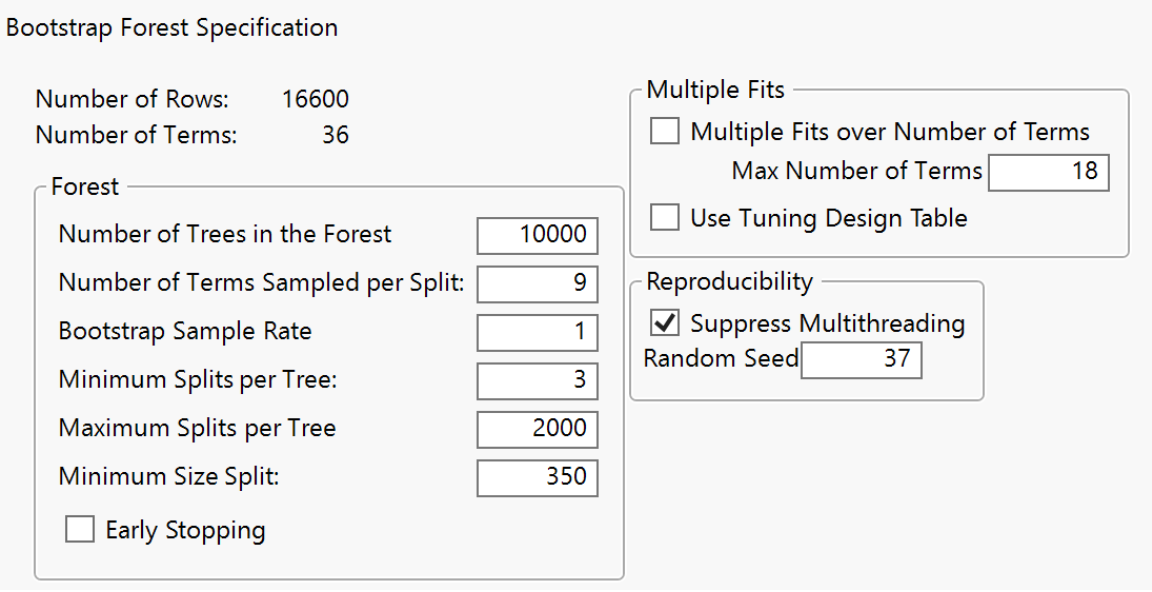

Figure 4.1: The BSF Menu for Criteria Chosen in the Predictive Model for AD Group "All" Subset

A validation column window was also created based on JMP pro 13 settings to facilitate the partitioning of the data subset by 80%, 10%, and 10% for Training, Test, and Validation sets, respectively (Figure.4.2).

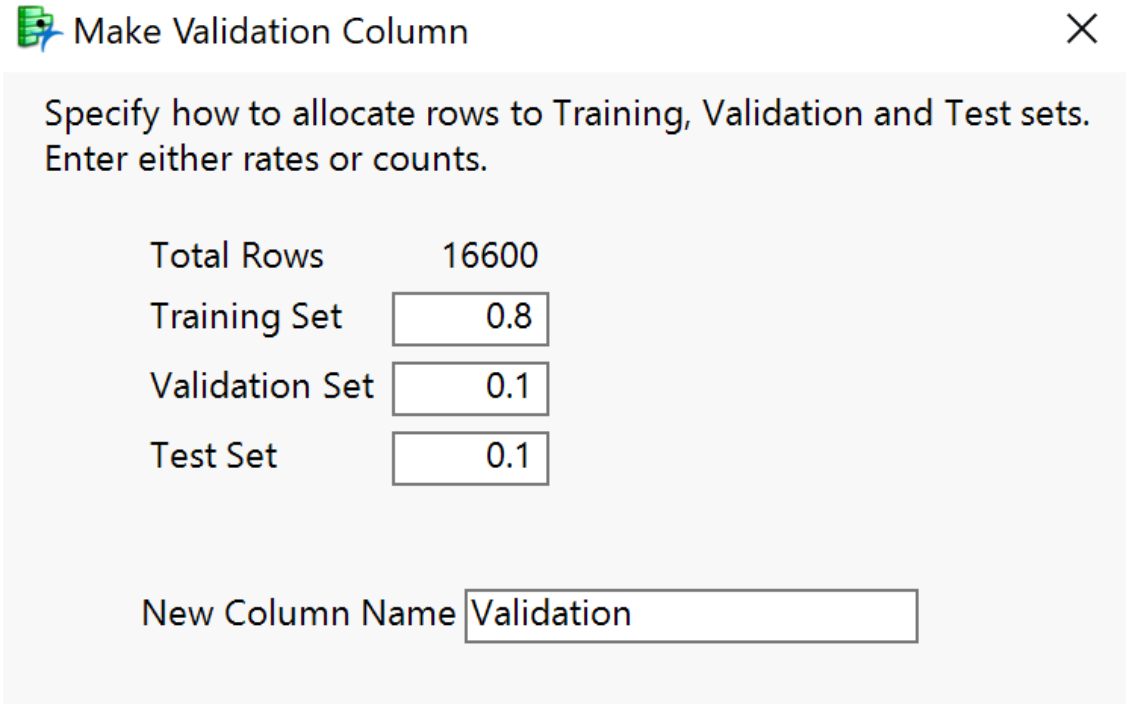

Figure 4.2: The BSF Validation Column Window for AD Group "All" Subset

The results of BSF analyses for group "All" are reported in multiple outputs, created by BSF platform of JMP pro 13 as follows:

A. <u>Confusion Matrix:</u> The values that resulted from the data-subset Test to create the evaluating criteria are summarized in a table called Confusion Matrix, specified for categorical data (Figure 4.3).

Confusion Matrix

Training

Actual	Predicted Count	
GYN Cancer	0	1
0	5536	1146
1	2734	3941

Validation

Actual	Predicted Count	
GYN Cancer	0	1
0	681	135
1	344	463

Test

Actual	Predicted Count	
GYN Cancer	0	1
0	676	126
1	319	499

Figure 4.3: Confusion Matrix for AD Group "All" Subset

The Confusion Matrix is a 2x2 table of all the possible situation (no = 0, yes = 1) in a prediction, and is a mean to measure how well the predictive model predicted the binary outcome of the study (GYNC negative = 0, GYN positive = 1). The Confusion Matrix table has 4 possibilities of true negative (TN) 0,0; false negative (FN) 1,0; false positive (FP) 0,1; and true positive (TP) 1,1.

The common measure for accuracy of a prediction model are sensitivity and specificity. Sensitivity is the proportion of observed positives that were also predicted positive, and specificity is the proportion of observed negatives that were also predicted as negative.

In Figure 4.3, the Test outcomes of TP = 499, FP = 126, FN = 319, and TN = 676 were used to calculate the sensitivity (true positive rate, TPR), specificity (true negative rate, TNR), false positive rate (FPR, the rate of negative test identified as positive), wrongly predicted value (WPV), and overall accurate prediction (OAP) of GYNC using patient-level predictors among patients with AD, group "All" subset. The resulted predictive values are shown in Table 4.8, using the following formulas:[95]

Sensitivity = TP/(TP+FN): Which is the true positive rate for GYNC prediction, and is the probability of GYNC (+) that was correctly predicted positive

Specificity = TN/(TN+FP): Which is the true negative rate for GYNC prediction, and is the probability of GYNC (-) that was correctly predicted negative.

False Positive Rate = FP/(FP+TN) = 1-Specificity: Which is the rate of GYNC (-) that were wrongly identified as GYNC (+).

Wrongly Predicted Values=(FP+FN)/(TN+TP+FP+FN): Which is the values that shows where GYNC (+) were identified as GYNC (-) and vice versa.

Overall Accurate Prediction = 1-Wrongly Predicted Value: Which is the probability of GYNC (+) were identified positive, and GYNC (-) were identified negative.

The sensitivity for the group "All" subset was approximately 61%, with specificity of 84%, and WPV of 27% (Table 4.8). This result indicates that the prediction model for group "All" subset had 61% accuracy in predicting true GYNC positives, and 84% accuracy in predicting true GYNC negatives when using the predictor variables. The 27% WPV indicates the percentage of wrong predictions of GYNC positives and GYNC negatives.

B. **Column Contribution:** Another measurement of the relationship between the 36 predictors and GYNC is shown in a table called Column Contribution (CC) which was created automatically by JMP pro13 platform during the analysis of AD group "All" subset data (Figure 4.4). The CC table in Figure 4.4 shows the predictors in the model that best predict GYNC. The results of the BSF analysis demonstrated that the predictor variable Hysterectomy is the strongest predictor of GYNC (61%), followed by Coagulopathy (11%), Age (7.8%), Chronic Pulmonary Disease (3.8%), Congestive Heart Failure (3.5%), Weight Loss (2.2%), Pay Source (2.0%), and Other Neurological Disorders (1.5%). As it is displayed, the Hysterectomy predictor also has the highest G^2

of 776.30 which is indicative of the reduction in the sum of squares, which means the largest value is the better fit of the predictor in the model.

C**. <u>Receiver Operating Characteristic (ROC):</u>** Another prediction measurement used was ROC, a methodology used to measure the FPR (1-specificity) against the TPR (sensitivity) by which the quality of the predictor is measured, using the area under the curve (AUC). AUC measures the 2-dimensional area underneath the ROC curve from (0,0), to (1,1).[99,100] In general AUC = 0.5 is considered no discrimination. Any value between 0.5-1.0 is subject to interpretation, where a value between 0.7-0.79 is considered acceptable, and the values between 0.8-0.89 and 0.90-1.00 indicate excellent and outstanding results respectively.[91] Figure 4.5 demonstrates the ROC for the AD group "All" subset, where the AUC is over 80%, which is indicative of an excellent prediction of GYNC by the predictors shown in CC table (Figure 4.5).

Column Contributions

Term	Number of Splits	G^2	Portion
Hysterectomy	8370	776.303521	0.6096
Coagulopathy	10395	139.253108	0.1094
Age	13950	99.2740444	0.0780
Chronic Pulmonary Disease	8477	47.9515461	0.0377
Congestive Heart Failure	6543	45.0006429	0.0353
Weight Loss	4341	28.1751161	0.0221
Pay Source	9996	25.6795487	0.0202
Other Neurological Disorders	4268	19.1280302	0.0150
Fluid and Electrolyte Disorders	7299	10.7830666	0.0085
Diabetes Complicated	2206	9.33880399	0.0073
Depression	4229	8.58305604	0.0067
Peripheral Vascular Disorders	1578	7.84595103	0.0062
Hypertension Uncomplicated	6753	6.5924298	0.0052
Length of Hospital Stay	2964	6.48693348	0.0051
Hypertension Complicated	3082	6.1415437	0.0048
Valvular Disease	1953	5.06702573	0.0040
Liver Disease	1871	4.96463968	0.0039
Race	4546	4.0793036	0.0032
Houshold Income	5707	3.96620122	0.0031
Cardiac Arrhythmia	3756	3.79213065	0.0030
Hypothyroidism	3403	3.01304299	0.0024
Renal Failure	2492	2.92932111	0.0023
Drug Abuse	508	2.66448053	0.0021
Obesity	2491	2.46260279	0.0019
Vaginectomy	118	1.91512421	0.0015
Diabetes Uncomplicated	2729	1.25459694	0.0010
Deficiency Anemia	341	0.46707354	0.0004
Pulmonary Circulation Disorders	265	0.20631332	0.0002
Alcohol Abuse	15	0.08468289	0.0001
Blood Loss Anemia	5	0.00441779	0.0000
Vulvectomy	0	0	0.0000

Figure 4.4: The Column Contribution for AD Group "All" Subset, the Measurement of the Relationship Between the Predictor Variables and GYNC

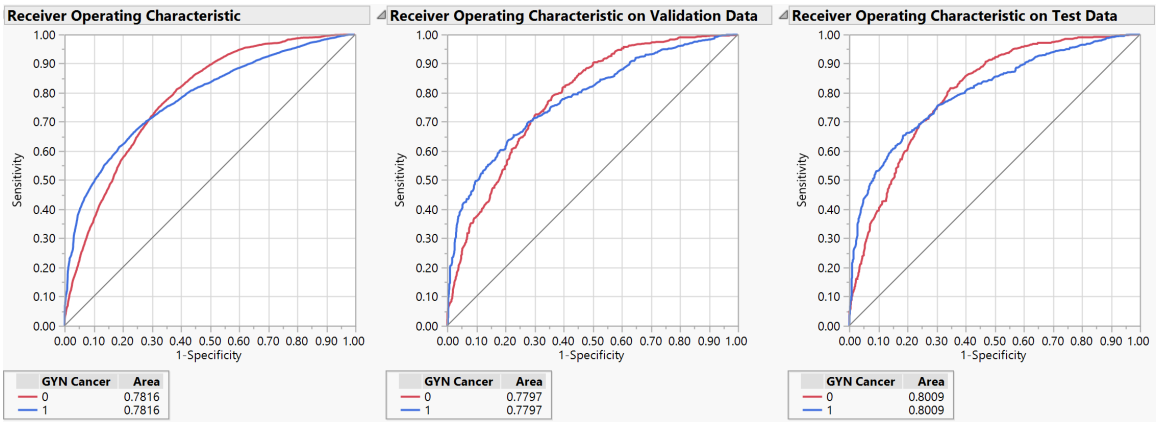

Figure 4.5: The Receiver Operating Characteristic for AD Group "All" Subset, the Measurement of the Quality of Predictors of GYNC

2. BSF analysis for patients with AD groups "A-D" subsets: Similarly, a random sample of 1:1 ratio of GYNC (+):GYNC (-) subjects were selected among the study population using SAS, to create the data subsets "A-D", called groups "A", "B", "C", and "D" subsets. The resulted sample included 9,140 AD patients with and without GYNC for group "A" subset, 2,200 for group "B" subset, 2,660 for group "C" subset, and 3,920 for group "D" subset (Table 4.8). All 36 categorical transformed predictor variables were used in BSF platform of JMP pro 13 to conduct the BSF analyses. A 10,000-decision tree model was created for each group "A-D" subset. Likewise, the Confusion Matrix was created for all 4 subsets, and were used to calculate the sensitivity, specificity, false positive rate, wrongly predicted value, and overall accurate prediction of GYNC using patient-level predictors among the patients with AD groups "A-D" subsets. The resulted predictive values are shown in Table 4.8. As it is reported in Table 4.8, the respective sensitivity, specificity, and WPV for each subset are as follows: subset "A": sensitivity = 65%, specificity = 73% with WPV = 31%; "B": sensitivity =48%, specificity = 73% with WPV = 40%; "C": sensitivity = 71%, specificity = 37% with WPV = 47%; and "D": sensitivity = 68%, specificity = 82% with WPV = 24%. Overall, the sensitivity of the subsets "A-D" was 48%-71%, with specificity of 37%-82%, and WPV of 24%-47%. Among all the AD groups "A-D" subsets, the "D" subset demonstrated the best combination of sensitivity (68%), specificity (82%), with lowest WPV (24%), along with AUC value (81%) which translates into a higher value for prediction of GYNC.

The results of BSF analyses for groups "A-D" subsets are also reported in multiple outputs, created by BSF platform of JMP pro 13. Because the group "D" subset showed the best sensitivity and specificity combination with a low WPV, and a high

AUC value, it is chosen as the representative model to demonstrate all the predictive analytics results for groups "A-D" subsets throughout the dissertation. The output created by BSF analysis platform of JMP pro 13 for AD groups "A-C" subsets are provided in the Tables and Figures section of the current chapter. The results of BSF analyses for group "D" subset are as follows:

A**. Confusion Matrix:** The values that resulted from the data-subset Test to create the evaluating criteria are summarized in a table called Confusion Matrix, specified for categorical data. The Confusion Matrix created for group "D" subset is displayed in Figure 4.6.

