3-2021

An Examination into Retention Behavior of Air Force Female Officers

Jessica M. Astudillo

An Examination into Retention Behavior of Air Force Female Officers

THESIS

Jessica M. Astudillo, Capt, USAF

AFIT-ENS-MS-21-M-142

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

DISTRIBUTION STATEMENT A
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

AFIT-ENS-MS-21-M-142

AN EXAMINATION INTO RETENTION BEHAVIOR OF AIR FORCE FEMALE OFFICERS

THESIS

Presented to the Faculty

Department of Operational Sciences

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of Master of Operations Research

Jessica M. Astudillo, BS

Capt, USAF

March 2021

AFIT-ENS-MS-21-M-142

AN EXAMINATION INTO RETENTION BEHAVIOR OF AIR FORCE FEMALE OFFICERS

THESIS

Jessica M. Astudillo, BS
Capt, USAF

Committee Membership:

Dr. Raymond R. Hill
Chair

Dr. Bruce A. Cox
Member

AFIT-ENS-MS-21-M-142

Abstract

Female retention rates in the US military have been considerably lower than that of their male counterparts for numerous years. In the Air Force, women represent 14 percent of officer ranks from O-5 level and above. Comparatively, the overall rate of women officers in service is 20 percent. Understanding the negative factors associated with the attrition rate of this group can help the Air Force leverage positive change. It may also influence adjustments that will increase the number of women serving, and improve diversity throughout both the officer and enlisted ranks.

In this study, logistic regression and survival analysis are applied to model retention and some understanding of how to diversify the Air Force, through increasing our female officer population. Demographic, organizational, and political elements are considered to ensure all affecting issues are measured. Programs that have gone into effect in the past five years, such as the Force of the Future, and Blended Retirement, are also considered to determine their statistical significance.

Applying logistic regression determines potential factors affecting retention rates. All elements are include in survival analysis to characterize female officer retention behavior. Implementing and providing such analysis will help generate a prediction model for retention rates amongst female officers, and how to further amplify diversity.

This research is dedicated to the people working in the background to improve diversity and equity in our Armed Forces.

Acknowledgements

First, and foremost, I would like to express my deep and sincere gratitude to my advisor, Dr. Raymond R. Hill for providing invaluable guidance throughout this process. Your patience, encouragement, and expertise are truly inspirational. I would also like to thank my 21M colleagues for their support and camaraderie throughout our time at AFIT. To my best friend, thank you for reminding me to take time for some self-care and helping me decompress when I needed it most. Finally, I am extremely grateful to my partner for his constant understanding, sacrifice, and love since we began our journey. Your endless support and dedication are the motivation that keeps me going.

Jessica M. Astudillo

Table of Contents

Page

Abstract iv

Dedication v

Acknowledgements vi

List of Figures ix

List of Tables xi

I. Introduction 1

1.1 Motivation 1
1.2 Background 3
1.3 Problem Statement 3
1.4 Issues, Needs, and Limitations 4
1.5 Organization of the Thesis 5

II. Literature Review 6

2.1 Significant Factors 6
2.1.1 Demographic 6
2.1.2 Organizational 8
2.1.3 Political 11
2.2 Previous Military Studies 13
2.3 Summary 16

III. Data Source 17

3.1 Demographic and Organizational Data 17
3.2 Political Data 18

IV. Results and Analysis 20

4.1 Logistic Regression 20
4.1.1 Introduction 20
4.1.2 Data 20
4.1.3 Calculations 23
4.1.4 Analysis at the Cohort Level 23
4.1.5 Analysis at CYOS Level 24
4.1.6 Analysis at the Developmental Category Level 31
4.2 Survival Analysis 42
4.2.1 Introduction 42
4.2.2 Data 43

Page

4.2.3 Calculations 44
4.2.4 Analysis at the Cohort Level 45
4.2.5 Analysis at the Developmental Category Level 52
4.2.6 Cox Proportional Hazards Regression 58

V. Conclusions and Recommendations 66

5.1 Limitations 66
5.2 Follow-On Research 66
5.3 Conclusion 67
5.4 Recommendations 69

Appendix A: R Code for Logistic Regression 71
Appendix B: R Code for Survival Analysis 162
Bibliography 180

List of Figures

Figure Page

1 Odds Ratio of Retention for Distinguished Graduates 26

2 Odds Ratio of Retention for Female Officers with Dependents 27

3 Odds Ratio of Retention for Number of Deployments 29

4 Odds Ratio of Retention for Marital Status 30

5 Odds Ratio of Retention for Marital Status 31

6 Air Ops & Special Warfare Odds Ratio of Retention with Dependents 34

7 Information Warfare Odds Ratio of Retention for Spouse's Career Category 36

8 Combat Support Odds Ratio of Retention with Deployments 38

9 Combat Support Odds Ratio of Retention with Deployments 39

10 Combat Support Odds Ratio of Retention for Marital Status 40

11 Force Modernization Odds Ratio of Retention for Marital Status 42

12 Kaplan-Meier Curve for Distinguished Graduate 46

13 Kaplan-Meier Curve for Dependents 47

14 Kaplan-Meier Curve for Deployments 48

15 Kaplan-Meier Curve for Marital Status 49

16 Kaplan-Meier Curve for Prior Service 50

17 Kaplan-Meier Curve for Race 51

18 Kaplan-Meier Curve for Spouse's Career Category 52

Figure | Page

19 Kaplan-Meier Curves for Deployments . . . 55

20 Kaplan-Meier Curve for Dependents (AOSW and CS) . . . 56

21 Kaplan-Meier Curve for Spouse's Career Category (IW and FM) . . . 57

22 Kaplan-Meier Curve for Marital Status (CS) . . . 58

23 Hazard Ratios for Covariates at the Cohort Level . . . 60

24 Hazard Ratios for Covariates for AOSW . . . 62

25 Hazard Ratios for Covariates for IW . . . 63

26 Hazard Ratios for Covariates for CS . . . 64

27 Hazard Ratios for Covariates for FM . . . 65

List of Tables

Table Page

1 Analysis of Effects by Cohort Summary of P-Values. Entries with yellow highlight are found significant in the analysis. . . . 24

2 Analysis of Effects by CYOS Summary of P-Values. Entries with yellow highlight are found significant in the analysis. . . . 25

3 Developmental Categories with associated AFSCs 32

4 Analysis of Effects for Air Ops & Special Warfare Summary of P-Values. Entries with yellow highlight are found significant in the analysis. . . . 33

5 Analysis of Effects for Information Warfare Summary of P-Values. Entries with yellow highlight are found significant in the analysis. . . . 35

6 Analysis of Effects for Combat Support Summary of P-Values. Entries with yellow highlight are found significant in the analysis. . . . 37

7 Analysis of Effects for Force Modernization Summary of P-Values. Entries with yellow highlight are found significant in the analysis. . . . 41

8 Factors Significant to a Developmental Category's Regression Model 53

9 Cox Proportional Covariate P-Values for Cohort level. Entries with yellow highlight are found significant in the analysis. . . . 59

10 Cox Proportional Covariate p-values for Developmental Categories. Entries with yellow highlight are found significant in the analysis. . . . 61

AN EXAMINATION INTO RETENTION BEHAVIOR OF AIR FORCE FEMALE OFFICERS

I. Introduction

The United States (US) military is working on improving diversity, equity, and inclusion among its ranks. This thesis examines one aspect of diversity in the Air Force (AF). This chapter provides the motivation for which this thesis is built on, followed by the research problem statement and limitations of this study. Finally, the organization of this thesis is summarized.

1.1 Motivation

Female retention rates in the United States military have been considerably lower than those of their male counterparts for numerous years. The low rate of women in the military and the retention of those in the military could be "robbing the service of the potential to improve innovation, agility and performance"(Matthews, 2018). In the past five years, each branch of the military has increasingly recognized the negative impact and taken steps towards diversifying the ranks (Keller et al., 2018). In the AF, women represent 14 percent of officers above the rank of Lieutenant Colonel (grade of O-5). Comparatively, the overall rate of women officers in service is 20 percent.

After a decade of force restructure, the Pentagon has been challenged with increasing its overall personnel strength. To bolster retention rates and increase appeal, the Pentagon has implemented a series of programs designed to increase flexibility, and provide incentives to specifically target members of Generations Y and Z (Mahoney, 2018). Some examples of which include Force of the Future, created to improve its demographic and geographic reach (Carter, 2016). The Blended Retirement Program was also enacted to replace the legacy retirement plan with a system containing bonus pay. While these programs have had some success, they have not been effective in recruiting and retaining the intended population (McMahon and Bernard, 2019). Unlike previous generations, the monetary gain between military and civilian pay is not a top priority for the younger workforce.

Diversity in an organization is important because of the value that multiple perspectives can bring to any enterprise. The diversity focus in this research is on the Air Force active duty female officer workforce. Research increasingly shows that companies with a significant number of women-executives will financially outperform their competitors (Johnson-Freese et al., 2014). The concept of critical mass is defined by Johnson-Freese et al. (2014) as an organization having at least one third women in leadership roles. Kanter (2008) argues that once an organization reaches a critical mass of women, "people would stop seeing them as women and start evaluating their work as manager. In short, they would be regarded equally". Understanding factors associated with the attrition rate of female officers will help the AF leverage change. According to the Bureau of Labor Statistics (2020), women account for 50% of jobs in America as of December 2019, excluding self-employed and farm workers. Although this percentage has increased from previous years, women remain underrepresented in leadership positions, at just 29% of senior management roles. By comparison, women in the AF serving at general officer ranks account for roughly 7.5% (Keller et al., 2018). A commitment to increase diversity and equity in the United States is needed to correct this situation.

1.2 Background

Statistical methods such as logistic regression analysis, and survival analysis have been applied to model retention rates for officer and enlisted alike (Zimmermann, 2017). Other studies have applied similar analyses on non-rated officers, and the enlisted force in other military branches. However, not much consideration has been focused on the retention rate of female military members.

Additionally, the AF recently reconfigured its single Line of the AF category into six developmental groups to ensure officers are competing against other members with similar career progressions (Secretary of the Air Force Public Affairs, 2019). Prior to this decision, officers from 40 different AFSCs (with various experiences, milestones, and missions) competed for promotion. Examining female officer retention based on reconfigured promotion groups will also provide insight to specific factors affecting each group.

1.3 Problem Statement

The focus of this research is determining potential factors affecting female officer retention based on demographic data provided by Headquarters Air Force Directorate of Personnel (HAF/A1). By recognizing these elements, analysis is conducted to examine retention behavior of female officers. Results will help improve current prediction models by incorporating factors directly affecting this population. Finally, conclusions from this study can support on-going efforts in improving diversity, inclusion, and equity.

1.4 Issues, Needs, and Limitations

Data for this study were provided by extracts take from Military Personnel Data System (MilPDS) via HAF/A1. Major updates have been made to mitigate loss of data and improve system performance. Unfortunately, MilPDS is still susceptible to errors as military records are added and/or updated constantly. For example, if a female officer changed her status from single to married, it is updated in the system but takes time to reflect in all records. Therefore, inaccurate data could present issues in analysis.

Contrary to previous research conducted on retention behavior (Schofield, 2015; Zens, 2016; Zimmermann, 2017), algorithms implemented for analysis in this study were computed using R. SAS has been the software program of choice for HAF/A1 analysts and preceding studies have used it to ensure products can be transferred and reproducible by them. Although data were provided in SAS format it was converted to CVS files and uploaded to R. This software, along with Python, are becoming the program of choice for statistical and mathematical analysis in the AF. Consequently, R has not been made available at all units and therefore, is challenging to provide code for reproducibilty if it is not accessible.

Factors associated with programs such as extended maternity leave, Blended Retirement System (BRS), and rescission of the combat exclusion policy were not included in this study. Issues associated with accessing factors for these programs are explained in III.

Finally, AF leadership introduced developmental categories that grouped AFSCs similar in mission focus, experiences and career milestones. However, each AFSC within these new promotion groups still differ in deployment rates, cultural aspects, and career progression. The data provided does not account for these characteristics and therefore may be present in final analysis.

1.5 Organization of the Thesis

Chapter two reviews literature encompassing retention in the military, and significant factors that may have an impact on the female officer population. Chapter three reviews the the data source applicable to factors presented in chapter 2. Chapter 4 presents the methodologies used, analyses conducted, and test results. Chapter 5 summarizes the study and provides recommendations for future research.

II. Literature Review

This literature review explores factors applicable to the female population in the Air Force. It also examines previous research conducted on different groups within the military. Due to the immense amount of literature surrounding female attrition in the workforce, this research centralizes on practical aspects and methodologies to provide a better understanding of factors that pertain to retaining female officers.

2.1 Significant Factors

Demographic, organizational, and political elements have been found to affect retention rates within the military; however, these impacts vary by member and member group. To ensure all affecting issues are represented, past research is considered to develop a list of measurable factors impacting female retention. Among these analyses are studies containing subjective data via questionnaires. However, such subjective data are problematic when producing a predictive model. For this reason, this data are not discussed in length.

2.1.1 Demographic

Keller et al. (2018) performed a qualitative research addressing barriers to female officer retention in the AF. Focus group interviews were conducted across twelve installations to provide insight on factors affecting women, and their choices to separate or remain on active duty. After the group interviews were completed, Keller and her team coded transcripts from each discussion to "identify key themes common across the groups" (Keller et al., 2018). A final analysis determined four components that were important when considering to stay in or leave: "personal and family issues, career, work environment, and broader Air Force and military issues". A few of other factors such as childcare, deployment status, and number of moves were also discussed. However, the main reason married female officers planned to separate from the Air Force was "compatibility with spouse's career or job," which is not considered a top priority for male officers (Keller et al., 2018). Married women also noted they did not feel "Air Force programs and policies adequately supported modern families with two working parents or females breadwinners who have stay-at-home husbands". They also expressed that available resources were "largely designed to support a 1950s family model that included a stay-at-home wife and mother." Female officers who were unmarried or single suggested issues related to dating as an element affecting retention decisions.

Although the factors found in this study are helpful, there are issues when attempting to use them in a predictive model. The focus group was not continuously surveyed, and did not include a longitudinal component to account for long term change; therefore, the survey data are not consistent nor practical to include in a model (Franzen, 2017).

According to Bissonnette (2012), demographic factors such as marital status and number of dependents are also areas of interest. This information captures characteristics impacting a female officer's life that might influence her decision to leave the AF. Bissonnette focuses on female naval aviators and organization factors that involved characteristics related to a specific unit's culture. For example, "the helicopter community is different from the FA-18 community, both in mission and organization, which creates conditions for a difference in culture." A logistic regression model is applied with most of the eight variables deemed categorical. Although cultural aspects were considered, results showed the most statistically significant factor was the number of dependents. Notably, 45% of the studied population who remained in the Navy were female aviators with children. Whereas those with children who "attrited" represented only 27% of the population.

Using logistic regression and survival analysis, Zimmermann (2017) developed prediction rates for AF enlisted retention, focusing on gender and marital status of an Airman as key predictors. The study examined four Air Force Career Specialty Codes (AFSC), or career fields, one of which contained a Selective Retention Bonus (SRB). Variables used in the final models for each AFSC were race, sex, grade, marital status, dependents, and years of service. Analysis from the logistic regression models became problematic due to the noisiness of the data. Zimmermann mitigates this issue by applying survival analysis to examine retention trends. Results demonstrate the percentage of female retention decreases around the fourth year of service. However, reasons for this decline were beyond the scope of the research. Issues found in other studies suggest family concerns as a contributing factors to a woman's early separation when compared to their male colleagues (Keller et al., 2018; United States Government Accountability Office, 2020).

Zimmermann (2017) also examined marital status to determine its affect on attrition. The six categories were legal separation, married, single, widowed, annulled, and divorced. The results show retention rates for single service members are lower than those under other status categories. This result indicates that a service member's marital status impacts their decision to either stay in the military or leave.

2.1.2 Organizational

Demirel (2002) discusses the effects of an officer's source of commission on retention rates. His research focused on the behavior of officers at the end of their minimum service requirements, and at ten years of service. Three of the most common ways in which a person commissions are: Officer Training School (OTS), Reserve Officer Training Corps (ROTC), and the United States Air Force Academy (USAFA). Each path requires a specific initial commitment starting at four years (OTS and ROTC) to at most five (USAFA). In addition, the ten year mark captured the end of service requirement for pilots in particular. His methodology involves separating each branch individually before evaluating them as a whole to find variations among the services. A binary logit regression model is used for each data set. The results suggest that there are "significant differences in officer retention across commissioning sources," for all branches (Demirel, 2002). However, these differences became insignificant after the initial service commitment. Given the initial findings, source of commissioning seems an essential factor to include when studying female officer retention rates.

In 2015, the Air Force formed a Barrier Analysis Working Group (BAWG) whose main focus was "identifying and eliminating workplace hiring barriers" (Kolano, 2019). Expanding on the importance of improving diversity and inclusion, Kolano provides a multivariate analysis on hiring and retention trends at Wright-Patterson Air Force Base (WPAFB), Ohio. This research identifies possible barriers existing within different organizations affecting hiring rates among women and minorities. Although the percentage of women in the workforce has been on an upward trend in the past decade, the population of civilian women serving on WPAFB has decreased between 2008 and 2018 (Bureau of Labor Statistics, 2020).

To determine the cause of this decline, Kolano separated the data based on science, technology, engineering, and math (STEM) fields and non-STEM fields. A large number of positions support research and development missions on base, which require a higher number of STEM positions. Surprisingly, the percentage of women in STEM positions grew more than men from 2008 to 2018. The decline of females working on WPAFB attributed to women working in non-STEM fields. This finding supports the claim by Bissonnette (2012), in that organizational factors influence female retention. To expand on these results, this study will include a female officer's AFSC as a factor to provide a greater understanding of possible differences between career fields, and possibly organizational structures.

Kolano (2019) also found a significant difference in pay between female and male civilian employees. For the year 2018, women were receiving an average salary of \$86,406, while their male coworkers received an average of \$98,846. The average salaries were then partitioned into five educational levels to determine any variation among the groups. Conclusions from this study shows the gap in pay grew as the level of education increased; men received \$12,239 more on average at the doctorate level, as opposed to \$3,022 more at the high school level. Results from Kolano's analysis did not involve normalizing the data to address the difference in male to female ratios for each education level. However, analysis from Chamberlain (2016) found that after adjusting for factors such as education there still remains a pay gap between men and women, with females earning an average of 94.6 cents per dollar a male makes.

Gender-based pay inequality has been a significant problem throughout the history of the United States and abroad. On average, a woman earns only 79% of what their male counterparts do (Schieder and Gould, 2016). That disparity can add up to half a million dollars throughout lifetime earnings of a person. While the pay gap has recently begun to narrow significantly, at the current rate, it is not projected to close until 2059 (Joint Economic Committee, 2016). The disproportion grows exponentially when race is considered. According to a report by the Pew Research Center, white women have narrowed the hourly wage gap by 22 cents from 1980 to 2015. In comparison, during the same period, black women have narrowed the gap by 9 cents, while Hispanic women only 5 cents (Patten, 2016).

A large proportion of the pay gap can be attributed to differences in education, experience, occupation, or other similar factors. The other element not explained by any measurable factor is systemic discrimination. Additionally, even some of the known elements like differences in education can, in fact, be ultimately attributed to structural oppression such as women lacking support in STEM fields (Patten, 2016). While gender-based pay discrimination can arise in multiple ways, one major contributing factor is the discriminatory trends of employers, customers, and even coworkers; the expected value of productivity (or reliability) due to prejudiced assumptions, leads to discriminatory practices (Blau and Kahn, 2007).

At the surface, the structure of military pay does not allow for a gender-based pay gap amongst peers in the same grade. An in-depth look reveals a difference exists within the average promotion rates of women and men. For example white women have a 31% likelihood of retaining to the rank of major, but white men have a likelihood of retaining to major at 45% (Asch et al., 2012). Based on these percentages, it can be deduced that the average pay of men would be higher than women when accounting for all men and women in the AF. The purpose of discussing the gender pay gap serves as a means of presenting underlying issues affecting the civilian sector that may also impact the female military population. It is imperative to understand that issues such as systemic discrimination affect all organizations, civilian and military alike. However, to detect and measure such underlying factors would require an in depth examination beyond the scope of this thesis.

2.1.3 Political

The military as a whole has implemented changes, such as expansion for maternity leave to entice more women to join the ranks (Losey, 2018). Other changes include the Force of the Future and BRS. Prior to 2016, maternity leave did not exceed six weeks in the AF. Under the Pentagon's Force of the Future initiative, Secretary of Defense Carter (2016) declared "12 weeks of maternity leave across all the forces establishes the right balance between offering a highly competitive leave policy while also maintaining the readiness of our total force." The initiative also extended paternity leave from 10 to 21 days. These expansions aim at providing a better family to service balance, which has been known to negatively affect military members (United States Government Accountability Office, 2020).

In 2013, the Pentagon rescinded the combat exclusion policy preventing women from serving in certain combat roles (Jomana, 2020). Three years later all remaining restrictions were removed, opening all combat positions to women. This change eliminated the "final institutional barrier to women's integration" in an effort to improve recruitment and retention rates among women (Smith and Rosenstein, 2017). The addition of these career fields widens the list of AFSCs available to female service members.

In fiscal year 2016, Congress passed the National Defense Authorization Act (NDAA), to include significant changes in the military retirement plan. Under the legacy system, service members received an "immediate annuity based on years of service and basic pay using a 2.5-percent multiplier" (Asch et al., 2017). The BRS combined the defined-benefit system with a defined-contribution system that allowed for government matching contributions (Jansen et al., 2015). Asch et al. (2017) constructed a dynamic retention model (DRM) based on retention decisions made by active and reserve members. The dynamic retention model is a type of simulation used to provide cost and retention estimates for the four military branches.

Under the NDAA, members who entered on or after January 1, 2018 are automatically enrolled in the BRS. Members who entered the military before January 1, 2018 were given the option to choose between the BRS or legacy plan. This change eliminated the problem of separating with no retirement funds prior to 20 years. Additionally, this retirement program offers a continuation bonus between 8-12 years of service which is between 2.5-13 times the monthly base pay. This incentive may encourage more members to remain as the legacy retirement system did not offer anything until 20 years of service. It also provides monetary incentive for AFSCs who are not offered an SRB; thus, increasing incentives to those thinking of separating. Previous studies by Demirel (2002), Schofield (2015), Zens (2016), and Zimmermann (2017) assume that service members would remain in the military after 10 years, since they will not receive any retirement pay if they separated before 20 years. However, the assumption can no longer be made and may require the data be analyzed in BRS and non-BRS groups.

2.2 Previous Military Studies

Understanding retention and attrition rates has been a top priority for all military branches. Previous research covered in this review concentrates on military services as a whole or on specific career fields; such as pilots (both Navy and Air Force) and non-rated positions (both officer and enlisted). Each approach presents statistical tools applicable to analyzing the attainment of female service members.

Schofield (2015) conducts an analysis on the attrition rates among four non-rated officer career fields, using logistic regression analysis to determine significant factors to create a prediction model. In the early 2000s, the Air Force produced sustainment lines to determine the "optimal" number of officers required for each accessions year group. The sustainment line is based on "comparing historical attrition rates to current manpower requirements for each career field" (Schofield, 2015). Upon finding the significant factors, Schofield performs survival analysis to provide a more precise model. This specific analysis is a statistical technique used to analyze data until the time of a specific event has occurred: such as time to failure, time to survival, or time to a specific event (Kleinbaum and Klein, 2010).

Final results by Schofield (2015) contains unique indicators for each career field dealing with attrition and retention. Overall, career field and commissioning source were significant retention indicators. The author's methodology heavily relied on historical data and did not account for variation throughout the years, which is problematic when trying to forecast retention rates. "While historical data certainly provides insight, it is desirable to utilize additional information and methods to provide improved predictions" (Zens, 2016). Although Schofield's study focuses on non-rated line officers and attrition rates, the statistical approach taken is a possible application to a subgroup of her study.

Zens (2016) study contained data provided by MilPDS for non-rated officers from 2002-2015. Her analysis continues Schofield's study and focuses on four Air Force Specialty Codes in specific, and provides survival analysis for each career field. A survival rate is calculated to apply a survival curve to each respective job. This curve is then accumulated to examine a group's behavior and end-strength. The regression results from her study revealed two factors significant in predicting retention rates: commissioning source and prior enlisted service. The statistical approach taken by Zens and Schofield provide helpful information to the AF in attrition and retention rates among some of the non-rated career fields.

Similar to Schofield (2015) and Zens (2016), Franzen (2017) uses logistic regression and survival analysis to determine a retention rate for rated officers in the Air Force. This population is smaller in comparison to previous studies conducted. Given this smaller data set, the analysis used to create a prediction model for this population can also be applied to female officers. Unlike the two previous studies, Franzen did not strictly limit the data to demographic information in determining a prediction model for retention. Her approach consisted of economic, political, and demographic data. The results of her study found "six demographic factors and one economic indicator that are statistically significant factors in modeling the retention behavior of rated officers" (Franzen, 2017).

Zais and Zhang (2016) address manpower issues in the United States Army using a Markov chain model. Forecasting models are used to project the number of enlisted members eligible for reenlistment. However, the current method lacked the ability to understand personnel dynamics to implement future incentive programs. Contrary to statistical methods, such as logistic regression, the Markov chain allows for mathematical advancements done on two levels. The individual level is defined as the "probabilistic progression for military personnel at a given career stage" Zais and Zhang (2016). The aggregate level contains continuous service and separation attitudes. Only two factors were used for the model: grade and time in grade. Results contained a model with higher retention prediction rates than other classification approaches. Despite their successful approach, characteristics chosen by Zais and Zhang do not include factors unique to the female population.

Allen (2018) explores effects early mentorship has on female officer retention in the Marine Corps. Her study involved one-on-one interviews containing active duty and prior active duty female officers. Thematic analysis, "a form of pattern recognition used in content analysis whereby themes (or codes) that emerge from the data become the categories for analysis," is used to analyze the interview data (Roberts et al., 2019). Results revealed a "lack of female mentors decreases female Marines' perceptions of future career options," more specifically in areas focusing on balancing family and work life. Similarly, Keller et al. (2018) reported one third of survey participants stressed the importance of receiving mentorship from successful females. Furthermore, decreasing numbers of successful female officers in higher ranks limits access junior female officers have to a mentor who can provide insights specific to their experience in the military (Allen, 2018; Keller et al., 2018). Although examination of mentorship on female retention is beyond the scope of this research, it is important to note that its impact is anything but negligible.

2.3 Summary

The studies examined for this survey indicate that the most notable indicators affecting women are commissioning sources, marital status, prior enlisted service, and number of dependents. These elements, as well as time-in-service and AFSCs, are included in this study. Although this thesis does not include pay as a factor, it is critical to note that female retention rates are affected by underlying factors outside the control of this research. One of which is the difference in pay based on promotion rates. Programs that have gone into effect in the past five years have not been applied to prior studies. While the time-frame of these programs are short, these changes may have altered the significance of previously identified indicators.

In most cases, retention-based research use a combination of logistic regression and survival analysis to determine attrition and retention patterns. Applying logistic regression determines the substantial factors affecting the retention rates. Survival analysis is used to develop a predictive model to indicate whether a military member will remain on active duty. Combining logistic regression and survival analysis to determine female officer retention provides a more statistically sound model that is capable of reliable prediction.

III. Data Source

This chapter reviews data sources applicable to the factors presented in chapter two. As discussed in that literature review, demographic, organizational, and political factors may be helpful in improving officer female retention.

3.1 Demographic and Organizational Data

MilPDS is the primary resource for demographic and organizational data used in this study. MilPDS is a computerized records database containing Air Force personnel data and events that occur throughout an Airman's career (Gildea, 2013). Information such as a service member's name, social security number, commissioning source, number of dependents, and promotions are managed on the system.

MilPDS has undergone major updates to mitigate loss of data and enhance system performance however, it still remains susceptible to occasional errors. Manual updates to current records and new information are added constantly by technicians across the AF. These changes prove challenging and may result in an incorrect input, deletion, or manipulation by accident. HAF/A1 has developed processes to automatically fix some errors, by scanning previous extracts to fill in missing information. To also help minimize and/or correct inaccurate information, notifications to review records and request updates, if necessary, are sent yearly to service members. If a member fails to submit a request for correction then the database remains inaccurate until it is discovered. Data back-ups are performed frequently as a precaution, and referenced, in the event of an unexpected malfunction. If a back-up occurs after an error is induced into the system, and not corrected by the next backup, then the obtained data still contains incorrect information (Schofield, 2015).

Extracts were provided by HAF/A1 in SAS format, converted to CVS files, and uploaded to R for analysis. The data consist of active duty officer personnel records for all AFSCs from September 2009 to September 2019. Demographic factors provided in the files include number of dependents, race description, marital status, spouse's career field status, and age. Organizational factors pulled from the data are represented by the duty AFSC and source of commissioning. For this study, the data set was refined to rated and non-rated line officers. This means attorneys, chaplains, and medical officers (dentists, nurses, doctors, etc.) were not included.

3.2 Political Data

Unlike demographic or organizational data, political data are not as apparent and becomes difficult to quantify. Aspects representing political influences can be seen in the approved NDAA each year. Some of these changes include extending maternity leave, allowing women to join combat related jobs, and altering retirement benefits. Other affects involve changes in operation tempos that are influenced by an electorate.

Factors related to extended maternity leave are not tracked in MilPDS. As a result, they are not provided in the SAS extracts. Information linked to an active duty mother's maternity leave requires access to records beyond the scope of this study. Although extending maternity leave from 6 to 12 weeks for active duty mothers have shown to have a positive affect on breastfeeding, it is still unknown if the expansion has improved female retention or job satisfaction (Delle Donne et al., 2019). Further research in female retention rates by including data such as maternity leave may provide a deeper understanding of our female service members.

Since the lifting of all gender-based job restrictions, the AF has worked towards integrating women in all military career fields. This opened seven previously male-only combat related jobs in the AF to female service members; of the seven career fields, two are coded as officer-only positions. Despite these additions, data points associated with these two jobs are low. Therefore, a significant influence due to these specific data sources will not be present. Additionally, the time-frame in which these jobs have been available to women expands to just four years. Future studies in female retention, encompassing a longer time-span, may employ a larger, more complete, dataset, allowing for an in depth analysis on possible affects of combat related career fields on female officer retention.

The new retirement system, BRS, was enacted by all military branches in January 2018. It combines benefits from the civilian 401(k) and the legacy ("high 3") military retirement system. Service members who joined prior to 1 January 2018 were able to choose between the new retirement plan and the legacy; all members joining after 1 January 2018 are automatically enrolled into the BRS. Unfortunately, the time-frame in which the BRS has been in place is too short to actually provide significant influence on female retention. Service members who may have chosen to opt into the BRS in 2018 and decided to separate in 2019 yield just one year of data in this study. One year of data is insufficient to provide insight to female retention behavior.

Other areas of interest that may best represent political influences are operations tempo (OPSTEMPO) and number of deployments. OPSTEMPO is defined as the rate in which a unit is involved in military activities (e.g. exercises, training deployments, and contingency operations) (Karmarck, 2020). Unfortunately, this type of factor fluctuates based on each military base and its mission, making it problematic for regression or survival analysis; Due to the dynamic nature of the data associated with OPSTEMPO, it is not included in this study. However, political influences can be represented by the number of deployments in a female officer's career. This information is provided and directly extracted from MilPDS.

IV. Results and Analysis

This chapter presents the methodologies used and analyses conducted in this study. Logistic regression is applied to identify factors significant to female officer retention. Odds ratios are then calculated to compare the likelihood of occurrence between variables. Finally, survival analysis is performed to develop a model for predicting retention behavior.

4.1 Logistic Regression

4.1.1 Introduction

Logistic regression is conducted to detect critical factors in relation to female officer retention. Once the significant variables are found, odds ratios for each significant element are gathered.

4.1.2 Data

Data provided by HAF/A1 are used for logistic regression. The time period for the data spans September 2009 to September 2019. A binary response variable, called "retain", is assigned to each service member with '0' signifying the member separated and '1' signifying they stayed in. Variables considered are marital status ('1' = Single, '2' = Married, '3' = Divorced/Annulled/Legally Separated, etc.), commissioning source ('1' = Academy, '2' = ROTC, '3' = OTS, '4' = Other), race ('1' = American Indian/Native Alaskan, '2' = Asian, '3' = Black/African American, '4' = Declined to Respond, '5' = More Than One Race, '6' = Native Hawaiian/Other Pacific Islander, '7' = White), spouse's career category ('1' = Active duty, '2'= Reserve/Air National Guard, '3' = Other, '4' = Not Applicable), number of dependents ('0' = No Dependents, '1' = One or More Dependents), number of deployments ('0' = Has Not

Deployed, ‘1’ = One Deployment, ‘2’ = Two or more Deployments), prior enlisted service (‘0’ = No Prior Service, ‘1’ = Has Prior Service), and distinguished graduate (DG) status at source of commission (‘0’ = regular graduate, ‘1’= DG).

Extracts containing officer-only service members are refined using SAS filtering options to sub-group the data into female only information encompassing fields such as marital status, AFSC, etc. Once the data sets were reduced to include all necessary factors, the data were saved in CSV format, which was used to clean the data and upload to R. Assumptions were made during the cleaning process such as placing a female (if here marital status = married) member’s spouse in the “other” category if their career type was not provided in MilPDS. A code of “not applicable” in the spouse’s career variable included all women identified as divorced/annulled/legally separated to reflect their marital status. Further refinement in R (see Appendix 5.4 for R code) included redefining variables and deleting duplicate records. In situations involving duplicate records, the last record is saved. It is assumed the last entry obtained is the most accurate and “stagnant” information (e.g. commissioning source, prior service, etc.) does not change over a female’s career.