Confusion Matrix

Training

Actual	Predicted Count	
GYN cancer	0	1
0	1330	234
1	506	1075

Validation

Actual	Predicted Count	
GYN cancer	0	1
0	175	27
1	64	146

Test

Actual	Predicted Count	
GYN cancer	0	1
0	159	35
1	54	115

Figure 4.6: Confusion Matrix for AD group "D" Subset

B. **Column Contribution:** The CC table in Figure 4.7 shows the predictors in the model that best predict GYNC for AD group "D" subset, among all 36 predictor variables. Similar to group "All" subset, the result of BSF analysis demonstrated that the independent variable Hysterectomy is the highest predictor of GYNC (76%). Other significant predictors in subset "D" were Age (7.4%), Fluid and Electrolyte Disorders (3.4%), Complicated and Uncomplicated Diabetes (3.3%, and 2.8%), Pay Source (1.7%), Obesity (1.6%), and Chronic Pulmonary Disease (1.5%) (Figure 4.7).

Column Contributions

Term	Number of Splits	G^2	Portion
Hysterectomy	4496	217.926817	0.7607
Age	4162	21.270253	0.0742
Fluid and Electrolyte Disorders	2669	9.59206516	0.0335
Diabetes Complicated	1153	9.35617803	0.0327
Diabetes Uncomplicated	2554	8.14417092	0.0284
Pay source	2798	4.95599688	0.0173
Obesity	1275	4.70133086	0.0164
Chronic Pulmonary Disease	1285	4.27895252	0.0149
Race	1758	2.29379043	0.0080
Depression	547	1.21980863	0.0043
Hypertension Uncomplicated	1633	1.19233983	0.0042
Household Income	1675	0.80443586	0.0028
Hypothyroidism	477	0.73171891	0.0026
Cardiac Arrhythmia	7	0.00137808	0.0000
Hospital Length of Stay	6	0.00103897	0.0000
Vaginectomy	0	0	0.0000
Vulvectomy	0	0	0.0000

Figure 4.7: The Column Contribution for AD Group "D" Subset, the Measurement of the Relationship between the Predictor Variables and GYNC

C. **Receiver Operating Characteristic:** The AUC created in ROC for group "D" subset (Figure 4.8) also confirms the significance of the independent variables shown in Figure 4.7 as predictors of GYNC in AD patients of group "D" subset (81%).

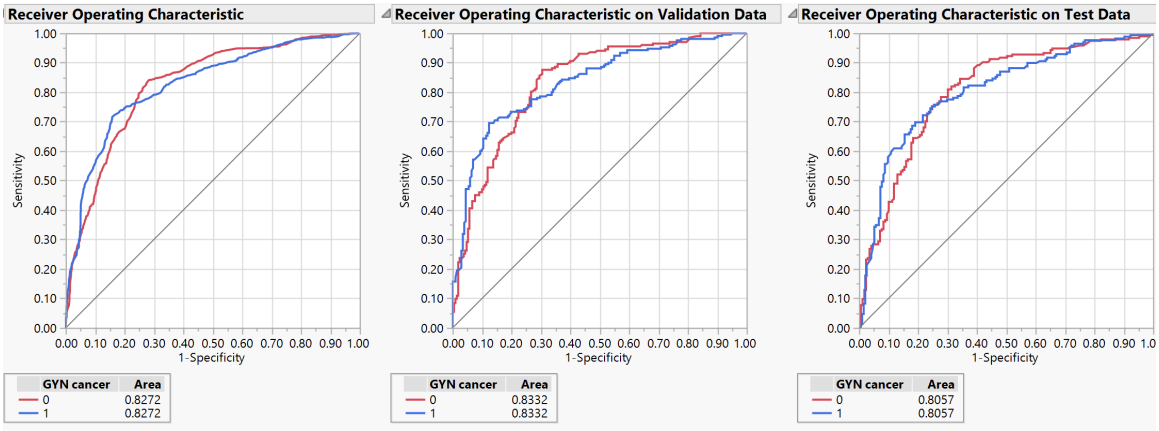

Figure 4.8: The Receiver Operating Characteristic for AD Group "D" Subset, the Measurement of the Quality of Predictors of GYNC

4.3.1.2 Adjusted BSF results, without the predictor "Hysterectomy"

The CC in both groups "All" subset and group "D" subset showed that Hysterectomy was the highest contributor in predicting GYNC. To demonstrate the

importance of other predictor variables in predicting GYNC in cases where there is not any Hysterectomy procedure involved, the variable Hysterectomy was eliminated in a subsequent BSF analysis. In this modified model, the process of creating the BSF model was repeated only without using the variable Hysterectomy for both groups “All” and “D” subsets. The results for the AD group “All” and “D” subsets are demonstrated in Figure 4. 9 and Figure 4.10 respectively.

The modified model of group “All” subset without the predictor Hysterectomy demonstrated that Age, became the most important predictor for GYNC (19.3%), followed by the comorbidities Coagulopathy (18.8%), Congestive Heart Failure (13.4%), and Chronic Pulmonary Diseases (12.8%) (Figure 4.9). As a result, the significance of predictors in terms of priority changed, in comparison to the CC table when the Hysterectomy variable was present (priorities were Coagulopathy, then Age, and later Chronic Pulmonary Diseases). Similarly, with group “D” subset, priorities were changed as well. When Hysterectomy was removed Age became the strongest predictor of GYNC (33.4%), followed by uncomplicated Diabetes (15.8%) and Fluids and Electrolyte Disorders (13.5%) (priorities were Age, then Fluids and Electrolyte Disorders, and later Diabetes, Figure.4.10).

Column Contributions

Term	Number of Splits	G^2	Portion
Age	9755	77.9282763	0.1934
Coagulopathy	8469	75.6675312	0.1878
Congestive Heart Failure	6461	53.9160777	0.1338
Chronic Pulmonary Disease	7763	51.6791646	0.1283
Other Neurological Disorders	4459	25.2524764	0.0627
Weight Loss	3226	16.3805869	0.0407
Diabetes Complicated	2580	12.7874807	0.0317
Depression	4415	12.6099063	0.0313
Pay Source	6330	11.9904265	0.0298
Hypertension Complicated	3522	9.51005044	0.0236
Peripheral Vascular Disorders	1625	9.1884974	0.0228
Valvular Disease	2025	5.90910694	0.0147
Renal Failure	2790	4.81648112	0.0120
Length of Hospital Stay	2607	4.80650104	0.0119
Cardiac Arrhythmia	3678	4.11253288	0.0102
Drug Abuse	647	3.62618912	0.0090
Liver Disease	1423	3.33971681	0.0083
Hypothyroidism	3209	3.2484807	0.0081
Obesity	2628	3.02629838	0.0075
Fluid and Electrolyte Disorders	4888	2.92029893	0.0072
Vaginectomy	141	2.3008891	0.0057
Hypertension Uncomplicated	4456	2.19582594	0.0054
Race	3271	1.91910049	0.0048
Houshold Income	4125	1.75353762	0.0044
Diabetes Uncomplicated	2320	0.82380636	0.0020
Deficiency Anemia	489	0.7538884	0.0019
Pulmonary Circulation Disorders	356	0.31389473	0.0008
Alcohol Abuse	23	0.12425201	0.0003
Blood Loss Anemia	19	0.01685252	0.0000
Vulvectomy	0	0	0.0000

Figure 4.9: The Adjusted Column Contribution for AD Group "All" Subset, when Variable Hysterectomy is Eliminated

Column Contributions

Term	Number of Splits	G^2	Portion
Age	3825	22.1780117	0.3345
Diabetes Uncomplicated	2691	10.4610359	0.1578
Fluid and Electrolyte Disorders	2757	8.9768823	0.1354
Chronic Pulmonary Disease	1747	7.1632548	0.1080
Diabetes Complicated	571	4.40031059	0.0664
Obesity	1489	4.38766569	0.0662
Depression	1003	3.23624039	0.0488
Pay source	2126	2.74179495	0.0414
Hypertension Uncomplicated	1516	1.1618764	0.0175
Hypothyroidism	713	1.01769448	0.0153
Race	1161	0.45826368	0.0069
Household Income	1313	0.1155769	0.0017
Cardiac Arrhythmia	5	0.00083267	0.0000
Hospital Length of Stay	0	0	0.0000
Vaginectomy	0	0	0.0000
Vulvectomy	0	0	0.0000

Figure 4.10: The Adjusted Column Contribution for AD Group "D" Subset, when Variable Hysterectomy is Eliminated

4.3.2 Decision Tree Results, using all Predictor Variables

In order to have a smaller, comprehensive partition tree that showed the relationship of the GYNC binary response with the 36 predictors, the Decision Tree (DT) Analysis technique from JMP® Pro 13 was used.

1. DT results for patients with AD group "All" subset: The DT output by JMP pro 13 demonstrated in Figure 4.11 was created for AD group "All" subset, using all 36 predictors. The DT analysis method searches all possible splits of the independent variables and chooses the best splits to predict the outcome response. In Figure 4.11, the first split shows that only 17.2% of patients (n=2,291) have had Hysterectomy with the majority having GYNC (first split on left node, blue color). In turn, among these patients with Hysterectomy, 76.8% are 45 years of age or older. This makes predictor variable Age a strong predictor of GYNC among patients with Hysterectomy.

However, looking at the right node on the first split, 82.8% of patients have no Hysterectomy. Among those with no Hysterectomy, Coaglupathy comorbidity demonstrates significance, followed by Chronic Pluminary Disease, and Age of 25 years or older.

On the opposite side of the split, nearly 30% of patients with no Hysterectomy and no Coaglupathy have Vulvectomy due to GYNC. However among those without vulvectomy, Age of 45 years or older, Weight Loss, and Congestive Heart Failure were good predictors of GYNC (Figure 4.11).

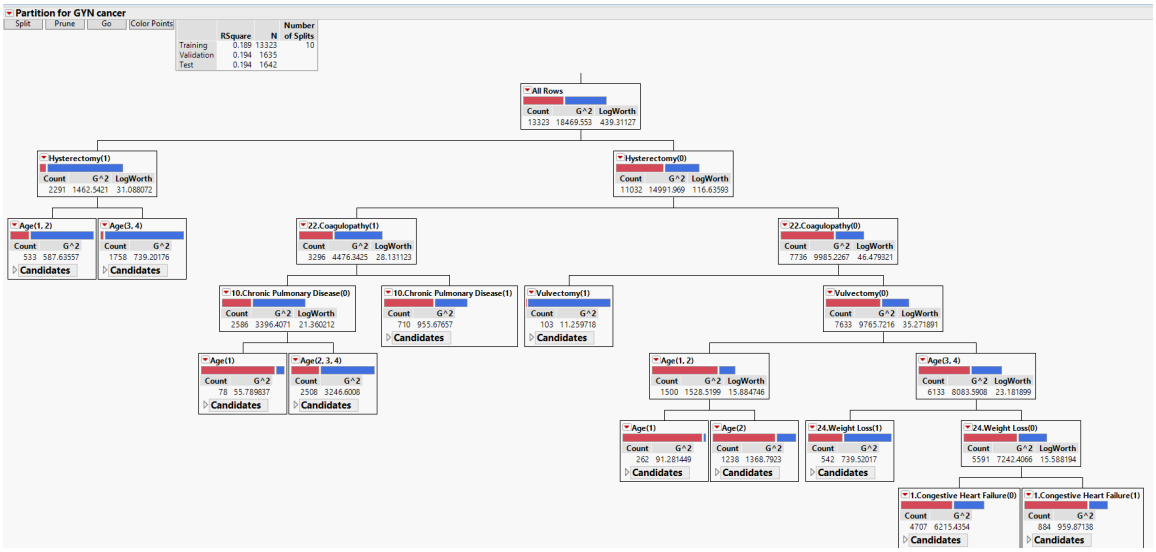

Figure 4.11: The Decision Tree Output for AD Group "All" Subset

2. DT results for patients with AD group "D" subset: Similarly, the DT output by JMP pro 13 demonstrated in Figure 4.12 was created for AD group "D" subset, using all 36 predictors. In Figure 4.12, the first split shows that 41% of patients have had Hysterectomy with the majority having GYNC (on left node, blue color). In turn, among patients with Hysterectomy (n = 1,285), 65% are 45 years of age or older. Among this population, uncomplicated Dibetes and Fluid and Electrolyte Disorders are significant.

However, looking at the right node on the first split, 59% of patients have no Hysterectomy (n=1,846). Likewise, among those with no Hysterectomy, the variable Age 45-65 years old demonstrats significance, followed by Chronic Plumonary Disease. In the opposite side of the split, nearly 73% of AD patients aged 45-65 years (n=812) had no Hysterectomy and no Chronic Plumonary Disorder. Among this later population, 84% of patients (n=682) who belonged to any Income Level (Q1-Q4) had a strong chance of having GYNC (G^2=871.9). This finding showes that the patients in group "D" subset with autoimmunity conditions involved glandular systems, and being 45-65 years old, with any income level are at risk for GYNC (Figure 4.12).

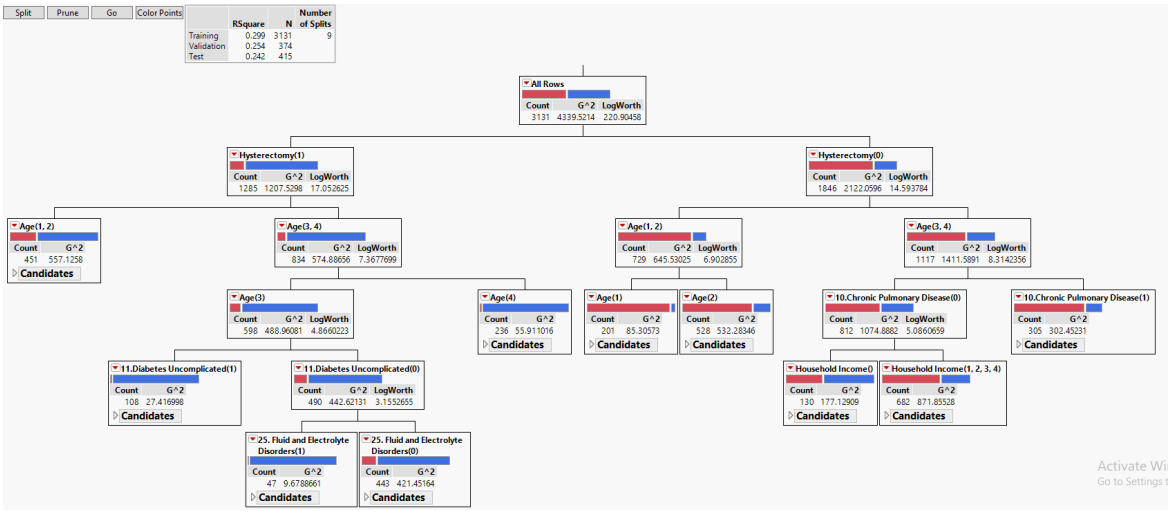

Figure 4.12: The Decision Tree Output for AD Group "D" Subset

Table 4.1 - Frequencies and Percentages of the Study Population, Florida HCUP-SID 2007-2013

	Autoimmune Disorder Groups Frequencies N (%)				
	"All"	**A**	**B**	**C**	**D**
Study Population	1,165,061 (100)	593,069 (50.90)	238,568 (20.47)	295,580 (25.37)	190,979 (16.39)
OUTCOME VARIABLE					
Patients With GYNC	8,316 (0.71)	4,578 (0.77)	1,109 (0.46)	1,340 (0.45)	1,963 (1.03)
Patients Without GYNC	1,156,745 (99.29)	588,491 (99.23)	237,459 (99.54)	294,240 (99.55)	189,016 (98.97)

Table 4.2 - Frequencies and Percentages of Demographic Characteristics of Patients with and without GYNC Among Women with AD Group "All", Florida HCUP- SID 2007-2013

Characteristics	AD N (%)	With GYNC N (%)	Without GYNC N (%)
	1,165,061 (100)	8,316 (0.71)	1,156,745 (99.3)
AGE (Years Old)			
1 (<25)	61,168 (5.25)	40 (0.48)	61,128 (5.28)
2 (25-45)	214,679 (18.43)	1,135 (13.65)	21,3544 (18.46)
3 (45-65)	390,296 (33.50)	3,642 (43.80)	386,654 (33.43)
4 (>65)	498,918 (42.82)	3,499 (42.08)	495,419 (42.83)
Race			
1 (White)	761,775(65.92)	5,391 (65.50)	756,384 (65.93)
2 (Black)	162,977 (14.10)	1,042 (12.66)	161,935 (14.11)
3 (Hispanic)	136,368 (11.80)	1,064 (12.93)	135,304 (11.79)
4 (Other)	94,407 (8.17)	734 (8.92)	93,673 (8.16)
Pay Source (Principle Payer)			
1 (Medicare)	620,955 (57.52)	3,902(50.79)	617,053 (57.57)
2 (Medicaid)	133,552 (12.37)	960 (12.50)	132,592 (12.37)
3 (Private Insurance)	238,520 (22.09)	2,156 (28.06)	236,364 (22.05)
4 (Self Pay)	428,10 (3.97)	247 (3.21)	42,563 (3.97)
5 (Other)	437,41 (4.05)	418 (5.44)	43,323 (4.04)
Median Household Income			
Quartile 1 (Q1)	362,197 (34.24)	2,480 (32.97)	359,717 (34.25)
Quartile 2 (Q2)	340,748 (32.21)	2,314 (30.76)	338,434 (32.22)
Quartile 3 (Q3)	254,673 (24.08)	1,886 (25.07)	252,787 (24.07)
Quartile 4 (Q4)	100,206 (9.47)	843 (11.21)	99,363 (9.46)
Length of Stay (Days)			
1 (< 1)	15,821 (1.36)	79 (0.95)	15,742 (1.36)
2 (1-10)	992,986 (85.23)	6,734 (80.98)	986,252 (85.26)
3 (11-20)	112,591 (9.66)	1,056 (12.70)	111,535 (9.64)
4 (21-30)	26,688 (2.29)	273 (3.28)	26,415 (2.28)
5 (31-40)	8,603 (0.74)	91 (1.09)	8,512 (0.74)
6 (>40)	8,372 (0.72)	83 (1.00)	8,289 (0.72)
Procedures*			
Hysterectomy	36,966 (3.17)	2,579 (31.01)	34,387 (2.97)
Vaginectomy	4,140 (0.36)	354 (4.26)	3,786 (0.33)
Vulvectomy	1,429 (0.12)	210 (2.53)	1,219 (0.11)

* The information for procedures in this table doesn't add to 100% due to not including the patients with none of the procedures done.

Table 4.2.1- Frequencies and Percentages of Demographic Characteristics of Patients with and without GYNC Among Women with AD Groups A-D, Florida HCUP-SID 2007-2013

AD Group	**A** **N (%)** 593,069 (6.18)		**B** **N (%)** 238,568 (2.49)		**C** **N (%)** 295,580 (3.08)		**D** **N (%)** 190,979 (1.99)	
Characteristics	**With GYNC**	**Without GYNC**	**With GYNC**	**Without GYNC**	**With GYNC**	**Without GYNC**	**With GYNC**	**Without GYNC**
Age (Years Old)								
1 (<25)	15 (0.33)	26,268 (4.46)	3 (0.27)	12,673 (5.34)	1 (0.07)	6,430 (2.19)	22 (1.12)	22,236 (11.76)
2 (25-45)	456 (9.96)	84,427 (14.35)	84 (7.57)	39,875 (16.79)	135 (10.07)	49,625 (16.87)	534 (27.20)	68,357 (36.16)
3 (45-65)	1,858 (40.59)	178,129 (30.27)	459 (41.39)	82,688 (34.82)	712 (53.13)	129,865 (44.14)	935 (47.63)	59,363 (31.41)
4 (>65)	2,249 (49.13)	299,667 (50.92)	563 (50.77)	102,223 (43.05)	492 (36.72)	108,320 (36.81)	472 (24.04)	39,060 (20.66)
Race								
1 (White)	2,982 (65.68)	376,399 (64.51)	723 (65.79)	146,356 (62.14)	1,036 (78.19)	223,844 (76.59)	1,128 (58.26)	111,302 (59.42)
2 (Black)	562 (12.38)	88,331 (15.14)	132 (12.01)	37,119 (15.76)	65 (4.91)	23,123 (7.91)	338 (17.46)	34,882 (18.62)
3 (Hispanic)	595 (13.11)	71,715 (12.29)	146 (13.28)	32,501 (13.80)	107 (8.08)	22,014 (7.53)	292 (15.08)	25,704 (13.72)
4 (Other)	401 (8.83)	47,020 (8.06)	98 (8.92)	19,568 (8.31)	117 (8.83)	23,291 (7.97)	178 (9.19)	15,433 (8.24)
Pay Source (Principle Payer)								
1 (Medicare)	2,373 (56.10)	345,717 (63.28)	62 (60.23)	128,381 (58.40)	704 (56.96)	164,549 (60.30)	550 (30.35)	62,196 (35.59)
2 (Medicaid)	550 (13.00)	62,056 (11.36)	126 (12.22)	28,824 (13.11)	124 (10.03)	28,141 (10.31)	235 (12.97)	31,069 (17.78)
3 (Private)	1,026 (24.26)	102,252 (18.72)	221 (21.44)	45,602 (20.74)	312 (25.24)	61,322 (22.47)	757 (41.78)	58,434 (33.44)
4 (Self Pay)	95 (2.25)	17,834 (3.29)	23 (2.23)	8,219 (3.74)	37 (2.99)	8,753 (3.21)	115 (6.35)	12,018 (6.88)
5 (Other)	186 (4.40)	18,436 (3.37)	40 (3.88)	8,806 (4.01)	59 (4.77)	10,126 (3.71)	155 (8.55)	11,023 (6.31)
Median Household Income								
Quartile 1 (Q1)	1,375 (33.22)	182,331 (34.05)	318 (31.27)	76,311 (35.43)	395 (32.56)	89,360 (33.41)	588 (33.22)	58,976 (34.50)
Quartile 2 (Q2)	1,253 (30.27)	171,775 (32.08)	328 (32.25)	68,402 (31.76)	368 (30.34)	90,141 (33.70)	543 (30.68)	53,945 (31.56)
Quartile 3 (Q3)	1,053 (25.44)	130,107 (24.30)	268 (26.35)	50,607 (23.50)	311 (25.64)	63,749 (23.83)	412 (23.28)	41,195 (24.10)
Quartile 4 (Q4)	458 (11.07)	51,288 (9.58)	103 (10.13)	20,057 (9.31)	139 (11.46)	24,210 (9.05)	227 (12.82)	16,805 (9.83)

Continue On next Page

Length Of Stay (Los)(Days)								
1 (< 1)	44 (0.96)	7,337 (1.25)	9 (0.81)	2,788 (1.17)	14 (1.04)	4,627 (1.57)	19 (0.97)	2,694 (1.43)
2 (1-10)	3,488 (76.19)	491,236 (83.47)	844 (76.10)	195,706 (82.42)	1,170 (87.31)	259,000 (88.02)	1,756 (89.45)	165,511 (87.56)
3 (11-20)	721 (15.75)	63,768 (10.84)	187 (16.86)	28,111 (11.84)	111 (8.28)	23,073 (7.84)	141 (7.18)	14,482 (7.66)
4 (21-30)	198 (4.33)	15,783 (2.68)	43 (3.88)	6,847 (2.880)	28 (2.09)	4,905 (1.67)	30 (1.53)	3,552 (1.88)
5 (31-40)	63 (1.38)	5,314 (0.90)	12 (1.08)	2,104 (0.89)	13 (0.97)	1,424 (0.48)	8 (0.41)	1,206 (0.64)
6 (>40)	64 (1.40)	5,053 (0.86)	14 (1.26)	1,903 (0.80)	4 (0.30)	12,11 (0.41)	9 (0.46)	1,571 (0.83)
Procedures*								
Hyster-ectomy	842 (18.39)	3,766 (0.64)	230 (20.74)	1,214 (0.51)	375 (27.99)	2,612 (0.89)	1,329 (67.70)	28,793 (15.23)
Vagin-ectomy	56 (1.22)	484 (0.08)	18 (1.62)	195 (0.08)	18 (1.34)	432 (0.15)	276 (14.06)	2,912 (1.54)
Vulv-ectomy	123 (2.69)	421 (0.07)	40 (3.61)	183 (0.08)	32 (2.39)	206 (0.07)	47 (2.39)	534 (0.28)

***** The information for procedures in this table doesn't add to 100% due to not including the information for the patients with none of the procedures done.

Table 4.3 - Frequencies and Percentages of Comorbidities of Patients with and without GYNC Among Women with AD Group "All", Florida HCUP-SID 2007-2013

Comorbidities	Total AD with CO* N (%)	With GYNC N (%)	Without GYNC N (%)
Congestive Heart Failure	191,632 (16.45)	681 (8.19)	190,951(16.51)
Cardiac Arrhythmia	250,981 (21.54)	1,431 (0.57)	249,550 (21.57)
Valvular Disease	117,758 (10.11)	486 (0.41)	117,272 (10.14)
Pulmonary Circulation Disorders	69,624 (5.98)	343 (0.49)	69,281 (5.99)
Peripheral Vascular Disorders	82,643 (7.09)	273 (0.33)	82,370 (7.12)
Hypertension Uncomplicated	487,250 (41.82)	3,472 (0.71)	483,778 (41.82)
Hypertension Complicated	185,160 (15.89)	863 (0.47)	184,297 (15.93)
Paralysis	17,293 (1.48)	76 (0.44)	17,217 (1.49)
Other Neurological Disorders	147,122 (12.63)	591 (7.11)	146,531 (12.67)
Chronic Pulmonary Disease	339,038 (29.10)	1,459 (17.54)	337,579 (29.18)
Diabetes Uncomplicated	264,501 (22.70)	1,650 (19.80)	262,851 (22.72)
Diabetes Complicated	111,249 (9.55)	399 (4.80)	110,850 (9.58)
Hypothyroidism	230,725 (19.80)	1,353(16.27)	229,372 (19.83)
Renal Failure	195,639 (16.79)	994 (11.95)	194,645 (16.83)
Liver Disease	116,927 (10.04)	729 (8.77)	116,198 (10.05)
Peptic Ulcer Disease Excluding Bleeding	17,072 (1.47)	71 (0.85)	17,001 (1.47)
AIDS/HIV	7,042 (0.60)	65(0.78)	6,977 (0.60)
Lymphoma	14,196 (1.22)	30 (0.36)	14,166 (1.22)
Coagulopathy	240,069 (20.61)	2,715 (32.65)	237,354 (20.52)
Obesity	168,114 (14.43)	1,308 (15.73)	166,806 (14.42)
Weight Loss	72,835 (6.25)	981 (11.80)	718,54(6.21)
Fluid and Electrolyte Disorders	394,742 (33.88)	3,117 (37.48)	391,625 (33.86)
Blood Loss Anemia	23,966 (2.06)	284 (3.42)	23,682 (2.05)
Deficiency Anemia	59,433 (5.10)	267 (3.21)	59,166 (5.11)
Alcohol Abuse	41,930 (3.60)	104 (1.25)	41,826 (3.62)
Drug Abuse	51,811 (4.45)	160 (1.92)	51,651 (4.47)
Psychoses	25,064 (2.15)	97 (1.17)	24,967 (2.16)
Depression	245,523 (21.07)	1,279 (15.38)	244,244 (21.11)

*Co=Comorbidity

Table 4.3.1- Frequencies and Percentages of Comorbidities of Patients with and without GYNC Among Women with AD Groups A-D, Florida HCUP-SID 2007-2013

AD Group	A N (%) 593,069 (6.18)		B N (%) 238,568 (2.49)		C N (%) 295,580 (3.08)		D N (%) 190,979 (1.99)	
Comorbidity	**With GYNC**	**Without GYNC**	**With GYNC**	**Without GYNC**	**With GYNC**	**Without GYNC**	**With GYNC**	**Without GYNC**
Congestive Heart Failure	428 (9.35)	108,222 (18.39)	172 (15.51)	55,908 (23.54)	77 (5.75)	34,274 (11.65)	70 (3.57)	17,886 (9.46)
Cardiac Arrhythmia	922 (20.14)	149,271 (25.37)	236 (21.28)	54,080 (22.77)	205 (15.30)	53,224 (18.09)	185 (9.42)	24,148 (12.78)
Valvular Disease	282 (6.16)	68,019 (11.56)	103 (9.29)	32,711 (13.78)	78 (5.82)	23,054 (7.84)	68 (3.46)	9,265 (4.90)
Pulmonary Circulation Disorders	235 (5.13)	42,969 (7.30)	61 (5.50)	18,961 (7.98)	49 (3.66)	12,715 (4.32)	27 (1.38)	5,459 (2.89)
Peripheral Vascular Disorders	158 (3.45)	45,345 (7.71)	66 (5.95)	20,551 (8.65)	50 (3.73)	16,824 (5.72)	27 (1.38)	9,828 (5.20)
Hypertension Uncomplicated	1,971 (43.05)	265,326 (45.09)	440 (39.68)	77,896 (32.80)	631 (47.09)	142,322 (48.37)	715 (36.42)	59,681 (31.57)
Hypertension Complicated	536 (11.71)	98,197 (16.69)	242 (21.82)	64,060 (26.98)	71 (5.30)	26,745 (9.09)	101 (5.15)	23,991 (12.69)
Paralysis	53 (1.16)	8,728 (1.48)	9 (0.81)	2,074 (0.87)	15 (1.12)	6,951 (2.36)	7 (0.36)	1,704 (0.90)
Other Neurological Disorders	319 (6.97)	63,896 (10.86)	65 (5.86)	22,106 (9.31)	218 (16.27)	66,883 (22.73)	46 (2.34)	15,897 (8.41)
Chronic Pulmonary Disease	812 (17.74)	183,584 (31.20)	240 (21.64)	68,205 (28.72)	337 (25.15)	98,863 (33.60)	225 (11.46)	37,109 (19.63)
Diabetes Uncomplicated	889 (19.42)	122,477 (20.81)	240 (21.64)	44,509 (18.74)	228 (17.01)	58,769 (19.97)	440 (22.41)	71,581 (37.87)
Diabetes Complicated	133 (2.91)	30,022 (5.10)	179 (16.14)	52,916 (22.28)	48 (3.58)	15,667 (5.32)	95 (4.84)	31,757 (16.80)
Hypothyroidism	806 (17.61)	122,491 (20.81)	199 (17.94)	44,359 (18.68)	273 (20.37)	65,854 (22.38)	236 (12.02)	32,364 (17.12)
Renal Failure	628 (13.72)	103,426 (17.57)	277 (24.98)	69,919 (29.44)	79 (5.90)	26,681 (9.07)	111 (5.65)	26,604 (14.07)
Liver Disease	260 (5.68)	50,632 (8.60)	503 (45.36)	71,757 (30.22)	38 (2.84)	13,003 (4.42)	56 (2.85)	11,190 (5.92)
Peptic Ulcer Disease, No Bleeding	44 (0.96)	8,573 (1.46)	17 (1.53)	4,327 (1.82)	12 (0.90)	4,482 (1.52)	6 (0.31)	2,344 (1.24)
AIDS/HIV	39 (0.85)	3,888 (0.66)	10 (0.90)	2,153 (0.91)	5 (0.37)	594 (0.20)	15 (0.76)	1,176 (0.62)
Lymphoma	18 (0.39)	10,517 (1.79)	6 (0.54)	2,067 (0.87)	2 (0.15)	2,282 (0.78)	6 (0.31)	1,147 (0.61)
Coagulopathy	2,618 (57.19)	224,490 (38.15)	179 (16.14)	31,008 (13.06)	73 (5.45)	10,821 (3.68)	65 (3.31)	7,966 (4.21)
	Continue on next page							