Prior to model production, verification of logistic regression assumptions are checked. The assumption of a binary dependent variable is met as the response variable for each model is 0 = female officer separated and 1 = female officer is retained. Logistic regression also requires observations to be independent, so a check of multicollinearity is conducted. If variables are shown to have a near-linear dependence then the ”problem of multicollinearity exists” (Montgomery et al., 2021). Dependencies between variables cause inferences in the regression model to become misleading and must be addressed before model creation. During examination, covariates such as age and grade revealed relationships were highly correlated and were removed to meet the independence assumption.

The first logistic regression model built is at an aggregate level containing all records from 2009-2019. Each female officer had exactly one entry for this data set, creating a complete cohort. Analysis for this model are discussed in section 4.1.4.

Data are then partitioned into five commissioned years of service (CYOS) groups. Service members are assigned to a CYOS subset if their time is service spans the entire CYOS range. Due to the time-frame of the data captured, each record is susceptible to truncation and censoring. Truncation occurs when data are observed only if it covers a particular range and values that "fall outside a certain range" are not observed (Meeker and Escobar, 2014). Censoring occurs when "response values cannot be observed for some or all of the units under study" (Meeker and Escobar, 2014). This occurs because the data extracts are collected from a certain timeline that may not contain an officer's completed military record. The middle, beginning, or end of a record may not be observed during the 12 year span of records under examination. For example, if a female officer separated after 14 years then her record would be included in the 0-6 CYOS, 4-8 CYOS, and 8-14 CYOS. However, her record is not observed in the 12-19 CYOS since her record does not span the entire range of that subset. For each data set, there was only one entry per female officer whose time in service potentially spanned the entirety of that CYOS subset.

Finally, the data were organized into officer developmental categories (provided in Section 4.1.6, Figure 3) by CYOS to reflect recent changes to future officer promotion boards. In October 2020, the AF reconfigured its single Line of the AF category into six developmental groups. Prior to this decision, officers from 40 different AFSCs with various experiences, milestones, and missions, competed for promotion. Re-categorizing officers into these subgroups give officer's a chance to compete against other members with similar career progressions when they reach their promotion boards (Secretary of the Air Force Public Affairs, 2019).

The categories examined in this study are Air Operations and Special Warfare (Rated officers), Information Warfare, Combat Support, and Force Modernization. Developmental groups involving nuclear and missile operations and space operations are not inspected in this study. Data collected for rated officers did not distinguish the difference between student and non-student rated officers. Therefore, determining retention patterns between training and non-training environments will require additional research.

4.1.3 Calculations

Once the data are refined for each iteration of building a logistic regression model, R's glm() command was used to generate a model for female officer retention. All characteristics utilized in the model (listed in section 4.1.2) are categorical or binary variables. Wald Chi-Square p-values for each covariate are summarized and provided in each of sections presented below 4.1.4-4.1.6. With 95% confidence level, values with p-values under 0.05 are significant and highlighted in yellow in each analysis of effects table.

Odds ratios for significant indicators are also examined to analyze the likeliness of retaining over a baseline case. The baseline case for each instance is set to 1 to measure the association between variables. If a variable is greater than one than it is n times more likely to retain over the baseline (Hosmer Jr et al., 2013). If it is under 1 then it is n times less likely to retain over the baseline. All R code for this analysis is included in Appendix A.

4.1.4 Analysis at the Cohort Level

Table 1 displays the covariates sorted from most significant to least based on the p-values for this logistic regression model. The R-Square value produced was 0.0434, which signifies the model is not a good fit in predicting female officer retention. This is expected as the data contains potential errors due real-world extracts taken from MilPDS.

Table 1. Analysis of Effects by Cohort Summary of P-Values. Entries with yellow highlight are found significant in the analysis.

Variable	**P-Value**
Dependents	<0.0001
Marital Status	<0.0001
Deployments	<0.0001
Prior Service	0.0002
Race	0.0018
Spouse's Career	0.0153
DG	0.0291
Commissioning Source	0.0753
Obs	**7017**

In past research, commissioning source has been shown to have an affect on officer retention (Schofield, 2015; Franzen, 2017); however, when solely examining female officers this is not the case. Logistic regression performed at the cohort level indicates all variables except commissioning source are significant indicators for female officer retention. For this reason, this covariate is removed for logistic regression analysis performed at the CYOS and Developmental Category levels.

4.1.5 Analysis at CYOS Level

When analyzing female officers at the CYOS level, factors vary based on the range of years examined. Table 2 displays the covariates from most commonly significant to least with the given Wald Chi-square p-values. The most significant characteristic across all CYOS ranges is DG status, followed by number of dependents, number of deployments, marital status, and prior enlisted service. Odds ratios for these indicators are examined to analyze the likelihood to retain based on CYOS.

Table 2. Analysis of Effects by CYOS Summary of P-Values. Entries with yellow highlight are found significant in the analysis.

Variable	**CYOS**				
	0-6	**4-8**	**8-14**	**12-19**	**20-22**
DG	0.0277	0.0018	0.0002	0.0092	0.0184
Dependents	<0.0001	<0.0001	<0.0001	0.0031	0.4162
Deployments	<0.0001	<0.0001	<0.0001	0.2441	0.0006
Marital Status	<0.0001	0.0119	0.0003	0.0027	0.7998
Prior Service	0.0002	0.0022	0.1280	0.0061	0.0944
Race	0.0018	0.7642	0.1280	<0.0001	0.5675
Spouse's Career	0.0073	0.1283	0.0539	0.0881	0.0445
Obs	**7017**	**5392**	**2657**	**1373**	**517**

Odds ratio of retention based on DG status is shown in Figure 1. Female officers who did not graduate with DG status are set as the baseline of comparison with a value set to one. DGs graduate at the top 10% of their class are recognized to have a high level of leadership in comparison to their peers. This covariate has two CYOS bins with odds ratio confidence intervals that cross 1, meaning there is insufficient data to suggest a statistical difference between DG and regular graduates. These occur at the 8-14 and 20-22 CYOS.

Overall, those who graduate as DG have a higher likelihood of retention across all years of service ranges. The closest relationship between DGs and regular graduates occurs in the 0-6 CYOS group with an odds ratio of 1.44. However, the largest difference between female graduate status is at the 20-22 year mark. DGs are 2.05

times more likely of retaining to 20 years of service in comparison to their classmates who commissioned as regular graduates.

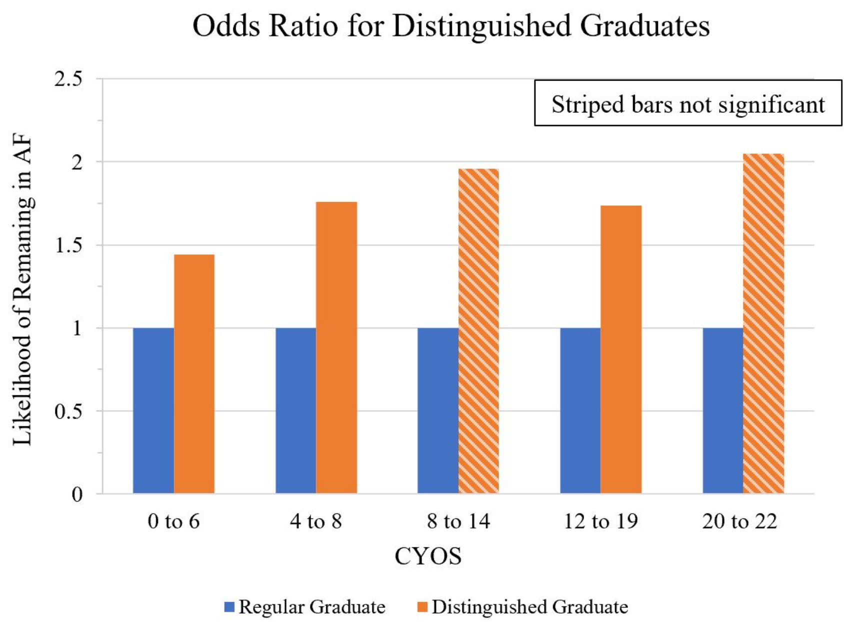

Figure 1. Odds Ratio of Retention for Distinguished Graduates

Figure 2 shows the odds ratio of retention for female officers with dependents. The baseline of comparison for each CYOS range is female officers without dependents. The 4-8 CYOS bin has an odds ratio with a confidence interval crossing 1, indicating there is insufficient evidence to conclude statistical differences exist between those with and without dependents (signified by striped bar). All other CYOS ranges are shown to signify there is a difference in retention between females with dependents and those without.

For commissioning years between 0-6 and 12-19 CYOS, female officers with dependents are consistently more likely to retain than those without dependents. These odds ratios are more distinct in the 0-6 and 4-8 CYOS groups. Females with dependents are 2.04 times more likely to retain than those without from 0-6 CYOS. Between 4-8 CYOS, females with dependents are 2.54 times more likely to retain than those without. The delta between these two groups does decrease from 8-14 and 12-19 CYOS. Although reasons cannot be provided based on this data, further research could be conducted to determine possible underlying factors.

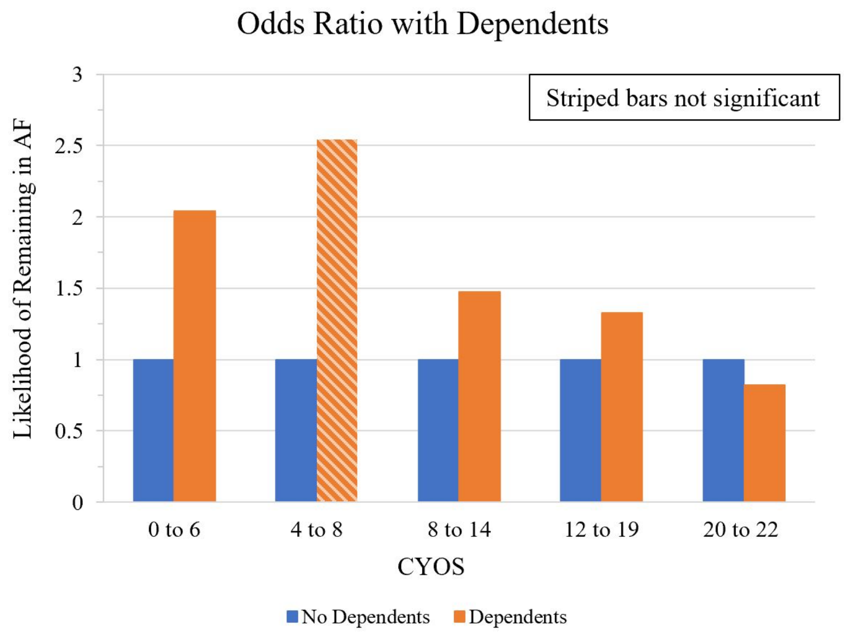

Figure 2. Odds Ratio of Retention for Female Officers with Dependents

The odds ratio for retention based on the number of deployments is shown in Figure 3. Female officers with no deployments are set as the baseline. All CYOS ranges were considered significant with exception to the 0-6 and 4-8 CYOS bins. Overall, women with two or more deployments were more likely to retain than those with one or no deployments. The greatest disparity between those with two or more deployments and those with one or less are in the 0-6 and 4-8 CYOS groups.

Although this data cannot provide the exact reason why, these high retention likelihoods may be influenced by initial service commitments required for AF officers. For example, those who commission from ROTC and OTS have a commitment of four years of service. While those who commission through the USAFA have a commitment of five years before they are eligible to separate. Additionally, pilots incur a 10-year service commitment and navigators incur a 6-year service commitment from the date they complete training. Furthermore, each of the 27 different AFSCs analyzed in this data set have different operations tempos. For example, female officers in security forces deploy at a higher rate in the first 8 years of service than those serving as operations analysts.

The variation between those with deployments and those without, decreases significantly by 8-14 CYOS. At 12-19 YOS, those with one or no deployments are equally likely to retain; those with 2 or more are 1.29 times more likely to retain. During this time an officer has reached an important point in her career: to continue to retirement or separate. This is also around the same time pilots and navigators complete their service commitment affecting the smaller gap in likelihoods between the three groups under investigation. The data also covers the period during the pilot shortage that affected the AF, which may be reflected in the 8-14 CYOS range. Conversely, the odds ratio for retention in female officers with one or more deployments increases by the 20-22 CYOS. This may likely be due to the perception that officers with deployment experience have a higher probability of being promoted to the General Officers corps.

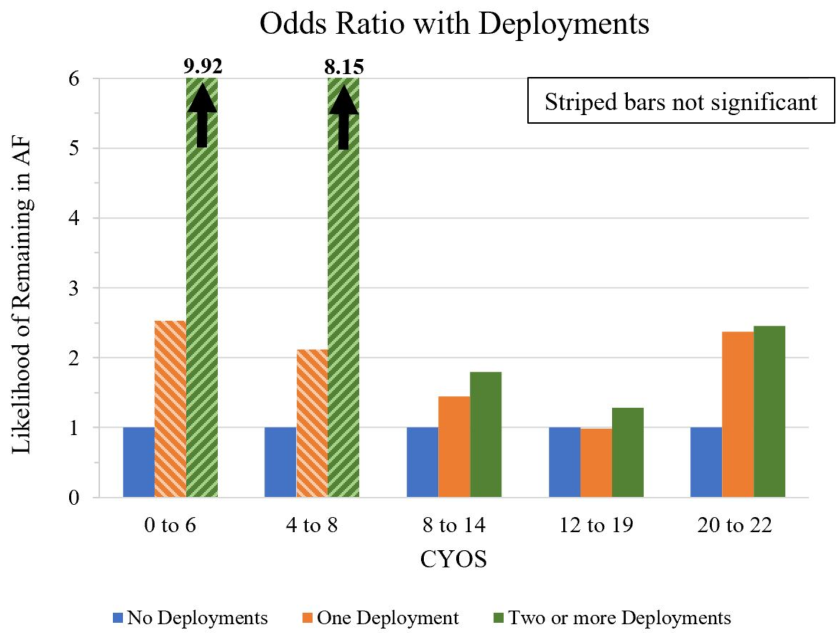

Figure 3. Odds Ratio of Retention for Number of Deployments

Figure 4 displays the odds ratio of retention based on marital status. The baseline for marital status is single female officers, with a value kept at one. Female officers who were previously married were more likely to retain when compared to single or married female officers in the 0-6 CYOS range (1.21 odds ratio). This disparity increases in the 4-8 CYOS range making them 1.58 times more likely to retain. Surprisingly, results for this population has the lowest odds ratios between 8-14 and 12-19 CYOS (0.54 and 0.50, respectively).

Married female officers are 0.65 times less likely to retain than single females in their first 6 years of service. The likelihood of retaining for this population becomes slightly higher than single females with an odds ratio of 1.06 between 4-8 CYOS. However, this likelihood decreases significantly between 8-14 and 12-19 CYOS. These two CYOS bins are significant to a female officer's career as she: 1) reaches roughly the half-way point to retirement and must decided to continue service and 2) approaches promotional boards for O-5 (Lieutenant Colonel). Both events are imperative to retention since this is when the percentage of female representation drops.

While this data cannot provide specific indicators explaining lower retention ratios for married and previously married females, previous research suggests underlying factors related to family or personal matters may contribute to the lower likelihoods (Keller et al., 2018). It should also be noted that single female officer are also sacrificing their personal lives (dating, children, etc.) to continue to 20 years of service.

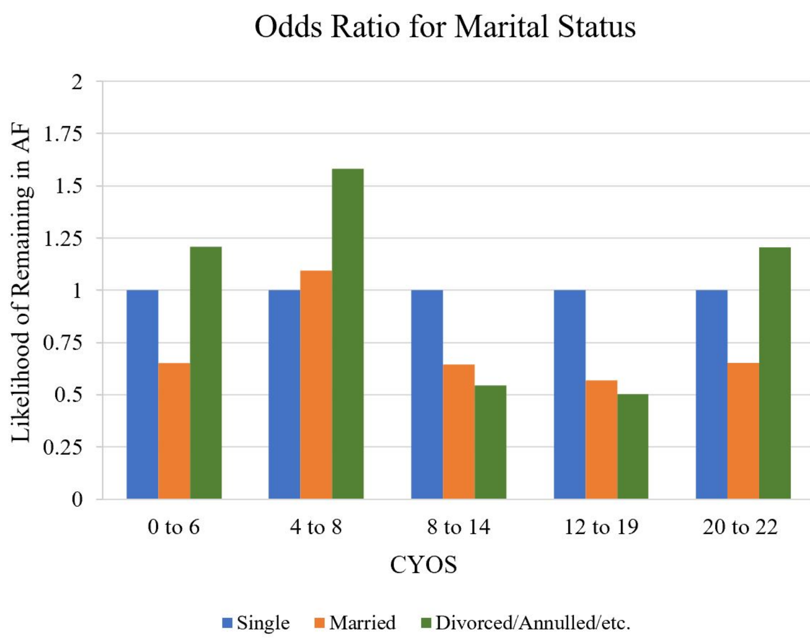

Figure 4. Odds Ratio of Retention for Marital Status

Odds ratio of retention based on prior years of service is shown in Figure 5. Female officers with no prior years of service are used as the baseline. Even though two CYOS bins (12-19 and 20-22) are considered insignificant, female officers with prior service are more likely to retain than females without prior service. This conclusion is similar to previous studies conducted on non-rated and rated officers (Schofield, 2015; Franzen, 2017).

In general, officers must serve 20 years (cumulative) in the military to become eligible for retirement. As part of the voluntary force management program enacted in 2012, the AF offered a 10-8 waiver on retirement for prior enlisted members. This meant prior service officers with 10 years of service were able to apply for retirement if they had 8 CYOS. Female officers with at least 18 total active duty years of service fall within the 8-14 CYOS. Females who also have 11 or more years of enlisted service are also eligible for retirement during this time. Both situations may explain the drop in retention for prior service from 4-8 to 8-14 CYOS. Although considered insignificant, 12-19 CYOS suggests prior enlisted females are almost twice more likely to retain than non-prior female officers.

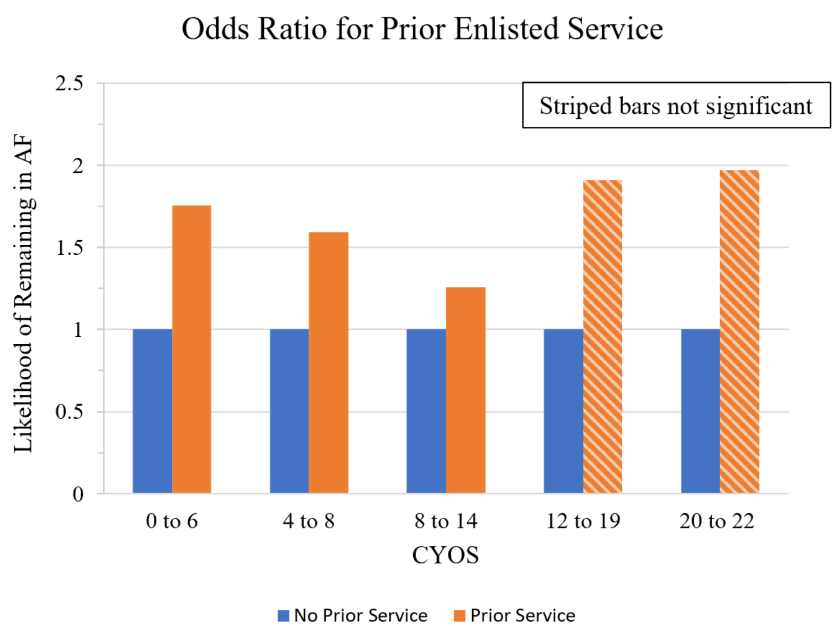

Figure 5. Odds Ratio of Retention for Marital Status

4.1.6 Analysis at the Developmental Category Level

Logistic regression models for the four developmental categories (noted in section 4.1.2) are analyzed using the same categorical variables presented in section 4.1.5. Although results at the CYOS level determined Race and Spouse's Career category significant for 2 of 5 CYOS, they remain in the regression models. Retaining the factors helps examine the sensitivity of each developmental category in relation to each other covariate. Table 3 shows the AFSCs associated with each development category analyzed in this study.

Table 3. Developmental Categories with associated AFSCs

Developmental Category	AFSCs
Air Operations and Special Warfare	Pilot (11x), Combat Systems(12x), Air Battle Manager (13B), Special Tactics (13C), Combat Rescue (13D), Tactical Air Control Party(13L), Remotely Piloted Aircraft Pilot(18x)
Information Warfare	Information Operations (14F), Intelligence (14N) , Weather (15W) Cyber Operations (17X), Public Affairs (35X), Operations Research Analyst (15A), Special Investigation (71S)
Combat Support	Airfield Operations (13M), Aircraft Maintenance (21A), Munitions and Missile Maintenance (21M), Logistics Readiness (21R), Security Forces (31P), Civil Engineering (32E), Force Support (38F), Contracting (64P), Financial Management (65X)
Force Modernization	Chemist (61C), Nuclear Engineer/Physicist (61D) Developmental Engineer (62E), Acquisition Management (63A)

Female officers serving in the Air Operations and Special Warfare category are inspected first. Table 4 shows the Wald Chi-square p-values for each variable based on CYOS. The most significant variables affecting retention rates for women in this category are number of dependents and number of deployments. Both covariates are significant in 4-8, 8-14, and 12-19 CYOS. The 0-6 and 20-22 CYOS do not contain any significant variables. The 0-6 CYOS may be affected by service commitment requirements for pilots and navigators (Combat Systems officers). Pilots incur a 10 year active duty service commitment (ADSC), while navigators serve 6 years the date they complete training. By 4-8 CYOS, navigators have completed their commitment and now have the option of separating, thus impacting retention beginning within this rang. The 20-22 CYOS may not consist of any significant factors due to the lack of data (60 data points total).

DG, marital status, race, and spouse's career category contain p-values less than 0.05 in one CYOS bin. Unlike results provided in section 4.1.5, prior service for rated officers are not shown to be significant at all. Odds ratios for dependents are reviewed to identify if females with or without dependents are more likely to retain. Likelihoods for deployments are not discussed, as all odds ratios were deemed insignificant.

Table 4. Analysis of Effects for Air Ops & Special Warfare Summary of P-Values. Entries with yellow highlight are found significant in the analysis.

Variable	**CYOS**				
	0-6	**4-8**	**8-14**	**12-19**	**20-22**
DG	0.6750	0.4400	0.0929	0.0294	0.2480
Dependents	0.3950	0.0226	<0.0001	<0.0001	0.4930
Deployments	0.5310	0.0102	<0.0001	0.0032	0.1600
Marital Status	0.1950	0.4933	0.0556	0.0365	0.5350
Prior Service	0.5620	0.8866	0.3927	0.1547	0.6110
Race	0.3370	0.0837	0.0263	0.2862	0.4700
Spouse's Career	0.7770	0.0602	0.4124	0.0326	0.5400
Obs	**1587**	**1262**	**542**	**248**	**60**

Figure 6 shows the odds ratio of retention for rated officers with dependents. The baseline of comparison is female officers without dependents. All CYOS bins, with exception to 0-6 CYOS, has odds ratios with confidence intervals crossing 1, indicating there is insufficient evidence to conclude statistical differences exist between those with and without dependents (signified by striped bar).

For 0-6 CYOS, female officers with dependents are 0.63 less likely to retain than those without dependents. However, the likelihood of retaining for rated with dependents increases from 4-8 to 20-22 CYOS. Females with dependents are between 2.16 to 4.84 times more likely to retain than those without. Future research analyzing this developmental category at the AFSC level may provide more understanding of the effects of dependents on retention.

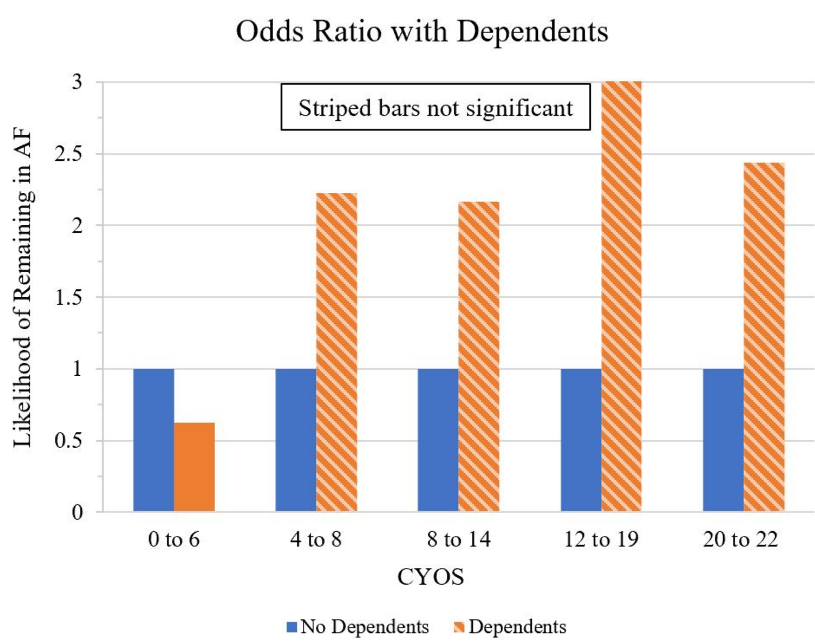

Figure 6. Air Ops & Special Warfare Odds Ratio of Retention with Dependents

Table 5 shows the Wald Chi-square p-values for female officers serving in the Information Warfare category. Performing logistic regression at the developmental category confirmed there exists differences in significant factors when compared to results provided in section 4.1.5. The most commonly significant variable affecting retention rates for women in this category is a spouse's career category. This variable is significant in 0-6, 4-8, and 20-22 CYOS. Unlike section 4.1.4, race is considered insignificant across all CYOS ranges. Odds ratios for spouse's career category are reviewed.

Table 5. Analysis of Effects for Information Warfare Summary of P-Values. Entries with yellow highlight are found significant in the analysis.

Variable	**CYOS**				
	0-6	**4-8**	**8-14**	**12-19**	**20-22**
DG	0.0611	0.0257	0.0540	0.6185	0.0127
Dependents	<0.0001	<0.0001	0.7635	0.7176	0.3956
Deployments	<0.0001	<0.0001	0.1395	0.4557	0.2929
Marital Status	0.0010	0.2617	0.1576	0.1073	0.2307
Prior Service	0.0047	0.0644	0.8874	0.0928	0.0551
Race	0.1203	0.4139	0.9733	0.0676	0.2689
Spouse's Career	0.0077	0.0210	0.9245	0.1031	0.0180
Obs	**2118**	**1561**	**800**	**442**	**190**

Figure 7 shows the likelihood of retaining based on a spouse's career category for Information Warfare officers. The baseline category for spouse's career category is a female officer married to an active duty member (also known as "dual military"). Female officers married to partners serving in a Reserve or Air National Guard component are 1.28 times more likely to retain than dual military members from 0-6 CYOS. This likelihood increases to 1.88 between 4-8 CYOS. Surprisingly the likelihood of retaining decreased drastically after 4-8 CYOS. These women reach the lowest likelihood of retaining between 12-19 CYOS, with a 0.48 odds ratio. Lower retention likelihoods for this category may be influenced by a provision passed in the 2012 NDAA (Kapp and Salazar Torreon, 2020). Under this provision, reserve members were able to be involuntarily activated in support of combatant command missions. The Global War on Terrorism has also activated thousands of reserve and national guard members since 2001.

Women with spouses serving in "other" occupations (those in the civilian sector and non-DoD federal workers) are 0.88 times less likely to retain than dual military members between 0-6. However, the likelihood of retaining positively increases from the 4-8 to 20-22 CYOS bins. The likelihood of these female officers are between 1.19 and 2.45 times more likely to retain than dual military female officers. One reason why higher likelihoods are linked to a female officer with a civilian partner may be the re-location process during each permanent change of station (PCS). Although civilian partners can relocate with their spouse at each PCS event, their careers are negatively impacted (Keller et al., 2018). Issues related to lower likelihoods of retaining for dual military (when compared to those married to civilian member) are separate assignments and back-to-back deployments resulting in extended periods of time spent separated.

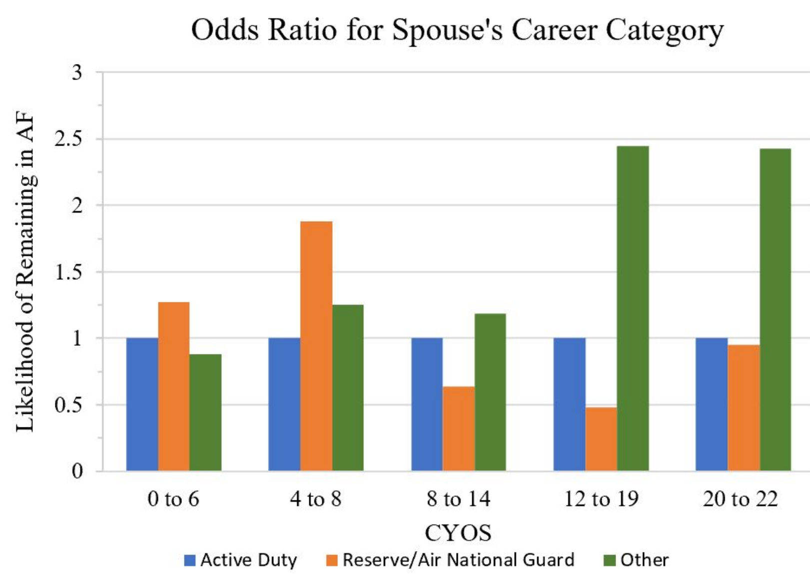

Figure 7. Information Warfare Odds Ratio of Retention for Spouse's Career Category

Analysis of effects for female officers serving in the Combat Support category are shown in Table 6 with Wald Chi-square p-values. When compared with the other divisions, women in this section tend to have more factors affecting them from 0-6 to 8-14 CYOS. The most commonly significant variables affecting retention rates for women in this category are number of dependents, number of deployments, marital status, and prior service. This developmental category displays similar results discussed in section 4.1.5 when examining odds ratios for deployments and marital status.

Table 6. Analysis of Effects for Combat Support Summary of P-Values. Entries with yellow highlight are found significant in the analysis.

Variable	**CYOS**				
	0-6	**4-8**	**8-14**	**12-19**	**20-22**
DG	0.4435	0.0668	0.0163	0.8524	0.5160
Dependents	<0.0001	<0.0001	0.0002	0.8551	0.9060
Deployments	<0.0001	<0.0001	0.0503	0.8369	0.1040
Marital Status	0.0060	0.0125	0.0051	0.3556	0.0562
Prior Service	0.0098	0.0016	0.0176	0.2417	0.4850
Race	0.0044	0.0091	0.5890	0.9381	0.4160
Spouse's Career	0.0028	0.9495	0.1337	0.0001	0.5740
Obs	**2394**	**1848**	**947**	**483**	**194**

Odds ratios of retention based on dependents are displayed in Figure 8. Women without dependents are the baseline case with a ratio set to 1. Women with dependents are more likely to retain than women without dependents between 0-6 and 8-14 CYOS. Unlike outcomes presented in Figure 2, odds of retaining decreases below those without dependents between 12-19 CYOS. The shift in retention likelihoods between 8-14 and 12-19 CYOS reveals a critical decision point for a female officer. At this point she has served about half the time required for retirement. She is also approaching promotional boards for O-5 (Lieutenant Colonel). Both, or one, of these events are negatively impacting female officer retention for this developmental category as the number of women serving at this point decreases.

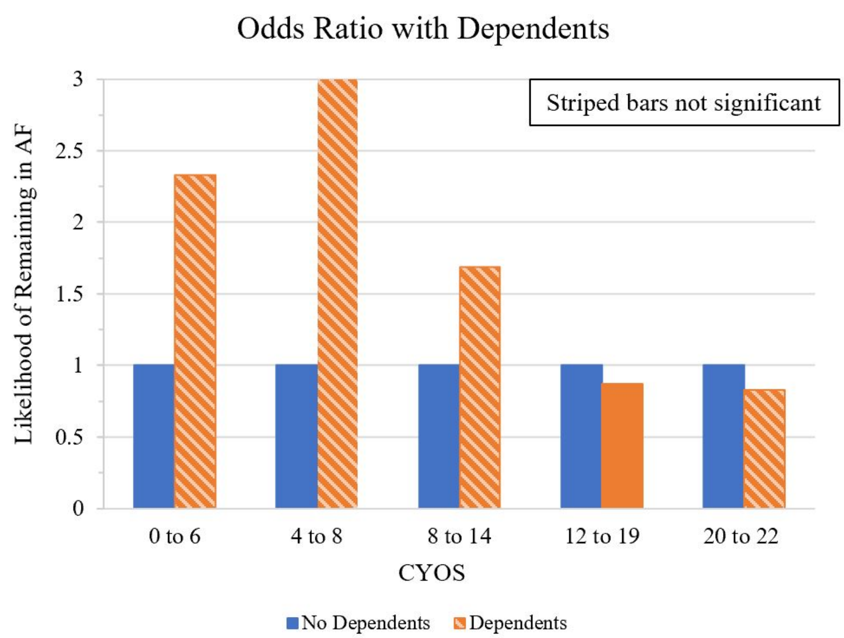

Figure 8. Combat Support Odds Ratio of Retention with Deployments

Figure 9 shows the odds ratio of retention based deployments. The baseline case for deployments are females with no deployment experience. Again, results for this developmental category are very similar to ratios displayed in Figure 3. The only difference is a higher likelihood of retaining from 20-22 CYOS. Women with one deployment are more likely to retain (odds ratio of 8.77) than women with 2 or no deployments. Female officers serving in the Combat Support sector tend to have a higher deployment-to-dwell rate due to their direct support in combat operations. This places a strain not only the female officer but their family members as well; Keller et al. (2018) concluded 78% of women expressed concerns on the affects deployments have on spouses and children.