Obesity	579 (12.65)	80,376 (13.66)	199 (17.94)	35,703 (15.04)	212 (15.82)	53,823 (18.29)	438 (22.31)	22,429 (11.87)
Weight Loss	673 (14.70)	41,793 (7.10)	163 (14.70)	19,102 (8.04)	117 (8.73)	11,984 (4.07)	112 (5.71)	10,606 (5.61)
Fluid and Electrolyte Disorders	2098 (45.83)	211,418 (35.93)	485 (43.73)	93,812 (39.51)	404 (30.15)	80,840 (27.47)	411 (20.94)	64,031 (33.88)
Blood Loss Anemia	152 (3.32)	12,524 (2.13)	47 (4.24)	6,762 (2.85)	40 (2.99)	4,002 (1.36)	66 (3.36)	3,596 (1.90)
Deficiency Anemia	179 (3.91)	35,248 (5.99)	43 (3.88)	13,905 (5.86)	30 (2.24)	12,183 (4.14)	39 (1.99)	6,983 (3.69)
Alcohol Abuse	64 (1.40)	22,378 (3.80)	20 (1.80)	6,731 (2.83)	13 (0.97)	7,840 (2.66)	18 (0.92)	9,781 (5.17)
Drug Abuse	87 (1.90)	20,956 (3.56)	20 (1.80)	9,276 (3.91)	41 (3.06)	19,624 (6.67)	34 (1.73)	10,216 (5.40)
Psychoses	59 (1.29)	11,913 (2.02)	10 (0.90)	4,454 (1.88)	21 (1.57)	8,119 (2.76)	16 (0.82)	3,540 (1.87)
Depression	693 (15.14)	108,476 (18.43)	163 (14.70)	43,176 (18.18)	375 (27.99)	98,867 (33.60)	199 (10.14)	33,583 (17.77)

Table 4.4 - Unadjusted Association Between Patient Characteristics and GYNC Among Women with AD Group "All", Florida HCUP-SID 2007-2013

Characteristics	OR	95% CI
Age (Year)		
1 (<25)	0.12	0.09-0.17
2 (25-45)	1.00	Referent
3 (45-65)	1.77	1.66-1.90
4 (>65)	1.33	1.24-1.42
Race		
1 (White)	1.00	Referent
2 (Black)	0.90	0.84-0.96
3 (Hispanic)	1.10	1.03-1.18
4 (Other)	1.10	1.02-1.20
Pay (Insurance Type)		
1 (Medicare)	1.00	Referent
2 (Medicaid)	1.14	1.07-1.23
3 (Private)	1.44	1.37-1.52
4 (Self Pay)	0.92	0.81-1.04
5 (Other)	1.53	1.40-1.69
Median Household Income Level		
Quartile 1 (Q1)	0.81	0.75-0.88
Quartile 2 (Q2)	0.81	0.74-0.87
Quartile 3 (Q3)	0.88	0.81-0.95
Quartile 4 (Q4)	1.00	Referent
Length of Stay (LOS)(Days)		
1 (< 1)	0.50	0.38 - 0.62
2 (1-10)	0.66	0.58—0.75
3 (11-20)	0.92	0.80-1.05
4 (21-30)	1.00	Referent
5 (31-40)	1.03	0.82-1.31
6 (>40)	1.00	0.76-1.24
Procedures		
Hysterectomy (Yes vs. No)	14.68	13.99-15.40
Vaginectomy (Yes vs. No)	13.54	12.12-15.13
Vulvectomy (Yes vs. No)	24.56	21.18-28.48

OR, Odd Ratios; CI, Confidence Interval

Table 4.4.1 - Unadjusted Association Between Patient Characteristics and GYNC Among Women with AD Groups A-D, Florida HCUP-SID 2007-2013

AD Group	**A** **N (%)** 593,069 (6.18)		**B** **N (%)** 238568 (2.49)		**C** **N (%)** 295,580 (3.08)		**D** **N (%)** 190979(1.99)	
Characteristics	**OR**	**CI**	**OR**	**CI**	**OR**	**CI**	**OR**	**CI**
Age (Years Old)								
1 (<25)	0.11	0.06-0.18	0.11	0.04-0.36	0.06	0.01-0.41	0.13	0.08-0.19
2 (25-45)	1.00	Ref	1.00	Ref	1.00	Ref	1.00	Ref
3 (45-65)	1.93	1.74-2.14	2.64	2.09-3.33	2.02	1.68-2.42	2.02	1.81-2.24
4 (>65)	1.39	1.25-1.54	2.61	2.08-3.29	1.67	1.38-2.02	1.55	1.37-1.75
Race								
1 (White)	1.00	Ref	1.00	Ref	1.00	Ref	1.00	Ref
2 (Black)	0.80	0.73-0.88	0.72	0.60-0.87	0.61	0.47-0.78	0.96	0.85-1.08
3 (Hispanic)	1.05	0.96-1.14	0.91	0.76-1.09	1.05	0.86-1.28	1.12	0.98-1.28
4 (Other)	1.08	0.97-1.19	1.01	0.82-1.25	1.08	0.90-1.32	1.14	0.97-1.33
Pay Source (Principle Payer)								
1 (Medicare)	1.00	Ref	1.00	Ref	1.00	Ref	1.00	Ref
2 (Medicaid)	1.29	1.18-1.42	0.90	0.75-1.09	1.03	0.85-1.25	0.85	0.73-1.0
3 (Private)	1.46	1.36-1.57	1.00	0.86-1.17	1.19	1.04-1.36	1.47	1.31-1.64
4 (Self-Pay)	0.78	0.63-0.95	0.58	0.38-0.88	0.99	0.71-1.38	1.08	0.88-1.32
5 (Other)	1.47	1.26-1.71	0.94	0.68-1.29	1.36	1.04-1.78	1.59	1.33-1.90
Median Household Income								
Quartile 1 (Q1)	0.84	0.76-0.94	0.81	0.65-1.01	0.77	0.63-0.93	0.74	0.63-0.86
Quartile 2 (Q2)	0.82	0.73-0.91	0.93	0.75-1.16	0.71	0.58-0.86	0.75	0.64-0.87
Quartile 3 (Q3)	0.91	0.81-1.01	1.03	0.82-1.29	0.85	0.69-1.04	0.74	0.63-0.87
Quartile 4 (Q4)	1.00	Ref	1.00	Ref	1.00	Ref	1.00	Ref
Length of Stay (LOS)(Days)								
1 (< 1)	0.48	0.34-0.66	0.51	0.25-1.05	0.53	0.28-1.00	0.84	0.47-1.49
2 (1-10)	0.56	0.49-0.65	0.69	0.51-0.93	0.79	0.54-1.15	1.26	0.87-1.81
3 (11-20)	0.90	0.77-1.05	1.06	0.76-1.48	0.84	0.55-1.28	1.15	0.78-1.71
4 (21-30)	1.00	Ref	1.00	Ref	1.00	Ref	1.00	Ref
5 (31-40)	0.94	0.71-1.26	0.91	0.48-1.73	1.60	0.83-3.10	0.78	0.36-1.72
6 (>40)	1.01	0.76-1.34	1.17	0.64-2.15	0.58	0.20-1.65	0.68	0.32-1.43
Procedures								
Hysterectomy (Yes Vs. No)	34.99	32.3-38.0	50.92	43.57-59.50	43.38	38.3-49.2	11.67	10.6-12.8
Vaginectomy (Yes Vs. No)	15.04	11.4-19.9	20.08	12.4-32.7	9.26	5.76-14.9	10.46	9.16-11.9
Vulvectomy (Yes Vs. No)	38.56	31.5-47.2	48.52	34.3-68.7	34.92	24.0-50.9	8.66	6.40-11.7

OR, Odd Ratios; CI, Confidence Interval; Ref, Referent

Table 4.5 - Unadjusted Association Between Patient Comorbidities and GYNC Among Women with AD Group "All", Florida HCUP-SID 2007-2013

Variable Category	OR	95% CI
Congestive Heart Failure	0.45	0.42-0.49
Cardiac Arrhythmia	0.76	0.71-0.80
Valvular Disease	0.55	0.50-0.60
Pulmonary Circulation Disorders	0.68	0.61-0.75
Peripheral Vascular Disorders	0.44	0.40-0.50
Hypertension Uncomplicated	1.00	0.95-1.04
Hypertension Complicated	0.61	0.57-0.66
Paralysis	0.61	0.49-0.77
Other Neurological Disorders	0.53	0.48-0.68
Chronic Pulmonary Disease	0.52	0.49-0.55
Diabetes Uncomplicated	0.84	0.80-0.89
Diabetes Complicated	0.50	0.43-0.53
Hypothyroidism	0.79	0.74-0.83
Renal Failure	0.67	0.63-0.72
Liver Disease	0.86	0.80-0.93
Peptic Ulcer Disease, no Bleeding	0.60	0.46-0.73
AIDS/HIV	1.30	1.02-1.70
Lymphoma	0.29	0.20-0.42
Coagulopathy	1.90	1.80-1.97
Obesity	1.11	1.04-1.17
Weight Loss	2.02	1.89-2.16
Fluid and Electrolyte Disorders	1.17	1.12-1.22
Blood Loss Anemia	1.69	1.50-1.90
Deficiency Anemia	0.61	0.54-0.69
Alcohol Abuse	0.34	0.28-0.41
Drug Abuse	0.42	0.36-0.49
Psychoses	0.54	0.44-0.65
Depression	0.68	0.64-0.72

OR, Odd Ratios; CI, Confidence Interval

Table 4.5.1 - Unadjusted Association Between Patient Comorbidities and GYNC Among Women with AD. Groups A-D, Florida HCUP-SID 2007-2013

AD Group	**A** **N (%)** 593,069 (6.18)		**B** **N (%)** 238568 (2.49)		**C** **N** (%) 295,580 (3.08)		**D** **N (%)** 190979 (1.99)	
VARIABLE	**OR**	**CI**	**OR**	**CI**	**OR**	**CI**	**OR**	**CI**
Congestive Heart Failure	0.46	0.41-0.51	0.60	0.51-0.70	0.46	0.37-0.58	0.35	0.28-0.45
Cardiac Arrhythmia	0.74	0.69-0.80	0.92	0.79-1.06	0.82	0.71-0.95	0.71	0.61-0.83
Valvular Disease	0.50	0.44-0.57	0.64	0.52-0.78	0.73	0.58-0.92	0.70	0.55-0.89
Pulmonary Circulation Disorders	0.69	0.60-0.78	0.67	0.52-0.87	0.84	0.63-1.12	0.45	0.32-0.69
Peripheral Vascular Disorders	0.43	0.36-0.50	0.67	0.52-0.86	0.64	0.48-0.85	0.25	0.17-0.37
Hypertension Uncomplicated	0.92	0.87-0.98	1.35	1.20-1.52	0.95	0.85-1.06	1.24	1.13-1.36
Hypertension Complicated	0.66	0.60-0.72	0.75	0.65-0.87	0.56	0.44-0.71	0.37	0.31-0.46
Paralysis	0.78	0.59-1.02	0.93	0.48-1.79	0.47	0.28-0.78	0.39	0.19-0.83
Other Neurological Disorders	0.62	0.55-0.69	0.61	0.47-0.78	0.66	0.57-0.76	0.26	0.19-0.35
Chronic Pulmonary Disease	0.48	0.44-0.51	0.68	0.59-0.79	0.66	0.59-0.75	0.53	0.46-0.61
Diabetes Uncomplicated	0.92	0.85-0.99	1.20	1.04-1.38	0.82	0.71-0.95	0.47	0.43-0.53
Diabetes Complicated	0.56	0.47-0.66	0.67	0.57-0.79	0.66	0.49-0.88	0.25	021-0.31
Hypothyroidism	0.81	0.75-0.88	0.95	0.82-1.11	0.89	0.77-1.01	0.66	0.58-0.76
Renal Failure	0.75	0.69-0.81	0.80	0.70-0.91	0.63	0.50-0.79	0.37	0.30-0.44
Liver Disease	0.64	0.56-0.72	1.92	1.70-2.16	0.63	0.46-0.87	0.47	0.36-0.61
Peptic Ulcer Disease, No Bleeding	0.66	0.49-0.88	0.84	0.52-1.36	0.58	0.33-1.03	0.24	0.11-0.55
AIDS/HIV	1.29	0.94-1.77	0.99	0.53-1.86	1.85	0.77-4.47	1.23	0.74-2.05
Lymphoma	0.22	0.14-0.35	0.62	0.28-1.38	0.19	0.05-0.77	0.50	0.22-1.12
Coagulopathy	2.17	2.04-2.30	1.28	1.01-1.50	1.51	1.20-1.91	0.78	0.61-1.00
Obesity	0.92	0.84-1.00	1.24	1.06-1.44	0.84	0.73-0.97	2.13	1.92-2.40
Weight Loss	2.25	2.08-2.45	1.97	1.67-2.33	2.25	1.86-2.73	1.02	0.84-1.23
Fluid and Electrolyte Disorders	1.51	1.42-1.60	1.19	1.06-1.34	1.14	1.01-1.28	0.52	0.46-0.58
Blood Loss Anemia	1.58	1.34-1.86	1.51	1.13-2.02	2.23	1.63-3.06	1.80	1.40-2.30
Deficiency Anemia	0.64	0.55-0.74	0.65	0.48-0.88	0.53	0.37-0.76	0.53	0.39-0.73
Alcohol Abuse	0.36	0.28-0.46	0.63	0.40-0.98	0.36	0.21-0.62	0.17	0.11-0.27
Drug Abuse	0.52	0.42-0.65	0.45	0.29-0.70	0.44	0.32-0.60	0.31	0.22-0.43
Psychoses	0.63	0.49-0.82	0.48	0.26-0.89	0.56	0.36-0.86	0.43	0.26-0.71
Depression	0.79	0.73-0.86	0.78	0.66-0.92	0.77	0.68-0.87	0.52	0.45-0.61

OR, Odd Ratios; CI, Confidence Interval

Table 4.6 - Adjusted ORs and 95% CIs of the Association Between Selected Demographic Variables and GYNC Among Women with AD Groups "All" using Patient Characteristics as Predictor Variables, Florida HCUP-SID 2007-2013

Variable	OR*	95% CI
Age (Year)		
1 (<25)	0.23	0.16-0.33
2 (25-45)	1.00	Referent
3 (45-65)	0.90	0.82-0.99
4 (>65)	0.79	0.70-0.88
Pay (Insurance Type)		
1 (Medicare)	1.00	Referent
2 (Medicaid)	1.36	1.23-1.51
3 (Private)	0.90	0.82-0.98
4 (Self Pay)	1.39	1.15-1.67
5 (Other)	1.22	1.06-1.40
Procedures		
Hysterectomy (Yes vs No)	41.38	37.40-45.78
Vaginectomy (Yes vs No)	1.82	1.47-2.25
Vulvectomy (Yes vs No)	15.70	12.60-19.54

OR, Odd Ratios; CI; Confidence Interval;
* ORs adjusted for all other variables in table

Table 4.6.1 - Adjusted ORs and 95% CIs of the Association Between Selected Demographic Variables and GYNC Among Women with AD Groups A-D using Patient Characteristics as Predictor Variables, Florida HCUP-SID 2007-2013

AD Group	A N (%) 593,069 (6.18)		B N (%) 238568 (2.49)		C N (%) 295,580 (3.08)		D N (%) 190979 (1.99)	
VARIABLE	**OR***	**CI**	**OR**	**CI**	**OR**	**CI**	**OR**	**CI**
Procedures								
Hysterectomy (Yes Vs. No)	44.61	37.6-53.0	44.6	33.5-59.3	67.87	51.6-89.3	35.43	29.3-42.80
Vaginectomy (Yes Vs. No)	1.60	0.94-2.69	---	---	---	---	2.21	1.73-2.82
Vulvectomy (Yes Vs. No)	39.00	27.0-56.3	89.8	45.6-177	34.37	19.1-61.9	4.00	2.56-6.19

OR, Odd Ratios; CI, Confidence Interval
* ORs adjusted for all other variables in table

Table 4.7 - Adjusted ORs and 95% CIs of the Association Between Selected Demographic Variables and GYNC Among Women with AD Group "All" using Comorbidities as Predictor Variables, Florida HCUP-SID 2007-2013

Variable	OR*	95% CI
Congestive Heart Failure	0.91	0.82-1.00
Cardiac Arrhythmia	0.91	0.84-0.98
Pulmonary Circulation Disorders	1.12	0.98-1.28
Peripheral Vascular Disorders	0.69	0.59-0.80
Hypertension Uncomplicated	0.94	0.89-1.00
Hypertension Complicated	0.86	0.73-1.02
Paralysis	0.66	0.50-0.86
Other Neurological Disorders	0.73	0.66-0.81
Chronic Pulmonary Disease	0.51	0.48-0.55
Hypothyroidism	0.90	0.84-0.98
Renal Failure	1.48	1.26-1.72
Liver Disease	0.64	0.58-0.71
AIDS/HIV	1.97	1.42-2.74
Lymphoma	0.34	0.22-0.53
Coagulopathy	1.26	1.18-1.34
Obesity	1.49	1.36-1.62
Weight Loss	1.12	1.02-1.21
Fluid and Electrolyte Disorders	1.11	1.04-1.18
Deficiency Anemia	0.80	0.69-0.93
Alcohol Abuse	0.71	0.56-0.90
Drug Abuse	1.22	1.01-1.48
Depression	1.06	0.98-1.14

OR, Odd Ratios; CI, Confidence Interval

* ORs adjusted for all other variables in table

Table 4.7.1 - Adjusted ORs and 95% CIs of the Association Between Selected Demographic Variables and GYNC Among Women with AD Groups A-D using Comorbidities as Predictor Variables, Florida HCUP-SID 2007-2013

AD Group	**A** **N (%)** 593,069 (6.18)		**B** **N (%)** 238,568 (2.49)		**C** **N (%)** 295,580 (3.08)		**D** **N (%)** 190,979 (1.99)	
Variable	**OR***	**CI**	**OR**	**CI**	**OR**	**CI**	**OR**	**CI**
Congestive Heart Failure	0.88	0.77-0.99	---	---	0.75	0.56-1.00	---	---
Cardiac Arrhythmia	0.86	0.79-0.95	1.23	1.03-1.47	---	---	0.81	0.64-1.02
Pulmonary Circulation Disorders	---	---	1.24	0.91-1.69	---	---	---	---
Peripheral Vascular Disorders	0.67	0.55-0.81	---	---	0.76	0.53-1.10	0.47	0.28-0.79
Hypertension Uncomplicated	0.87	0.81-0.95	---	---	---	---	---	---
Paralysis	0.62	0.45-0.85	---	---	---	---	0.27	0.08-0.86
Other Neurological Disorders	0.70	0.61-0.81	0.74	0.55-0.99	0.76	0.63-0.92	0.60	0.41-0.88
Chronic Pulmonary Disease	0.43	0.40-0.48	0.72	0.60-0.86	0.56	0.47-0.65	0.70	0.58-0.86
Diabetes Uncomplicated	---	---	0.88	0.74-1.05	---	---	1.14	0.97-1.36
Diabetes Complicated	---	---	---	---	1.46	1.02-2.08	1.40	1.05-1.85
Hypothyroidism	0.93	0.85-1.03	---	---	0.85	0.72-1.02	0.86	0.71-1.04
Renal Failure	1.42	1.17-1.72	1.70	1.43-2.02	---	---	1.23	0.93-1.61
Liver Disease	0.52	0.45-0.61	1.15	0.99-1.34	0.51	0.34-0.76	---	---
Peptic Ulcer Disease, No Bleeding		---		---	---	---	0.28	0.09-0.83
AIDS/HIV	1.56	1.01-2.40	---	---	4.82	1.39-16.8	6.35	3.05-13.21
Lymphoma	0.29	0.17-0.51	---	---	0.06	0.01-0.50	---	---
Coagulopathy	1.25	1.20-1.35	0.81	0.67-0.98	---	---	0.72	0.51-1.01
Obesity	1.47	1.31-1.66	1.61	1.32-1.97	1.22	1.0-1.50	1.51	1.25-1.82
Weight Loss	1.10	0.99-1.23	---	---	1.31	1.02-1.68	1.21	0.92-1.58
FLUID and Electrolyte Disorders	1.14	1.05-1.23	---	---	1.18	1.01-1.36	1.16	0.49-1.39
Blood Loss Anemia	---	---	1.37	0.96-1.97	---	---	---	---
Deficiency Anemia	0.82	0.69-0.99	0.70	0.47-1.02	0.54	0.34-0.86	---	---
Alcohol Abuse	0.79	0.59-1.07	---	---	0.64	0.33-1.2	0.52	0.28-0.93
Drug Abuse	1.29	0.99-1.69	---	---	---	---	1.54	0.99-2.39
Depression	---	---	---	---	1.11	0.95-1.29	1.18	0.96-1.44

OR, Odd Ratios; CI, Confidence Interval

* ORs adjusted for all other variables in table

Table 4.8 - Predictive values of GYNC Among Women with AD Groups "All", "A", "B", "C", "D", where Ratio of GYNC (+) to GYNC (-) is 1:1

Organ & System Affected	AD Group Subsets	Sample # (N)	Wrongly Predicted (WPV)	Overall Accurate Prediction	(Sensitivity) Accurate Prediction For GYNC (+)	(Specificity) Accurate Prediction For GYNC (-)	False Positive Rate (FPR)	AUC of Test Data
All	"All"	16,600	0.27	0.73	0.61	0.84	0.16	0.80
Systematic & Vascular; Rheumatoid; Dermatologic; Sensory	"A"	9,140	0.31	0.69	0.65	0.73	0.27	0.76
Major Organs; Digestive	"B"	2,200	0.40	0.60	0.48	0.73	0.27	0.66
Neuro-Muscular	"C"	2.660	0.47	0.53	0.71	0.37	0.63	0.53
Glandular	"D"	3,920	0.24	0.76	0.68	0.82	0.18	0.81

Confusion Matrix

Training

Actual	Predicted Count	
GYN cancer	0	1
0	2763	873
1	1364	2323

Validation

Actual	Predicted Count	
GYN cancer	0	1
0	364	113
1	188	267

Test

Actual	Predicted Count	
GYN cancer	0	1
0	334	123
1	151	277

Figure 4.13: Confusion Matrix for AD Group "A" Subset

Column Contributions

Term	Number of Splits	G^2	Portion
Hysterectomy	5090	240.407655	0.4994
Coagulopathy	8069	82.0274343	0.1704
Chronic Pulmonary Disease	6875	51.5541609	0.1071
Age	6399	24.5013952	0.0509
Congestive Heart Failure	3695	18.9987918	0.0395
Fluid and Electrolyte Disorders	5103	12.9759241	0.0270
Weight Loss	2483	12.9356717	0.0269
Pay Source	4579	9.02243553	0.0187
Hospital Length of Stay	2138	4.84736513	0.0101
Valvular Disease	1143	4.52762787	0.0094
Hypertension Complicated	1708	2.80546225	0.0058
Peripheral Vascular Disorders	579	2.69644176	0.0056
Cardiac Arrhythmia	2629	2.52398333	0.0052
Liver Disease	801	2.42310749	0.0050
Hypertension Uncomplicated	3837	2.35067039	0.0049
Renal Failure	1364	1.50231601	0.0031
Race	2199	1.1929236	0.0025
Other Neurological Disorders	639	1.15720707	0.0024
Depression	1436	0.88665817	0.0018
Household Income	2814	0.78830552	0.0016
Hypothyroidism	1420	0.55507329	0.0012
Diabetes Uncomplicated	1465	0.49556675	0.0010
Obesity	455	0.11438122	0.0002
Pulmonary Circulation Disorders	118	0.07990538	0.0002
Deficiency Anemia	27	0.03482737	0.0001
Vaginectomy	0	0	0.0000
Vulvectomy	0	0	0.0000

Figure 4.14: Column Contribution for AD Group "A" Subset, the Measurement of the Relationship Between the Predictor Variables and GYNC

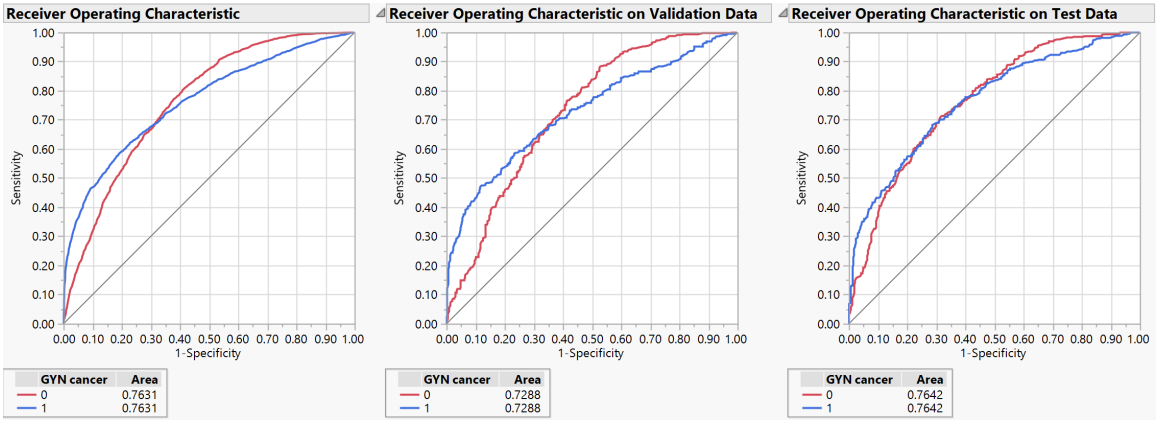

Figure 4.15: The Receiver Operating Characteristic for AD Group "A" Subset, the Measurement of the Quality of Predictors of GYNC

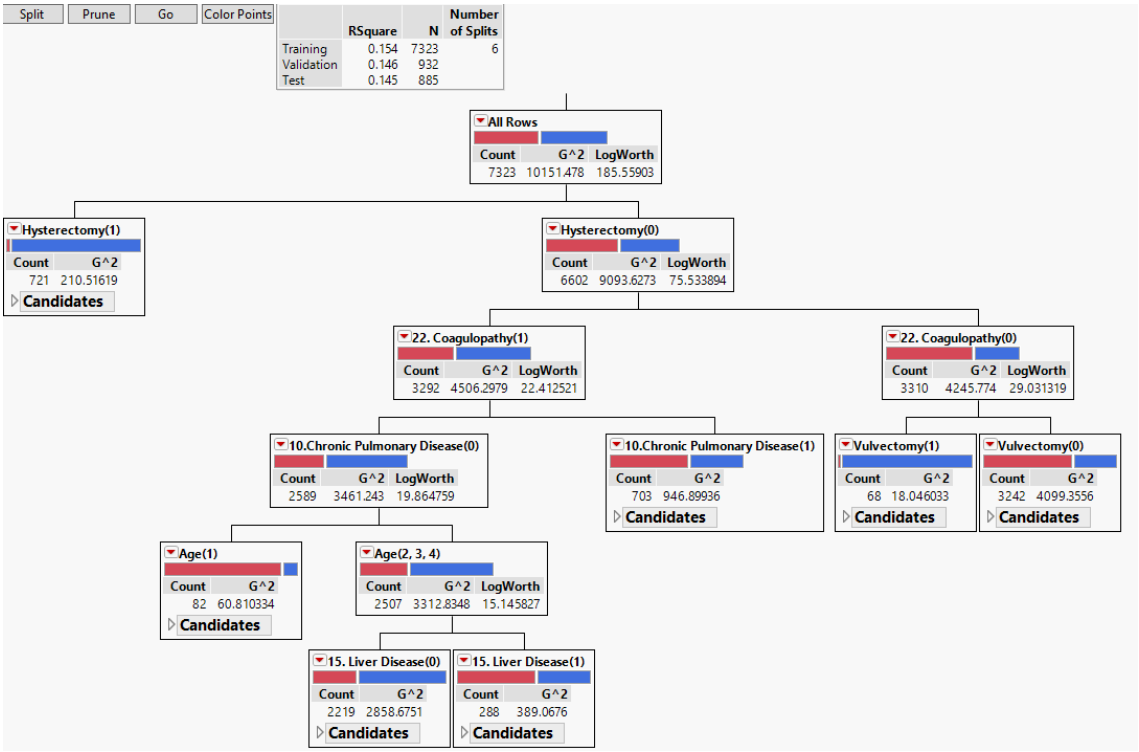

Figure 4.16: The Decision Tree Output for AD Group "A" Subset

Confusion Matrix

Training

Actual	Predicted Count	
GYN Cancer	0	1
0	629	269
1	469	399

Validation

Actual	Predicted Count	
GYN Cancer	0	1
0	67	35
1	64	53

Test

Actual	Predicted Count	
GYN Cancer	0	1
0	73	27
1	60	55

Figure 4.17: Confusion Matrix for AD Group "B" Subset

Column Contributions

Term	Number of Splits	G^2	Portion
Liver Disease	3273	7.08770875	0.4932
Hypertension Complicated	1404	1.16662265	0.0812
Age	1792	1.15183781	0.0802
Congestive Heart Failure	782	1.02187939	0.0711
Hypertension Uncomplicated	1690	0.85078757	0.0592
Household Income	1506	0.69252366	0.0482
Chronic Pulmonary Disease	911	0.58872188	0.0410
Renal Failure	985	0.46936933	0.0327
Diabetes Complicated	506	0.44513378	0.0310
Race	654	0.30625462	0.0213
Pay Source	950	0.27712207	0.0193
Fluid and Electrolyte Disorders	1259	0.16895048	0.0118
Hospital Length of Stay	109	0.1178215	0.0082
Diabetes Uncomplicated	170	0.01452267	0.0010
Cardiac Arrhythmia	144	0.01020268	0.0007
Hypothyroidism	2	0.00005413	0.0000
Hysterectomy	0	0	0.0000
Vaginectomy	0	0	0.0000
Vulvectomy	0	0	0.0000

Figure 4.18: Column Contribution for AD Group" B" Subset, the Measurement of the Quality of Predictors of GYNC

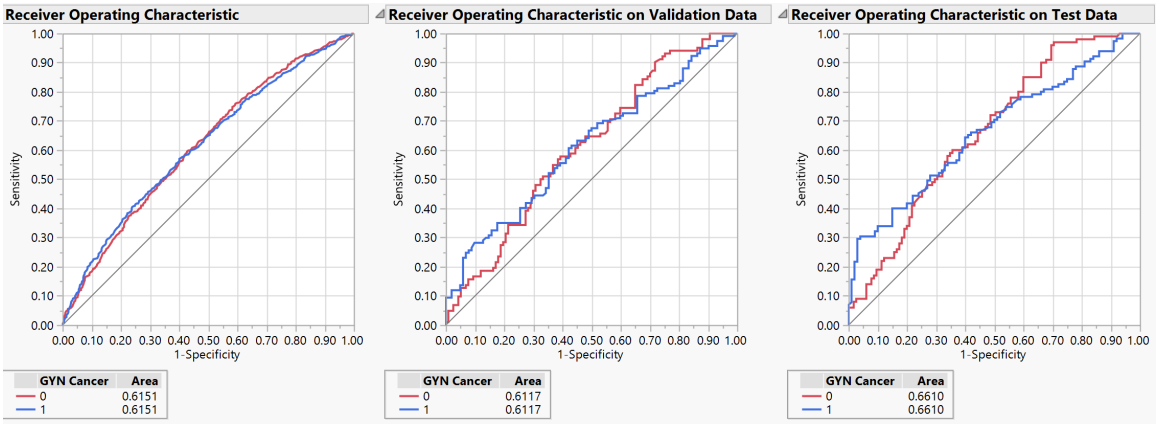

Figure 4.19: The Receiver Operating Characteristic for AD Group "B" Subset, the Measurement of the Quality of Predictors of GYNC