Lower likelihoods of retention occurring between 8-14 and 12-19 CYOS continue to reveal a critical time-frame for female officers. During these years a female officer must decide to continue service or separate. This decision must be made roughly 10 years from retirement and before competing to achieve the rank of O-5.

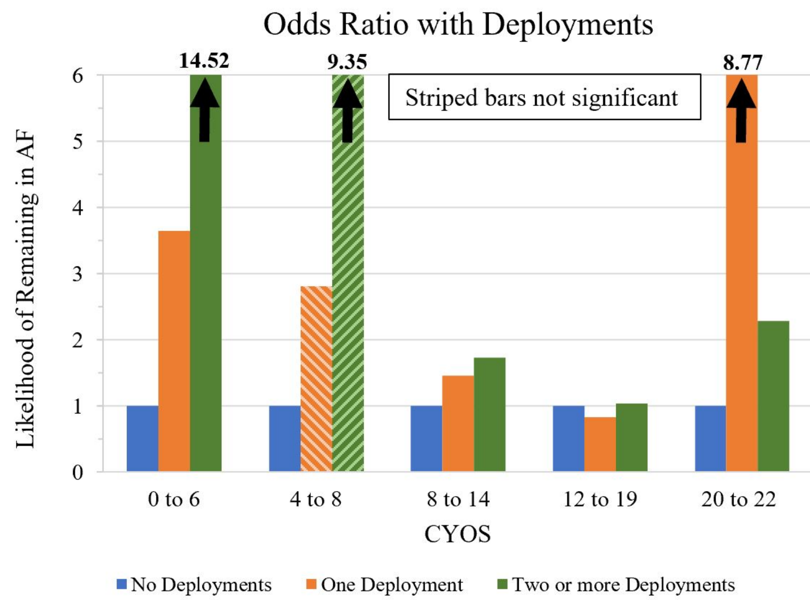

Figure 9. Combat Support Odds Ratio of Retention with Deployments

The odds ratios of retention based on marital status are shown in Figure 10. Single females are the baseline case. There are slight differences in odds ratios when comparing results in Figure 4 and the Combat Support group. Women who were previously married are 2.47 more likely to retain than single female from 0-6 CYOS and 2.26 more likely to retain between 4-8 CYOS. This is an increase of 1.26 and 0.68, respectively, when compared to Figure 4.

The ratios of previously married women may be related to their dependent status and number of deployments. When analyzing these three covariates together, previously married women with two or more deployments who have dependents are more likely to retain from 0-6 and 4-8 CYOS. When looking at the same three groups in the 12-19 CYOS the retention ratios decrease significantly. This means previously married with children are affected by higher deployment rates. These women are faced with balancing work-to-family life. They are separated from their children each time they deploy or travel, which means each time they deploy, they must depend on family or friends to become care providers for their children in their absence. Deployments may also be a contributing factor to higher rates of divorce in this developmental category. Future research including changes in marital status for a female officer may provide more insight.

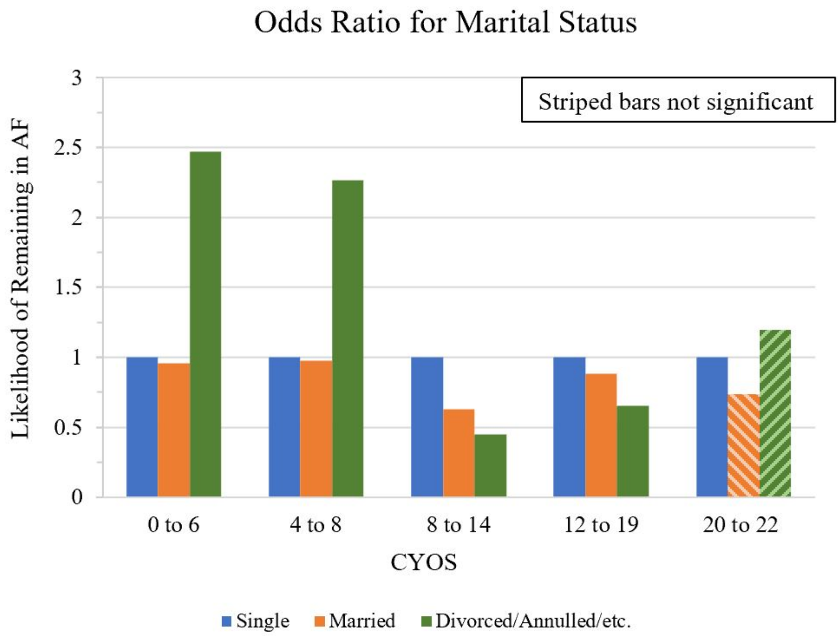

Figure 10. Combat Support Odds Ratio of Retention for Marital Status

Table 7 shows the Wald Chi-square p-values for female officers serving in the Force Modernization category. The most commonly significant variable affecting retention rates for women in this category is marital status. This variable is significant in 0-6, 8-14, and 12-19 CYOS. Unlike section 4.1.4, prior service is considered insignificant across all CYOS ranges. Reasons for this may be a lower population of females in this category or bachelor degree requirements for the AFSCs in this group. Specifically, Force Modernization contains more Science, Technology, Engineering, and Mathematics (STEM) related AFSCs in comparison to other categories. These types of bachelors degrees may be more challenging to obtain as a prior service member if she is attending school while serving full time in the military.

Table 7. Analysis of Effects for Force Modernization Summary of P-Values. Entries with yellow highlight are found significant in the analysis.

Variable	**CYOS**				
	0-6	**4-8**	**8-14**	**12-19**	**20-22**
DG	0.3114	0.1682	0.3625	0.0337	0.7897
Dependents	0.8734	0.0021	0.9838	0.2033	0.6089
Deployments	0.0257	0.0065	0.0517	0.3826	0.9484
Marital Status	0.0087	0.2821	0.0246	0.0431	0.6760
Prior Service	0.5270	0.7850	0.4964	0.4719	0.4016
Race	0.9367	0.8136	0.5928	0.3926	0.0369
Spouse's Career	0.0010	0.6205	0.0885	0.0025	0.2300
Obs	**918**	**721**	**368**	**200**	**73**

Figure 11 shows the odds ratio of retention for Force Modernization based on marital status. Single women are the baseline of comparison. Contrary to Figures 4 and 10, previously married women have the highest likelihood of retaining at 0-6 and 20-22 CYOS. This population also has the lowest likeliness of retaining at 8-14 and 12-19 CYOS, with odds ratios of 0.19 and 0.22, respectively. Exact reasons why this category has the highest retention rates for previously married women are unknown. Deployments may be a direct affect to higher divorce rates to this population, resulting in higher retention. It is also unknown why the likelihood of retention at 20-22 CYOS is drastically higher than compared to other AFSC categories.

Married women also have the highest likelihood of retaining at 0-6 and 4-8 CYOS when compared to Figures 4 and 10. Additionally, the lowest likelihoods of retaining for married women are also represented in this category. Married women are 0.37 less likely to retain at 8-14, 0.31 less likely at 12-19, and 0.30 less likely at 20-22 CYOS.

Declining likelihoods of retention occurring between 8-14 and 12-19 CYOS are also critical time-frames for female officers in this developmental category. Reasons for these low retention ratios are the same proposed for Combat Support odds ratios of retention for deployments and dependents.

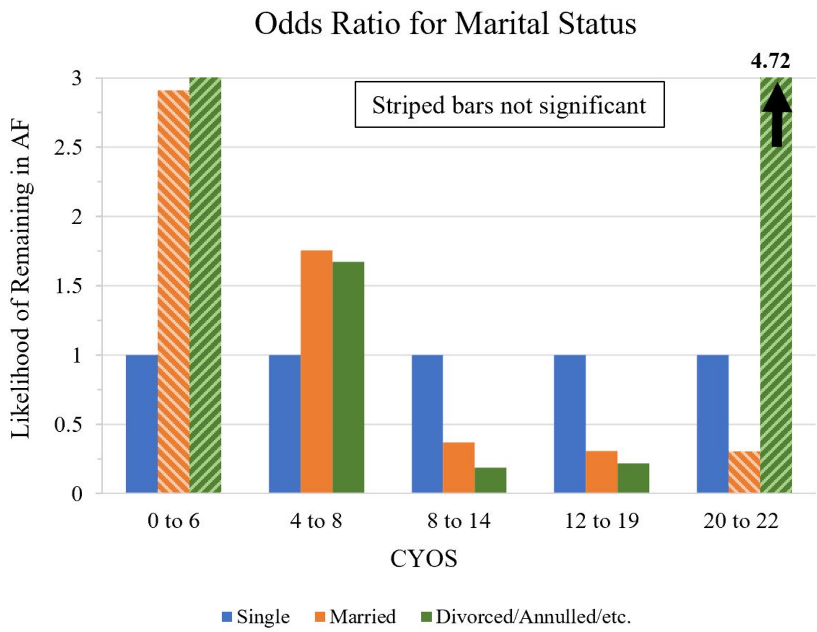

Figure 11. Force Modernization Odds Ratio of Retention for Marital Status

4.2 Survival Analysis

4.2.1 Introduction

Survival analysis is used to create survival curves with retention behavior for female officers. This analysis approach accommodates the censored data provided in this study. The random variable in this model represents time to an event, specifically time until a female officer leaves the military.

Cox proportional hazards regression are then used to examine the relationship between the survival time of a female officer and her predictor variables. This type of regression also accommodates censored data and does not require a normal distribution assumption on the data.

4.2.2 Data

Extracts provided by HAF/A1 were aggregated at a cohort level, which contains all female officers in the dataset. Each female officer has exactly one entry included number of dependents, marital status, deployments, prior service, race, spouse's career, and DG status. Refining data in R included redefining variable and deleting duplicate records. In situations involving duplicated records, the last record is saved. It is assumed the last record obtained is the most accurate and ”stagnant” information (e.g. commissioning source, prior service, etc.) does not change over a female's career. Examination at the CYOS level is not covered in this analysis because survival functions are graphical representations of an event over the span of a 30-year career. Since each CYOS covers at most 7 years it would not provide any insights to retention behavior. R code for survival analysis is provided in Appendix B.

Once the data are analyzed at the cohort level it is divided into four developmental categories. The categories examined in this study are Air Operations and Special Warfare (Rated officers), Information Warfare, Combat Support, and Force Modernization. Figure 3 shows the AFSCs associated with each category. Developmental groups involving nuclear and missile operations and space operations are not inspected in this study. This research also excludes non-line female officers serving as attorneys, chaplains, and medical officers (dentists, nurses, doctors, etc.).

Due to the time-frame of the data captured, each record is susceptible to truncation and censoring. Truncation occurs when data are observed only if it covers a particular range and values that “fall outside a certain range” are not observed (Meeker and Escobar, 2014). Specifically, left-truncation occurs when the data is observed but the observation period (time a female officer separates the military) is not captured. Censoring occurs when "response values cannot be observed for some or all of the units under study" (Meeker and Escobar, 2014). This is because the data extracts are collected from a certain timeline that may not contain an officer's completed military record. Therefore, for the portion of this analysis it is assumed that a female is a commissioned officers at the time she is observed, and her start time begins at time zero. For example, if a female officer is observed with 6 CYOS in 2010 then it is assumed she commissioned at 0 CYOS. Since 5 CYOS are not provided within the observation period (2009-2019). However, with this assumption she remains in the data set for analysis.

4.2.3 Calculations

Once the data are refined for each survival function, R's survfit() function is used to compute the Kaplan-Meier estimates for truncated and censored data. This nonparametric estimate is used to analyze time to a specific event; in this case, time to separate. R's surv(time = x, event = y) to represent the failure time (CYOS a female officer has completed in her record) and the censoring variable (0 = separated, 1 = retained). Cox proportional hazards are computed using R's coxph() function to test the impact of each explanatory variable. Assessing the proportional hazards help determine whether a Cox regression model adequately represents the data. Hazard ratios (HR) are then graphed using ggforest() to determine the association between each covariate and the event probability (event = separated/retained). A variable with a HR greater than 1 increases the hazard of the event. An HR equal to 1 signifies no effect and an HR less than 1 decreases the hazard of the event.

4.2.4 Analysis at the Cohort Level

Logistic regression at the cohort level determined commissioning source was not a significant factor to female officer retention. This variable is removed when analyzing survival analysis at the cohort level for survival analysis. Explanatory variables analyzed in this section are DG, dependents, deployments, marital status, prior service, race, and spouse's career category.

Kaplan-Meier curves for DG status at the cohort level is shown in Figure 12. Shaded regions around each line represent 95% confidence intervals. Although odds ratios for DG status (shown in Figure 1) were graphed by CYOS bins, likelihood of retaining for DG were consistently higher. Retention curves for DGs and regular graduates show similar results. The curves do not cross at any point, with DGs having the dominant retention curve from roughly 8 to 30 years of service. Regular graduates have a gradual decline in retaining over the course of their careers. This supports the perception that DGs have a higher level of leadership and therefore, have higher retention behaviors.

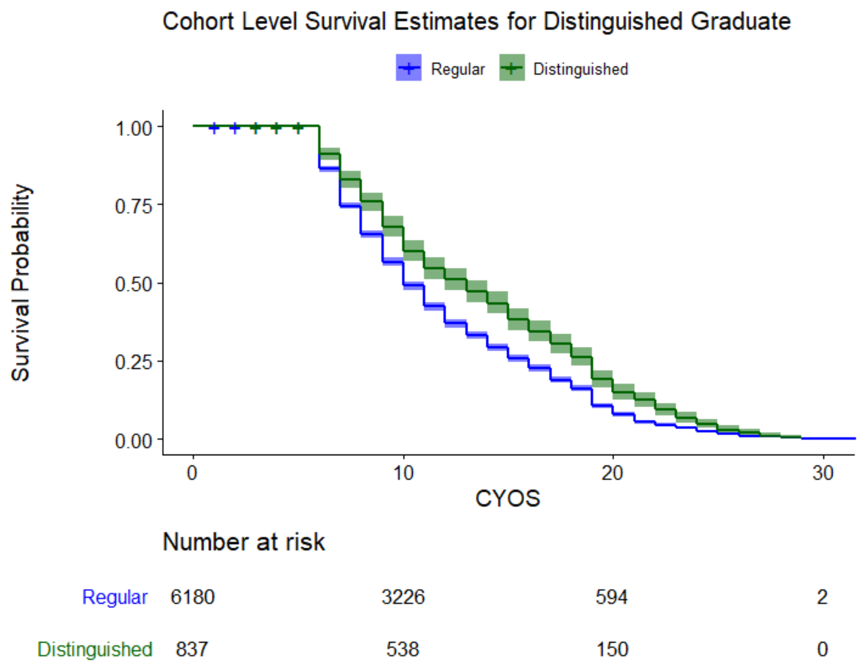

Figure 12. Kaplan-Meier Curve for Distinguished Graduate

Estimates for dependents are displayed in Figure 13. Similar to odds ratio results (reference Figure 2), female officers with dependents have a higher retention rate. The curves do not cross at any point in the timeline, with dependents serving as the dominant curve. Females with no dependents have a pronounced decline at roughly 7 CYOS and continues until about 12 CYOS. This is important as the survival probability drops from about 80% to 25% in this time-frame. Additionally, this behavior is not depicted in the odds ratios. Further research into the differences affecting women without dependents may uncover a cause of their low retention behavior.

Survival probabilities for females with dependents are fairly linear until just before 20 CYOS. Programs such as extended maternity leave, physical training exemptions, and deployment deferments are additional variables that may help provide a more accurate survival estimate.

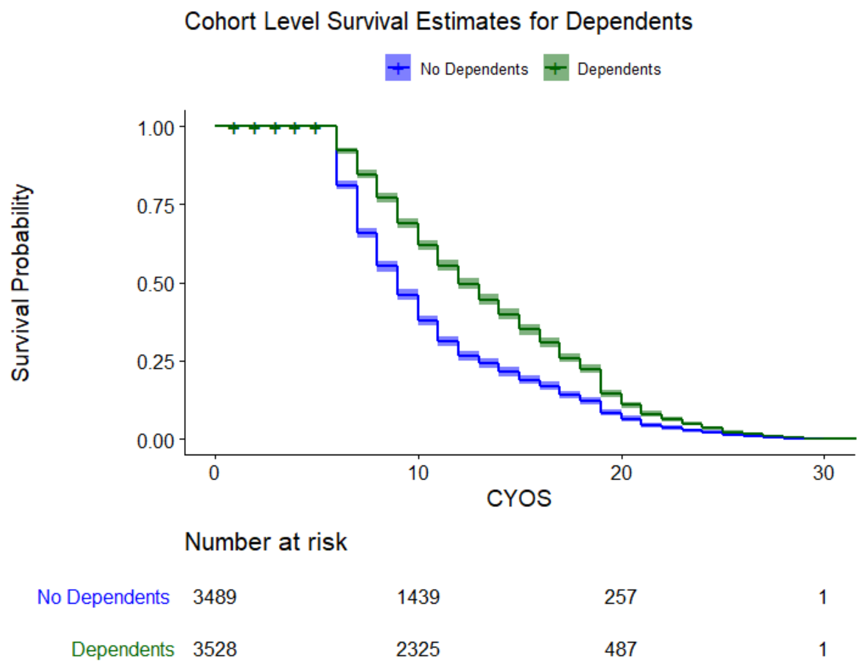

Figure 13. Kaplan-Meier Curve for Dependents

Figure 14 shows the Kaplan-Meier curves for number of deployments. Female officers with no deployment experience = 0, one deployment = 1, and two or more = 2+. Odds ratios for deployments displayed similar results with higher retention likelihoods for female officers with 2 or more deployments. Kaplan-Meier estimates provide additional information such as the wide gaps between each level and the number of women lost throughout the 30-year span.

It is important to note, results for deployments may be influenced by a female officer's AFSC type. Confounded elements are variables "whose presence affects the variables being studied so that the results do not reflect the actual relationship" (Pourhoseingholi et al., 2012). The number of deployments for each AFSC differs due to career and mission requirements. Therefore, it becomes challenging to minimize the affect between deployments and AFSC type. Results showing possible relationships between these variables are reviewed in Section 4.2.5.

There is a distinct difference in retention rates between women with no deployments and those with two or more. The gap between those with 0 and 1 are closer yet still show a definite gap in retention. The survival curve for those with no deployment experience has a steep decline at roughly 8 CYOS and continues until 18 CYOS. Females with one deployment have a gradual decline with no distinct plateaus. The curve for females with two or more deployments is the best, dominating 0 and 1. When comparing the numbers at risk those with 2+ lose 15% by 10 CYOS, 1 loses 35%, and 0 loses 59%.

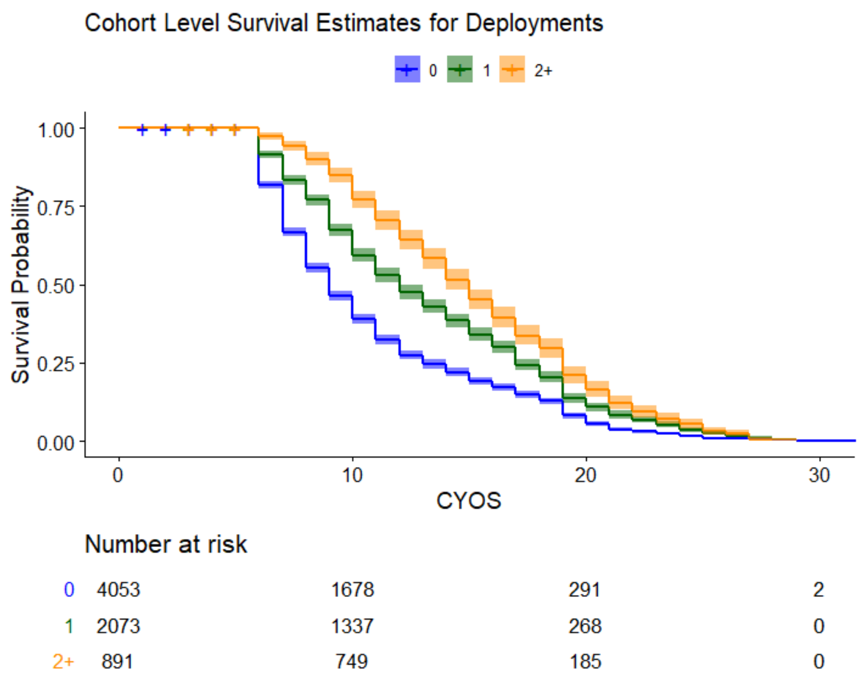

Figure 14. Kaplan-Meier Curve for Deployments

Marital status survival estimates are captured in Figure 15. Married female estimates weakly dominate estimates for women who were previously married. Both survival curves dominate single female officers. When analyzing risk numbers, the percentage of single females who separate from the AF is 58.2%. Roughly 43% of married women separate by 10 CYOS and approximately 38% leave who were previously married. Overall, single female officers are leaving at a higher rate than others. Understanding the retention behavior of single female officers may provide insight to create programs aimed at attracting and retaining them.

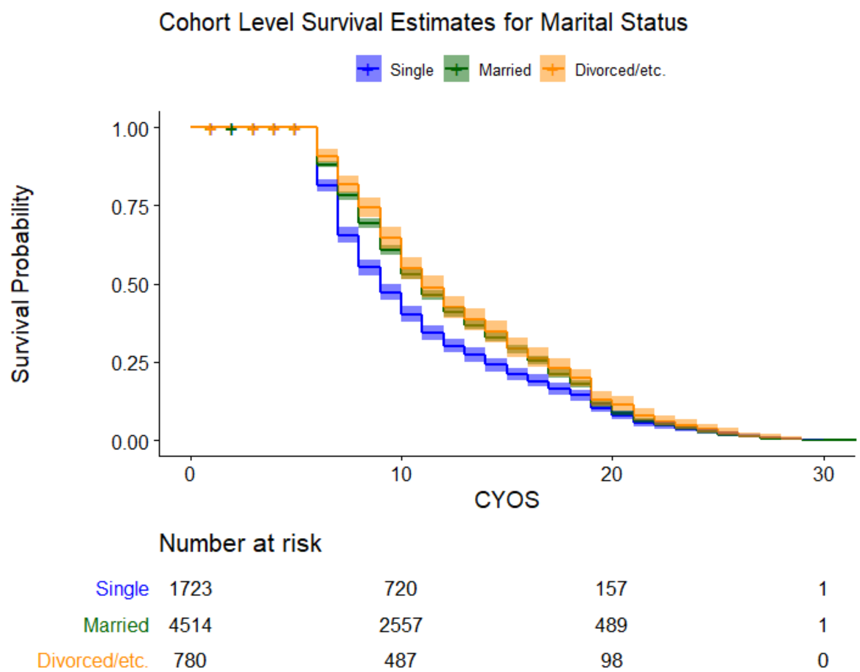

Figure 15. Kaplan-Meier Curve for Marital Status

The Kaplan-Meier estimates for prior service are shown in Figure 16. There are no distinct differences in retention behaviors between prior service and non-prior service members. Women with no prior service have slightly higher estimates. Women with 10 years of prior service would begin retiring at 10 CYOS, which may be impacting the lower probabilities shown from 10 to 20 CYOS.

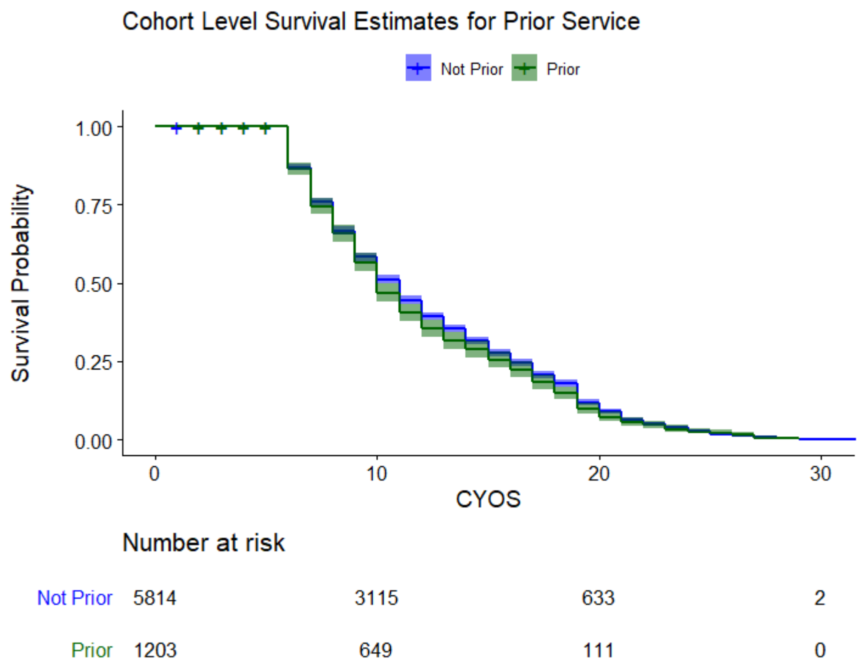

Figure 16. Kaplan-Meier Curve for Prior Service

Figure 17 shows the survival probabilities based on race. Confidence intervals are not provided for this estimate to provide a clear display of all retention rates. It should be noted that confidence intervals were tight around each survival curve, signifying sufficient data is provided for analysis. Female officers in this group with the lowest retention rate are those of Asian decent between 10-12 CYOS. From rougly 13 to 18 CYOS black females are shown to have the highest retention rates. The most notable information provided by this estimate are the risk numbers provided over the course of 30 CYOS. At 30 CYOS, two females achieved ranks within the general officer section. Roughly 8% of women represent officers "at brigadier general (O-7) or higher" (Keller et al., 2018). Although the data does not contain every female officer's career record, the low number of women serving at 30 CYOS shows how many remain past 20 CYOS.

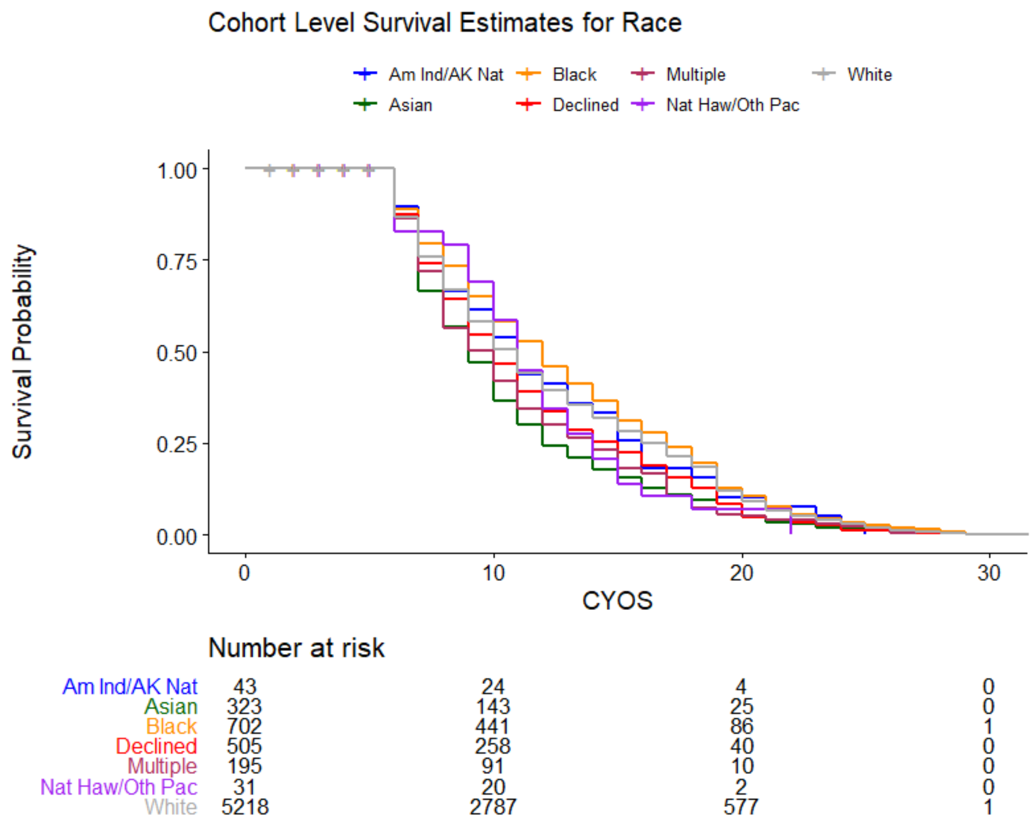

Figure 17. Kaplan-Meier Curve for Race

Kaplan-Meier estimates for a spouse's career category are represented in Figure 18. Overall, women married to spouses working in "other" occupations (those in the civilian sector and non-DoD federal workers) cross with women married to Reserve/Air National Guard (noted as RANG on the plot) members at about the 8 CYOS. However, those married to "other" become the dominant curve until roughly 26 CYOS.

Dual military women have the lowest retention behavior from 10 CYOS to 20 CYOS. This group has a lower retention when compared to the other married members. Roughly 53% of dual military separate from the military between approximately 7 to 10 CYOS. Some issues related to lower retention rates for dual military females are affected by extended periods of time spent separated from their spouses (and children); separate assignment locations, back-to-back deployments, and numerous business trips are just a few.

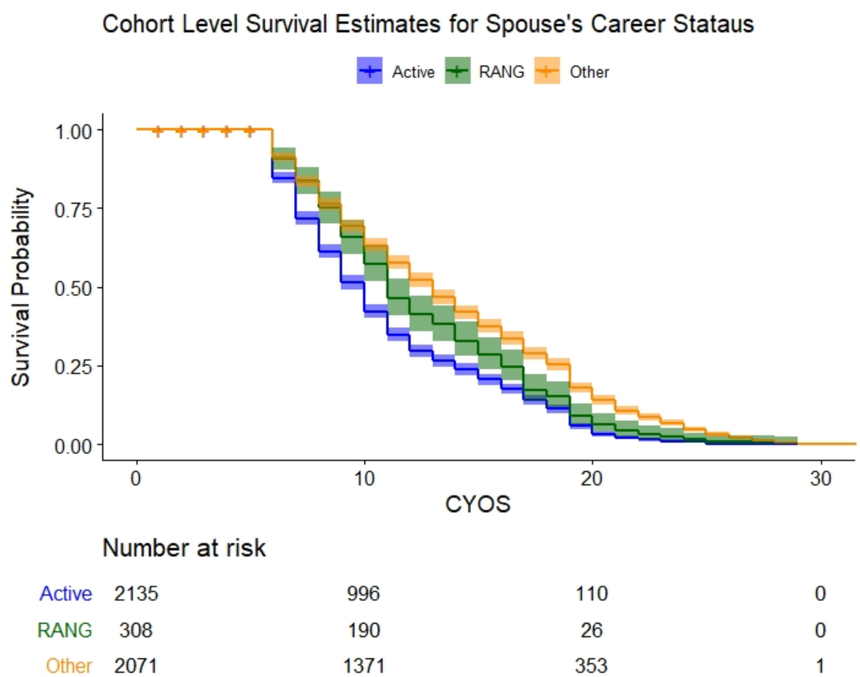

Figure 18. Kaplan-Meier Curve for Spouse's Career Category

4.2.5 Analysis at the Developmental Category Level

Table 8 displays the explanatory variables used for survival analysis by developmental category. The number of deployments tends to be the most prominent element across all categories. Estimates based on deployments for all categories are discussed first as they carry similar behaviors. Kaplan-Meier estimates for all other variables will be discussed individually next. This study did not distinguish the difference between student and non-student rated officers. Therefore, analysis conducted on the rated officers does not identify specific retention behaviors of those in or out of training.

Titles for each category are shortened to acronyms for ease of use in this section to: Air Operations and Special Warfare = AOSW, Information Warfare = IW, Force

Management = FM, and Combat Support = CS.

Table 8. Factors Significant to a Developmental Category's Regression Model

Developmental Category	Significant Factors
Air Ops & Special Warfare	Dependents, Deployments
Information Warfare	Spouse's Career, Deployments
Combat Support	Dependents, Deployments, Marital Status, Prior Service
Force Modernization	Deployments, Marital Status, Spouse's Career

Kaplan-Meier estimates based on deployments for all developmental categories are shown in Figure 19. No deployments = 0, one deployment = 1, and two or more = 2+. Women with more than 2 deployments are considered to have higher retention rates across all AFSC categories. However, confidence bands for AOSW and FM pose issues related to estimate accuracy. Generally, wider confidence intervals are related to lack of data. When reviewing risk numbers in these groups, 46 data points are in AOSW and 24 in FM. Additionally, results for deployments may be confounded with AFSC types included in each developmental category; this type of relationship was discussed in Section 4.2.4. Finally, including more categories to determine if a break in retention behavior exists seems intractable due to the high correlation between number of deployments and AFSC type.

Confidence bands for IW and CS related to 2+ deployments do not present the same characteristics, implying there are no issues with accuracy. Women with 2+ deployments in IW have the highest retention rates overall with over 75% survival probability from approximately 8-12 CYOS.