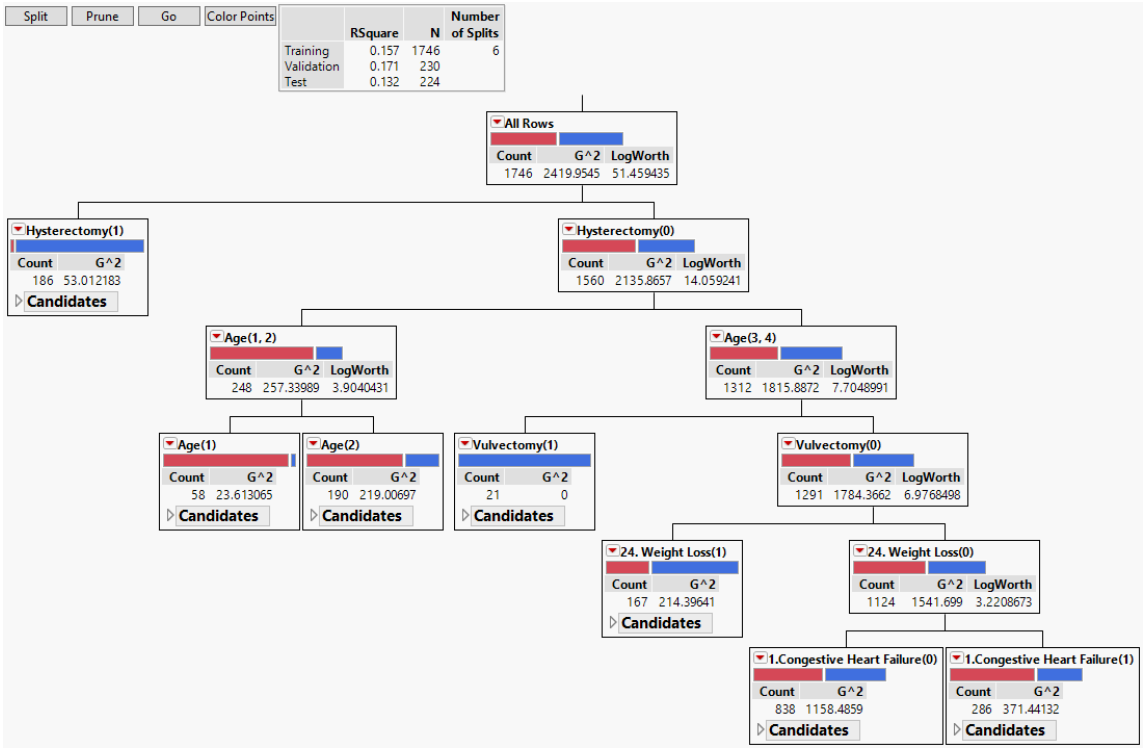

Figure 4.20: The Decision Tree Output for AD Group "B" Subset

Confusion Matrix

Training

Actual	Predicted Count	
GYN Cancer	0	1
0	363	673
1	255	824

Validation

Actual	Predicted Count	
GYN Cancer	0	1
0	60	87
1	19	96

Test

Actual	Predicted Count	
GYN Cancer	0	1
0	54	93
1	40	96

Figure 4.21: Confusion Matrix for AD Group "C" Subset

Column Contributions

Term	Number of Splits	G^2	Portion
Chronic Pulmonary Disease	2896	3.49153301	0.3590
Depression	2343	1.62537971	0.1671
Hysterectomy	45	1.32954772	0.1367
Household Income	1942	0.84204382	0.0866
Other Neurological Disorders	1224	0.66767739	0.0686
Diabetes Uncomplicated	1021	0.47572703	0.0489
Pay	1394	0.41518526	0.0427
Fluid and Electrolyte Disorders	1237	0.31469566	0.0324
Age	1219	0.23403349	0.0241
Hypertension Uncomplicated	1479	0.18215114	0.0187
Cardiac Arrhythmia	243	0.05884615	0.0061
Race	184	0.05800036	0.0060
Hypothyroidism	339	0.01843531	0.0019
Obesity	154	0.01323791	0.0014
Hospital Length of Stay	0	0	0.0000
Vaginectomy	0	0	0.0000

Figure 4.22: Column Contribution for AD Group "C" Subset, the Measurement of the Relationship Between the Predictor Variables and GYNC

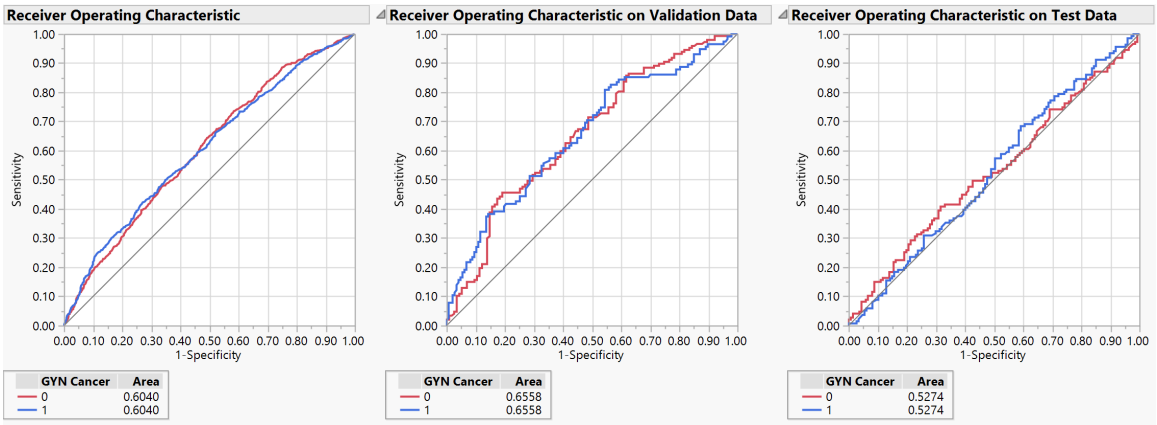

Figure 4.23: The Receiver Operating Characteristic for AD Group "C" Subset, the Measurement of the Quality of Predictors of GYNC

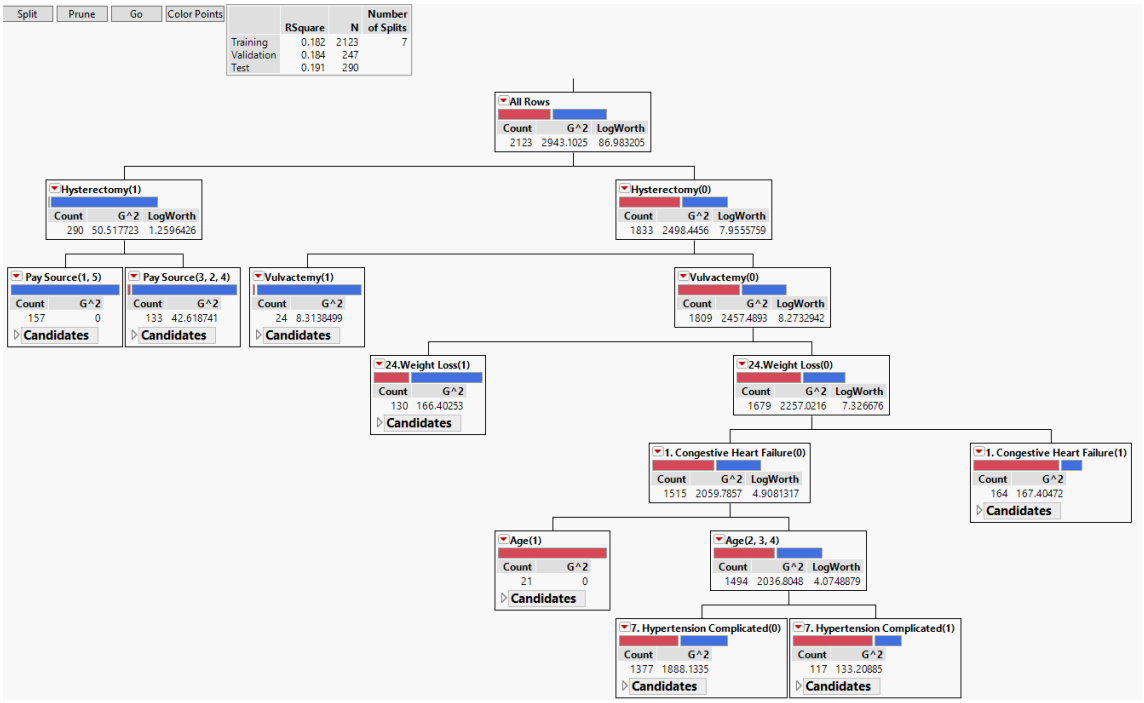

Figure 4.24: The Decision Tree Output for AD Group “C” Subset

CHAPTER 5: DISCUSSION

5.1 Overall Findings

This dissertation research was conducted to investigate the association between patient-level predictors and GYNC among U.S. women with AD. Based on previous studies, and knowing that cancer and ADs are both the accumulative effect of genetics and environmental exposures, and the fact that[24-27] 90-95% of all cancers are linked to environmental factors[27] and lifestyle,[101] this investigation was imperative to recognize the profile of comorbidities and how they relate to patients' characteristics. The underlying goal of this study was to uncover novel GYNC risk factors.

In this study of women with AD, older women had decreased odds of having GYNC. Another notifiable finding was that Medicaid and self-pay holders had increased odds of GYNC, whereas private insurance had decreased odds of GYNC among patients with AD. In addition, GYN related procedures, and in particular hysterectomy, had strong associations with GYNC among women with AD. With regards to comorbidities, AIDS/HIV, coagulopathy, weight loss, fluid and electrolyte disorders, renal failure, and obesity were found to have strong associations with GYNC among women with AD.

Classification Tree Analysis indicated that the variables Hysterectomy and Age are strong predictors of GYNC among subgroups of patients with autoimmune arthritis like, dermatologic, systemic & vascular disorders (group A), and patients with autoimmune glandular disorders (group D). Thus, the findings indicate that the subpopulation of women with autoimmune arthritis like, dermatologic, systemic & vascular, disorders (group A) who had a history of hysterectomy, and had comorbidities such as coagulopathy, and chronic pulmonary disorders have increased odds of getting

GYNC. Similarly, the subpopulation of women with autoimmune glandular disorders (group D) who had a hysterectomy, and had comorbidities such as diabetes and fluid and electrolyte disorders had increased odds of GYNC.

With regards to patient characteristics, one of the hypotheses of this study was that there would be an association between age and GYNC. Although the study findings showed an association between age and GYNC, it was not in the direction anticipated. In this study, older age was associated with decreased odds of GYNC among women with AD. The expectation for an association whereby older age would be associated with increased odds of GYNC was based on the findings in the literature where older age is a risk factor for most GYNC types among women.[1,102-104]

Although prior literature demonstrates that older age is a risk factor for GYNC,[1,104] in the current study the existence of autoimmunity among the study population may be the reason for the difference in our finding in comparison with prior findings. Other studies also have shown that the process of aging is the strongest risk factor for development of many cancers.[4,103] While the underlying mechanism for this association is not clearly known, an epigenetic study conducted by Xu et al. demonstrated that accumulation of age-associated changes in biochemical processes of DNA that control gene activities, may be responsible for age-associated risk factors.[104] This biochemical process is called DNA methylation which is the binding of chemical tags (methyl groups) onto DNA molecules. In this process, DNA methylation activates or silences some genes and interferes with the cell protein synthesis machinery, resulting in tumorigenesis.[104]

One explanation for our study findings (i.e. older age being associated with decreased odds of GYNC) being different with previous evidence in the literature may be the existence of autoimmunity among GYNC subjects of the current study, in which autoimmunity might have interfered with normal maintenance of DNA methylation. The biological reason for the decreased association of age with AD is demonstrated by previous studies, where it is shown that DNA methylation also plays an important role in normal function of immune system, and failure to maintain DNA methylation levels can result in autoimmunity condition in vivo.[105]

With regards to race/ethnicity, I further hypothesized that there would be an association between race/ethnicity and GYNC among women with AD. However, this hypothesis was not supported by the study findings. The expectation was based on evidence in the literature that indicates that Non-Hispanic Black women have higher cervical and ovarian cancer incidence rates[1,102,106] and higher cervical and breast cancer mortality rates than Non-Hispanic White women.[1] However, the difference in findings of this study might have been due to the relatively small sample size of women with GYNC among the AD population. Based on the literature, the prevalence of different ADs is different among races/ethnicities.[10-13]

Another study hypothesis under the first objective was that the type of insurance AD patients hold, could have an association with GYNC. This hypothesis was supported by the study findings. Medicaid, private insurance holders, and self-pay patients were compared with Medicare beneficiaries. Private insurance holders had decreased odds of GYNC compared to Medicare beneficiaries, whereas self-pay patients and Medicaid holders had increased odds of GYNC among the AD population. One explanation for the

increased association of Medicaid and self-pay with GYNC can be due to the global association of cancer with disparities and lack of access to treatment and preventive care.[1, 45,57] Although Medicaid provides a level of access to a wide variety of individuals with need, patients may still be less likely to get the treatment needed. Furthermore, patients without any insurance (i.e. self-pay) are also unlikely to get timely treatment. Roetzheim and colleagues in a cancer related study among Florida patients, demonstrated that Patients insured by Medicaid, and those who were uninsured were at greater risk of late stage diagnosis for breast and prostate cancer in Florida.[107]

Another explanation for uninsured and Medicaid holders having increased odds of having GYNC is that there are insurance based discriminations in healthcare that may result in delay in receiving necessary care and in getting access to optimal services.[108] Trivedi, et al. in a study related to healthcare discrimination demonstrated that the type of insurance a person holds determine the type and the degree of care the person receives. This is due to either presence of discrimination, or perceived discrimination by patients,[109] which both are the result of disparity. In their study, the authors reported that lack of insurance, or having Medicaid were significant positive predictors of reporting discrimination.[109] An example of discrimination of access to all health care services for non-traditional Medicaid holders — in the state of Florida — is that the Mayo Clinic at Florida does not accept patients with Medicaid Managed Care Plan.[110] Although the Medicaid in Florida covers a wide range of eligibilities,[111] however Florida is among the states that did not get Medicaid expansion with the application of the Patient Protection and Affordable Care Act (ACA) of 2010. With high expenses of healthcare in the U.S., neither Medicaid holders, nor self-pay individuals will be able to get the needed care in

full dimension without universal health care coverage. In the case of self-pay patients in the current study, the explanation is that the patients without contracted health insurance have higher out-of-pocket expenses, which in turn becomes an obstacle for receiving proper health care services and treatment. Consequently, the self-pay individuals become the high-risk patients for the diseases like GYNC. The provisions of important structures of ACA, such as Medicaid expansion, and health insurance exchange,[112] in all states may reduce the health disparity among Medicaid holders, and self-pay patients.

In regards to the decreased association of private insurance holders and GYNC, these beneficiaries are less likely to have disparities. Evidence in the literature demonstrates that median income among private insurance holders is 2.9 times and median wealth is 23.2 times more than uninsured people.[113] This translates into the possibility that private insurance holders may have higher education levels, more access to preventive care, and healthier lifestyles which in turn decreases the likelihood of them having cancer.[3,45,51]

With regards to the association of patients' hospital LOS with GYNC among patients with AD, this study found no association. It was thought that AD patients with GYNC would have more ailments, poorer health conditions, and thus longer hospital LOS.[33] However, the findings may support results from previous studies where the effect of ADs on risks and survival in female cancers was investigated.[12] The aforementioned study found that as the number of AD types increased, the risk of breast, ovarian, and endometrial cancer decreased.[12] This finding may suggest that cancer morbidity and mortality decrease due to the treatment related factors associated with certain ADs, and may also depend on the number and types of ADs.[12]

Another study hypothesis within the first objective was that patients with GYN related procedures have increased odds of having GYNC. This hypothesis was strongly supported by the study's findings and is somewhat consistent with the literature. About 10% of hysterectomy is associated with GYNC.[65] In addition, hysterectomy is a common treatment for other GYN related complications like endometriosis, [67] as well as risk reducing and elective procedures such as bilateral salpingo-oophorectomy (BSO) when cancer genes BRCA1 and BRCA2 are present.[63,64] Endometriosis alone accounts for 11% of all hysterectomies performed.[65] Evidence suggests that endometriosis is associated with ADs, and several types of cancer, including endometrial and epithelial ovarian cancer.[58,68] Thus, this study's finding support that hysterectomy may be considered a risk factor/predictor of GYNC.