The retention curve for females with no deployments in the AOSW group are shown to have a steep decline from about 7 to 10 CYOS. A loss of approximately 52% of the population occurs between 0-10 CYOS. The trend continues to drop with a loss of 90% between 10 to 20 CYOS. Females in IW and CS have similar retention curve behaviors, but probabilities for IW are slightly higher. For example between 0-10 the percentage of those leaving without deployment experience in IW is 64% and 67% for CS. When looking at the time between 10 to 20 CYOS these percentages increase to approximately 78% and 77%, respectively. This means women in combat support positions are leaving at higher rates than those in non-combat support positions. Women in the FM career fields with no deployment status have higher retention rates than the other developmental categories. Roughly 48% of female officers separate in FM between 0-10 CYOS. The percent of those leaving the AF is higher than IW and CS between 10-20 CYOS, at 80%. AFSCs in this category may not have high deployment rates, leading to lower separation percentages in this time-frame.

Possible life events affecting retention behaviors for females with no deployment experience may be pregnancies and perceptions associated with it. Keller et al. (2018) found "perceived stigma associated with pregnancy in the Air Force." Female officers within the survey group "described a perception by leadership and peers that female officer are not pulling their weight and others will have to pick up the slack of their workload when they are on maternity leave." Additional research including survey data and maternity leave are helpful to determine exact reasons why women without deployment experience have generally lower retention behaviors.

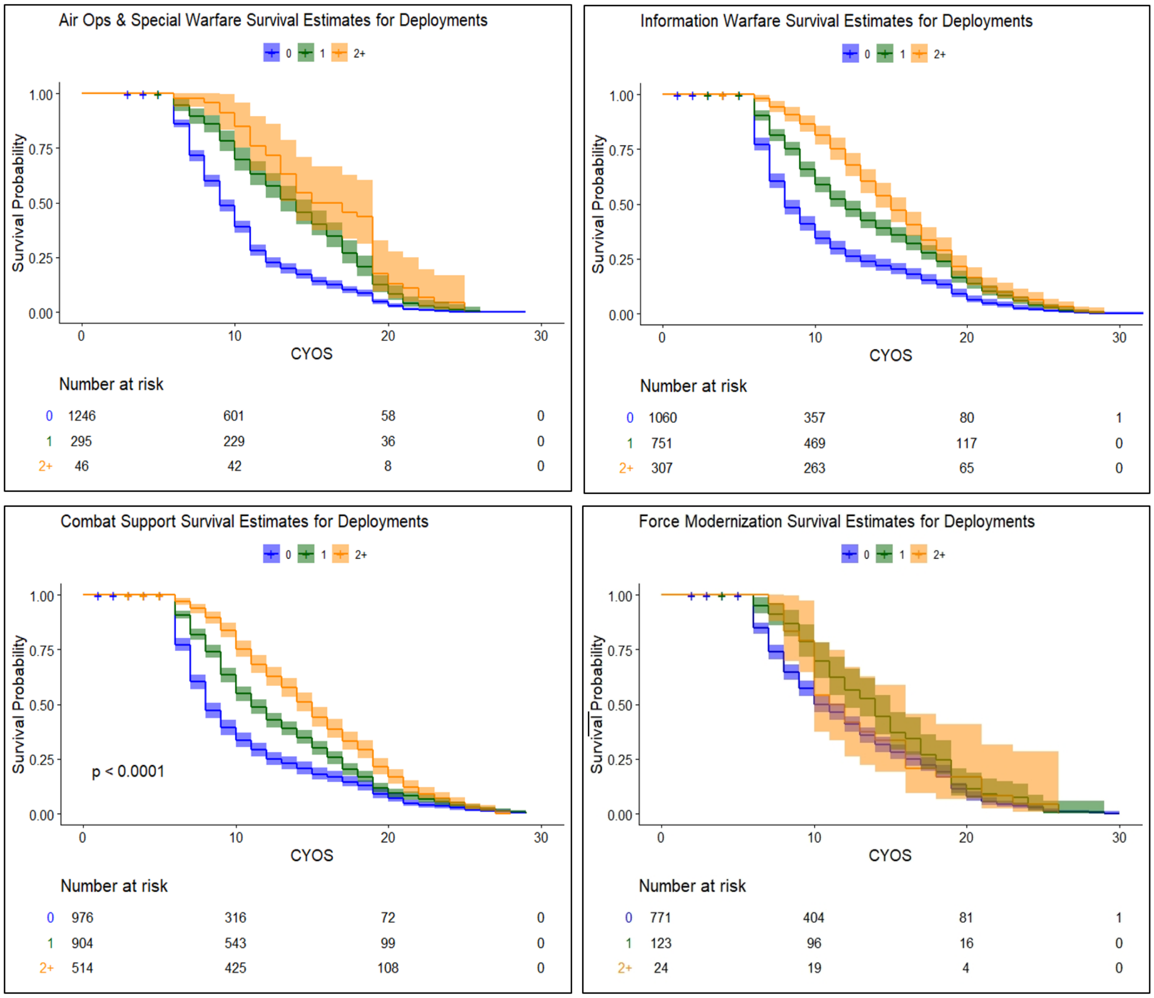

Figure 19. Kaplan-Meier Curves for Deployments

Figure 20 shows survival curves based on dependents for AOSW and CS. This result reflects conclusions based on the information in figure 6. Female officers with dependents have a dominant retention curves to those without in both cases. Women without dependents have a steep decline from 7 CYOS to about 10 CYOS. By 10 CYOS, their retention probability drops below 25%. This behavior was also recognized at the cohort level with odds ratios and survival analysis (reference Figures 2 and 13).

A lack of interest in health care benefits and job security may not be as appealing to this population; both of which may be the reason for higher retention rates in those with dependents. Women with dependents may be more inclined to stay to continue providing for their families. Those without dependents have the ability to search for jobs in the civilian sector with less financial or healthcare concerns.

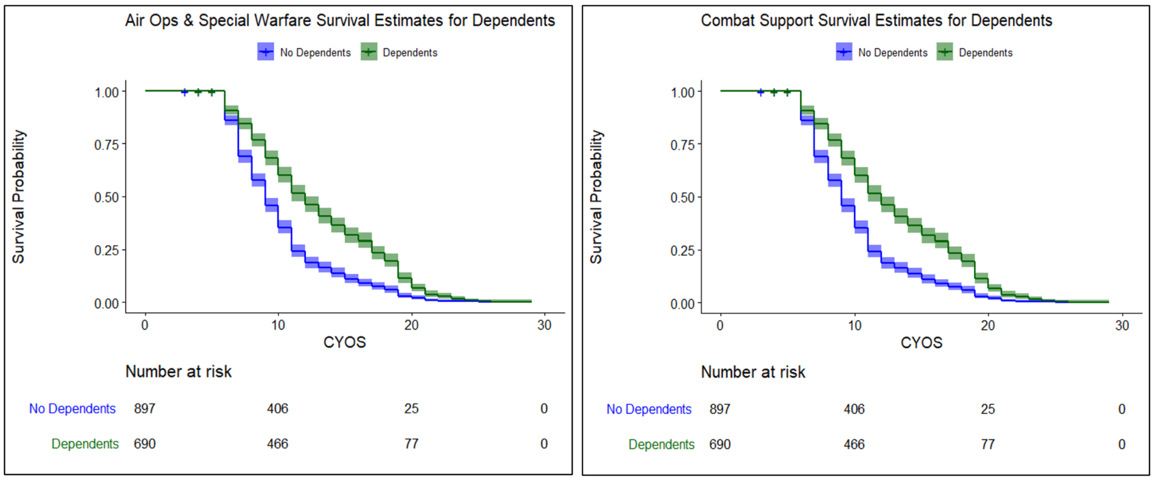

Figure 20. Kaplan-Meier Curve for Dependents (AOSW and CS)

Survival probabilities in Spouse's Career Category for IW and FM are presented in Figure 21. Confidence intervals are not provided for this estimate to provide a clear display of all retention rates. None of the confidence bands signified inaccuracy to the data. Analysis of effects provided at the CYOS level (covered in section 4.1.5) determined Spouse's Career Category was significant at 0-6 and 20-22 CYOS. Furthermore, when examined at the developmental category for IW, RANG had the best likelihoods of retaining in 0-6 and 4-8 CYOS. Women married to "other" had the best likelihoods from 12-19 to 20-22 CYOS. Retention curves for IW corroborate these conclusions. The retention curve for RANG crosses over "other" between 8 and 10 CYOS but "other" dominates from 10 to approximately 25 CYOS.

Women with spouse's in "other" jobs for FM are also shown to have the highest retention rates. Their curve dominates all other levels from about 8 to 25 CYOS. Survival curves for RANG and dual military cross at multiple points between 7 and 19 CYOS. At about 19 to 22 CYOS dual military are shown to separate, or retire, more than RANG. This may suggest dual military females are choosing to retire to allow their spouse's career to continue. When asked about the influence their spouse's career has on their decision to separate, dual military female officers felt, "the spouse whose military career generally suffered was the female rather than the male" (Keller et al., 2018). This also leads to lower representation of female officers in the general officer ranks.

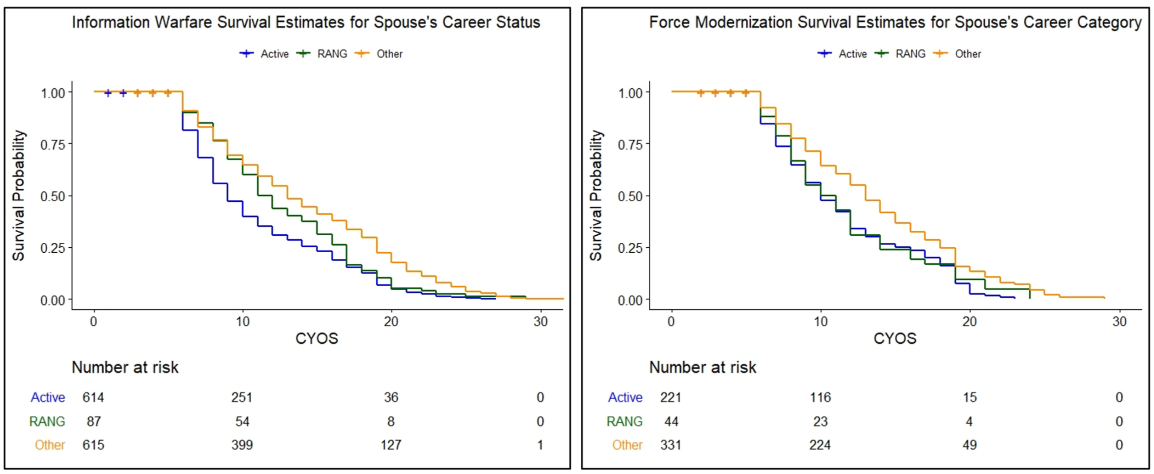

Figure 21. Kaplan-Meier Curve for Spouse's Career Category (IW and FM)

Figure 22 shows retention curves based on marital status for female officers in CS. Contrary to results provided in Figure 10, single female officers are shown to have the worst retention behavior. By 10 CYOS 59% of single female officers are shown to attrite. In comparison, 44% of married women and 37% of previously married women separate by 10 CYOS. Retention curves for married and previously married officers cross at multiple points between 10-20 CYOS, but remain higher than their single colleagues.

Overall, single female officers are leaving at a higher rate than others. Single female officers have raised concerns related to delaying their personal lives (e.g. marriage, children, etc.) as reasons to leave the military. They also express the lack of programs for single military members, as most programs provided on bases are family oriented

(Keller et al., 2018). Further research examining underlying issues affecting the single female population may aid in developing programs aimed at retaining them.

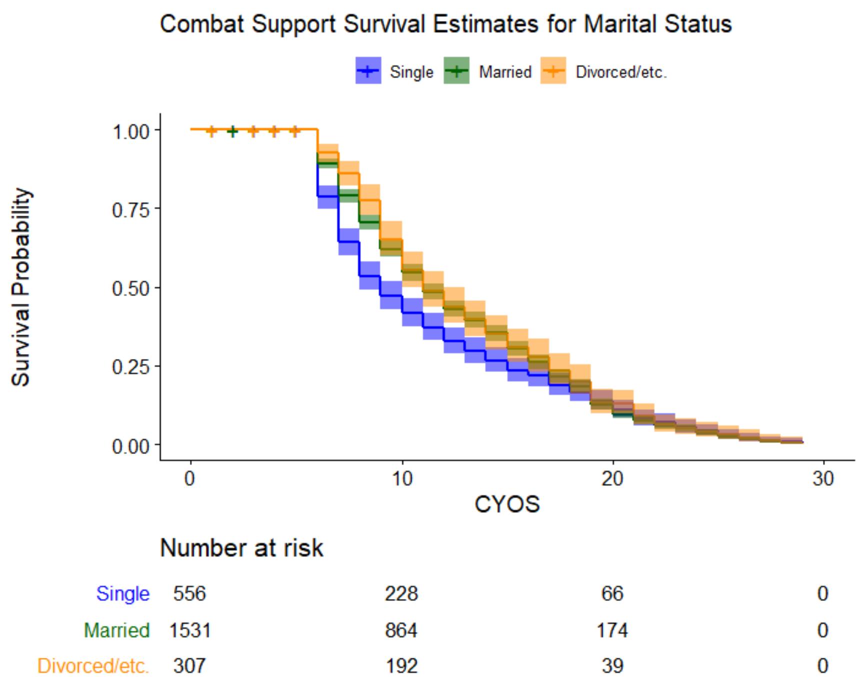

Figure 22. Kaplan-Meier Curve for Marital Status (CS)

4.2.6 Cox Proportional Hazards Regression

Cox proportional hazards regression is used to assess the effects of the explanatory indicators on female officer survival rates. Hazard ratios for significant variables are also examined at the cohort and developmental category levels. Step-wise regression from the coxph() function determined prior service was not significant. Table 9 shows the Wald Chi-Square p-values for significant factors at the cohort level.

Table 9. Cox Proportional Covariate P-Values for Cohort level. Entries with yellow highlight are found significant in the analysis.

Covariate	P-Value
DG	<0.0001
Dependents	<0.0001
Deployments	<0.0001
Marital Status	0.0406
Spouse's Career	<0.0001
Race	0.0004
Obs	**7017**

Hazard ratios are graphed to indicate whether a covariate is positively or negatively associated with the event probability. Covariates with ranges identified as "reference" are equivalent to variables set as baseline cases in the Odds Ratio analysis. Figure 23 shows married females, females of Asian decent and those identified to be more than one race have higher hazards of separating from the AF. Women with the lowest hazard ratio are those with 2+ deployments followed by those with one deployment, and women with dependents. Confidence intervals for each race are wide, signifying issues in the accuracy of this estimated covariate.

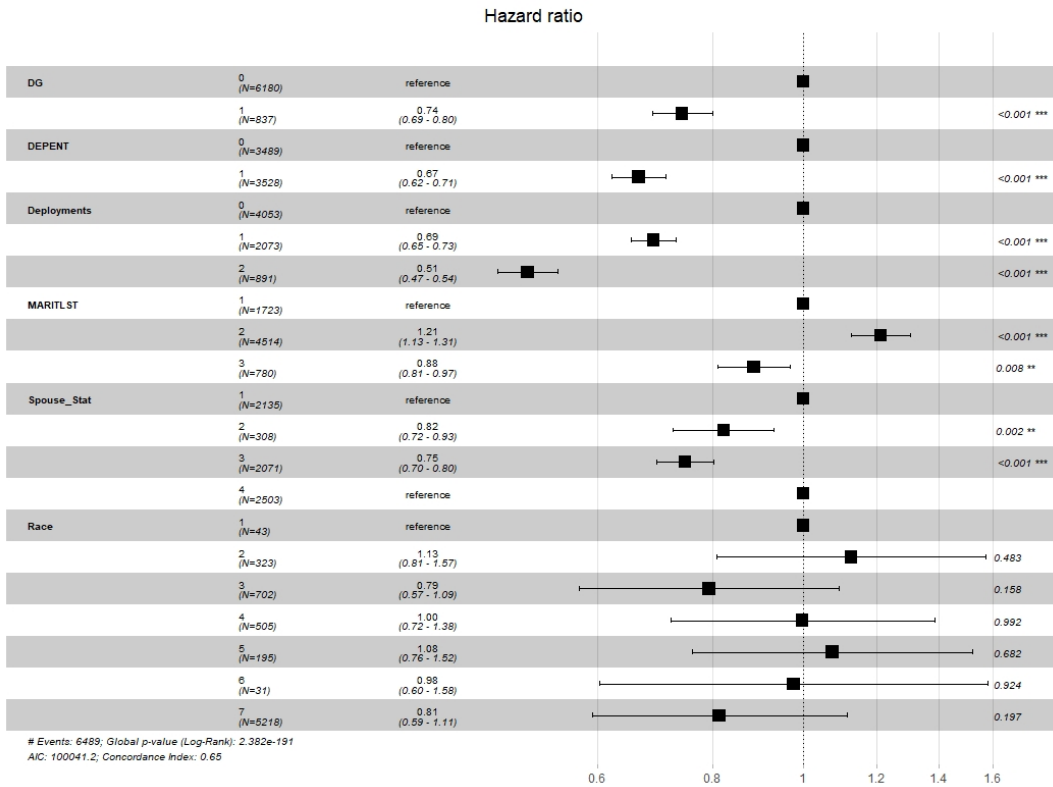

Figure 23. Hazard Ratios for Covariates at the Cohort Level

Explanatory variables for each developmental category are shown in Table 10. Analysis began with the most commonly significant variables resulting from Logistic Regression (reference Section 4.1.6 and Table 8). Results from the step-wise regression shows the final variables shown to be significant for each category.

Table 10. Cox Proportional Covariate p-values for Developmental Categories. Entries with yellow highlight are found significant in the analysis.

Covariate	**Developmental Categories**			
	AOSW	**IW**	**CS**	**FM**
Dependents	<0.001	—	<0.0001	—
Deployments	<0.001	0.0002	<0.0001	0.0343
Marital Status	—	—	—	0.0004
Spouse's Career	—	<0.0001	—	0.0123
Prior Service	—	—	—	—
Obs	**1587**	**2118**	**2394**	**918**

Figure 24 shows the hazard ratios for Air Ops and Special Warfare. Confidence intervals are wide for female officers with 2+ deployments, which supports the results from the Kaplan-Meier curves. This confirms there are issues with the accuracy of the data provided.

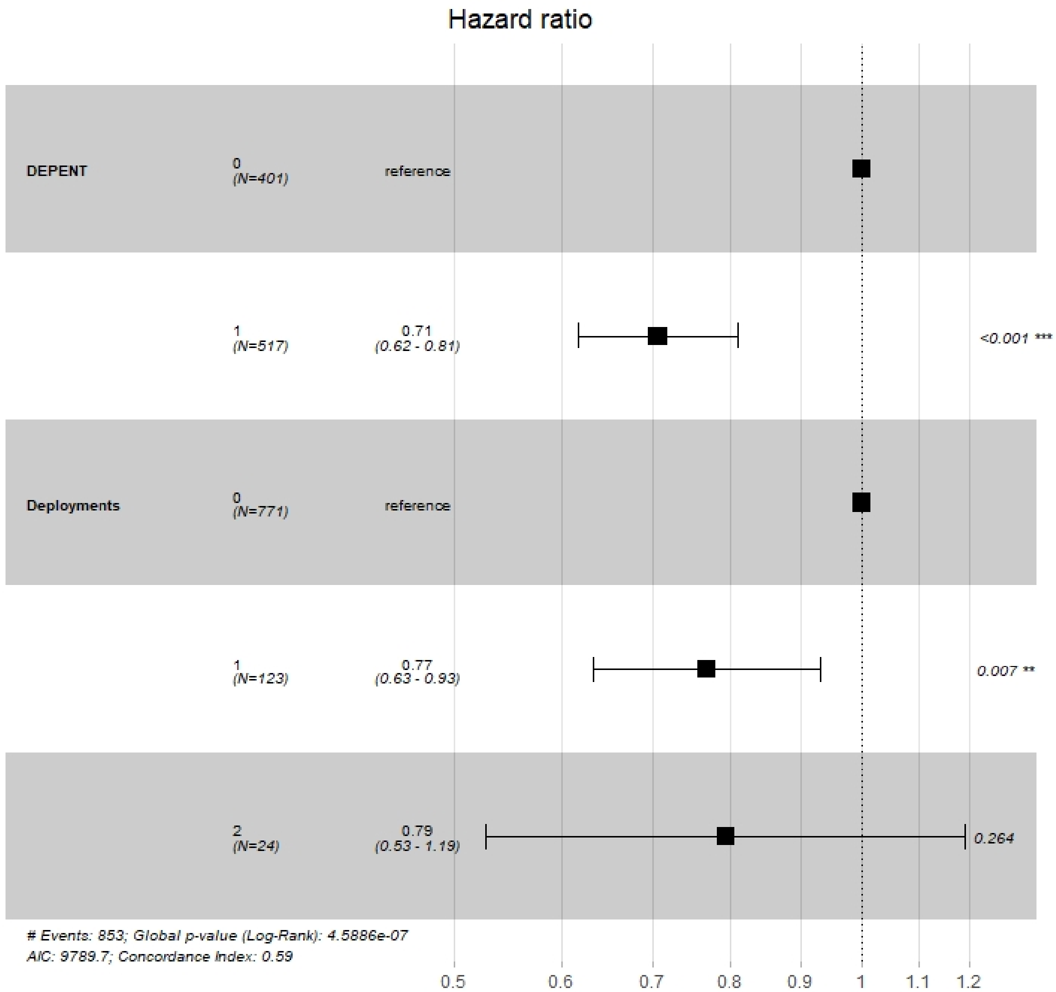

Figure 24. Hazard Ratios for Covariates for AOSW

Figure 25 captures the hazard ratios for Information Warfare. Reference variables are not present in the figure since the p-values for the other levels are significantly less than 0.0001. Women with 2+ deployments are shown to have the best retention ratio.

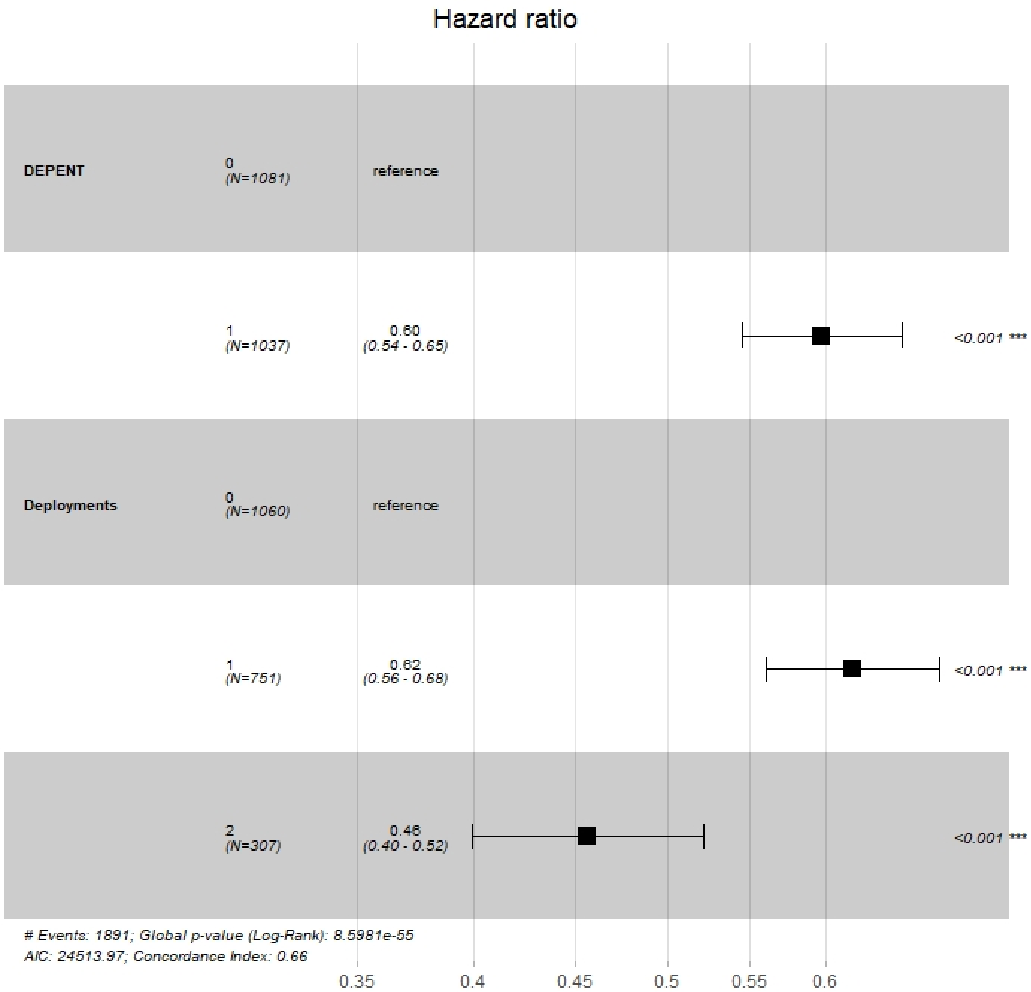

Figure 25. Hazard Ratios for Covariates for IW

Hazard ratios for Combat Support are shown in Figure 26. There are no issues with accuracy for any of the covariates present in the graph. Results for this category coincide with results provided in the Odds Ratios and Kaplan-Meier estimates. Female officers with 2+ deployments are a good prognostic factor (HR<1), followed by with females with dependents and 1 deployment. Bad prognostic factors are prior enlisted and married female officers.

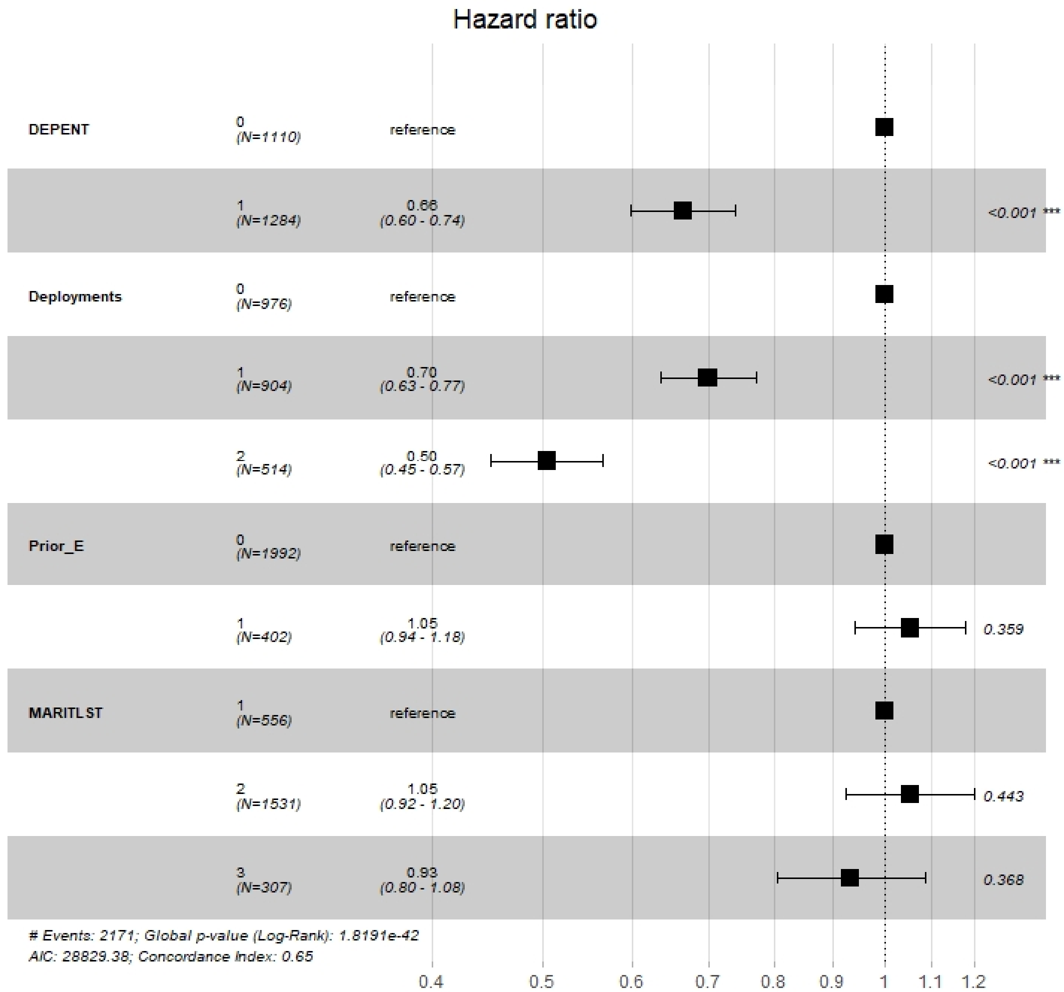

Figure 26. Hazard Ratios for Covariates for CS

Finally, ratios for Force Modernization are displayed in 27. Confidence intervals for women with 2+ deployments confirm accuracy issues do exist with the element, supporting results found in Figure 19. Married women (noted under "Spouse_Stat") also display concerns with accuracy of the data. Previously married women have the best likelihoods of retaining.

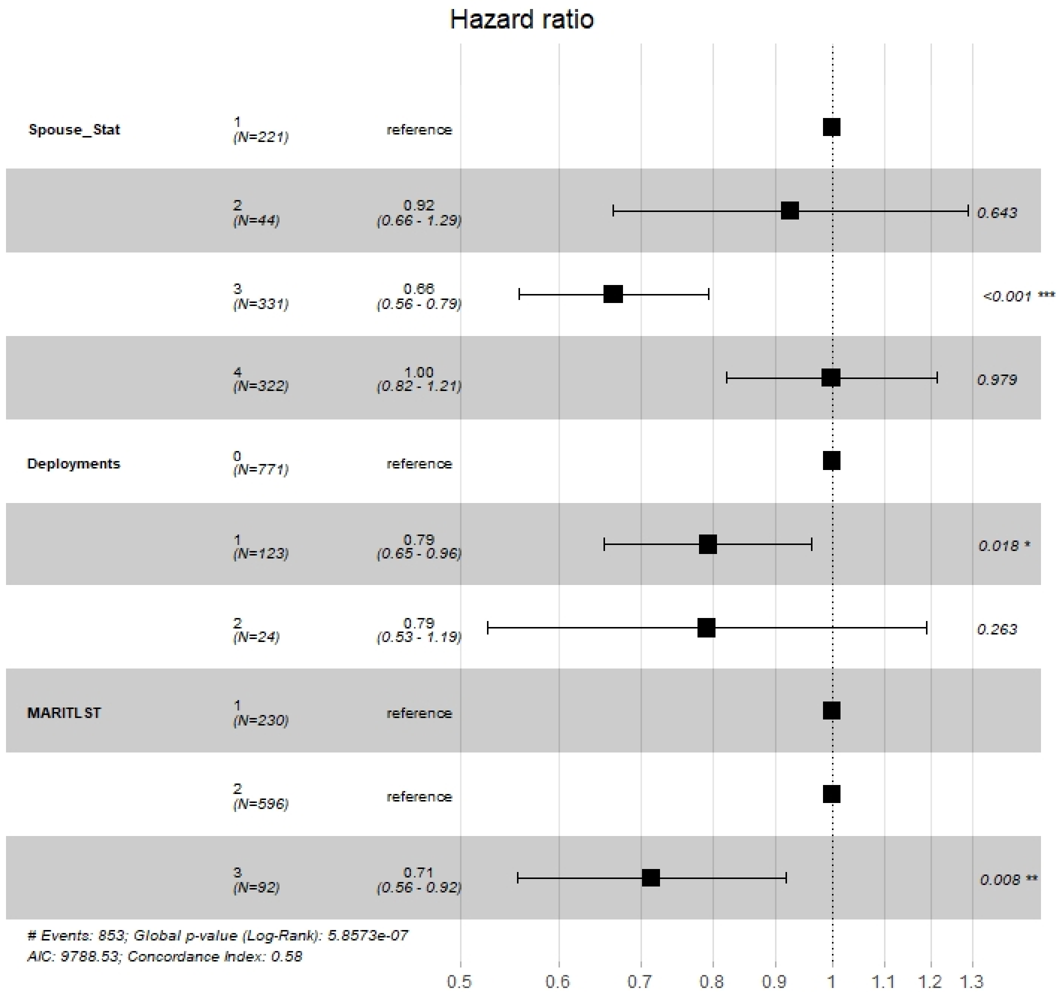

Figure 27. Hazard Ratios for Covariates for FM

V. Conclusions and Recommendations

5.1 Limitations

This study was limited to characterizing retention behavior based on MilPDS demographic data. Access to additional data (e.g. duration of maternity leave, age of dependents, BRS opt-in, etc.) will provide more insight on factors influencing women to separate. The findings in this research has created a foundation for further analysis in female officer retention.

Additionally, analysis of developmental categories did not capture unique nuances related to each AFSC. For example, deployment requirements for an operational research analyst (15A) differs from a cyber analyst (17X). Conclusions for each category have the possibility of being skewed if contains a higher population of any one AFSC, dominating the remaining AFSCs in the group. Therefore, findings are concluded on the entire category. To understand what affects a specific AFSC will require analysis to be conducted at that AFSC level.

Finally, the MilPDS system is prone to errors such as incorrect inputs, deletions, or glitches. Backups of the system is conducted on a routine basis to mitigate these faults, along with processes set in place to fill in missing data. Assumptions were also made to fill in missing information such as the spouse's career category of a married female.

5.2 Follow-On Research

This research is the first conducted on solely focusing on AF female officer retention behavior through the use of logistic regression and survival analysis. Future research involving other methodologies (e.g. forecasting, simulation, etc.) on female officer retention could be used to predict future behavior or explain trends. Including economic, political, and AFSC specific factors in analysis will provide more sufficient results. Factors reflecting recent changes to AF policy, such as extended maternity leave and combat related AFSCs, will also inform leadership of the affects these programs have on female officers.

Another program to be included in follow-on research would be the BRS. This change was not covered as it was placed in affect January 2018. This only left one year of data to analyze, which is not enough to provide an in-depth affect on retention behaviors. Future analysis can compare attrition rates between BRS to legacy retirement members. Including the BRS may have positive influence as it introduces a mid-career bonus pay. This information is imperative as results from this study have shown retention decline between 8-12 CYOS.

Studying the female enlisted force is another approach that continues focus on improving female retention. It is unknown whether elements affecting the officer population also affect enlisted members. Thus, examining this population will aid in determining significant factors and their effect on retention.

5.3 Conclusion

Women remain underrepresented in leadership positions in the Air Force. Diversity in an organization is important because of the value that multiple perspectives can bring to the group. Antecedent studies have analyzed officer and enlisted sectors as a whole to understand retention behaviors. However, statistical analysis focused solely on the female population, at the officer or enlisted levels, has not been performed. This study provides further understanding of what influences low retention rates of female officer members.

The purpose of this research was to determine factors significant to female officers and their retention behavior based on these elements. Results have shown that number of dependents, number of deployments, marital status, and spouse's career category are the most influential on female retention. These factors vary when analyzed at the developmental category level. Dependents affected the rated career fields and spouse's career category affecting information warfare. Combat support was affected the most by number of dependents, marital status, and prior service. Force Modernization was also influenced by marital status and spouse's career category. The most prominent factors affecting all categories was number of deployments. It should be noted the number of deployments may be confounded with AFSC type, thus influencing results for each developmental category. Finally, contrary to previous studies including male and females (Schofield, 2015; Franzen, 2017), source of commissioning had no bearing on female attrition.

Common patterns were recognized throughout analysis conducted at the cohort level. Most notably, survival probabilities for populations with the lowest retention rates tended to separate between the 8-15 CYOS. This suggests incentives currently in place aimed for those between these years of service may not be influencing female officers to remain. Single female officers tended to separate from the military after their initial service commitment at a higher rate than those married or previously married. This conclusion also applies to women without dependents, all of which are recognized as single in MilPDS. When analyzing marital status, dual military females attrited more than those married to partners in the reserve or civilian sectors. Those with no deployment experience also left the military at a higher rate.

Results for each of the developmental categories contained all the same factors negatively affecting the same populations found at the cohort level. The same can be said for those with the highest retention behavior.

Populations with the highest retention behaviors were women with two or more deployments, and have dependents. Women married to those in the non-DoD or civilian sector and reserves had very similar retention behaviors. However, those married to civilians showed to remain in more than others past 20 YOS.

Findings from this study supports HAF/A1's work to increase the number of women serving, improve diversity, inclusion, and equity. Results from this study also supports the diversity and inclusion council, enacted by the Air Force chief of staff, who are currently implementing changes to improve recruitment and retainment.

5.4 Recommendations

Results provided in this study revealed the 8-12 CYOS is a critical time-span in which separation occurs the most for female officers. SRB opportunities should be evaluated by HAF/A1 to determine whether members closer to the middle of their careers are taking advantage of these programs. Extending SRB service commitments should also be re-evaluated that focuses on retaining members during the time attrition rates are highest. Data related to the BRS continuation pay should also be examined to discern its affects on female officer retention. This information can also be used to compare attrition rates between those who opt-ed into the BRS vs those with the legacy retirement.

Policies supporting the military family is another area that should be under review. Literature found resources made available to military members and their families do not cater towards the military female member. Dynamics of today's modern family differs from that of older models containing a working father and stay-at-home mother. Although women married to members in the civilian sectors displayed higher retention rates, this does not conclude that female representation will increase. It is recommended HAF/A1 conduct research on military members and male military spouses specifically to assess their needs to create effective programs that will improve marital stability, life, and female retention. Additionally, needs of female members married to other military members may vary from females with male spouses. Research conducted on dual military members and their needs is also recommended to provide insight on creating programs to support females married to other military members.