Lastly, within the first hypothesis, I hypothesized that comorbidities would be associated with GYNC among AD patients and this hypothesis was supported by the findings. Comorbidities such as Chronic pulmonary disease, Renal failure, Obesity, and AIDS were among the comorbid conditions that were found to be significantly associated with GYNC. These findings are somewhat consistent with previous studies where evidences have shown that factors such as inflammation, hormonal activity, metabolism and immunological factors are all associated with tumorigenesis.[26,30] For example, AIDS-associated cancers such as Kaposi's Sarcoma, Non-Hodgkin Lymphoma, and invasive cervical cancers are as a result of suppressed immunity.[55,35] Also, Obesity is a known risk factor for GYNC and may be involved with the alteration of metabolism and hormonal activity.[30,70]

The second aim of this study was to determine the various subpopulations of AD patients with increased likelihood of GYNC based on patients' characteristics and comorbid conditions. Using predictive analytics, the predictive variable Hysterectomy was consistently the highest predictor of GYNC among AD groups involving glandular (group D), and systemic and vascular systems disorders (group A). Previous studies have shown the association of hysterectomy with GYNC and endometriosis where the latter is considered as a comorbid condition for ADs and as an autoimmune-like disease itself.[78] Similarly, the predictor Age was also found to be a significant predictor of GYNC in patients with autoimmune systemic and vascular systems (group A), and glandular disorders (group D). As mentioned before, prior literature has demonstrated that Age is a risk factor for GYNC.[9,102]

In the predictive model of this study, 4 subpopulations of AD patients at risk for GYNC, based on unique combinations of risk factors were established: Subpopulation A) Hysterectomy, Age, Coagulopathy, and Chronic pulmonary disease were associated with GYNC among this subgroup of patients with systemic and vascular systems disorders; Subpopulation B) Liver diseases, Age, Hypertension, and Median Income Level were associated with GYNC among this subgroup of patients with major organs AD related disorders; Subpopulation C) Chronic Pulmonary Disease, Other Neurological Disorders, Median Income Level, and Pay Source were associated with GYNC among this subgroup of patients with neuromuscular related ADs; Subpopulation D) Hysterectomy, Age, Diabetes, and Fluid and Electrolytes disorders were associated with GYNC among this subgroup of patients with autoimmune glandular system diseases. It should be noted that these findings related to identification of predictors of GYNC in

these subpopulations are consistent with findings from the first objective of the current study. The unique combinations of characteristics and comorbidities that described subpopulations of patients at risk for GYNC may be used as a potential risk assessment for GYNC as well as for early detection and/or prevention tools.

5.2 Strength and Limitations

This study had a number of strengths and limitations. Many of the limitations are related to the use of administrative data, where the accuracy required for research may not always be guaranteed.[71] Although HCUP data hold a known degree of accuracy,[114] coding errors may be possible. ICD-9-CM codes for ADs and other diagnostics and procedures used in this study could be recorded incorrectly. Such errors could have resulted in non-differential misclassification. Also, the use of HCUP data limits obtaining a complete clinical profile for each patient. For example, factors such as obesity, smoking, alcohol consumption, use of hormone replacement therapy (HRT) and immunosuppressant drugs, diet and nutrition, and other environmental influences are not included in HCUP data. In addition, information such as cancer stage and the severity of comorbidities are also not available in HCUP data. As a result, it was not possible to investigate whether these variables were confounders in this study. While selection bias was somewhat limited due to the fact that most hospitals in Florida (> 95%),[75] were represented in the dataset, individuals who did not seek or did not have access to care would not be included. Furthermore, given the cross-sectional nature of the data, causality could not be determined. Thus, the true temporal sequence of events was unknown given that the data did not have information on the date of diagnosis of ADs or GYNC.

However, despite these limitations, the study also had a number of strengths. The strengths of this study include the large sample size, and information on patient demographics, diagnoses, and procedures.[74,76] Although the data were from the inpatient discharges of only one state, the results of this study could be generalizable to large segments of the U.S. population, due to the diverse and heterogeneous nature of residents in the state of Florida. Most importantly, to date there has not been any study conducted to evaluate GYNC in relation to comorbidities and patient characteristics among U.S. women with ADs. Because of the diverse nature of ADs and GYNC, and the multiple risk factors involved with both diseases, it was important to recognize the profile of comorbidities and how they relate to patients' characteristics to discover novel risk factors among AD patients with GYNC. Considering cancer and ADs are both the accumulative effect of genetics and environmental exposures,[24,25,27] and that 90-95% of all cancers are linked to lifestyle and environmental factors,[27,101] it was imperative to conduct a comprehensive study to investigate the association between patient-level predictors and several GYNC among U.S. women with AD.

5.3 Implications and Future Studies

This research has several implications. The significant impact of this research will be through the possible improvement of women's health in general by setting priorities for recognizing autoimmunity as a serious disease, and cancer prevention and management through healthier living. Discovery of novel risk factors associated with cancer among women with autoimmunity disorders will have public health and public policy relevance. The strong correlation between patient-level characteristics and GYNC

may lead women with ADs to be recommended for yearly cancer screening for early detection, and better management of possible GYNC.

Moreover, the findings from this research study may be applied to other cancers in relation to ADs. The interference of autoimmunity in DNA methylation processes of cell worth further investigation toward possible treatments such as gene therapy or drug targeting. In addition, the finding about hysterectomy as a strong risk factor is novel, and may be used as a screening marker for GYNC. If Hysterectomy was used as a risk factor for cancer, a large percentage of women going through non-cancer related GYN related procedures (90%) could be screened for GYNC management and prevention, and thus women's health could be improved.

Furthermore, considering only 10% of hysterectomies are due to GYNC, it is also important to determine patients at risk for hysterectomy for non-cancer reasons. This can be done by evaluating the association between patient-level predictors and hysterectomy using statistical procedures similar to those used in the current study. This could help determine the characteristics of patients who get the hysterectomy for non-cancer related medical reasons (90% of hysterectomies), and may help to discover novel risk factors for hysterectomy among AD patients for management of GYNC.

This study, although comprehensive, can be extended in the future by applying the study objectives to patient record based data or longitudinal data, with a larger sample size for each AD category. These types of data would allow for a more complete clinical profile for each patient. Thus, information on influencing factors — such as family history, obesity, smoking, alcohol consumption, use of HRT and immunosuppressant drugs, diet and nutrition, and other environmental influences and previous infections,

cancer stage, severity of comorbidities — and temporality of cancer and ADs would be available. Through predictive analytics such as machine learning and BSF methodology, a thorough prediction of risk factors could be performed, with the goal of personalizing medicine for cancer prevention, diagnostics, and treatment.

5.4. Concluding Remarks

Because cancer and autoimmunity are two major chronic diseases among women, this dissertation research was conducted to investigate the association between patient-level predictors and GYNC among U.S. women with AD. The study population included women with any AD diagnosis. In this study, it was found that older age was associated with decreased odds of having GYNC among women with AD. Medicaid holders and self-pay patients were found to have increased odds of having GYNC. This strong association could be due to lack of complete access to health care services. The lack of ACA Medicaid expansion in Florida could have a large contribution in the health care disparity of some Florida residents of particular socioeconomic statuses. The provisions of important structures of ACA, such as Medicaid expansion, and health insurance exchange, in all states may reduce the health disparity among Medicaid holders, and self-pay patients.

Hysterectomy was found to have a very strong association with GYNC, and can be used as predictive marker for GYNC among patient with AD. Further investigation in regards to hysterectomy's risk factors will determine the characteristics of patients who receive the hysterectomy procedure for non-cancer related medical reasons (90% of hysterectomies), and can help to discover novel risk factors for hysterectomy toward GYNC management.

Comorbidities such as AIDS/HIV, coagulopathy, weight loss, fluid and electrolyte disorders, renal failure, and obesity were found to have strong associations with GYNC among women with AD. Using predictive analytics some comorbidities such as coagulopathy and chronic pulmonary disease along with hysterectomy and age were shown to be strong predictors of GYNC among specific populations of ADs. The unique combinations of characteristics that described subgroups of patients at risk for GYNC can be used as a potential risk assessment for GYNC as well as for early detection and/or prevention tools. The strong correlation between potential predictors and GYNC may lead the women with AD to be recommended for yearly GYN cancer screening for early detection and thus improve women's health.

REFERENCES

1. Siegel RL, Miller KD, Ahmedin, J; Cancer Statistics, 2019. CA: Cancer Journal for Clinicians. 2019; 69(1):7-34. doi:10.3322/caac.21551. Epub 2019 jan 8.

2. Walsh, S.J., Rau, LM. Autoimmune Diseases: A Leading Cause of Death Among Young and Middle-Aged Women in The United States. American Journal of Public Health. 2000;90 (9): 1463-1465.

3. Anand, P, Kunnumakara, A B, Sundaram, C, et al. Cancer is a Preventable Disease That Requires Major Lifestyle Changes. Pharmaceutical Research. 2008; Vol. 25, No. 9. DOI: 10.1007/s11095-008-9661-9.

4. McCormick, P J. Cancer Tsunami: Emerging Trends, Economic Burden, and Perioperative Implications. 2018; Current anes. Reports.8:348-354.

5. Weir HK, Anderson RN, Coleman King SM, et al. Heart Disease and Cancer Deaths — Trends and Projections in the United States, 1969–2020. Prev Chronic Dis 2016;13:160211. DOI: http://dx.doi.org/10.5888/pcd13.160211

6. The Top 10 Causes of Death. World Health Organization. 2019. Available at: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death

7. Murphy SL, Xu JQ, Kochanek KD, Arias E. Mortality in the United States, 2017. NCHS Data Brief. Hyattsville, MD: National Center for Health Statistics. 2018; no 328:1-8. https://www.cdc.gov/nchs/data/databriefs/db328-h.pdf

8. Xu JQ, Murphy SL, Kochanek KD, et al. Final data for 2016. National Vital Statistics Reports. Hyattsville, MD: National Center for Health Statistics.2018; 67(5):1-76. https://www.cdc.gov/nchs/data/nvsr/nvsr67/nvsr67_05.pdf

9. Insight Knowledge 2017, Facts about Gynecological Cancer. CDC Publication #22-0098. U.S. Department of Health & Human Services, CDC. https://www.cdc.gov/cancer/knowledge/pdf/cdc_gyn_comprehensive_brochure.pdf

10. Rosenblum, MD, Gratz, IK, Paw1, JS et al. Treating Human Autoimmunity: Current Practice and Future Prospects. Sci Transl Med. 2012; 4(125): 125sr1. doi:10.1126/scitranslmed.3003504.

11. American Autoimmune Related Diseases Association. Retrieved 2017. Available at: https://www.aarda.org/autoimmune-information/autoimmune-disease-in-women.

12. Hemminki, K, Liu, X, Ji, J et al. Effect of Autoimmune Diseases on Risk and Survival in Female Cancers. Gynecologic Oncology 127. 2012; 180–185.

13. Rose, N R. Autoimmune Disease 2002: An overview. J Invest Dermatol Symp Proc. 2004; 9: 1-4.

14. Quintero, OL, Amador-Patarroyo, MJ, Montoya-Ortiz, G, et al. Autoimmune Disease and Gender: Plausible Mechanisms for the Female Predominance of Autoimmunity. J Autoimmun. 2012; 38(2-3): J109-19. doi: 10.1016/j.jaut.2011.10.003. Epub 2011 Nov 12.

15. Forabosco, P, Bouzigon, E, Ng, MY et al. Meta-Analysis of Genome-Wide Linkage Studies Across Autoimmune Diseases. European Journal of Human Genetics. 2009; 17, 236–243.

16. Tait, D L, Bahrani-Mostafavi, Z. et al. Downregulation of HOXC6 in Serous Ovarian Cancer. Cancer Investigation. 2015; Volume 33, Issue 7. doi: 10.3109/07357907.2015.1041641.

17. American Cancer Society. Revised 2019. Overview of endometrial cancer. Retrieved from: https://www.cancer.org/cancer/endometrial-cancer/about/what-is-endometrial-cancer.html

18. Kaaks, R., Lukanova, A., Kurzer, M S. Obesity, Endogenous Hormones, and Endometrial Cancer Risk: A Synthetic Review. 2002; Cancer Epidemiology, Biomarkers & Prevention. 11:1531-1543.

19. National Cancer Institute. Vaginal Cancer Treatment (PDQ®)–Patient Version. October 13, 2017. Accessed at: www.cancer.gov/types/vaginal/patient/vaginal-treatment-pdq on March 6, 2018.

20. American Cancer Society. Revised 2019. Overview of vulvar cancer. Retrieved from: https://www.cancer.org/cancer/vulvar-cancer.html

21. American Cancer Society. Revised 2019. Overview of vaginal cancer. Retrieved from: https://www.cancer.org/cancer/vaginal-cancer/about.html

22. American Cancer Society. Revised 2019. Overview of ovarian cancer. Retrieved from: https://www.cancer.org/cancer/ovarian-cancer.html

23. Collins, Y, Holcomb, K, Chapman-Davis, E, et al. Gynecologic Cancer Disparities: A Report from the Health Disparities Taskforce of the Society of Gynecologic Oncology. Gynecologic Oncology. 2014; 133: 353–361.

24. Wilson S, Jones L, Couseens C, et al. Cancer and the Environment: Gene-Environment Interaction. Washington (DC): National Academies Press (US). 2002; 3.

25. Sansone P, Bromberg J. Environment, Inflammation, and Cancer. Curr Opin Genet Dev. 2011;21: 80-85.

26. Franks, A L, Slansky, J E. Multiple Associations Between a Broad Spectrum of Autoimmune Diseases, Chronic Inflammatory Diseases, and Cancer. Anticancer Research. 2012; 32: 1119-1136.

27. Aggarwal, B B., Vijayalekshmi, RV., Sung, B. Targeting Inflammatory Pathways for Prevention and Therapy of Cancer: Short-Term Friend, Long- Term Foe. Clin. Cancer Res. 2009;15: 425–430.

28. Dooley, M A., Hogan, S L., Environmental Epidemiology and Risk Factors for Autoimmune Disease. Curr. Opin. Rheumatol. 2003; 15(2).

29. Yoneda, M., Imamura, R., Nitta, H. et al. Enhancement of Cancer Invasion and Growth Via the C5a-C5a Receptor System: Implications for Cancer Promotion by Autoimmune Diseases and Association with Cervical Cancer Invasion. 2019. Oncology Letters;17:913-920.

30. Grivennikov S I., Greten F R., Karin M. Immunity, Inflammation, and Cancer. Cell. 2010; 140: 883-89935.

31. Alberts, B., Johnson, A., Lewis, J., et al. The Preventable Causes of Cancer. Molecular Biology of the Cell. 4th Edition. 2002; Garland Science, New York.

32. Parikh-Patel, A., White, R H., Allen, M., et al. Risk of Cancer Among Rheumatoid Arthritis Patients in California. Cancer Causes Control. 2009; 20:1001-10.

33. Ogle, K. S., et al. Cancer and Comorbidity: Redefining Chronic Diseases. Cancer. 2000; 88(3) :653-663.

34. Stare, S M., and Jozefowicz, J J. The Effects of Environmental Factors on Cancer Prevalence Rates and Specific Cancer Mortality Rates in a Sample of OECD Developed Countries. Int. J. Appl. Econ. 2008; 5(2), 92-115.