Appendix A: R Code for Logistic Regression

```
#The following code contains logistic regression analysis code used for
#this research
#Jessica Astudillo

#download necessary packages for calculations
install.packages("rsq")
install.packages("oddsratio")

#Call libraries needed for analysis
require(readxl)
library(dplyr)
library(tidyverse)
library(readxl)
library(modelr)
library(broom)
library(oddsratio)
library(MASS)
library(rsq)

#-----------------------------------------------------------------------

#Set the work space for the data
getwd() #This is to get the working directory

#Now set the working directory to the data file
setwd("C:/Users/User/Desktop/Astudillo_Backup/Data/Data in R")

#check if directory exists on your system
file.exists("C:/Users/User/Desktop/Astudillo_Backup/Data/Data in
Excel/offinv_annual/Female Only/Fiscal Year/Final Thesis Data")

#------------------------------------------------------------------------
#Read in necessary files and perform analysis on 2009 data

data_2009 = read_excel("C:/Users/User/Desktop/Astudillo_Backup/Data/
Data in Excel/offinv_annual/Female Only/Fiscal Year/Final Thesis
Data/FY2009_Thesis_v2.xlsx")

data_2010 = read_excel("C:/Users/User/Desktop/Astudillo_Backup/Data/
Data in Excel/offinv_annual/Female Only/Fiscal Year/Final Thesis
```

```
Data/FY2010_Thesis_v2.xlsx")

data_2011 = read_excel("C:/Users/User/Desktop/Astudillo_Backup/Data/
Data in Excel/offinv_annual/Female Only/Fiscal Year/Final Thesis
Data/FY2011_Thesis_v2.xlsx")

data_2012 = read_excel("C:/Users/User/Desktop/Astudillo_Backup/Data/
Data in Excel/offinv_annual/Female Only/Fiscal Year/Final Thesis
Data/FY2012_Thesis_v2.xlsx")

data_2013 = read_excel("C:/Users/User/Desktop/Astudillo_Backup/Data/
Data in Excel/offinv_annual/Female Only/Fiscal Year/Final Thesis
Data/FY2013_Thesis_v2.xlsx")

data_2014 = read_excel("C:/Users/User/Desktop/Astudillo_Backup/Data/
Data in Excel/offinv_annual/Female Only/Fiscal Year/Final Thesis
Data/FY2014_Thesis_v2.xlsx")

data_2015 = read_excel("C:/Users/User/Desktop/Astudillo_Backup/Data/
Data in Excel/offinv_annual/Female Only/Fiscal Year/Final Thesis
Data/FY2015_Thesis_v2.xlsx")

data_2016 = read_excel("C:/Users/User/Desktop/Astudillo_Backup/Data/
Data in Excel/offinv_annual/Female Only/Fiscal Year/Final Thesis
Data/FY2016_Thesis_v2.xlsx")

data_2017 = read_excel("C:/Users/User/Desktop/Astudillo_Backup/Data/
Data in Excel/offinv_annual/Female Only/Fiscal Year/Final Thesis
Data/FY2017_Thesis_v2.xlsx")

data_2018 = read_excel("C:/Users/User/Desktop/Astudillo_Backup/Data/
Data in Excel/offinv_annual/Female Only/Fiscal Year/Final Thesis
Data/FY2018_Thesis_v2.xlsx")

data_2019 = read_excel("C:/Users/User/Desktop/Astudillo_Backup/Data/
Data in Excel/offinv_annual/Female Only/Fiscal Year/Final Thesis
Data/FY2019_Thesis_v2.xlsx")

#-----------------------------------------------------------------------
#-----------------------------------------------------------------------
#-----------------------------------------------------------------------
#-----------------------------------------------------------------------
```

```
#Now perform logistic regression on each commissioning year group:
0 to 20+ based on each CYOS group

#--------------------------------------------------------------------

# Years 0 to 6

#2009---------------------------------------------------------------
YG1_Data_2009_1 = subset(data_2009,CYOS_EFY >= 6 & Retain >= 0, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

YG1_Data_2009_2 = subset(data_2009,CYOS_EFY <= 5 & Retain < 1, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#combine both files
CYOS_0to6_2009_Data = rbind(YG1_Data_2009_1,YG1_Data_2009_2)

#2010---------------------------------------------------------------

YG1_Data_2010_1 = subset(data_2010,CYOS_EFY >= 6 & Retain >= 0, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

YG1_Data_2010_2 = subset(data_2010,CYOS_EFY <= 5 & Retain < 1, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#combine both files
CYOS_0to6_2010_Data= rbind(YG1_Data_2010_1,YG1_Data_2010_2)

#2011---------------------------------------------------------------
YG1_Data_2011_1 = subset(data_2011,CYOS_EFY >= 6 & Retain >= 0, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

YG1_Data_2011_2 = subset(data_2011,CYOS_EFY <= 5 & Retain < 1, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#combine both files
CYOS_0to6_2011_Data = rbind(YG1_Data_2011_1,YG1_Data_2011_2)

#2012---------------------------------------------------------------
```

```
YG1_Data_2012_1 = subset(data_2012,CYOS_EFY >= 6 & Retain >= 0, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

YG1_Data_2012_2 = subset(data_2012,CYOS_EFY < 6 & Retain < 1, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#combine both files
CYOS_0to6_2012_Data = rbind(YG1_Data_2012_1,YG1_Data_2012_2)

#2013------------------------------------------------------------
YG1_Data_2013_1 = subset(data_2013,CYOS_EFY >= 6 & Retain >= 0, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

YG1_Data_2013_2 = subset(data_2013,CYOS_EFY < 6 & Retain < 1, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#combine both files
CYOS_0to6_2013_Data = rbind(YG1_Data_2013_1,YG1_Data_2013_2)


#2014------------------------------------------------------------
YG1_Data_2014_1 = subset(data_2014,CYOS_EFY >= 6 & Retain >= 0, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

YG1_Data_2014_2 = subset(data_2014,CYOS_EFY < 6 & Retain < 1, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#combine both files
CYOS_0to6_2014_Data = rbind(YG1_Data_2014_1,YG1_Data_2014_2)

#2015------------------------------------------------------------
YG1_Data_2015_1 = subset(data_2015,CYOS_EFY >= 6 & Retain >= 0, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

YG1_Data_2015_2 = subset(data_2015,CYOS_EFY < 6 & Retain < 1, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
```

```
Deployments:Prior_E))

#combine both files
CYOS_0to6_2015_Data= rbind(YG1_Data_2015_1,YG1_Data_2015_2)


#2016-----------------------------------------------------------
YG1_Data_2016_1 = subset(data_2016,CYOS_EFY >= 6 & Retain >= 0, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

YG1_Data_2016_2 = subset(data_2016,CYOS_EFY < 6 & Retain < 1, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#combine both files
CYOS_0to6_2016_Data = rbind(YG1_Data_2016_1,YG1_Data_2016_2)

#2017-----------------------------------------------------------
YG1_Data_2017_1 = subset(data_2017,CYOS_EFY >= 6 & Retain >= 0, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

YG1_Data_2017_2 = subset(data_2017,CYOS_EFY < 6 & Retain < 1, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#combine both files
CYOS_0to6_2017_Data = rbind(YG1_Data_2017_1,YG1_Data_2017_2)

#2018-----------------------------------------------------------
YG1_Data_2018_1 = subset(data_2018,CYOS_EFY >= 6 & Retain >= 0, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

YG1_Data_2018_2 = subset(data_2018,CYOS_EFY < 6 & Retain < 1, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#combine both files
CYOS_0to6_2018_Data = rbind(YG1_Data_2018_1,YG1_Data_2018_2)
```

```
#2019-----------------------------------------------------------
YG1_Data_2019_1 = subset(data_2019,CYOS_EFY >= 6 & Retain >= 0,select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

YG1_Data_2019_2 = subset(data_2019,CYOS_EFY < 6 & Retain < 1, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#combine both files
CYOS_0to6_2019_Data = rbind(YG1_Data_2019_1,YG1_Data_2019_2)

#Combine all data frames containing CYOS 0-6 now-------------------
CYOS_0to6_Combined = rbind(CYOS_0to6_2009_Data,CYOS_0to6_2010_Data,
CYOS_0to6_2011_Data,CYOS_0to6_2012_Data,CYOS_0to6_2013_Data,
CYOS_0to6_2014_Data,CYOS_0to6_2015_Data,CYOS_0to6_2016_Data,
CYOS_0to6_2017_Data,CYOS_0to6_2018_Data,CYOS_0to6_2019_Data)

#Reverse order of data to ensure the most recent record is obtained
when removing duplicates
revOrd_CYOS_0to6_Comb = CYOS_0to6_Combined[order(
CYOS_0to6_Combined$SSAN,CYOS_0to6_Combined$CYOS_EFY,
decreasing=TRUE),]

#Now remove the duplicated information (records containing older
information) by identified number:
CYOS_0to6_NoDups = revOrd_CYOS_0to6_Comb
[!duplicated(revOrd_CYOS_0to6_Comb$SSAN),]

#Add a new column called 'Retained' with values = 0
CYOS_0to6_NoDups$Retained = 0

#Change all retained values to 1 if CYOS >= 6. Currently have
#0 values.
CYOS_0to6_Final = CYOS_0to6_NoDups %>% mutate(Retained=replace
(Retained,CYOS_EFY >=6, 1)) %>% as.data.frame()

#Now remove the 'Retain' column since 'Retained' column now has the
#updated information..
CYOS_0to6_Final$Retain = NULL

transform(CYOS_0to6_Final, Retained = as.numeric(Retained))
#transform to numeric value if necessary
```

```
#subset the data so it does not contain SSAN:
CYOS_0to6_LogRegData = CYOS_0to6_Final[c(3,4,6:9,11,12)]
write.csv(CYOS_0to6_LogRegData,"C:/Users/User/Desktop/Astudillo_Backup
/Data/Data in Excel/offinv_annual/Female Only/Fiscal Year/
Final Thesis Data/CYOS Info/CYOS_0to6_LogRegData.csv",row.names = TRUE)

#Perform logistic regression on the Final dataframe containing
#all fiscal years for CYOS 0 to 6:
LogReg_CYOS_0to6 = glm(Retained ~ ., data = CYOS_0to6_LogRegData,
family = "binomial")
summary(LogReg_CYOS_0to6)

rsq(LogReg_CYOS_0to6, adj = FALSE, type = 'kl')
rsq(LogReg_CYOS_0to6, adj = TRUE, type = 'kl')
confint(LogReg_CYOS_0to6)


#Perform Logistic Regression on the Developmental Categories
#based on CYOS 0 to 6--------------

#Air Operations and Special Warfare
AOSWData_Final = subset(CYOS_0to6_Final,Dev_Cat =="1", select =
c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,Prior_E,Retained))
AOSWData_CYOS_0to6 = AOSWData_Final[c(1:8)]
AOSWLR_CYOS_0to6 = glm(Retained ~ ., data = AOSWData_CYOS_0to6,
family = "binomial")
summary(AOSWLR_CYOS_0to6)
write.csv(AOSWData_CYOS_0to6,"C:/Users/User/Desktop/Astudillo_Backup/
Data/Data in Excel/offinv_annual/Female Only/Fiscal Year/Final Thesis
Data/Developmental Categories/AOSWDATA_CYOS_0to6.csv",row.names = TRUE)
confint(AOSWLR_CYOS_0to6)

#Info Warfare
IWData_Final = subset(CYOS_0to6_Final,Dev_Cat =="2", select =
c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,Prior_E,Retained))
IWData_CYOS_0to6 = IWData_Final[c(1:8)]
IWLR_CYOS_0to6 = glm(Retained ~ ., data = IWData_CYOS_0to6,
family = "binomial")
summary(IWLR_CYOS_0to6)
write.csv(AOSWData_CYOS_0to6,"C:/Users/User/Desktop/Astudillo_Backup/
Data/Data in Excel/offinv_annual/Female Only/Fiscal Year/Final Thesis
Data/Developmental Categories/IWDATA_CYOS_0to6.csv",row.names = TRUE)
```

```
confint(IWLR_CYOS_0to6)

#Combat Support
CSData_Final = subset(CYOS_0to6_Final,Dev_Cat =="3", select =
c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,Prior_E,Retained))
CSData_CYOS_0to6 = CSData_Final[c(1:8)]
CSLR_CYOS_0to6 = glm(Retained ~ ., data = CSData_CYOS_0to6,
family = "binomial")
summary(CSLR_CYOS_0to6)
write.csv(CSData_CYOS_0to6,"C:/Users/User/Desktop/Astudillo_Backup/
Data/Data in R/CSDATA_CYOS_0to6.csv",row.names = TRUE)
confint(CSLR_CYOS_0to6)

#Force Modernization
FMData_Final = subset(CYOS_0to6_Final,Dev_Cat =="4", select =
c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,Prior_E,Retained))
FMData_CYOS_0to6 = FMData_Final[c(1:8)]
FMLR_CYOS_0to6 = glm(Retained ~ ., data = FMData_CYOS_0to6,
family = "binomial")
summary(FMLR_CYOS_0to6)
write.csv(AOSWData_CYOS_0to6,"C:/Users/User/Desktop/Astudillo_Backup/
Data/Data in Excel/offinv_annual/Female Only/Fiscal Year/Final Thesis
Data/Developmental Categories/FMDATA_CYOS_0to6.csv",row.names = TRUE)
confint(FMLR_CYOS_0to6)

#----------------------------------------------------------------------
#----------------------------------------------------------------------

# Years 4 to 8


#2009------------------------------------------------------------

#If data is greater than or equal to 5 YOS then keep
YG2_Data_2009_1 = subset(data_2009,CYOS_EFY >= 5, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values greater than 8 YOS and retain
#values are 0 or 1
YG2_Data_2009_2 = subset(YG2_Data_2009_1,CYOS_EFY >= 8 & Retain >= 0,
select =c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
```

```
#Subset the data with values less than 7 YOS and retain = 0
YG2_Data_2009_3 = subset(YG2_Data_2009_1, CYOS_EFY <=7 & Retain < 1,
select =c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,
Race,Deployments:Prior_E))

#Combine both files
CYOS_4to8_2009_Data= rbind(YG2_Data_2009_2,YG2_Data_2009_3)

#2010-----------------------------------------------------------------

#If data is greater than or equal to 5 YOS then keep
YG2_Data_2010_1 = subset(data_2010,CYOS_EFY >= 5, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values greater than 8 YOS and retain
#values are 0 or 1
YG2_Data_2010_2 = subset(YG2_Data_2010_1,CYOS_EFY >= 8 & Retain >= 0,
select=c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values less than 7 YOS and retain = 0
YG2_Data_2010_3 = subset(YG2_Data_2010_1, CYOS_EFY <=7 & Retain < 1,
select=c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Combine both files
CYOS_4to8_2010_Data= rbind(YG2_Data_2010_2,YG2_Data_2010_3)


#2011-----------------------------------------------------------------

#If data is greater than or equal to 5 YOS then keep
YG2_Data_2011_1 = subset(data_2011,CYOS_EFY >= 5, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values greater than 8 YOS and retain values
#are 0 or 1
YG2_Data_2011_2 = subset(YG2_Data_2011_1,CYOS_EFY >= 8 & Retain >= 0,
select=c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,
Race,Deployments:Prior_E))
```

```
#Subset the data with values less than 7 YOS and retain = 0
YG2_Data_2011_3 = subset(YG2_Data_2011_1, CYOS_EFY <=7 & Retain < 1,
select=c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,
Race,Deployments:Prior_E))

#Combine both files
CYOS_4to8_2011_Data= rbind(YG2_Data_2011_2,YG2_Data_2011_3)

#2012-----------------------------------------------------------------

#If data is greater than or equal to 5 YOS then keep
YG2_Data_2012_1 = subset(data_2012,CYOS_EFY >= 5, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values greater than 8 YOS and retain values
#are 0 or 1
YG2_Data_2012_2 = subset(YG2_Data_2012_1,CYOS_EFY >= 8 & Retain >= 0,
select =c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values less than 7 YOS and retain = 0
YG2_Data_2012_3 = subset(YG2_Data_2012_1, CYOS_EFY <=7 & Retain < 1,
select =c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Combine both files
CYOS_4to8_2012_Data= rbind(YG2_Data_2012_2,YG2_Data_2012_3)


#2013-----------------------------------------------------------------

#If data is greater than or equal to 5 YOS then keep
YG2_Data_2013_1 = subset(data_2013,CYOS_EFY >= 5, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values greater than 8 YOS and retain values
#are 0 or 1
YG2_Data_2013_2 = subset(YG2_Data_2013_1,CYOS_EFY >= 8 & Retain >= 0,
select=c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,
Race,Deployments:Prior_E))
```

```
#Subset the data with values less than 7 YOS and retain = 0
YG2_Data_2013_3 = subset(YG2_Data_2013_1, CYOS_EFY <=7 & Retain < 1,
select=c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,
Race,Deployments:Prior_E))

#Combine both files
CYOS_4to8_2013_Data= rbind(YG2_Data_2013_2,YG2_Data_2011_3)

#2014-----------------------------------------------------------------

#If data is greater than or equal to 5 YOS then keep
YG2_Data_2014_1 = subset(data_2014,CYOS_EFY >= 5, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race
,Deployments:Prior_E))

#Subset the data with values greater than 8 YOS and retain values
#are 0 or 1
YG2_Data_2014_2 = subset(YG2_Data_2014_1,CYOS_EFY >= 8 & Retain >= 0,
select=c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values less than 7 YOS and retain = 0
YG2_Data_2014_3 = subset(YG2_Data_2014_1, CYOS_EFY <=7 & Retain < 1,
select=c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,
Race,Deployments:Prior_E))

#Combine both files
CYOS_4to8_2014_Data= rbind(YG2_Data_2014_2,YG2_Data_2014_3)


#2015-----------------------------------------------------------------

#If data is greater than or equal to 5 YOS then keep
YG2_Data_2015_1 = subset(data_2015,CYOS_EFY >= 5, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values greater than 8 YOS and retain values
#are 0 or 1
YG2_Data_2015_2 = subset(YG2_Data_2015_1,CYOS_EFY >= 8 & Retain >= 0,
select=c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,
Race,
```

```
Deployments:Prior_E))

#Subset the data with values less than 7 YOS and retain = 0
YG2_Data_2015_3 = subset(YG2_Data_2015_1, CYOS_EFY <= 7 & Retain < 1,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race
,Deployments:Prior_E))

#Combine both files
CYOS_4to8_2015_Data= rbind(YG2_Data_2015_2,YG2_Data_2015_3)


#2016-----------------------------------------------------------------

#If data is greater than or equal to 5 YOS then keep
YG2_Data_2016_1 = subset(data_2016,CYOS_EFY >= 5, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values greater than 8 YOS and retain values
#are 0 or 1
YG2_Data_2016_2 = subset(YG2_Data_2016_1,CYOS_EFY >= 8 & Retain >= 0,
select=c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,
Race,Deployments:Prior_E))

#Subset the data with values less than 7 YOS and retain = 0
YG2_Data_2016_3 = subset(YG2_Data_2016_1, CYOS_EFY <=7 & Retain < 1,
select=c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,
Race,Deployments:Prior_E))

#Combine both files
CYOS_4to8_2016_Data= rbind(YG2_Data_2016_2,YG2_Data_2016_3)

#2017-----------------------------------------------------------------

#If data is greater than or equal to 5 YOS then keep
YG2_Data_2017_1 = subset(data_2017,CYOS_EFY >= 5, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values greater than 8 YOS and retain values
#are 0 or 1
YG2_Data_2017_2 = subset(YG2_Data_2017_1,CYOS_EFY >= 8 & Retain >= 0,
select=c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,
```

```
Race,
Deployments:Prior_E))

#Subset the data with values less than 7 YOS and retain = 0
YG2_Data_2017_3 = subset(YG2_Data_2017_1, CYOS_EFY <=7 & Retain < 1,
select=c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,
Race,Deployments:Prior_E))

#Combine both files
CYOS_4to8_2017_Data= rbind(YG2_Data_2017_2,YG2_Data_2017_3)

#2018-----------------------------------------------------------------

#If data is greater than or equal to 5 YOS then keep
YG2_Data_2018_1 = subset(data_2018,CYOS_EFY >= 5, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values greater than 8 YOS and retain values
#are 0 or 1
YG2_Data_2018_2 = subset(YG2_Data_2018_1,CYOS_EFY >= 8 & Retain >= 0,
select=c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,
Race,Deployments:Prior_E))

#Subset the data with values less than 7 YOS and retain = 0
YG2_Data_2018_3 = subset(YG2_Data_2018_1, CYOS_EFY <=7 & Retain < 1,
select =c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Combine both files
CYOS_4to8_2018_Data= rbind(YG2_Data_2018_2,YG2_Data_2018_3)

#2019-----------------------------------------------------------------

#If data is greater than or equal to 5 YOS then keep
YG2_Data_2019_1 = subset(data_2019,CYOS_EFY >= 5, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values greater than 8 YOS and retain values
#are 0 or 1
YG2_Data_2019_2 = subset(YG2_Data_2019_1,CYOS_EFY >= 8 & Retain >= 0,
select=c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
```

```
Deployments:Prior_E))

#Subset the data with values less than 7 YOS and retain = 0
YG2_Data_2019_3 = subset(YG2_Data_2019_1, CYOS_EFY <=7 & Retain < 1,
select=c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Combine both files
CYOS_4to8_2019_Data= rbind(YG2_Data_2019_2,YG2_Data_2019_3)


#Combine all data frames containing CYOS 4-8 now-------------------
CYOS_4to8_Combined = rbind(CYOS_4to8_2009_Data,CYOS_4to8_2010_Data,
CYOS_4to8_2011_Data,CYOS_4to8_2012_Data,CYOS_4to8_2013_Data,
CYOS_4to8_2014_Data,CYOS_4to8_2015_Data,CYOS_4to8_2016_Data,
CYOS_4to8_2017_Data,CYOS_4to8_2018_Data,CYOS_4to8_2019_Data)

#Reverse order of data to ensure the most recent record is obtained when
#removing duplicates
revOrd_CYOS_4to8_Comb =CYOS_4to8_Combined[order(CYOS_4to8_Combined$SSAN,
CYOS_4to8_Combined$CYOS_EFY,decreasing=TRUE),]

#Now remove the duplicated information (records containing
#older information) by identified number:
CYOS_4to8_NoDups = revOrd_CYOS_4to8_Comb[!duplicated
(revOrd_CYOS_4to8_Comb$SSAN),]

#Add a new column called 'Retained' with values = 0
CYOS_4to8_NoDups$Retained = 0

#Change all retained values to 1 if CYOS >= 8. Some currently
#have 0 values.
CYOS_4to8_Final = CYOS_4to8_NoDups %>% mutate(Retained=replace
(Retained,CYOS_EFY >= 8, 1)) %>% as.data.frame()

#Now remove the 'Retain' column since 'Retained' column now has the
#updated information..
CYOS_4to8_Final$Retain = NULL

transform(CYOS_4to8_Final, Retained = as.numeric(Retained))
#transform to numeric value if necessary

#subset the data so it does not contain SSAN:
```

```
CYOS_4to8_LogRegData = CYOS_4to8_Final[c(3,4,6:9,11,12)]
write.csv(CYOS_0to6_LogRegData,"C:/Users/User/Desktop/Astudillo_Backup/
Data/Data in Excel/offinv_annual/Female Only/Fiscal Year/Final Thesis Data/
CYOS Info/CYOS_4to8_LogRegData.csv",row.names = TRUE)

#Perform logistic regression on the Final dataframe containing all fiscal
#years for CYOS 4 to 8:
LogReg_CYOS_4to8 = glm(Retained ~ ., data = CYOS_4to8_LogRegData,
family = "binomial")
summary(LogReg_CYOS_4to8)
confint(LogReg_CYOS_4to8)


#Perform Logistic Regression on the Developmental Categories
#based on CYOS 4 to 8--------------

#Air Operations and Special Warfare
AOSWData_CYOS_4to8 = subset(CYOS_4to8_Final,Dev_Cat =="1", select =
c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,Prior_E,Retained))
AOSWLR_CYOS_4to8 = glm(Retained ~ ., data = AOSWData_CYOS_4to8 ,
family = "binomial")
summary(AOSWLR_CYOS_4to8)
write.csv(AOSWData_CYOS_4to8,"C:/Users/User/Desktop/Astudillo_Backup/
Data/Data in Excel/offinv_annual/Female Only/Fiscal Year/Final Thesis Data/
Developmental Categories/AOSWDATA_CYOS_4to8.csv",row.names = TRUE)
confint(AOSWLR_CYOS_4to8)

#Info Warfare
IWData_CYOS_4to8 = subset(CYOS_4to8_Final,Dev_Cat =="2", select =
c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,Prior_E,Retained))
IWLR_CYOS_4to8 = glm(Retained ~ ., data = IWData_CYOS_4to8,
family = "binomial")
summary(IWLR_CYOS_4to8)
write.csv(IWData_CYOS_4to8,"C:/Users/User/Desktop/Astudillo_Backup/Data/
Data in Excel/offinv_annual/Female Only/Fiscal Year/Final Thesis Data/
Developmental Categories/IWData_CYOS_4to8.csv",row.names = TRUE)
confint(IWLR_CYOS_4to8)

#Combat Support
CSData_CYOS_4to8 = subset(CYOS_4to8_Final,Dev_Cat =="3", select =
c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,Prior_E,Retained))
CSLR_CYOS_4to8 = glm(Retained ~ ., data = CSData_CYOS_4to8,
family = "binomial")
```

```
summary(CSLR_CYOS_4to8)
write.csv(CSData_CYOS_4to8,"C:/Users/User/Desktop/Astudillo_Backup/
Data/Data in Excel/offinv_annual/Female Only/Fiscal Year/
Final Thesis Data/Developmental Categories/CSData_CYOS_4to8.csv",
row.names = TRUE)
confint(CSLR_CYOS_4to8)

#Force Modernization
FMData_CYOS_4to8 = subset(CYOS_4to8_Final,Dev_Cat =="4", select =
c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,Prior_E,Retained))

FMLR_CYOS_4to8 = glm(Retained ~ ., data = FMData_CYOS_4to8,
family = "binomial")
summary(FMLR_CYOS_4to8)
write.csv(FMData_CYOS_4to8,"C:/Users/User/Desktop/Astudillo_Backup/Data/
Data in Excel/offinv_annual/Female Only/Fiscal Year/Final Thesis Data/
Developmental Categories/FMData_CYOS_4to8.csv",row.names = TRUE)
confint(FMLR_CYOS_4to8)

# Years 8 to 14

#2009------------------------------------------------------------

#If data is greater than or equal to 9 YOS then keep
YG3_Data_2009_1 = subset(data_2009,CYOS_EFY >= 9, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values greater than 14 YOS and retain
#values are 0 or 1
YG3_Data_2009_2 = subset(YG3_Data_2009_1,CYOS_EFY >= 14 & Retain >= 0,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,
Race,Deployments:Prior_E))

#Subset the data with values less than 13 YOS and retain = 0
YG3_Data_2009_3 = subset(YG3_Data_2009_1, CYOS_EFY <= 13 & Retain < 1,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,
Race,Deployments:Prior_E))

#Combine both files
CYOS_8to14_2009_Data= rbind(YG3_Data_2009_2,YG3_Data_2009_3)

#2010------------------------------------------------------------
```

```
#If data is greater than or equal to 9 YOS then keep
YG3_Data_2010_1 = subset(data_2010,CYOS_EFY >= 9, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values greater than 14 YOS and retain values
#are 0 or 1
YG3_Data_2010_2 = subset(YG3_Data_2010_1,CYOS_EFY >= 14 & Retain >= 0,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,
Race,Deployments:Prior_E))

#Subset the data with values less than 13 YOS and retain = 0
YG3_Data_2010_3 = subset(YG3_Data_2010_1, CYOS_EFY <= 13 & Retain < 1,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,
Race,Deployments:Prior_E))

#Combine both files
CYOS_8to14_2010_Data= rbind(YG3_Data_2010_2,YG3_Data_2010_3)

#2011------------------------------------------------------------

#If data is greater than or equal to 9 YOS then keep
YG3_Data_2011_1 = subset(data_2011,CYOS_EFY >= 9, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,
Race,Deployments:Prior_E))

#Subset the data with values greater than 14 YOS and retain values
#are 0 or 1
YG3_Data_2011_2 = subset(YG3_Data_2011_1,CYOS_EFY >= 14 &
Retain >= 0, select =c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,
Spouse_Stat,Race,Deployments:Prior_E))

#Subset the data with values less than 13 YOS and retain = 0
YG3_Data_2011_3 = subset(YG3_Data_2011_1, CYOS_EFY <= 13 & Retain < 1,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,
Race,Deployments:Prior_E))

#Combine both files
CYOS_8to14_2011_Data= rbind(YG3_Data_2011_2,YG3_Data_2011_3)

#2012------------------------------------------------------------
```

```
#If data is greater than or equal to 9 YOS then keep
YG3_Data_2012_1 = subset(data_2012,CYOS_EFY >= 9, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values greater than 14 YOS and retain
#values are 0 or 1
YG3_Data_2012_2 = subset(YG3_Data_2012_1,CYOS_EFY >= 14 & Retain >= 0,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values less than 13 YOS and retain = 0
YG3_Data_2012_3 = subset(YG3_Data_2012_1, CYOS_EFY <= 13 & Retain < 1,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Combine both files
CYOS_8to14_2012_Data= rbind(YG3_Data_2012_2,YG3_Data_2012_3)

#2013-------------------------------------------------------------

#If data is greater than or equal to 9 YOS then keep
YG3_Data_2013_1 = subset(data_2013,CYOS_EFY >= 9, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values greater than 14 YOS and retain values
#are 0 or 1
YG3_Data_2013_2 = subset(YG3_Data_2013_1,CYOS_EFY >= 14 & Retain >= 0,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values less than 13 YOS and retain = 0
YG3_Data_2013_3 = subset(YG3_Data_2013_1, CYOS_EFY <= 13 & Retain < 1,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Combine both files
CYOS_8to14_2013_Data= rbind(YG3_Data_2013_2,YG3_Data_2013_3)

#2014-------------------------------------------------------------

#If data is greater than or equal to 9 YOS then keep
```

```
YG3_Data_2014_1 = subset(data_2014,CYOS_EFY >= 9, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values greater than 14 YOS and retain values
#are 0 or 1
YG3_Data_2014_2 = subset(YG3_Data_2014_1,CYOS_EFY >= 14 & Retain >= 0,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values less than 13 YOS and retain = 0
YG3_Data_2014_3 = subset(YG3_Data_2014_1, CYOS_EFY <= 13 & Retain < 1,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Combine both files
CYOS_8to14_2014_Data= rbind(YG3_Data_2014_2,YG3_Data_2014_3)

#2015------------------------------------------------------------

#If data is greater than or equal to 9 YOS then keep
YG3_Data_2015_1 = subset(data_2015,CYOS_EFY >= 9, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values greater than 14 YOS and retain values
#are 0 or 1
YG3_Data_2015_2 = subset(YG3_Data_2015_1,CYOS_EFY >= 14 & Retain >= 0,
select =c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values less than 13 YOS and retain = 0
YG3_Data_2015_3 = subset(YG3_Data_2015_1, CYOS_EFY <= 13 & Retain < 1,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Combine both files
CYOS_8to14_2015_Data= rbind(YG3_Data_2015_2,YG3_Data_2015_3)

#2016------------------------------------------------------------

#If data is greater than or equal to 9 YOS then keep
YG3_Data_2016_1 = subset(data_2016,CYOS_EFY >= 9, select =
```

```
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E)

#Subset the data with values greater than 14 YOS and retain values
#are 0 or 1
YG3_Data_2016_2 = subset(YG3_Data_2016_1,CYOS_EFY >= 14 & Retain >= 0,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values less than 13 YOS and retain = 0
YG3_Data_2016_3 = subset(YG3_Data_2016_1, CYOS_EFY <= 13 & Retain < 1,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Combine both files
CYOS_8to14_2016_Data= rbind(YG3_Data_2016_2,YG3_Data_2016_3)

#2017-------------------------------------------------------------

#If data is greater than or equal to 9 YOS then keep
YG3_Data_2017_1 = subset(data_2017,CYOS_EFY >= 9, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values greater than 14 YOS and retain values
#are 0 or 1
YG3_Data_2017_2 = subset(YG3_Data_2017_1,CYOS_EFY >= 14 & Retain >= 0,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values less than 13 YOS and retain = 0
YG3_Data_2017_3 = subset(YG3_Data_2017_1, CYOS_EFY <= 13 & Retain < 1,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Combine both files

CYOS_8to14_2017_Data= rbind(YG3_Data_2017_2,YG3_Data_2017_3)

#2018-------------------------------------------------------------

#If data is greater than or equal to 9 YOS then keep
YG3_Data_2018_1 = subset(data_2018,CYOS_EFY >= 9, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
```

```
Deployments:Prior_E))

#Subset the data with values greater than 14 YOS and retain values
#are 0 or 1
YG3_Data_2018_2 = subset(YG3_Data_2018_1,CYOS_EFY >= 14 & Retain >= 0,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values less than 13 YOS and retain = 0
YG3_Data_2018_3 = subset(YG3_Data_2018_1, CYOS_EFY <= 13 & Retain < 1,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Combine both files
CYOS_8to14_2018_Data= rbind(YG3_Data_2018_2,YG3_Data_2018_3)

#2019-----------------------------------------------------------

#If data is greater than or equal to 9 YOS then keep
YG3_Data_2019_1 = subset(data_2019,CYOS_EFY >= 9, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values greater than 14 YOS and retain values
#are 0 or 1
YG3_Data_2019_2 = subset(YG3_Data_2019_1,CYOS_EFY >= 14 & Retain >= 0,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,
Race,Deployments:Prior_E))

#Subset the data with values less than 13 YOS and retain = 0
YG3_Data_2019_3 = subset(YG3_Data_2019_1, CYOS_EFY <= 13 & Retain < 1,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,
Race,Deployments:Prior_E))

#Combine both files
CYOS_8to14_2019_Data= rbind(YG3_Data_2019_2,YG3_Data_2019_3)

#Combine all data frames containing CYOS 8-14 now-------------------
CYOS_8to14_Combined = rbind(CYOS_8to14_2009_Data,CYOS_8to14_2010_Data,
CYOS_8to14_2011_Data,CYOS_8to14_2012_Data,CYOS_8to14_2013_Data,
CYOS_8to14_2014_Data,CYOS_8to14_2015_Data,CYOS_8to14_2016_Data,
CYOS_8to14_2017_Data,CYOS_8to14_2018_Data,CYOS_8to14_2019_Data)
```

```
#Reverse order of data to ensure the most recent record is obtained
#when removing duplicates
revOrd_CYOS_8to14_Comb = CYOS_8to14_Combined[order
(CYOS_8to14_Combined$SSAN,CYOS_8to14_Combined$CYOS_EFY,decreasing=TRUE),]

#Now remove the duplicated information (records containing
#older information) by identified number:
CYOS_8to14_NoDups = revOrd_CYOS_8to14_Comb
[!duplicated(revOrd_CYOS_8to14_Comb$SSAN),]

#Add a new column called 'Retained' with values = 0
CYOS_8to14_NoDups$Retained = 0

#Change all retained values to 1 if CYOS >= 8. Some currently have
#0 values.
CYOS_8to14_Final = CYOS_8to14_NoDups %>% mutate(Retained=replace
(Retained,CYOS_EFY >= 14, 1)) %>% as.data.frame()

#Now remove the 'Retain' column since 'Retained' column now has the
#updated information..
CYOS_8to14_Final$Retain = NULL
transform(CYOS_8to14_Final, Retained = as.numeric(Retained))
#transform to numeric value if necessary

#subset the data so it does not contain SSAN:
CYOS_8to14_LogRegData = CYOS_8to14_Final[c(3,4,6:9,11,12)]
write.csv(CYOS_8to14_LogRegData,"C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only
/Fiscal Year/Final Thesis Data/CYOS Info/
CYOS_8to14_LogRegData.csv",row.names = TRUE)

#Perform logistic regression on the Final dataframe containing all
#fiscal years for CYOS 8 to 14:
LogReg_CYOS_8to14 = glm(Retained ~ ., data = CYOS_8to14_LogRegData,
family = "binomial")
summary(LogReg_CYOS_8to14)
confint(LogReg_CYOS_8to14)

#Perform Logistic Regression on the Developmental Categories
#based on CYOS 8 to 14--------------

#Air Operations and Special Warfare
AOSWData_CYOS_8to14 = subset(CYOS_8to14_Final,Dev_Cat =="1",
```

```
select =c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,
Prior_E,Retained))
AOSWLR_CYOS_8to14 = glm(Retained ~ ., data = AOSWData_CYOS_8to14,
family = "binomial")
summary(AOSWLR_CYOS_8to14)
write.csv(AOSWData_CYOS_8to14,"C:/Users/User/Desktop/Astudillo_Backup/
Data/Data in Excel/offinv_annual/Female Only/Fiscal Year/
Final Thesis Data/Developmental Categories/AOSWData_CYOS_8to14.csv",
row.names = TRUE)
confint(AOSWLR_CYOS_8to14)

#Info Warfare
IWData_CYOS_8to14 = subset(CYOS_8to14_Final,Dev_Cat =="2", select =
c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,Prior_E,Retained))
IWLR_CYOS_8to14 = glm(Retained ~ ., data = IWData_CYOS_8to14,
family = "binomial")
summary(IWLR_CYOS_8to14)
write.csv(IWData_CYOS_8to14,"C:/Users/User/Desktop/Astudillo_Backup/
Data/Data in Excel/offinv_annual/Female Only/Fiscal Year/
Final Thesis Data/Developmental Categories/
IWData_CYOS_8to14.csv",row.names = TRUE)
confint(IWLR_CYOS_8to14)

#Combat Support
CSData_CYOS_8to14 = subset(CYOS_8to14_Final,Dev_Cat =="3", select =
c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,Prior_E,Retained))
CSLR_CYOS_8to14 = glm(Retained ~ ., data = CSData_CYOS_8to14,
family = "binomial")
summary(CSLR_CYOS_8to14)
write.csv(CSData_CYOS_8to14,"C:/Users/User/Desktop/Astudillo_Backup/
Data/Data in Excel/offinv_annual/Female Only/Fiscal Year/
Final Thesis Data/Developmental Categories/
CSData_CYOS_8to14.csv",row.names = TRUE)
confint(CSLR_CYOS_8to14)

#Force Modernization
FMData_CYOS_8to14 = subset(CYOS_8to14_Final,Dev_Cat =="4", select =
c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,Prior_E,Retained))
FMLR_CYOS_8to14 = glm(Retained ~ ., data = FMData_CYOS_8to14,
family = "binomial")
summary(FMLR_CYOS_8to14)
write.csv(FMData_CYOS_8to14,"C:/Users/User/Desktop/Astudillo_Backup/
Data/Data in Excel/offinv_annual/Female Only/Fiscal Year/
```

```
Final Thesis Data/Developmental Categories/
FMData_CYOS_8to14.csv",row.names = TRUE)
confint(FMLR_CYOS_8to14)

#---------------------------------------------------------------------
#---------------------------------------------------------------------


# Years 12 to 19

#2009------------------------------------------------------------------

#If data is greater than or equal to 12 YOS then keep
YG4_Data_2009_1 = subset(data_2009,CYOS_EFY >= 12, select
=c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,
Race,Deployments:Prior_E))

#Subset the data with values greater than 19 YOS and retain
#values are 0 or 1
YG4_Data_2009_2 = subset(YG4_Data_2009_1,CYOS_EFY >= 19 & Retain >= 0,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values less than 18 YOS and retain = 0
YG4_Data_2009_3 = subset(YG4_Data_2009_1, CYOS_EFY <= 18 & Retain < 1,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Combine both files
CYOS_12to19_2009_Data= rbind(YG4_Data_2009_2,YG4_Data_2009_3)

#2010------------------------------------------------------------------

#If data is greater than or equal to 12 YOS then keep
YG4_Data_2010_1 = subset(data_2010,CYOS_EFY >= 12, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values greater than 19 YOS and retain values
#are 0 or 1
YG4_Data_2010_2 = subset(YG4_Data_2010_1,CYOS_EFY >= 19 & Retain >= 0,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
```

```
#Subset the data with values less than 18 YOS and retain = 0
YG4_Data_2010_3 = subset(YG4_Data_2010_1, CYOS_EFY <= 18 & Retain < 1,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Combine both files
CYOS_12to19_2010_Data= rbind(YG4_Data_2010_2,YG4_Data_2010_3)

#2011-----------------------------------------------------------------

#If data is greater than or equal to 12 YOS then keep
YG4_Data_2011_1 = subset(data_2011,CYOS_EFY >= 12, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values greater than 19 YOS and retain values
#are 0 or 1
YG4_Data_2011_2 = subset(YG4_Data_2011_1,CYOS_EFY >= 19 & Retain >= 0,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values less than 18 YOS and retain = 0
YG4_Data_2011_3 = subset(YG4_Data_2011_1, CYOS_EFY <= 18 & Retain < 1,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Combine both files
CYOS_12to19_2011_Data= rbind(YG4_Data_2011_2,YG4_Data_2011_3)

#2012-----------------------------------------------------------------

#If data is greater than or equal to 12 YOS then keep
YG4_Data_2012_1 = subset(data_2012,CYOS_EFY >= 12, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Subset the data with values greater than 19 YOS and retain values
#are 0 or 1
YG4_Data_2012_2 = subset(YG4_Data_2012_1,CYOS_EFY >= 19 & Retain >= 0,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values less than 18 YOS and retain = 0
```

```
YG4_Data_2012_3 = subset(YG4_Data_2012_1, CYOS_EFY <= 18 & Retain < 1,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Combine both files
CYOS_12to19_2012_Data= rbind(YG4_Data_2012_2,YG4_Data_2012_3)


#2013-----------------------------------------------------------------

#If data is greater than or equal to 12 YOS then keep
YG4_Data_2013_1 = subset(data_2013,CYOS_EFY >= 12, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values greater than 19 YOS and retain values
#are 0 or 1
YG4_Data_2013_2 = subset(YG4_Data_2013_1,CYOS_EFY >= 19 & Retain >= 0,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values less than 18 YOS and retain = 0
YG4_Data_2013_3 = subset(YG4_Data_2013_1, CYOS_EFY <= 18 & Retain < 1,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Combine both files
CYOS_12to19_2013_Data= rbind(YG4_Data_2013_2,YG4_Data_2013_3)

#2014-----------------------------------------------------------------

#If data is greater than or equal to 12 YOS then keep
YG4_Data_2014_1 = subset(data_2014,CYOS_EFY >= 12, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Subset the data with values greater than 19 YOS and retain values
#are 0 or 1
YG4_Data_2014_2 = subset(YG4_Data_2014_1,CYOS_EFY >= 19 & Retain >= 0,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))

#Subset the data with values less than 18 YOS and retain = 0
YG4_Data_2014_3 = subset(YG4_Data_2014_1, CYOS_EFY <= 18 & Retain < 1,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
```

```
Deployments:Prior_E))

#Combine both files
CYOS_12to19_2014_Data= rbind(YG4_Data_2014_2,YG4_Data_2014_3)

#2015-----------------------------------------------------------------

#If data is greater than or equal to 12 YOS then keep
YG4_Data_2015_1 = subset(data_2015,CYOS_EFY >= 12, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Subset the data with values greater than 19 YOS and retain values
#are 0 or 1
YG4_Data_2015_2 = subset(YG4_Data_2015_1,CYOS_EFY >= 19 & Retain >= 0,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Subset the data with values less than 18 YOS and retain = 0
YG4_Data_2015_3 = subset(YG4_Data_2015_1, CYOS_EFY <= 18 & Retain < 1,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Combine both files
CYOS_12to19_2015_Data= rbind(YG4_Data_2015_2,YG4_Data_2015_3)

#2016-----------------------------------------------------------------

#If data is greater than or equal to 12 YOS then keep
YG4_Data_2016_1 = subset(data_2016,CYOS_EFY >= 12, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Subset the data with values greater than 19 YOS and retain values
#are 0 or 1
YG4_Data_2016_2 = subset(YG4_Data_2016_1,CYOS_EFY >= 19 & Retain >= 0,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Subset the data with values less than 18 YOS and retain = 0
YG4_Data_2016_3 = subset(YG4_Data_2016_1, CYOS_EFY <= 18 & Retain < 1,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Combine both files
CYOS_12to19_2016_Data= rbind(YG4_Data_2016_2,YG4_Data_2016_3)


#2017-----------------------------------------------------------------
```

```
#If data is greater than or equal to 12 YOS then keep
YG4_Data_2017_1 = subset(data_2017,CYOS_EFY >= 12, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Subset the data with values greater than 19 YOS and retain
#values are 0 or 1
YG4_Data_2017_2 = subset(YG4_Data_2017_1,CYOS_EFY >= 19 & Retain >= 0,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Subset the data with values less than 18 YOS and retain = 0
YG4_Data_2017_3 = subset(YG4_Data_2017_1, CYOS_EFY <= 18 & Retain < 1,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Combine both files
CYOS_12to19_2017_Data= rbind(YG4_Data_2017_2,YG4_Data_2017_3)


#2018-----------------------------------------------------------

#If data is greater than or equal to 12 YOS then keep
YG4_Data_2018_1 = subset(data_2018,CYOS_EFY >= 12, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Subset the data with values greater than 19 YOS and retain values
#are 0 or 1
YG4_Data_2018_2 = subset(YG4_Data_2018_1,CYOS_EFY >= 19 & Retain >= 0,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Subset the data with values less than 18 YOS and retain = 0
YG4_Data_2018_3 = subset(YG4_Data_2018_1, CYOS_EFY <= 18 & Retain < 1,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Combine both files
CYOS_12to19_2018_Data= rbind(YG4_Data_2018_2,YG4_Data_2018_3)


#2019-----------------------------------------------------------

#If data is greater than or equal to 12 YOS then keep
YG4_Data_2019_1 = subset(data_2019,CYOS_EFY >= 12, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
```

```
#Subset the data with values greater than 19 YOS and retain values
#are 0 or 1
YG4_Data_2019_2 = subset(YG4_Data_2019_1,CYOS_EFY >= 19 & Retain >= 0,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Subset the data with values less than 18 YOS and retain = 0
YG4_Data_2019_3 = subset(YG4_Data_2019_1, CYOS_EFY <= 18 & Retain < 1,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Combine both files
CYOS_12to19_2019_Data= rbind(YG4_Data_2019_2,YG4_Data_2019_3)

#Combine all data frames containing CYOS 12-19 now------------------
CYOS_12to19_Combined = rbind(CYOS_12to19_2009_Data,
CYOS_12to19_2010_Data,CYOS_12to19_2011_Data,CYOS_12to19_2012_Data,
CYOS_12to19_2013_Data,CYOS_12to19_2014_Data,CYOS_12to19_2015_Data,
CYOS_12to19_2016_Data,CYOS_12to19_2017_Data,CYOS_12to19_2018_Data,
CYOS_12to19_2019_Data)

#Reverse order of data to ensure the most recent record is obtained
#when removing duplicates
revOrd_CYOS_12to19_Comb = CYOS_12to19_Combined[order
(CYOS_12to19_Combined
$SSAN,CYOS_12to19_Combined$CYOS_EFY,decreasing=TRUE),]

#Now remove the duplicated information (records containing older
#information) by identified number:
CYOS_12to19_NoDups = revOrd_CYOS_12to19_Comb
[!duplicated(revOrd_CYOS_12to19_Comb$SSAN),]

#Add a new column called 'Retained' with values = 0
CYOS_12to19_NoDups$Retained = 0

#Change all retained values to 1 if CYOS >= 8. Some currently
#have 0 values.
CYOS_12to19_Final = CYOS_12to19_NoDups %>% mutate(Retained=replace
(Retained,CYOS_EFY >= 19, 1)) %>% as.data.frame()

#Now remove the 'Retain' column since 'Retained' column now has the
#updated information..
CYOS_12to19_Final$Retain = NULL
transform(CYOS_12to19_Final, Retained = as.numeric(Retained))
#transform to numeric value if necessary
```

```
#subset the data so it does not contain SSAN:
CYOS_12to19_LogRegData = CYOS_12to19_Final[c(3,4,6:9,11,12)]
write.csv(CYOS_12to19_LogRegData,"C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/CYOS Info/CYOS_12to19_LogRegData.csv",
row.names = TRUE)


#Perform logistic regression on the Final dataframe containing all
#fiscal years for CYOS 12 to 19:
LogReg_CYOS_12to19 = glm(Retained ~ ., data = CYOS_12to19_LogRegData,
family = "binomial")
summary(LogReg_CYOS_12to19)
confint(LogReg_CYOS_12to19)

#Perform Logistic Regression on the Developmental Categories based on
CYOS 12 to 19--------------

#Air Operations and Special Warfare
AOSWData_CYOS_12to19 = subset(CYOS_12to19_Final,Dev_Cat =="1",
select =c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,Prior_E,
Retained))
AOSWLR_CYOS_12to19 = glm(Retained ~ ., data = AOSWData_CYOS_12to19,
family = "binomial")
summary(AOSWLR_CYOS_12to19)
write.csv(AOSWData_CYOS_12to19,"C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/Developmental Categories/
AOSWData_CYOS_12to19.csv",row.names = TRUE)
confint(AOSWLR_CYOS_12to19)

#Info Warfare
IWData_CYOS_12to19 = subset(CYOS_12to19_Final,Dev_Cat =="2",
select = c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,Prior_E,
Retained))
IWLR_CYOS_12to19 = glm(Retained ~ ., data = IWData_CYOS_12to19,
family = "binomial")
summary(IWLR_CYOS_12to19)
write.csv(IWData_CYOS_12to19,"C:/Users/User/Desktop/Astudillo_Backup/
Data/Data in Excel/offinv_annual/Female Only/Fiscal Year/
Final Thesis Data/Developmental Categories/IWData_CYOS_12to19.csv",
row.names = TRUE)
```

```
confint(IWLR_CYOS_12to19)

#Combat Support
CSData_CYOS_12to19 = subset(CYOS_12to19_Final,Dev_Cat =="3",
select = c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,
Prior_E,Retained))
CSLR_CYOS_12to19 = glm(Retained ~ ., data = CSData_CYOS_12to19,
family = "binomial")
summary(CSLR_CYOS_12to19)
write.csv(CSData_CYOS_12to19,"C:/Users/User/Desktop/Astudillo_Backup
/Data/Data in Excel/offinv_annual/Female Only/Fiscal Year/
Final Thesis Data/Developmental Categories/CSData_CYOS_12to19.csv",
row.names = TRUE)
confint(CSLR_CYOS_12to19)

#Force Modernization
FMData_CYOS_12to19 = subset(CYOS_12to19_Final,Dev_Cat =="4",
select = c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,
Prior_E,Retained))
FMLR_CYOS_12to19 = glm(Retained ~ ., data = FMData_CYOS_12to19,
family = "binomial")
summary(FMLR_CYOS_12to19)
write.csv(FMData_CYOS_12to19,"C:/Users/User/Desktop/Astudillo_Backup/
Data/Data in Excel/offinv_annual/Female Only/Fiscal Year/
Final Thesis Data/Developmental Categories/FMData_CYOS_12to19.csv",
row.names = TRUE)
confint(FMLR_CYOS_12to19)


#---------------------------------------------------------------------
#---------------------------------------------------------------------

# Years 20 to 22

#2009------------------------------------------------------------------

#If data is greater than or equal to 20 YOS then keep
YG5_Data_2009_1 = subset(data_2009,CYOS_EFY >= 20, select
=c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race
,Deployments:Prior_E))
#Subset the data with values greater than 22 YOS and retain
#values are 0 or 1
YG5_Data_2009_2 = subset(YG5_Data_2009_1,CYOS_EFY >= 22 & Retain >= 0,
```

```
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Subset the data with values less than 21 YOS and retain = 0
YG5_Data_2009_3 = subset(YG5_Data_2009_1, CYOS_EFY <= 21 & Retain < 1,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Combine both files
CYOS_20to22_2009_Data= rbind(YG5_Data_2009_2,YG5_Data_2009_3)


#2010-------------------------------------------------------------

#If data is greater than or equal to 20 YOS then keep
YG5_Data_2010_1 = subset(data_2010,CYOS_EFY >= 20, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Subset the data with values greater than 22 YOS and retain
#values are 0 or 1
YG5_Data_2010_2 = subset(YG5_Data_2010_1,CYOS_EFY >= 22 & Retain >= 0,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Subset the data with values less than 21 YOS and retain = 0
YG5_Data_2010_3 = subset(YG5_Data_2010_1, CYOS_EFY <= 21 & Retain < 1,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Combine both files
CYOS_20to22_2010_Data= rbind(YG5_Data_2010_2,YG5_Data_2010_3)


#2011-------------------------------------------------------------

#If data is greater than or equal to 20 YOS then keep
YG5_Data_2011_1 = subset(data_2011,CYOS_EFY >= 20, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Subset the data with values greater than 22 YOS and retain
#values are 0 or 1
YG5_Data_2011_2 = subset(YG5_Data_2011_1,CYOS_EFY >= 22 & Retain >= 0,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Subset the data with values less than 21 YOS and retain = 0
YG5_Data_2011_3 = subset(YG5_Data_2011_1, CYOS_EFY <= 21 & Retain < 1,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
```

```
Deployments:Prior_E))
#Combine both files
CYOS_20to22_2011_Data= rbind(YG5_Data_2011_2,YG5_Data_2011_3)


#2012-----------------------------------------------------------------

#If data is greater than or equal to 20 YOS then keep
YG5_Data_2012_1 = subset(data_2012,CYOS_EFY >= 20, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Subset the data with values greater than 22 YOS and retain
#values are 0 or 1
YG5_Data_2012_2 = subset(YG5_Data_2012_1,CYOS_EFY >= 22 & Retain >= 0,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Subset the data with values less than 21 YOS and retain = 0
YG5_Data_2012_3 = subset(YG5_Data_2012_1, CYOS_EFY <= 21 & Retain < 1,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Combine both files
CYOS_20to22_2012_Data= rbind(YG5_Data_2012_2,YG5_Data_2012_3)

#2013-----------------------------------------------------------------

#If data is greater than or equal to 20 YOS then keep
YG5_Data_2013_1 = subset(data_2013,CYOS_EFY >= 20, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Subset the data with values greater than 22 YOS and retain
#values are 0 or 1
YG5_Data_2013_2 = subset(YG5_Data_2013_1,CYOS_EFY >= 22 & Retain >= 0,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Subset the data with values less than 21 YOS and retain = 0
YG5_Data_2013_3 = subset(YG5_Data_2013_1, CYOS_EFY <= 21 & Retain < 1,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Combine both files
CYOS_20to22_2013_Data= rbind(YG5_Data_2013_2,YG5_Data_2013_3)


#2014-----------------------------------------------------------------
```

```
#If data is greater than or equal to 20 YOS then keep
YG5_Data_2014_1 = subset(data_2014,CYOS_EFY >= 20, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Subset the data with values greater than 22 YOS and retain
#values are 0 or 1
YG5_Data_2014_2 = subset(YG5_Data_2014_1,CYOS_EFY >= 22 & Retain >= 0,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Subset the data with values less than 21 YOS and retain = 0
YG5_Data_2014_3 = subset(YG5_Data_2014_1, CYOS_EFY <= 21 & Retain < 1,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Combine both files
CYOS_20to22_2014_Data= rbind(YG5_Data_2014_2,YG5_Data_2014_3)


#2015-----------------------------------------------------------------

#If data is greater than or equal to 20 YOS then keep
YG5_Data_2015_1 = subset(data_2015,CYOS_EFY >= 20, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Subset the d5ata with values greater than 22 YOS and retain
#values are 0 or 1
YG5_Data_2015_2 = subset(YG5_Data_2015_1,CYOS_EFY >= 22 & Retain >= 0,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Subset the data with values less than 21 YOS and retain = 0
YG5_Data_2015_3 = subset(YG5_Data_2015_1, CYOS_EFY <= 21 & Retain < 1,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Combine both files
CYOS_20to22_2015_Data= rbind(YG5_Data_2015_2,YG5_Data_2015_3)


#2016-----------------------------------------------------------------

#If data is greater than or equal to 20 YOS then keep
YG5_Data_2016_1 = subset(data_2016,CYOS_EFY >= 20, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
```

```
#Subset the d5ata with values greater than 22 YOS and retain
#values are 0 or 1
YG5_Data_2016_2 = subset(YG5_Data_2016_1,CYOS_EFY >= 22 & Retain >= 0,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Subset the data with values less than 21 YOS and retain = 0
YG5_Data_2016_3 = subset(YG5_Data_2016_1, CYOS_EFY <= 21 & Retain < 1,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Combine both files
CYOS_20to22_2016_Data= rbind(YG5_Data_2016_2,YG5_Data_2016_3)


#2017-----------------------------------------------------------------

#If data is greater than or equal to 20 YOS then keep
YG5_Data_2017_1 = subset(data_2017,CYOS_EFY >= 20, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Subset the d5ata with values greater than 22 YOS and retain
#values are 0 or 1
YG5_Data_2017_2 = subset(YG5_Data_2017_1,CYOS_EFY >= 22 & Retain >= 0,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Subset the data with values less than 21 YOS and retain = 0
YG5_Data_2017_3 = subset(YG5_Data_2017_1, CYOS_EFY <= 21 & Retain < 1,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Combine both files
CYOS_20to22_2017_Data= rbind(YG5_Data_2017_2,YG5_Data_2017_3)


#2018-----------------------------------------------------------------

#If data is greater than or equal to 20 YOS then keep
YG5_Data_2018_1 = subset(data_2018,CYOS_EFY >= 20, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Subset the d5ata with values greater than 22 YOS and retain
#values are 0 or 1
YG5_Data_2018_2 = subset(YG5_Data_2018_1,CYOS_EFY >= 22 & Retain >= 0,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
```

```
#Subset the data with values less than 21 YOS and retain = 0
YG5_Data_2018_3 = subset(YG5_Data_2018_1, CYOS_EFY <= 21 & Retain < 1,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Combine both files
CYOS_20to22_2018_Data= rbind(YG5_Data_2018_2,YG5_Data_2018_3)


#2019-----------------------------------------------------------------

#If data is greater than or equal to 20 YOS then keep
YG5_Data_2019_1 = subset(data_2019,CYOS_EFY >= 20, select =
c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Subset the d5ata with values greater than 22 YOS and retain
#values are 0 or 1
YG5_Data_2019_2 = subset(YG5_Data_2019_1,CYOS_EFY >= 22 & Retain >= 0,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Subset the data with values less than 21 YOS and retain = 0
YG5_Data_2019_3 = subset(YG5_Data_2019_1, CYOS_EFY <= 21 & Retain < 1,
select = c(SSAN,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Combine both files
CYOS_20to22_2019_Data= rbind(YG5_Data_2019_2,YG5_Data_2019_3)


#Combine all data frames containing CYOS 20-22 now-------------------
CYOS_20to22_Combined = rbind(CYOS_20to22_2009_Data,
CYOS_20to22_2010_Data,CYOS_20to22_2011_Data,CYOS_20to22_2012_Data,
CYOS_20to22_2013_Data,CYOS_20to22_2014_Data,CYOS_20to22_2015_Data,
CYOS_20to22_2016_Data,CYOS_20to22_2017_Data,CYOS_20to22_2018_Data,
CYOS_20to22_2019_Data)

#Reverse order of data to ensure the most recent record is obtained
#when removing duplicates
revOrd_CYOS_20to22_Comb = CYOS_20to22_Combined[order(
CYOS_20to22_Combined$SSAN,CYOS_20to22_Combined$CYOS_EFY,
decreasing=TRUE),]

#Now remove the duplicated information (records containing older
#information) by identified number:
CYOS_20to22_NoDups = revOrd_CYOS_20to22_Comb
```

```
[!duplicated(revOrd_CYOS_20to22_Comb$SSAN),]

#Add a new column called 'Retained' with values = 0
CYOS_20to22_NoDups$Retained = 0

#Change all retained values to 1 if CYOS >= 8. Some
#currently have 0 values.
CYOS_20to22_Final = CYOS_20to22_NoDups %>% mutate
(Retained=replace(Retained,CYOS_EFY >= 22, 1)) %>% as.data.frame()

#Now remove the 'Retain' column since 'Retained' column now has
#the updated information..
CYOS_20to22_Final$Retain = NULL
transform(CYOS_20to22_Final, Retained = as.numeric(Retained))
#transform to numeric value if necessary

#subset the data so it does not contain SSAN:
CYOS_20to22_LogRegData = CYOS_20to22_Final[c(3,4,6:9,11,12)]
write.csv(CYOS_20to22_LogRegData,"C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/CYOS Info/CYOS_20to22_LogRegData.csv",
row.names = TRUE)


#Perform logistic regression on the Final dataframe containing
#all fiscal years for CYOS 12 to 19:
LogReg_CYOS_20to22 = glm(Retained ~ ., data = CYOS_20to22_LogRegData,
family = "binomial")
summary(LogReg_CYOS_20to22)
confint(LogReg_CYOS_20to22)

#Perform Logistic Regression on the Developmental Categories based
#on CYOS 20 to 22--------------

#Air Operations and Special Warfare
AOSWData_CYOS_20to22 = subset(CYOS_20to22_Final,Dev_Cat =="1",
select = c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,Prior_E,
Retained))
AOSWLR_CYOS_20to22 = glm(Retained ~ ., data = AOSWData_CYOS_20to22,
family = "binomial")
summary(AOSWLR_CYOS_20to22)
write.csv(AOSWData_CYOS_20to22,"C:/Users/User/Desktop/Astudillo_Backup/
Data/Data in Excel/offinv_annual/Female Only/Fiscal Year/
```

```
Final Thesis Data/Developmental Categories/AOSWData_CYOS_20to22.csv",
row.names = TRUE)
confint(AOSWLR_CYOS_20to22)

#Info Warfare
IWData_CYOS_20to22 = subset(CYOS_20to22_Final,Dev_Cat =="2", select =
c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,Prior_E,Retained))
IWLR_CYOS_20to22 = glm(Retained ~ ., data = IWData_CYOS_20to22,
family = "binomial")
summary(IWLR_CYOS_20to22)
write.csv(IWData_CYOS_20to22,"C:/Users/User/Desktop/Astudillo_Backup/
Data/Data in Excel/offinv_annual/Female Only/Fiscal Year/
Final Thesis Data/Developmental Categories/IWData_CYOS_20to22.csv",
row.names = TRUE)
confint(IWLR_CYOS_20to22)

#Combat Support
CSData_CYOS_20to22 = subset(CYOS_20to22_Final,Dev_Cat =="3", select =
c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,Prior_E,Retained))
CSLR_CYOS_20to22 = glm(Retained ~ ., data = CSData_CYOS_20to22,
family = "binomial")
summary(CSLR_CYOS_20to22)
write.csv(CSData_CYOS_20to22,"C:/Users/User/Desktop/Astudillo_Backup/
Data/Data in Excel/offinv_annual/Female Only/Fiscal Year/
Final Thesis Data/Developmental Categories/CSData_CYOS_20to22.csv",
row.names = TRUE)
confint(CSLR_CYOS_20to22)

#Force Modernization
FMData_CYOS_20to22 = subset(CYOS_20to22_Final,Dev_Cat =="4", select =
c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,Prior_E,Retained))
FMLR_CYOS_20to22 = glm(Retained ~ ., data = FMData_CYOS_20to22,
family = "binomial")
summary(FMLR_CYOS_20to22)
#summary(FMLR_CYOS_20to22)$coefficients[,4] #Another way to calculate
#p-values
write.csv(FMData_CYOS_20to22,"C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/Developmental Categories/
FMData_CYOS_20to22.csv",row.names = TRUE)
confint(FMLR_CYOS_20to22)

#---------------------------------------------------------------------
```

```
#This section will combine all data into a cohort to determine
#significant values.

#Create the dataframe to perform logistic regression of the cohort:
cohortData = rbind(CYOS_0to6_Final,CYOS_4to8_Final,
CYOS_8to14_Final,CYOS_12to19_Final,CYOS_20to22_Final)

#Reverse order of data to reflect most recent record for each
#female officer:
revOrd_cohortData = cohortData[order(cohortData$SSAN,
cohortData$CYOS_EFY,decreasing=TRUE),]

#Now remove the duplicated information (records containing
#older information) by identified number:
cohortData_NoDups = revOrd_cohortData
[!duplicated(revOrd_cohortData$SSAN),]

#subset the data so it does not contain identification number:
cohort_LogRegData = cohortData_NoDups[c(3:9,11,12)]
write.csv(cohort_LogRegData,"C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only
/Fiscal Year/Final Thesis Data/CYOS Info/cohort_LogRegData.csv",
row.names = TRUE)


#Perform logistic regression on the Final dataframe
#containing all female officer records:
LogReg_cohortData = glm(Retained ~ ., data = cohort_LogRegData,
family = "binomial")
summary(LogReg_cohortData)

rsqValue = rsq(LogReg_cohortData,adj = FALSE)
print(rsqValue) #rsq = 0.0433967

rsqValueAdj = rsq(LogReg_cohortData,adj = TRUE)
print(rsqValue) #rsq = 0.0434

rsq_comb_Value = rsq(LogReg_cohortData,adj = FALSE,type=
c('v','kl','sse','lr','n'))
print(rsq_comb_Value) #rsq = 0.0433967

rsq_v_Value = rsq(LogReg_cohortData,adj = FALSE,type='v')
```

```
print(rsq_v_Value) #rsq = 0.0433967

rsq_kl_Value = rsq(LogReg_cohortData,adj = FALSE,type='kl')
print(rsq_kl_Value) #rsq = 0.0860785

rsq_sse_Value = rsq(LogReg_cohortData,adj = FALSE,type='sse')
print(rsq_sse_Value) #rsq = 0.04360268

rsq_lr_Value = rsq(LogReg_cohortData,adj = FALSE,type='lr')
print(rsq_lr_Value) #rsq = 0.04492576

rsq_n_Value = rsq(LogReg_cohortData,adj = FALSE,type='n')
print(rsq_n_Value) #rsq = 0.1085828


#The following code contains logistic regression and odds ratio,
#code used for this research
#Jessica Astudillo

#Create path for necessary Rtool
writeLines('PATH="S{RTOOLS40_HOME}\\usr\\bin;${PATH}"',con = "~/.
Renviron")

#Install ncessary packages for logistic regression, odds ratio,
and survival analysis
install.packages("oddsratio")
install.packages("survival")
install.packages("survminer")
install.packages("ranger")
install.packages("ggfortify")
install.packages("Rtools")
install.packages("ggsci")

#Call libraries needed for logistic regression analysis
library(readxl)
library(dplyr)
library(tidyverse)
library(readxl)
library(modelr)
library(broom)
library(oddsratio)
library(rsq)
#-------------------------------------------------------------------
```

```
#Set the work space for the data
getwd() #This is to get the working directory

#Now set the working directory to the data file
setwd("C:/Users/User/Desktop/Astudillo_Backup/Data/Data in R")

#check if directory exists on your system
file.exists("C:/Users/User/Desktop/Astudillo_Backup/Data/Data
in Excel/offinv_annual/Female Only/Fiscal Year/Final Thesis Data")

#---------------------------------------------------------------------

#Read in necessary files and perform analysis on 2009 data
data_2009_cat = read_excel("C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/FY2009_Thesis_v2.xlsx")

data_2010_cat = read_excel("C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/FY2010_Thesis_v2.xlsx")

data_2011_cat = read_excel("C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/FY2011_Thesis_v2.xlsx")

data_2012_cat = read_excel("C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/FY2012_Thesis_v2.xlsx")

data_2013_cat = read_excel("C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/FY2013_Thesis_v2.xlsx")

data_2014_cat = read_excel("C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/FY2014_Thesis_v2.xlsx")

data_2015_cat = read_excel("C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/FY2015_Thesis_v2.xlsx")
```

```
data_2016_cat = read_excel("C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/FY2016_Thesis_v2.xlsx")

data_2017_cat = read_excel("C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/FY2017_Thesis_v2.xlsx")

data_2018_cat = read_excel("C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/FY2018_Thesis_v2.xlsx")

data_2019_cat = read_excel("C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/FY2019_Thesis_v2.xlsx")

#-------------------------------------------------------------------
#-------------------------------------------------------------------
#-------------------------------------------------------------------

#Convert coded categorical variables into factors to analyze
#statistical significance of levels within each factor

data_2009_cat$MARITLST = as.factor(data_2009_cat$MARITLST)
data_2010_cat$MARITLST = as.factor(data_2010_cat$MARITLST)
data_2011_cat$MARITLST = as.factor(data_2011_cat$MARITLST)
data_2012_cat$MARITLST = as.factor(data_2012_cat$MARITLST)
data_2013_cat$MARITLST = as.factor(data_2013_cat$MARITLST)
data_2014_cat$MARITLST = as.factor(data_2014_cat$MARITLST)
data_2015_cat$MARITLST = as.factor(data_2015_cat$MARITLST)
data_2016_cat$MARITLST = as.factor(data_2016_cat$MARITLST)
data_2017_cat$MARITLST = as.factor(data_2017_cat$MARITLST)
data_2018_cat$MARITLST = as.factor(data_2018_cat$MARITLST)
data_2019_cat$MARITLST = as.factor(data_2019_cat$MARITLST)

data_2009_cat$DEPENT = as.factor(data_2009_cat$DEPENT)
data_2010_cat$DEPENT = as.factor(data_2010_cat$DEPENT)
data_2011_cat$DEPENT = as.factor(data_2011_cat$DEPENT)
data_2012_cat$DEPENT = as.factor(data_2012_cat$DEPENT)
data_2013_cat$DEPENT = as.factor(data_2013_cat$DEPENT)
data_2014_cat$DEPENT = as.factor(data_2014_cat$DEPENT)
data_2015_cat$DEPENT = as.factor(data_2015_cat$DEPENT)
data_2016_cat$DEPENT = as.factor(data_2016_cat$DEPENT)
```

```
data_2017_cat$DEPENT = as.factor(data_2017_cat$DEPENT)
data_2018_cat$DEPENT = as.factor(data_2018_cat$DEPENT)
data_2019_cat$DEPENT = as.factor(data_2019_cat$DEPENT)

data_2009_cat$Comm_Source = as.factor(data_2009_cat$Comm_Source)
data_2010_cat$Comm_Source = as.factor(data_2010_cat$Comm_Source)
data_2011_cat$Comm_Source = as.factor(data_2011_cat$Comm_Source)
data_2012_cat$Comm_Source = as.factor(data_2012_cat$Comm_Source)
data_2013_cat$Comm_Source = as.factor(data_2013_cat$Comm_Source)
data_2014_cat$Comm_Source = as.factor(data_2014_cat$Comm_Source)
data_2015_cat$Comm_Source = as.factor(data_2015_cat$Comm_Source)
data_2016_cat$Comm_Source = as.factor(data_2016_cat$Comm_Source)
data_2017_cat$Comm_Source = as.factor(data_2017_cat$Comm_Source)
data_2018_cat$Comm_Source = as.factor(data_2018_cat$Comm_Source)
data_2019_cat$Comm_Source = as.factor(data_2019_cat$Comm_Source)

data_2009_cat$DG = as.factor(data_2009_cat$DG)
data_2010_cat$DG = as.factor(data_2010_cat$DG)
data_2011_cat$DG = as.factor(data_2011_cat$DG)
data_2012_cat$DG = as.factor(data_2012_cat$DG)
data_2013_cat$DG = as.factor(data_2013_cat$DG)
data_2014_cat$DG = as.factor(data_2014_cat$DG)
data_2015_cat$DG = as.factor(data_2015_cat$DG)
data_2016_cat$DG = as.factor(data_2016_cat$DG)
data_2017_cat$DG = as.factor(data_2017_cat$DG)
data_2018_cat$DG = as.factor(data_2018_cat$DG)
data_2019_cat$DG = as.factor(data_2019_cat$DG)

data_2009_cat$Spouse_Stat = as.factor(data_2009_cat$Spouse_Stat)
data_2010_cat$Spouse_Stat = as.factor(data_2010_cat$Spouse_Stat)
data_2011_cat$Spouse_Stat = as.factor(data_2011_cat$Spouse_Stat)
data_2012_cat$Spouse_Stat = as.factor(data_2012_cat$Spouse_Stat)
data_2013_cat$Spouse_Stat = as.factor(data_2013_cat$Spouse_Stat)
data_2014_cat$Spouse_Stat = as.factor(data_2014_cat$Spouse_Stat)
data_2015_cat$Spouse_Stat = as.factor(data_2015_cat$Spouse_Stat)
data_2016_cat$Spouse_Stat = as.factor(data_2016_cat$Spouse_Stat)
data_2017_cat$Spouse_Stat = as.factor(data_2017_cat$Spouse_Stat)
data_2018_cat$Spouse_Stat = as.factor(data_2018_cat$Spouse_Stat)
data_2019_cat$Spouse_Stat = as.factor(data_2019_cat$Spouse_Stat)

data_2009_cat$Race = as.factor(data_2009_cat$Race)
data_2010_cat$Race = as.factor(data_2010_cat$Race)
data_2011_cat$Race = as.factor(data_2011_cat$Race)
```

```
data_2012_cat$Race = as.factor(data_2012_cat$Race)
data_2013_cat$Race = as.factor(data_2013_cat$Race)
data_2014_cat$Race = as.factor(data_2014_cat$Race)
data_2015_cat$Race = as.factor(data_2015_cat$Race)
data_2016_cat$Race = as.factor(data_2016_cat$Race)
data_2017_cat$Race = as.factor(data_2017_cat$Race)
data_2018_cat$Race = as.factor(data_2018_cat$Race)
data_2019_cat$Race = as.factor(data_2019_cat$Race)

data_2009_cat$Deployments = as.factor(data_2009_cat$Deployments)
data_2010_cat$Deployments = as.factor(data_2010_cat$Deployments)
data_2011_cat$Deployments = as.factor(data_2011_cat$Deployments)
data_2012_cat$Deployments = as.factor(data_2012_cat$Deployments)
data_2013_cat$Deployments = as.factor(data_2013_cat$Deployments)
data_2014_cat$Deployments = as.factor(data_2014_cat$Deployments)
data_2015_cat$Deployments = as.factor(data_2015_cat$Deployments)
data_2016_cat$Deployments = as.factor(data_2016_cat$Deployments)
data_2017_cat$Deployments = as.factor(data_2017_cat$Deployments)
data_2018_cat$Deployments = as.factor(data_2018_cat$Deployments)
data_2019_cat$Deployments = as.factor(data_2019_cat$Deployments)

data_2009_cat$Prior_E = as.factor(data_2009_cat$Prior_E)
data_2010_cat$Prior_E = as.factor(data_2010_cat$Prior_E)
data_2011_cat$Prior_E = as.factor(data_2011_cat$Prior_E)
data_2012_cat$Prior_E = as.factor(data_2012_cat$Prior_E)
data_2013_cat$Prior_E = as.factor(data_2013_cat$Prior_E)
data_2014_cat$Prior_E = as.factor(data_2014_cat$Prior_E)
data_2015_cat$Prior_E = as.factor(data_2015_cat$Prior_E)
data_2016_cat$Prior_E = as.factor(data_2016_cat$Prior_E)
data_2017_cat$Prior_E = as.factor(data_2017_cat$Prior_E)
data_2018_cat$Prior_E = as.factor(data_2018_cat$Prior_E)
data_2019_cat$Prior_E = as.factor(data_2019_cat$Prior_E)


#--------------------------------------------------------------------
#--------------------------------------------------------------------
#--------------------------------------------------------------------

#Now perform logistic regression on each commissioning year group:
#0 to 20+ based on each CYOS group

#-------------------------------------------------------------------
#-------------------------------------------------------------------
```

```
# Years 0 to 6

#2009-------------------------------------------------------------
YG1_catData_2009_1 = subset(data_2009_cat,CYOS_EFY >= 6 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))

YG1_catData_2009_2 = subset(data_2009_cat,CYOS_EFY <= 5 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#combine both files
CYOS_0to6_2009_catData = rbind(YG1_catData_2009_1,YG1_catData_2009_2)

#2010-------------------------------------------------------------

YG1_catData_2010_1 = subset(data_2010_cat,CYOS_EFY >= 6 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))

YG1_catData_2010_2 = subset(data_2010_cat,CYOS_EFY <= 5 &
Retain < 1,select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#combine both files
CYOS_0to6_2010_catData= rbind(YG1_catData_2010_1,YG1_catData_2010_2)


#2011-------------------------------------------------------------
YG1_catData_2011_1 = subset(data_2011_cat,CYOS_EFY >= 6 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))

YG1_catData_2011_2 = subset(data_2011_cat,CYOS_EFY <= 5 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#combine both files
CYOS_0to6_2011_catData = rbind(YG1_catData_2011_1,YG1_catData_2011_2)


#2012-------------------------------------------------------------
YG1_catData_2012_1 = subset(data_2012_cat,CYOS_EFY >= 6 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
```

```
YG1_catData_2012_2 = subset(data_2012_cat,CYOS_EFY < 6 & Retain < 1,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,
DG,Spouse_Stat,Race,Deployments:Prior_E))
#combine both files
CYOS_0to6_2012_catData = rbind(YG1_catData_2012_1,YG1_catData_2012_2)


#2013-----------------------------------------------------------
YG1_catData_2013_1 = subset(data_2013_cat,CYOS_EFY >= 6 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT
,Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))

YG1_catData_2013_2 = subset(data_2013_cat,CYOS_EFY < 6 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#combine both files
CYOS_0to6_2013_catData = rbind(YG1_catData_2013_1,YG1_catData_2013_2)


#2014-----------------------------------------------------------
YG1_catData_2014_1 = subset(data_2014_cat,CYOS_EFY >= 6 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))

YG1_catData_2014_2 = subset(data_2014_cat,CYOS_EFY < 6 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#combine both files
CYOS_0to6_2014_catData = rbind(YG1_catData_2014_1,YG1_catData_2014_2)


#2015-----------------------------------------------------------
YG1_catData_2015_1 = subset(data_2015_cat,CYOS_EFY >= 6 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))

YG1_catData_2015_2 = subset(data_2015_cat,CYOS_EFY < 6 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#combine both files
CYOS_0to6_2015_catData= rbind(YG1_catData_2015_1,YG1_catData_2015_2)
```

```
#2016-----------------------------------------------------------
YG1_catData_2016_1 = subset(data_2016_cat,CYOS_EFY >= 6 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))

YG1_catData_2016_2 = subset(data_2016_cat,CYOS_EFY < 6 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#combine both files
CYOS_0to6_2016_catData = rbind(YG1_catData_2016_1,YG1_catData_2016_2)


#2017-----------------------------------------------------------
YG1_catData_2017_1 = subset(data_2017_cat,CYOS_EFY >= 6 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))

YG1_catData_2017_2 = subset(data_2017_cat,CYOS_EFY < 6 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#combine both files
CYOS_0to6_2017_catData = rbind(YG1_catData_2017_1,YG1_catData_2017_2)


#2018-----------------------------------------------------------
YG1_catData_2018_1 = subset(data_2018_cat,CYOS_EFY >= 6 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))

YG1_catData_2018_2 = subset(data_2018_cat,CYOS_EFY < 6 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#combine both files
CYOS_0to6_2018_catData = rbind(YG1_catData_2018_1,YG1_catData_2018_2)


#2019-----------------------------------------------------------
YG1_catData_2019_1 = subset(data_2019_cat,CYOS_EFY >= 6 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))

YG1_catData_2019_2 = subset(data_2019_cat,CYOS_EFY < 6 &
```

```
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#combine both files
CYOS_0to6_2019_catData = rbind(YG1_catData_2019_1,YG1_catData_2019_2)


#Combine all data frames containing CYOS 0-6 now-------------------
CYOS_0to6_catCombined = rbind(CYOS_0to6_2009_catData,
CYOS_0to6_2010_catData,CYOS_0to6_2011_catData,CYOS_0to6_2012_catData,
CYOS_0to6_2013_catData,CYOS_0to6_2014_catData,CYOS_0to6_2015_catData,
CYOS_0to6_2016_catData,CYOS_0to6_2017_catData,CYOS_0to6_2018_catData,
CYOS_0to6_2019_catData)

#Reverse order of data to ensure the most recent record is
#obtained when removing duplicates
revOrd_CYOS_0to6_catComb = CYOS_0to6_catCombined
[order(CYOS_0to6_catCombined$SSAN,
CYOS_0to6_catCombined$CYOS_EFY,decreasing=TRUE),]

#Now remove the duplicated information (records containing
#older information) by identified number:
CYOS_0to6_catNoDups = revOrd_CYOS_0to6_catComb
[!duplicated(revOrd_CYOS_0to6_catComb$SSAN),]

#Add a new column called 'Retained' with values = 0
CYOS_0to6_catNoDups$Retained = 0

#Change all retained values to 1 if CYOS >= 6. Currently
#have 0 values.
CYOS_0to6_catFinal = CYOS_0to6_catNoDups %>% mutate(
Retained=replace(Retained,CYOS_EFY >=6, 1)) %>% as.data.frame()

#Now remove the 'Retain' column since 'Retained' column now
#has the updated information..
CYOS_0to6_catFinal$Retain = NULL
transform(CYOS_0to6_catFinal, Retained = as.numeric(Retained))
#transform to numeric value if necessary

#subset the data so it does not contain SSAN:
CYOS_0to6_catLogRegData = CYOS_0to6_catFinal[c(4,5,7:10,12,13)]
write.csv(CYOS_0to6_catLogRegData,"C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/Developmental
```

```
Categories/CYOS_0to6_LogRegData_Categorical.csv",row.names = TRUE)

#Perform logistic regression on the Final dataframe containing
#all fiscal years for CYOS 0 to 6:
catLogReg_CYOS_0to6 = glm(Retained ~ ., data = CYOS_0to6_catLogRegData,
family = "binomial")
summary(catLogReg_CYOS_0to6)


#Perform Logistic Regression on the Developmental Categories based
#on CYOS 0 to 6--------------

#Air Operations and Special Warfare
combAOSWData_CYOS_0to6 = subset(CYOS_0to6_catFinal,Dev_Cat =="1",
select = c(SSAN,DAFSCNR,MARITLST,DEPENT,DG,Spouse_Stat,Race,
Deployments,CYOS_EFY,Prior_E,Retained))
catAOSWData_CYOS_0to6 = subset(combAOSWData_CYOS_0to6, select =
c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,Prior_E,Retained))
catAOSWLR_CYOS_0to6 = glm(Retained ~ ., data = catAOSWData_CYOS_0to6,
family = "binomial")
summary(catAOSWLR_CYOS_0to6)
write.csv(catAOSWData_CYOS_0to6,"C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/Developmental
Categories/AOSWDATA_CYOS_0to6_Cat.csv",row.names = TRUE)
confint(catAOSWLR_CYOS_0to6)

#Info Warfare
combIWData_CYOS_0to6 = subset(CYOS_0to6_catFinal,Dev_Cat =="2",
select = c(SSAN,DAFSCNR,MARITLST,DEPENT,DG,Spouse_Stat,Race,
Deployments,CYOS_EFY,Prior_E,Retained))
catIWData_CYOS_0to6 = subset(combIWData_CYOS_0to6, select =
c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,Prior_E,Retained))
catIWLR_CYOS_0to6 = glm(Retained ~ ., data = catIWData_CYOS_0to6,
family = "binomial")
summary(catIWLR_CYOS_0to6)
write.csv(catIWData_CYOS_0to6,"C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/Developmental
Categories/IWDATA_CYOS_0to6_Cat.csv",row.names = TRUE)
confint(catIWLR_CYOS_0to6)
```

```
#Combat Support
combCSData_CYOS_0to6 = subset(CYOS_0to6_catFinal,Dev_Cat =="3",
select = c(SSAN,DAFSCNR,MARITLST,DEPENT,DG,Spouse_Stat,Race,
Deployments,CYOS_EFY,Prior_E,Retained))
catCSData_CYOS_0to6 = subset(combCSData_CYOS_0to6, select =
c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,Prior_E,Retained))
catCSLR_CYOS_0to6 = glm(Retained ~ ., data = catCSData_CYOS_0to6,
family = "binomial")
summary(catCSLR_CYOS_0to6)
write.csv(catCSData_CYOS_0to6,"C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/Developmental
Categories/CSDATA_CYOS_0to6_Cat.csv",row.names = TRUE)
confint(catCSLR_CYOS_0to6)

#Force Modernization
combFMData_CYOS_0to6 = subset(CYOS_0to6_catFinal,Dev_Cat =="4",
select = c(SSAN,DAFSCNR,MARITLST,DEPENT,DG,Spouse_Stat,Race,
Deployments,CYOS_EFY,Prior_E,Retained))
catFMData_CYOS_0to6 = subset(combFMData_CYOS_0to6, select =
c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,Prior_E,Retained))
catFMLR_CYOS_0to6 = glm(Retained ~ ., data = catFMData_CYOS_0to6,
family = "binomial")
summary(catFMLR_CYOS_0to6)
write.csv(catFMData_CYOS_0to6,"C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/DevelopmentalCategories/
FMDATA_CYOS_0to6_Cat.csv",row.names = TRUE)
confint(catFMLR_CYOS_0to6)


#---------------------------------------------------------------------
#---------------------------------------------------------------------

# Years 4 to 8

#2009---------------------------------------------------------------

#If data is greater than or equal to 5 YOS then keep
YG2_catData_2009_1 = subset(data_2009_cat,CYOS_EFY >= 5, select =
c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,Race,
Deployments:Prior_E))
#Subset the data with values greater than 8 YOS and retain values
```

```
#are 0 or 1
YG2_catData_2009_2 = subset(YG2_catData_2009_1,CYOS_EFY >= 8 &
Retain >= 0, select =c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,
DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 7 YOS and retain = 0
YG2_catData_2009_3 = subset(YG2_catData_2009_1, CYOS_EFY <=7 &
Retain < 1, select =c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_4to8_2009_catData= rbind(YG2_catData_2009_2,YG2_catData_2009_3)

#2010-----------------------------------------------------------------

#If data is greater than or equal to 5 YOS then keep
YG2_catData_2010_1 = subset(data_2010_cat,CYOS_EFY >= 5,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,
Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values greater than 8 YOS and retain
#values are 0 or 1
YG2_catData_2010_2 = subset(YG2_catData_2010_1,CYOS_EFY >= 8 &
Retain >= 0, select =c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 7 YOS and retain = 0
YG2_catData_2010_3 = subset(YG2_catData_2010_1, CYOS_EFY <=7 &
Retain < 1, select =c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_4to8_2010_catData= rbind(YG2_catData_2010_2,YG2_catData_2010_3)

#2011-----------------------------------------------------------------

#If data is greater than or equal to 5 YOS then keep
YG2_catData_2011_1 = subset(data_2011_cat,CYOS_EFY >= 5,
select =c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,
DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values greater than 8 YOS and retain values
#are 0 or 1
YG2_catData_2011_2 = subset(YG2_catData_2011_1,CYOS_EFY >= 8 &
Retain >= 0, select =c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,
Race,Deployments:Prior_E))
#Subset the data with values less than 7 YOS and retain = 0
YG2_catData_2011_3 = subset(YG2_catData_2011_1, CYOS_EFY <=7 &
```

```
Retain < 1, select =c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_4to8_2011_catData= rbind(YG2_catData_2011_2,YG2_catData_2011_3)


#2012-----------------------------------------------------------------

#If data is greater than or equal to 5 YOS then keep
YG2_catData_2012_1 = subset(data_2012_cat,CYOS_EFY >= 5,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,
DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values greater than 8 YOS and retain
#values are 0 or 1
YG2_catData_2012_2 = subset(YG2_catData_2012_1,CYOS_EFY >= 8 &
Retain >= 0, select =c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 7 YOS and retain = 0
YG2_catData_2012_3 = subset(YG2_catData_2012_1, CYOS_EFY <=7 &
Retain < 1, select =c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_4to8_2012_catData= rbind(YG2_catData_2012_2,YG2_catData_2012_3)

#2013-----------------------------------------------------------------

#If data is greater than or equal to 5 YOS then keep
YG2_catData_2013_1 = subset(data_2013_cat,CYOS_EFY >= 5,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,
DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values greater than 8 YOS and retain
#values are 0 or 1
YG2_catData_2013_2 = subset(YG2_catData_2013_1,CYOS_EFY >= 8 &
Retain >= 0, select =c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 7 YOS and retain = 0
YG2_catData_2013_3 = subset(YG2_catData_2013_1, CYOS_EFY <=7 &