35. Barbaro, G, and Barbarini G: HIV Infection and Cancer in the Era of Highly Active Antiretroviral Therapy (Review). Oncol Rep. 2007; 17: 1121-112636.

36. Wang, L., Wang, F.S., Gershwin, M.E. Human Autoimmune Diseases: A Comprehensive Update. J Intern Med. 2015; 278:369–395.

37. McGee, E E., Gastro, F A., Engels, E A., et al. Association between Autoimmune Conditions and Hepatobiliary Cancer Risk among Elderly U.S. Adult. 2019. Int. J. Cancer; 144:707-717.

38. Moss S F. and Blaser M J. Mechanisms of Disease: Inflammation and the Origins of Cancer. Nat Clin Pract Oncol. 2005; 2: 90-97.

39. Goldin, L.R. and Landgren, O. Autoimmunity and Lymphomagenesis. 2009; Int. J. Cancer 124: 1497 -1502.

40. Martin, D N., Mikhail, I S., and Landgren, O. Autoimmunity and Hematologic Malignancies: Associations and Mechanisms. Leuk Lymphoma. 2009; 50(4):541-50. doi: 10.1080/10428190902780677.

41. Krystonsson, S. Koshiol, J. Bejorkholm, L., et. al. Immune-Related and Inflammatory Conditions and Risk of Lymphoplasmacytic Lymphoma or Waldenstorm Macroglubolinemia. 2010, J of National Cancer Inst. 2010; 102(8):557-567.

42. Lin, W. and Karin, M. A Cytokine-Mediated Link between Innate Immunity, Inflammation, and Cancer. The Journal of Clinical Investigation. 2007; 117, (5) 1175-1183. doi.org/10.1172/JCI31537.

43. Clancy, R M., Ashok R A., Abramson, S B. The Role of Nitric Oxide in Inflammation and Immunity. Arthritis & Rheumatism. 1998. 41, 7: 1141-1151.

44. Thomsen, L L., Lawton, F G., Knowles, R G., et.al. Nitric Oxide Synthase Activity in Human Gynecological Cancer. Cancer Research. 1994; 54: 1352-1354.

45. Sigel R L. An Assessment of Progress in Cancer Control. 2018: CA Cancer J. Clinicians.68:329-339.

46. Holmqvist, A S., Olsen, J H., Mellemkjaer, L., et al. Autoimmune Diseases in Adult Life after Childhood Cancer in Scandinavia. Annals of the Rheumatic diseases. BMJ journals.2015; 75(9). http://ard.bmj.com/content/early/2015/10/08/annrheumdis-2015-207659.

47. Boruta II, D M., Gehrig, P A., Fader, A M. Management of Women with Uterine Papillary Serous Cancer: A Society of Gynecologic Oncology (SGO) review. Gynecologic Oncology. 2009; 115: 142-153.

48. Flemming G., Siedman, E L. Epithelial ovarian cancer. In: Barakat R, Berchuck A, Markman M, Randall M, eds. Principles and Practice of Gynecologic Oncology, 6th ed. Philadelphia, PA: Lippincott Williams &Wilkins; 2013;757–847.

49. American Cancer Society. Revised 2019. Overview of cervical cancer. Retrieved from: https://www.cancer.org/content/cancer/en/cancer/cervical-cancer.html

50. Watson, M. National Assessment of HPV and PAP Tests. Prev. Med; 2017; 100:243-247.

51. Egen, O., Beaty, K., Blackley, D. et al. Health and Social Conditions of the Poorest Versua Wealthiest Counties in the U.S. AJPH Research. 2017; 107(1):130-135.

52. Spencer, R J., Rice, L.W., Ye, C. et al. Disparities in the Allocation of Research Funding to Gynecologic Cancer by Funding to Lethality Scores. Gynecologic Oncology.2019; 152:106-111.

53. Romos-Casals, M., Brito-Zeron, P., Kostov, B. et al. Google-Drive Search for Big Data in Autoimmune Geoepidemiology: Analysis of 394,827 Patients with Systemic Autoimmune Diseases. Autoimmunity Review. 2015; 14:670-679.

54. Cutolo, M. Autoimmune Polyendocrine Syndromes. Autoimmunity reviews. 2014; 13: 85-89.

55. Gayed, M., Bernatsky, S., Ramsey-Goldman, R. et al. Lupus and Cancer. Lupus 2009; 18; 479.

56. Wahren-Herlenius, M., Dorner, T. Immuno-Pathogenic Mechanisms of Systemic Autoimmune Disease. Lancet. 2013; 382:819-831. doi: 10.1016/S0140-6736(13)60954-X.

57. Yost, K., Perkins, C., Cohen, R. et al. Socioeconomic Status and Breast Cancer Incident in California for Different Race/Ethnic Group. Cancer causes and control.2001; 12:703-711.

58. Farquhar, C M., Steiner C A. Hysterectomy Rates in the United States 1990-1997. Obstet Gynecol. 2002; 99:229-34.

59. Park, S. Y., Lee, J H., Choi, J S. et al. Laparoscopically Assisted Vaginal Hysterectomy for Women with Anterior Wall Adherence After Cesarean Section. JSLS; 2014.18:3e.2014.00315. doi: 10.4293/JSLS.2014.00315

60. Kjerulffi, K., Langenberg, P., Guzinsk, G. The Socioeconomic Correlates of Hysterectomies in the United States. 1993. American J. Public Health.; 83: 1.

61. Keshavarz, H. Hillis, S D., Burney A., et al. Hysterectomy Surveillance — United States, 1994–1999. 2002. CDC MMWR; 51: ss-5www.cdc.gov/mmwr.

62. Wright, K N. Jonsdottir, G M., Gorgensen, S.et al. Costs and Outcomes of Abdominal, Vaginal, Laparoscopic and Robotic Hysterectomies. 2012. JSLS: 16:519–524. doi: 10.4293/108680812X13462882736736

63. Kvaskoff, M., Horne, A W. Informing women with Endometriosis about Ovarian Cancer Risk. Lancet. 2017; 390: 2433-2434.

64. Vitonis, A F, Titus-Ernstoff, L., Cramer, D W. et al. Assessing Ovarian Cancer Risk When Considering Elective Oophorectomy at the time of Hysterectomy. Obes Gynecol.2011; 117(5):1042-1050.

65. Merrill, RM. Hysterectomy Surveillance in the United State, 1997-2005. Med. Sci. Moni;2008; 14(1):CR24-31.

66. Pearce, C.L., Templeman, C., Rossing, M.A., et al. Endometriosis and Risk of Histological Subtypes of Ovarian Cancer: A Pooled Analysis of Case-Control Studies. Lancet Oncol. 2012; 13:385-394.

67. Wang, C., Liang, Z., Liu, X.et al. The association between Endometriosis, Tubal ligation, Hysterectomy and Epithelial Ovarian Cancer: Meta- Analysis. International Journal of Environmental Research and Public Health.2016; 13:1138.1-12.

68. Kvaskoff, M., Mu, F. Terry, K L. et.al. Endometriosis: A High-Risk Population for Major Chronic Diseases? Human Reproduction update. 2015; 21:500-516.

69. Sarfati, D. Koczwara, B., Jacson, C. The Impact of Comorbidity on Cancer and its Treatment. Cancer J. Clinicians. 2016;66(4): 338-350.

70. Tetsche, M S, Dethlefsen, C. Pedersen, L. The impact of Comorbidity and Stage on Ovarian Cancer Mortality: A Nationwide Danish Cohort Study. BMC Cancer. 2008; 8:31.

71. Yi-sheng, J., Gong, T T., Wang, Y-L. et.al. Comorbidity and survival among women with ovarian cancer: evidence from prospective studies. 2015. Scientific Reports. 2015; 5:11720.

72. Bernatsky, S. Ramsy-Goldman, R. Clark, A. Malignancy and Autoimmunity. Curr Opin Rheumatol.2006; 18 (2):129–134.

73. Engel, G.L. The Need for a New Medical Model (a challenge for biomedicine). Science. 1977;196: 129-136.

74. HCUP-US Overview. Healthcare Cost and Utilization Project(HCUP). 2019; Available from: https://www.hcup-us.ahrq.gov/overview.jsp

75. HCUP-US Overview. Healthcare Cost and Utilization Project(HCUP). 2017. Number of hospitals by State; Available from: https://www.hcupus.ahrq.gov/db/state/siddist/siddist_hospital.jsp#2007

76. Studnicki, J; Ekezue, B.F.; Tsulukidze, M. et al. Classification Tree Analysis of Race-Specific Subgroups at Risk for a Central Venous Catheter–Related Bloodstream Infection. The Joint Commission Journal on Quality and Patient Safety.2014; 40 (3):134-143.

77. International Classification of Diseases, 9th Revision. Geneva, Switzerland: World Health Organization; 1977.

78. American Autoimmune Related Diseases Association, INC. Retrieved 2017. Available from: https://www.aarda.org/autoimmune-information/list-of-diseases/

79. Autoimmune Diseases ICD-9-Codes List. Quest Diagnostics. Retrieved 2017.Available from: https://questdiagnostics.com/dms/Documents/Other/CPT-2015/ICD_9-10_Codes_Autoimmune-MI4954.pdf

80. HCUP Median Income Level. Retrieved 2017. Available at: https://www.hcup-us.ahrq.gov/db/vars/zipinc_qrtl/nisnote.jsp

81. Barrett, M.L., Weiss, A.J., Stocks, C. et.al. Procedures to Treat Benign Uterine Fibroids in Hospital Inpatient and Hospital-Based Ambulatory Surgery Settings 2013. HCUP, AHRQ. 2016; Statistical Brief #200.

82. HCUP-US Overview. Healthcare Cost and Utilization Project(HCUP). Comorbidity Software. Available at: https://www.hcup us.ahrq.gov/reports/methods/ComorbiditySoftwareDocumentationFinal.pdf.

83. Elixhauser, A., Steiner, C., Harris, D R. Comorbidity Measures for Use with Administrative Data. Medical Care.1998; Vol. 36, No. 1.pp. 8-27.

84. Maldonado, G., Greenland, S. Simulation study of Confounder-Selection Strategies. Am J Epidemiol. 1993;138(11): 923-936.

85. Camp, N. J., Slattery, M.L. Classification Tree Analysis: A Statistical Tool to Investigate Risk Factor Interactions with an Example for Colon Cancer (United States). Cancer Causes & Control. 2002; 13 (9): 813–823. doi:10.1023/A:1020611416907.

86. Brinkley, J. Using JMP® to apply Decision Trees and Random Forests as Screening Tools for Limiting Candidate Predictors in Regression Models. SESUG. 2016; Paper RIV-121:1-10.

87. Liaw, A. and Wiener, M. "Classification and Regression by Random Forest." R news. 2002; 2.3:18-22.

88. Uesawa, Y. Rigorous selection of Random Forest Models for Identifying Compounds that activate Toxicity-Related Pathways. In Environmental Science. 2016. doi: 10.3389/fenvs.2016.00009.

89. Hurely, G.J. JMP pro Bootstrap Forest. Retrieved 2018. Available from: https://docplayer.net/30597247-Paper-jm-04-jmp-pro-bootstrap-forest-george-j-hurle...

90. Overview of Bootstrap Forest. JMP support. Retrieved 2018. Available from: https://www.jmp.com/support/help/14-2/overview-of-the-bootstrap-forest-platform.shtml

91. Alex YU, C H. Decision Tree and bootstrap Forest. Retrieved 2018. Available From: www.creativewisdom.com/teaching/462/.../decision_tree_n_bootstrap_forest.pptx

92. Muccini, S. Advanced and Predictive Analytics with JMP 12 PRO.2016; JMP. A Business Unit of SAS. SAS Campus Drive, Cary, NC 27513.

93. Breiman, L. Random Forest Manual.2001. Retrieved 2018. Available at: https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf

94. Statistics and Graphic Guide. JMP, release 7. JMP support. Retrieved 2018. Available from: https://www.jmp.com/support/downloads/pdf/jmp_stat_graph_guide.pdf

95. Gaudard, M. Pulmonary Embolism Classification using CT Images (and JMP Pro). 2010; Pages1-30. Retrieved 2018. Available from: https://community.jmp.com/kvoqx44227/attachments/...us.../Guadard_Paper.pdf

96. Shapiro, M. Using JMP Partition to grow decision Trees in Base SAS. Pharma SUG. 2012; Paper AD22. Retrieved 2019 https://analytics.ncsu.edu/sesug/2013/JMP-04.pdf

97. Modeling and Multivariate Methods. JMP Manual. version 13. JMP, A Business Unit of SAS. SAS Campus Drive, Cary, NC 27513.

98. Sall, J. Monte Carlo Calibration of Distributions of Partition Statistics. 2002; SAS Institute. SAS Campus Drive, Cary, NC 27513.

99. Bertenshaw, GP., Yip, P., Seshaia, P. et al. Multianalyte Profiling of Serum Antigens and Autoimmune and Infectious Disease Molecules to ID Biomarkers Dysregulation in Epithelial Ovarian Cancer. Cancer Epidemiol. Biomarkers Prev. 2008; 17(10).2872-2881.

100. Fowcett, T. An Introduction to ROC Analysis. Pattern Recognition Letters. 2006; 27:861-874.

101. Farhud, D D. Impact of Lifestyle on Health. Iran J Public Health. 2015; 44 (11):1442-1444.

102. Society of Gynecologic Oncology. Vaginal Cancer Risk Factors. Accessed at www.sgo.org/patients-caregivers-survivors/caregivers/vaginal-cancer-risk-factors/ on March 6, 2018.

103. Sahyoun, N R. Lentzner, H., Robinson, K N. Trends in causes of death among the elderly. Aging Trends. 2001; 1, 1–10.

104. Xu Z, Taylor J A. Genome-Wide Age-Related DNA Methylation Changes in Blood and Other Tissues Relate to Histone Modification, Expression, and Cancer. Carcinogenesis. 2014; 35(2):356-364. doi:10.1093/carcin/bgt391.

105. Richardson, B. DNA Methylation and Autoimmune disease. Clinical immunology.2003; 109(1):72-79.

106. Moorman, P., Palmieri, R T., Akushevich, L., et al. Ovarian Cancer in African-American and White Women. Am J Epidemiol. 2009; 170:598–606.

107. Roetzheim, R G., Pal, N., Tennant, C., et al. Effect of Health Insurance and Race on Early Detection of Cancer. J. Natl Cancer Inst 1999; 91 (61): 1409-15.

108. Han, X., Thiede Call, K., Pintor, J K. et. al. Reports of Insurance-Based Discrimination in Health Care and Its Association with Access to Care. Am J Public Health. 2015; 105 (Suppl. 3): S517–S525. doi:10.2105/AJPH.2015.302668.

109. Trivedi, A N., and Ayanian, J Z. Perceived Discrimination and Use of Preventive Health Services. J Gen Intern Med. 2006; 21(6): 553–558. doi: 10.1111/j.1525-1497. 2006.00413.x

110. Mayo Clinic. Retrieved 2019. Available from: https://www.mayoclinic.org/patient-visitor-guide/billing-insurance/insurance/accepted-insurance/medicai

111. Medicate program in Florida State. Retrieved 2019. Available from: https://www.myflfamilies.com/service-programs/access/medicaid.shtml

112. Brandon, W P., Medicaid Transformed: Why ACA Opponents Should Keep Expanded Medicaid. Journal of Health Care for the Poor and Underserved. Johns Hopkins University Press. 2012; 23(4): 1360-1382. doi:10.1353/hpu.2012.0176.

113. Bernard, D M., Banthin, J S., Encinosa, W E. Wealth, Income, and the Affordability of Health Insurance. Health Affairs.2009; 28 (3): 887-896.

114. Abilez, C A. Evaluation of Health Care Cost and Utilization Project Data (HCUP) in Healthcare Research. 2013. Fort Worth, Tx: University of North Texas Health Science Center. https://digitalcommons.hsc.unt.edu/theses/132