Retain < 1, select =c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_4to8_2013_catData= rbind(YG2_catData_2013_2,YG2_catData_2011_3)

#2014-----------------------------------------------------------------
```

```
#If data is greater than or equal to 5 YOS then keep
YG2_catData_2014_1 = subset(data_2014_cat,CYOS_EFY >= 5,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,
Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values greater than 8 YOS and retain
#values are 0 or 1
YG2_catData_2014_2 = subset(YG2_catData_2014_1,CYOS_EFY >= 8 &
Retain >= 0, select =c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 7 YOS and retain = 0
YG2_catData_2014_3 = subset(YG2_catData_2014_1, CYOS_EFY <=7 &
Retain < 1, select =c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_4to8_2014_catData= rbind(YG2_catData_2014_2,YG2_catData_2014_3)

#2015-----------------------------------------------------------

#If data is greater than or equal to 5 YOS then keep
YG2_catData_2015_1 = subset(data_2015_cat,CYOS_EFY >= 5,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,
DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values greater than 8 YOS and retain
#values are 0 or 1
YG2_catData_2015_2 = subset(YG2_catData_2015_1,CYOS_EFY >= 8 &
Retain >= 0, select =c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 7 YOS and retain = 0
YG2_catData_2015_3 = subset(YG2_catData_2015_1, CYOS_EFY <= 7 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_4to8_2015_catData= rbind(YG2_catData_2015_2,YG2_catData_2015_3)


#2016-----------------------------------------------------------

#If data is greater than or equal to 5 YOS then keep
YG2_catData_2016_1 = subset(data_2016_cat,CYOS_EFY >= 5,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,
DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values greater than 8 YOS and retain
```

```
#values are 0 or 1
YG2_catData_2016_2 = subset(YG2_catData_2016_1,CYOS_EFY >= 8 &
Retain >= 0, select =c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 7 YOS and retain = 0
YG2_catData_2016_3 = subset(YG2_catData_2016_1, CYOS_EFY <=7
& Retain < 1, select =c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_4to8_2016_catData= rbind(YG2_catData_2016_2,YG2_catData_2016_3)

#2017-----------------------------------------------------------------

#If data is greater than or equal to 5 YOS then keep
YG2_catData_2017_1 = subset(data_2017_cat,CYOS_EFY >= 5,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,
Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values greater than 8 YOS and retain
#values are 0 or 1
YG2_catData_2017_2 = subset(YG2_catData_2017_1,CYOS_EFY >= 8 &
Retain >= 0, select =c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 7 YOS and retain = 0
YG2_catData_2017_3 = subset(YG2_catData_2017_1, CYOS_EFY <=7 &
Retain < 1, select =c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_4to8_2017_catData= rbind(YG2_catData_2017_2,YG2_catData_2017_3)


#2018-----------------------------------------------------------------

#If data is greater than or equal to 5 YOS then keep
YG2_catData_2018_1 = subset(data_2018_cat,CYOS_EFY >= 5,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,
DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values greater than 8 YOS and retain
#values are 0 or 1
YG2_catData_2018_2 = subset(YG2_catData_2018_1,CYOS_EFY >= 8 &
Retain >= 0, select =c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 7 YOS and retain = 0
YG2_catData_2018_3 = subset(YG2_catData_2018_1, CYOS_EFY <=7 &
```

```
Retain < 1, select =c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_4to8_2018_catData= rbind(YG2_catData_2018_2,YG2_catData_2018_3)


#2019-----------------------------------------------------------

#If data is greater than or equal to 5 YOS then keep
YG2_catData_2019_1 = subset(data_2019_cat,CYOS_EFY >= 5,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,
DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values greater than 8 YOS and retain
#values are 0 or 1
YG2_catData_2019_2 = subset(YG2_catData_2019_1,CYOS_EFY >= 8 &
Retain >= 0, select =c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 7 YOS and retain = 0
YG2_catData_2019_3 = subset(YG2_catData_2019_1, CYOS_EFY <=7 &
Retain < 1, select =c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_4to8_2019_catData= rbind(YG2_catData_2019_2,YG2_catData_2019_3)


#Combine all data frames containing CYOS 4-8 now-------------------
CYOS_4to8_catCombined = rbind(CYOS_4to8_2009_catData,
CYOS_4to8_2010_catData,CYOS_4to8_2011_catData,CYOS_4to8_2012_catData,
CYOS_4to8_2013_catData,CYOS_4to8_2014_catData,CYOS_4to8_2015_catData,
CYOS_4to8_2016_catData,CYOS_4to8_2017_catData,CYOS_4to8_2018_catData,
CYOS_4to8_2019_catData)

#Reverse order of data to ensure the most recent record is obtained
#when removing duplicates
revOrd_CYOS_4to8_catComb = CYOS_4to8_catCombined[order(
CYOS_4to8_catCombined$SSAN,CYOS_4to8_catCombined$CYOS_EFY,
decreasing=TRUE),]

#Now remove the duplicated information (records containing
#older information) by identified number:
CYOS_4to8_catNoDups = revOrd_CYOS_4to8_catComb
[!duplicated(revOrd_CYOS_4to8_catComb$SSAN),]
```

```
#Add a new column called 'Retained' with values = 0
CYOS_4to8_catNoDups$Retained = 0

#Change all retained values to 1 if CYOS >= 8. Some currently
#have 0 values.
CYOS_4to8_catFinal = CYOS_4to8_catNoDups %>% mutate(
Retained=replace(Retained,CYOS_EFY >= 8, 1)) %>% as.data.frame()

#Now remove the 'Retain' column since 'Retained' column now
#has the updated information..
CYOS_4to8_catFinal$Retain = NULL
transform(CYOS_4to8_catFinal, Retained = as.numeric(Retained))
#transform to numeric value if necessary

#subset the data so it does not contain SSAN:
CYOS_4to8_catLogRegData = CYOS_4to8_catFinal[c(4,5,7:10,12,13)]
write.csv(CYOS_0to6_catLogRegData,"C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/Developmental
Categories/CYOS_4to8_LogRegData_Categorical.csv",
row.names = TRUE)

#Perform logistic regression on the Final dataframe containing
#all fiscal years for CYOS 4 to 8:
catLogReg_CYOS_4to8 = glm(Retained ~ ., data =
CYOS_4to8_catLogRegData,family = "binomial")
summary(catLogReg_CYOS_4to8)


#Perform Logistic Regression on the Developmental Categories
#based on CYOS 4 to 8--------------

#Air Operations and Special Warfare
combAOSWData_CYOS_4to8 = subset(CYOS_4to8_catFinal,Dev_Cat =="1",
select = c(SSAN,DAFSCNR,MARITLST,DEPENT,DG,Spouse_Stat,Race,
Deployments,CYOS_EFY,Prior_E,Retained))
catAOSWData_CYOS_4to8 = subset(combAOSWData_CYOS_4to8,
select = c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,
Prior_E,Retained))
catAOSWLR_CYOS_4to8 = glm(Retained ~ ., data = catAOSWData_CYOS_4to8,
family = "binomial")
summary(catAOSWLR_CYOS_4to8)
write.csv(catAOSWData_CYOS_4to8,"C:/Users/User/Desktop/
```

```
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/Developmental
Categories/AOSWDATA_CYOS_4to8_Cat.csv",row.names = TRUE)
confint(catAOSWLR_CYOS_4to8)

#Info Warfare
combIWData_CYOS_4to8 = subset(CYOS_4to8_catFinal,Dev_Cat =="2",
select = c(SSAN,DAFSCNR,MARITLST,DEPENT,DG,Spouse_Stat,Race,
Deployments,CYOS_EFY,Prior_E,Retained))
catIWData_CYOS_4to8 = subset(combIWData_CYOS_4to8, select =
c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,Prior_E,Retained))
catIWLR_CYOS_4to8 = glm(Retained ~ ., data = catIWData_CYOS_4to8,
family = "binomial")
summary(catIWLR_CYOS_4to8)
write.csv(catIWData_CYOS_4to8,"C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/Developmental
Categories/IWData_CYOS_4to8_Cat.csv",row.names = TRUE)
confint(catIWLR_CYOS_4to8)

#Combat Support
combCSData_CYOS_4to8 = subset(CYOS_4to8_catFinal,Dev_Cat =="3",
select = c(SSAN,DAFSCNR,MARITLST,DEPENT,DG,Spouse_Stat,Race,
Deployments,CYOS_EFY,Prior_E,Retained))
catCSData_CYOS_4to8 = subset(combCSData_CYOS_4to8, select =
c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,Prior_E,Retained))
catCSLR_CYOS_4to8 = glm(Retained ~ ., data = catCSData_CYOS_4to8,
family = "binomial")
summary(catCSLR_CYOS_4to8)
write.csv(catCSData_CYOS_4to8,"C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/Developmental
Categories/CSData_CYOS_4to8_Cat.csv",row.names = TRUE)
confint(catCSLR_CYOS_4to8)

#Force Modernization
combFMData_CYOS_4to8 = subset(CYOS_4to8_catFinal,Dev_Cat =="4",
select = c(SSAN,DAFSCNR,MARITLST,DEPENT,DG,Spouse_Stat,Race,
Deployments,CYOS_EFY,Prior_E,Retained))
catFMData_CYOS_4to8 = subset(combFMData_CYOS_4to8, select =
c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,Prior_E,Retained))
catFMLR_CYOS_4to8 = glm(Retained ~ ., data = catFMData_CYOS_4to8,
family = "binomial")
```

```
summary(catFMLR_CYOS_4to8)
write.csv(catFMData_CYOS_4to8,"C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/Developmental
Categories/FMData_CYOS_4to8_Cat.csv",row.names = TRUE)
confint(catFMLR_CYOS_4to8)


# Years 8 to 14

#2009------------------------------------------------------------------

#If data is greater than or equal to 9 YOS then keep
YG3_catData_2009_1 = subset(data_2009_cat,CYOS_EFY >= 9,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,
Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values greater than 14 YOS and retain
#values are 0 or 1
YG3_catData_2009_2 = subset(YG3_catData_2009_1,CYOS_EFY >= 14 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 13 YOS and retain = 0
YG3_catData_2009_3 = subset(YG3_catData_2009_1, CYOS_EFY <= 13 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_8to14_2009_catData= rbind(YG3_catData_2009_2,YG3_catData_2009_3)


#2010------------------------------------------------------------------

#If data is greater than or equal to 9 YOS then keep
YG3_catData_2010_1 = subset(data_2010_cat,CYOS_EFY >= 9,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,
Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values greater than 14 YOS and retain
#values are 0 or 1
YG3_catData_2010_2 = subset(YG3_catData_2010_1,CYOS_EFY >= 14 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 13 YOS and retain = 0
YG3_catData_2010_3 = subset(YG3_catData_2010_1, CYOS_EFY <= 13 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
```

```
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_8to14_2010_catData= rbind(YG3_catData_2010_2,YG3_catData_2010_3)


#2011-----------------------------------------------------------------

#If data is greater than or equal to 9 YOS then keep
YG3_catData_2011_1 = subset(data_2011_cat,CYOS_EFY >= 9,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,
Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values greater than 14 YOS and retain
#values are 0 or 1
YG3_catData_2011_2 = subset(YG3_catData_2011_1,CYOS_EFY >= 14 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 13 YOS and retain = 0
YG3_catData_2011_3 = subset(YG3_catData_2011_1, CYOS_EFY <= 13 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_8to14_2011_catData= rbind(YG3_catData_2011_2,YG3_catData_2011_3)


#2012-----------------------------------------------------------------

#If data is greater than or equal to 9 YOS then keep
YG3_catData_2012_1 = subset(data_2012_cat,CYOS_EFY >= 9,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,
Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values greater than 14 YOS and retain
#values are 0 or 1
YG3_catData_2012_2 = subset(YG3_catData_2012_1,CYOS_EFY >= 14 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 13 YOS and retain = 0
YG3_catData_2012_3 = subset(YG3_catData_2012_1, CYOS_EFY <= 13 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_8to14_2012_catData= rbind(YG3_catData_2012_2,YG3_catData_2012_3)
```

```
#2013-----------------------------------------------------------------

#If data is greater than or equal to 9 YOS then keep
YG3_catData_2013_1 = subset(data_2013_cat,CYOS_EFY >= 9,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,
Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values greater than 14 YOS and retain
#values are 0 or 1
YG3_catData_2013_2 = subset(YG3_catData_2013_1,CYOS_EFY >= 14 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 13 YOS and retain = 0
YG3_catData_2013_3 = subset(YG3_catData_2013_1, CYOS_EFY <= 13 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_8to14_2013_catData= rbind(YG3_catData_2013_2,YG3_catData_2013_3)


#2014-----------------------------------------------------------------

#If data is greater than or equal to 9 YOS then keep
YG3_catData_2014_1 = subset(data_2014_cat,CYOS_EFY >= 9,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,
Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values greater than 14 YOS and retain
#values are 0 or 1
YG3_catData_2014_2 = subset(YG3_catData_2014_1,CYOS_EFY >= 14 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 13 YOS and retain = 0
YG3_catData_2014_3 = subset(YG3_catData_2014_1, CYOS_EFY <= 13 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_8to14_2014_catData= rbind(YG3_catData_2014_2,YG3_catData_2014_3)


#2015-----------------------------------------------------------------

#If data is greater than or equal to 9 YOS then keep
YG3_catData_2015_1 = subset(data_2015_cat,CYOS_EFY >= 9,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,
```

```
Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values greater than 14 YOS and retain
#values are 0 or 1
YG3_catData_2015_2 = subset(YG3_catData_2015_1,CYOS_EFY >= 14 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 13 YOS and retain = 0
YG3_catData_2015_3 = subset(YG3_catData_2015_1, CYOS_EFY <= 13 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_8to14_2015_catData= rbind(YG3_catData_2015_2,YG3_catData_2015_3)


#2016-----------------------------------------------------------------

#If data is greater than or equal to 9 YOS then keep
YG3_catData_2016_1 = subset(data_2016_cat,CYOS_EFY >= 9,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,
Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values greater than 14 YOS and retain
#values are 0 or 1
YG3_catData_2016_2 = subset(YG3_catData_2016_1,CYOS_EFY >= 14 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 13 YOS and retain = 0
YG3_catData_2016_3 = subset(YG3_catData_2016_1, CYOS_EFY <= 13 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_8to14_2016_catData= rbind(YG3_catData_2016_2,YG3_catData_2016_3)


#2017-----------------------------------------------------------------

#If data is greater than or equal to 9 YOS then keep
YG3_catData_2017_1 = subset(data_2017_cat,CYOS_EFY >= 9,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,
Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values greater than 14 YOS and retain
#values are 0 or 1
YG3_catData_2017_2 = subset(YG3_catData_2017_1,CYOS_EFY >= 14 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
```

```
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 13 YOS and retain = 0
YG3_catData_2017_3 = subset(YG3_catData_2017_1, CYOS_EFY <= 13 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_8to14_2017_catData= rbind(YG3_catData_2017_2,YG3_catData_2017_3)


#2018-----------------------------------------------------------------

#If data is greater than or equal to 9 YOS then keep
YG3_catData_2018_1 = subset(data_2018_cat,CYOS_EFY >= 9,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,
Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values greater than 14 YOS and retain
#values are 0 or 1
YG3_catData_2018_2 = subset(YG3_catData_2018_1,CYOS_EFY >= 14 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 13 YOS and retain = 0
YG3_catData_2018_3 = subset(YG3_catData_2018_1, CYOS_EFY <= 13 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_8to14_2018_catData= rbind(YG3_catData_2018_2,YG3_catData_2018_3)


#2019-----------------------------------------------------------------

#If data is greater than or equal to 9 YOS then keep
YG3_catData_2019_1 = subset(data_2019_cat,CYOS_EFY >= 9, select =
c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,
Race,Deployments:Prior_E))
#Subset the data with values greater than 14 YOS and retain
#values are 0 or 1
YG3_catData_2019_2 = subset(YG3_catData_2019_1,CYOS_EFY >= 14 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 13 YOS and retain = 0
YG3_catData_2019_3 = subset(YG3_catData_2019_1, CYOS_EFY <= 13 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT
,Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
```

```
#Combine both files
CYOS_8to14_2019_catData= rbind(YG3_catData_2019_2,YG3_catData_2019_3)


#Combine all data frames containing CYOS 8-14 now-------------------
CYOS_8to14_catCombined = rbind(CYOS_8to14_2009_catData,
CYOS_8to14_2010_catData,CYOS_8to14_2011_catData,
CYOS_8to14_2012_catData,CYOS_8to14_2013_catData,
CYOS_8to14_2014_catData,CYOS_8to14_2015_catData,
CYOS_8to14_2016_catData,CYOS_8to14_2017_catData,
CYOS_8to14_2018_catData,CYOS_8to14_2019_catData)

#Reverse order of data to ensure the most recent record is
#obtained when removing duplicates
revOrd_CYOS_8to14_catComb = CYOS_8to14_catCombined[
order(CYOS_8to14_catCombined$SSAN,
CYOS_8to14_catCombined$CYOS_EFY,decreasing=TRUE),]

#Now remove the duplicated information (records containing
#older information) by identified number:
CYOS_8to14_catNoDups = revOrd_CYOS_8to14_catComb
[!duplicated(revOrd_CYOS_8to14_catComb$SSAN),]

#Add a new column called 'Retained' with values = 0
CYOS_8to14_catNoDups$Retained = 0

#Change all retained values to 1 if CYOS >= 8. Some
#currently have 0 values.
CYOS_8to14_catFinal = CYOS_8to14_catNoDups %>% mutate(
Retained=replace(Retained,CYOS_EFY >= 14, 1)) %>% as.data.frame()

#Now remove the 'Retain' column since 'Retained' column now
#has the updated information..
CYOS_8to14_catFinal$Retain = NULL
transform(CYOS_8to14_catFinal, Retained = as.numeric(Retained))
#transform to numeric value if necessary

#subset the data so it does not contain SSAN:
CYOS_8to14_catLogRegData = CYOS_8to14_catFinal[c(4,5,7:10,12,13)]
write.csv(CYOS_8to14_catLogRegData,"C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/Developmental
Categories/CYOS_8to14_LogRegData_Categorical.csv",row.names = TRUE)
```

```
#Perform logistic regression on the Final dataframe containing all
fiscal years for CYOS 8 to 14:
catLogReg_CYOS_8to14 = glm(Retained ~ ., data =
CYOS_8to14_catLogRegData,family = "binomial")
summary(catLogReg_CYOS_8to14)


#Perform Logistic Regression on the Developmental Categories
#based on CYOS 8 to 14--------------

#Air Operations and Special Warfare
combAOSWData_CYOS_8to14 = subset(CYOS_8to14_catFinal,Dev_Cat =="1",
select = c(SSAN,DAFSCNR,MARITLST,DEPENT,DG,Spouse_Stat,Race,
Deployments,CYOS_EFY,Prior_E,Retained))
catAOSWData_CYOS_8to14 = subset(combAOSWData_CYOS_8to14,select =
c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,Prior_E,Retained))
catAOSWLR_CYOS_8to14 = glm(Retained ~ ., data = catAOSWData_CYOS_8to14,
family = "binomial")
summary(catAOSWLR_CYOS_8to14)
write.csv(catAOSWData_CYOS_8to14,"C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/Developmental
Categories/AOSWData_CYOS_8to14_Cat.csv",row.names = TRUE)
confint(catAOSWLR_CYOS_8to14)

#Info Warfare
combIWData_CYOS_8to14 = subset(CYOS_8to14_catFinal,Dev_Cat =="2",
select = c(SSAN,DAFSCNR,MARITLST,DEPENT,DG,Spouse_Stat,Race,
Deployments,CYOS_EFY,Prior_E,Retained))
catIWData_CYOS_8to14 = subset(combIWData_CYOS_8to14, select =
c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,Prior_E,Retained))
catIWLR_CYOS_8to14 = glm(Retained ~ ., data = catIWData_CYOS_8to14,
family = "binomial")
summary(catIWLR_CYOS_8to14)
write.csv(catIWData_CYOS_8to14,"C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/Developmental
Categories/IWData_CYOS_8to14_Cat.csv",row.names = TRUE)
confint(catIWLR_CYOS_8to14)

#Combat Support
combCSData_CYOS_8to14 = subset(CYOS_8to14_catFinal,Dev_Cat =="3",
```

```
select = c(SSAN,DAFSCNR,MARITLST,DEPENT,DG,Spouse_Stat,Race,
Deployments,CYOS_EFY,Prior_E,Retained))
catCSData_CYOS_8to14 = subset(combCSData_CYOS_8to14, select =
c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,Prior_E,Retained))
catCSLR_CYOS_8to14 = glm(Retained ~ ., data = catCSData_CYOS_8to14,
family = "binomial")
summary(catCSLR_CYOS_8to14)
write.csv(catCSData_CYOS_8to14,"C:/Users/Jess/Desktop/Astudillo_Backup/
Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/Developmental
Categories/CSData_CYOS_8to14_Cat.csv",row.names = TRUE)
confint(catCSLR_CYOS_8to14)

#Force Modernization
combFMData_CYOS_8to14 = subset(CYOS_8to14_catFinal,Dev_Cat =="4",
select = c(SSAN,DAFSCNR,MARITLST,DEPENT,DG,Spouse_Stat,Race,
Deployments,CYOS_EFY,Prior_E,Retained))
catFMData_CYOS_8to14 = subset(combFMData_CYOS_8to14, select =
c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,Prior_E,Retained))
catFMLR_CYOS_8to14 = glm(Retained ~ ., data = catFMData_CYOS_8to14,
family = "binomial")
summary(catFMLR_CYOS_8to14)
write.csv(catFMData_CYOS_8to14,"C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/Developmental
Categories/FMData_CYOS_8to14_Cat.csv",row.names = TRUE)
confint(catFMLR_CYOS_8to14)


#--------------------------------------------------------------------
#--------------------------------------------------------------------

# Years 12 to 19

#2009------------------------------------------------------------

#If data is greater than or equal to 12 YOS then keep
YG4_catData_2009_1 = subset(data_2009_cat,CYOS_EFY >= 12,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,
Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values greater than 19 YOS and retain
#values are 0 or 1
YG4_catData_2009_2 = subset(YG4_catData_2009_1,CYOS_EFY >= 19 &
```

```
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 18 YOS and retain = 0
YG4_catData_2009_3 = subset(YG4_catData_2009_1, CYOS_EFY <= 18 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_12to19_2009_catData= rbind(YG4_catData_2009_2,YG4_catData_2009_3)


#2010------------------------------------------------------------

#If data is greater than or equal to 12 YOS then keep
YG4_catData_2010_1 = subset(data_2010_cat,CYOS_EFY >= 12,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,
Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values greater than 19 YOS and retain
#values are 0 or 1
YG4_catData_2010_2 = subset(YG4_catData_2010_1,CYOS_EFY >= 19 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 18 YOS and retain = 0
YG4_catData_2010_3 = subset(YG4_catData_2010_1, CYOS_EFY <= 18 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_12to19_2010_catData= rbind(YG4_catData_2010_2,YG4_catData_2010_3)


#2011------------------------------------------------------------

#If data is greater than or equal to 12 YOS then keep
YG4_catData_2011_1 = subset(data_2011_cat,CYOS_EFY >= 12,
select =c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,
DG,Spouse_Stat,
Race,Deployments:Prior_E))
#Subset the data with values greater than 19 YOS and retain values
#are 0 or 1
YG4_catData_2011_2 = subset(YG4_catData_2011_1,CYOS_EFY >= 19 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 18 YOS and retain = 0
YG4_catData_2011_3 = subset(YG4_catData_2011_1, CYOS_EFY <= 18 &
```

```
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_12to19_2011_catData= rbind(YG4_catData_2011_2,YG4_catData_2011_3)

#2012-----------------------------------------------------------------

#If data is greater than or equal to 12 YOS then keep
YG4_catData_2012_1 = subset(data_2012_cat,CYOS_EFY >= 12,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,
DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values greater than 19 YOS and retain
#values are 0 or 1
YG4_catData_2012_2 = subset(YG4_catData_2012_1,CYOS_EFY >= 19 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 18 YOS and retain = 0
YG4_catData_2012_3 = subset(YG4_catData_2012_1, CYOS_EFY <= 18 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_12to19_2012_catData= rbind(YG4_catData_2012_2,YG4_catData_2012_3)

#2013-----------------------------------------------------------------

#If data is greater than or equal to 12 YOS then keep
YG4_catData_2013_1 = subset(data_2013_cat,CYOS_EFY >= 12,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,
DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values greater than 19 YOS and retain values
#are 0 or 1
YG4_catData_2013_2 = subset(YG4_catData_2013_1,CYOS_EFY >= 19 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 18 YOS and retain = 0
YG4_catData_2013_3 = subset(YG4_catData_2013_1, CYOS_EFY <= 18 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_12to19_2013_catData= rbind(YG4_catData_2013_2,YG4_catData_2013_3)

#2014-----------------------------------------------------------------
```

```
#If data is greater than or equal to 12 YOS then keep
YG4_catData_2014_1 = subset(data_2014_cat,CYOS_EFY >= 12,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,
Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values greater than 19 YOS and retain
#values are 0 or 1
YG4_catData_2014_2 = subset(YG4_catData_2014_1,CYOS_EFY >= 19 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 18 YOS and retain = 0
YG4_catData_2014_3 = subset(YG4_catData_2014_1, CYOS_EFY <= 18 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_12to19_2014_catData= rbind(YG4_catData_2014_2,YG4_catData_2014_3)

#2015------------------------------------------------------------

#If data is greater than or equal to 12 YOS then keep
YG4_catData_2015_1 = subset(data_2015_cat,CYOS_EFY >= 12,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,
Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values greater than 19 YOS and retain
#values are 0 or 1
YG4_catData_2015_2 = subset(YG4_catData_2015_1,CYOS_EFY >= 19 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 18 YOS and retain = 0
YG4_catData_2015_3 = subset(YG4_catData_2015_1, CYOS_EFY <= 18 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_12to19_2015_catData= rbind(YG4_catData_2015_2,YG4_catData_2015_3)

#2016------------------------------------------------------------

#If data is greater than or equal to 12 YOS then keep
YG4_catData_2016_1 = subset(data_2016_cat,CYOS_EFY >= 12,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,
Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values greater than 19 YOS and retain
#values are 0 or 1
YG4_catData_2016_2 = subset(YG4_catData_2016_1,CYOS_EFY >= 19 &
```

```
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 18 YOS and retain = 0
YG4_catData_2016_3 = subset(YG4_catData_2016_1, CYOS_EFY <= 18 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_12to19_2016_catData= rbind(YG4_catData_2016_2,YG4_catData_2016_3)

#2017-----------------------------------------------------------

#If data is greater than or equal to 12 YOS then keep
YG4_catData_2017_1 = subset(data_2017_cat,CYOS_EFY >= 12,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,
DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values greater than 19 YOS and retain
#values are 0 or 1
YG4_catData_2017_2 = subset(YG4_catData_2017_1,CYOS_EFY >= 19 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 18 YOS and retain = 0
YG4_catData_2017_3 = subset(YG4_catData_2017_1, CYOS_EFY <= 18 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_12to19_2017_catData= rbind(YG4_catData_2017_2,YG4_catData_2017_3)

#2018-----------------------------------------------------------

#If data is greater than or equal to 12 YOS then keep
YG4_catData_2018_1 = subset(data_2018_cat,CYOS_EFY >= 12,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,
Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values greater than 19 YOS and retain
#values are 0 or 1
YG4_catData_2018_2 = subset(YG4_catData_2018_1,CYOS_EFY >= 19 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 18 YOS and retain = 0
YG4_catData_2018_3 = subset(YG4_catData_2018_1, CYOS_EFY <= 18 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
```

```
CYOS_12to19_2018_catData= rbind(YG4_catData_2018_2,YG4_catData_2018_3)

#2019------------------------------------------------------------

#If data is greater than or equal to 12 YOS then keep
YG4_catData_2019_1 = subset(data_2019_cat,CYOS_EFY >= 12,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,
DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values greater than 19 YOS and retain
#values are 0 or 1
YG4_catData_2019_2 = subset(YG4_catData_2019_1,CYOS_EFY >= 19 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 18 YOS and retain = 0
YG4_catData_2019_3 = subset(YG4_catData_2019_1, CYOS_EFY <= 18 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_12to19_2019_catData= rbind(YG4_catData_2019_2,YG4_catData_2019_3)

#Combine all data frames containing CYOS 12-19 now-------------------
CYOS_12to19_catCombined = rbind(CYOS_12to19_2009_catData,
CYOS_12to19_2010_catData,CYOS_12to19_2011_catData,
CYOS_12to19_2012_catData,CYOS_12to19_2013_catData,
CYOS_12to19_2014_catData,CYOS_12to19_2015_catData,
CYOS_12to19_2016_catData,CYOS_12to19_2017_catData,
CYOS_12to19_2018_catData,CYOS_12to19_2019_catData)

#Reverse order of data to ensure the most recent record is
#obtained when removing duplicates
revOrd_CYOS_12to19_catComb = CYOS_12to19_catCombined[order
(CYOS_12to19_catCombined$SSAN,CYOS_12to19_catCombined$CYOS_EFY,
decreasing=TRUE),]

#Now remove the duplicated information (records containing
#older information) by identified number:
CYOS_12to19_catNoDups =revOrd_CYOS_12to19_catComb
[!duplicated(revOrd_CYOS_12to19_catComb$SSAN),]

#Add a new column called 'Retained' with values = 0
CYOS_12to19_catNoDups$Retained = 0

#Change all retained values to 1 if CYOS >= 8. Some currently
```

```
#have 0 values.
CYOS_12to19_catFinal = CYOS_12to19_catNoDups %>%
mutate(Retained=replace(Retained,CYOS_EFY >= 19, 1))
%>% as.data.frame()

#Now remove the 'Retain' column since 'Retained' column now
#has the updated information..
CYOS_12to19_catFinal$Retain = NULL
transform(CYOS_12to19_catFinal, Retained = as.numeric(Retained))
#transform to numeric value if necessary

#subset the data so it does not contain SSAN:
CYOS_12to19_catLogRegData = CYOS_12to19_catFinal[c(4,5,7:10,12,13)]
write.csv(CYOS_12to19_catLogRegData,"C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/Developmental
Categories/CYOS_12to19_LogRegData_Categorical.csv",row.names = TRUE)


#Perform logistic regression on the Final dataframe containing all
fiscal years for CYOS 12 to 19:
catLogReg_CYOS_12to19 = glm(Retained ~ ., data =
CYOS_12to19_catLogRegData,family = "binomial")
summary(catLogReg_CYOS_12to19)


#Perform Logistic Regression on the Developmental
#Categories based on CYOS 12 to 19--------------

#Air Operations and Special Warfare
combAOSWData_CYOS_12to19 = subset(CYOS_12to19_catFinal,
Dev_Cat =="1", select = c(SSAN,DAFSCNR,MARITLST,DEPENT,DG,
Spouse_Stat,Race,Deployments,CYOS_EFY,Prior_E,Retained))
catAOSWData_CYOS_12to19 = subset(combAOSWData_CYOS_12to19,select =
c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,Prior_E,Retained))
catAOSWLR_CYOS_12to19 = glm(Retained ~ ., data =
catAOSWData_CYOS_12to19,family = "binomial")
summary(catAOSWLR_CYOS_12to19)
write.csv(catAOSWData_CYOS_12to19,"C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/Developmental
Categories/AOSWData_CYOS_12to19_Cat.csv",row.names = TRUE)
confint(catAOSWLR_CYOS_12to19)
```

```
#Info Warfare
combIWData_CYOS_12to19 = subset(CYOS_12to19_catFinal,Dev_Cat =="2",
select = c(SSAN,DAFSCNR,MARITLST,DEPENT,DG,Spouse_Stat,Race,
Deployments,CYOS_EFY,Prior_E,Retained))
catIWData_CYOS_12to19 = subset(combIWData_CYOS_12to19, select =
c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,Prior_E,Retained))
catIWLR_CYOS_12to19 = glm(Retained ~ ., data =
catIWData_CYOS_12to19,family = "binomial")
summary(catIWLR_CYOS_12to19)
write.csv(catIWData_CYOS_12to19,"C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/Developmental
Categories/IWData_CYOS_12to19_Cat.csv",row.names = TRUE)
confint(catIWLR_CYOS_12to19)

#Combat Support
combCSData_CYOS_12to19 = subset(CYOS_12to19_catFinal,
Dev_Cat =="3", select = c(SSAN,DAFSCNR,MARITLST,DEPENT,DG,
Spouse_Stat,Race,Deployments,CYOS_EFY,Prior_E,Retained))
catCSData_CYOS_12to19 = subset(combCSData_CYOS_12to19,
select = c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,
Prior_E,Retained))
catCSLR_CYOS_12to19 = glm(Retained ~ ., data = catCSData_CYOS_12to19,
family = "binomial")
summary(catCSLR_CYOS_12to19)
write.csv(catCSData_CYOS_12to19,"C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/Developmental
Categories/CSData_CYOS_12to19_Cat.csv",row.names = TRUE)
confint(catCSLR_CYOS_12to19)

#Force Modernization
combFMData_CYOS_12to19 = subset(CYOS_12to19_catFinal,Dev_Cat =="4",
select = c(SSAN,DAFSCNR,MARITLST,DEPENT,DG,Spouse_Stat,Race,
Deployments,CYOS_EFY,Prior_E,Retained))
catFMData_CYOS_12to19 = subset(combFMData_CYOS_12to19, select =
c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,Prior_E,Retained))
catFMLR_CYOS_12to19 = glm(Retained ~ ., data =
catFMData_CYOS_12to19,family = "binomial")
summary(catFMLR_CYOS_12to19)
write.csv(catFMData_CYOS_12to19,"C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
```

```
Fiscal Year/Final Thesis Data/Developmental
Categories/FMData_CYOS_12to19_Cat.csv",row.names = TRUE)
confint(catFMLR_CYOS_12to19)

#---------------------------------------------------------------------
#---------------------------------------------------------------------

# Years 20 to 22

#2009-------------------------------------------------------------

#If data is greater than or equal to 20 YOS then keep
YG5_catData_2009_1 = subset(data_2009_cat,CYOS_EFY >= 20, select
=c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,Spouse_Stat,
Race,Deployments:Prior_E))
#Subset the data with values greater than 22 YOS and retain
#values are 0 or 1
YG5_catData_2009_2 = subset(YG5_catData_2009_1,CYOS_EFY >= 22 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 21 YOS and retain = 0
YG5_catData_2009_3 = subset(YG5_catData_2009_1, CYOS_EFY <= 21 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_20to22_2009_catData= rbind(YG5_catData_2009_2,YG5_catData_2009_3)


#2010-------------------------------------------------------------

#If data is greater than or equal to 20 YOS then keep
YG5_catData_2010_1 = subset(data_2010_cat,CYOS_EFY >= 20,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,
Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values greater than 22 YOS and retain
#values are 0 or 1
YG5_catData_2010_2 = subset(YG5_catData_2010_1,CYOS_EFY >= 22 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 21 YOS and retain = 0
YG5_catData_2010_3 = subset(YG5_catData_2010_1, CYOS_EFY <= 21 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
```

```
#Combine both files
CYOS_20to22_2010_catData= rbind(YG5_catData_2010_2,YG5_catData_2010_3)


#2011-----------------------------------------------------------------

#If data is greater than or equal to 20 YOS then keep
YG5_catData_2011_1 = subset(data_2011_cat,CYOS_EFY >= 20,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,
DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values greater than 22 YOS and retain
#values are 0 or 1
YG5_catData_2011_2 = subset(YG5_catData_2011_1,CYOS_EFY >= 22 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 21 YOS and retain = 0
YG5_catData_2011_3 = subset(YG5_catData_2011_1, CYOS_EFY <= 21 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_20to22_2011_catData= rbind(YG5_catData_2011_2,YG5_catData_2011_3)


#2012-----------------------------------------------------------------

#If data is greater than or equal to 20 YOS then keep
YG5_catData_2012_1 = subset(data_2012_cat,CYOS_EFY >= 20,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,
DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values greater than 22 YOS and retain
#values are 0 or 1
YG5_catData_2012_2 = subset(YG5_catData_2012_1,CYOS_EFY >= 22 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 21 YOS and retain = 0
YG5_catData_2012_3 = subset(YG5_catData_2012_1, CYOS_EFY <= 21 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_20to22_2012_catData= rbind(YG5_catData_2012_2,YG5_catData_2012_3)

#2013-----------------------------------------------------------------
```

```
#If data is greater than or equal to 20 YOS then keep
YG5_catData_2013_1 = subset(data_2013_cat,CYOS_EFY >= 20,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,
DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values greater than 22 YOS and retain
#values are 0 or 1
YG5_catData_2013_2 = subset(YG5_catData_2013_1,CYOS_EFY >= 22 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 21 YOS and retain = 0
YG5_catData_2013_3 = subset(YG5_catData_2013_1, CYOS_EFY <= 21 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_20to22_2013_catData= rbind(YG5_catData_2013_2,YG5_catData_2013_3)


#2014-----------------------------------------------------------------

#If data is greater than or equal to 20 YOS then keep
YG5_catData_2014_1 = subset(data_2014_cat,CYOS_EFY >= 20,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,
Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values greater than 22 YOS and retain
#values are 0 or 1
YG5_catData_2014_2 = subset(YG5_catData_2014_1,CYOS_EFY >= 22 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 21 YOS and retain = 0
YG5_catData_2014_3 = subset(YG5_catData_2014_1, CYOS_EFY <= 21 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_20to22_2014_catData= rbind(YG5_catData_2014_2,YG5_catData_2014_3)

#2015-----------------------------------------------------------------

#If data is greater than or equal to 20 YOS then keep
YG5_catData_2015_1 = subset(data_2015_cat,CYOS_EFY >= 20,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG
,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the d5ata with values greater than 22 YOS and retain
#values are 0 or 1
```

```
YG5_catData_2015_2 = subset(YG5_catData_2015_1,CYOS_EFY >= 22 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 21 YOS and retain = 0
YG5_catData_2015_3 = subset(YG5_catData_2015_1, CYOS_EFY <= 21 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_20to22_2015_catData= rbind(YG5_catData_2015_2,YG5_catData_2015_3)

#2016-------------------------------------------------------------

#If data is greater than or equal to 20 YOS then keep
YG5_catData_2016_1 = subset(data_2016_cat,CYOS_EFY >= 20,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,
Spouse_Stat,Race,Deployments:Prior_E))
#Subset the d5ata with values greater than 22 YOS and retain
#values are 0 or 1
YG5_catData_2016_2 = subset(YG5_catData_2016_1,CYOS_EFY >= 22 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 21 YOS and retain = 0
YG5_catData_2016_3 = subset(YG5_catData_2016_1, CYOS_EFY <= 21 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_20to22_2016_catData= rbind(YG5_catData_2016_2,YG5_catData_2016_3)

#2017-------------------------------------------------------------

#If data is greater than or equal to 20 YOS then keep
YG5_catData_2017_1 = subset(data_2017_cat,CYOS_EFY >= 20,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,
DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the d5ata with values greater than 22 YOS and retain
#values are 0 or 1
YG5_catData_2017_2 = subset(YG5_catData_2017_1,CYOS_EFY >= 22 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 21 YOS and retain = 0
YG5_catData_2017_3 = subset(YG5_catData_2017_1, CYOS_EFY <= 21 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT
,Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
```

```
#Combine both files
CYOS_20to22_2017_catData= rbind(YG5_catData_2017_2,YG5_catData_2017_3)

#2018-------------------------------------------------------------

#If data is greater than or equal to 20 YOS then keep
YG5_catData_2018_1 = subset(data_2018_cat,CYOS_EFY >= 20,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,
Spouse_Stat,Race,Deployments:Prior_E))
#Subset the d5ata with values greater than 22 YOS and retain
#values are 0 or 1
YG5_catData_2018_2 = subset(YG5_catData_2018_1,CYOS_EFY >= 22 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 21 YOS and retain = 0
YG5_catData_2018_3 = subset(YG5_catData_2018_1, CYOS_EFY <= 21 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_20to22_2018_catData= rbind(YG5_catData_2018_2,YG5_catData_2018_3)

#2019-------------------------------------------------------------

#If data is greater than or equal to 20 YOS then keep
YG5_catData_2019_1 = subset(data_2019_cat,CYOS_EFY >= 20,
select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,Comm_Source,DG,
Spouse_Stat,Race,Deployments:Prior_E))
#Subset the d5ata with values greater than 22 YOS and retain
#values are 0 or 1
YG5_catData_2019_2 = subset(YG5_catData_2019_1,CYOS_EFY >= 22 &
Retain >= 0, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Subset the data with values less than 21 YOS and retain = 0
YG5_catData_2019_3 = subset(YG5_catData_2019_1, CYOS_EFY <= 21 &
Retain < 1, select = c(SSAN,DAFSCNR,Dev_Cat,MARITLST,DEPENT,
Comm_Source,DG,Spouse_Stat,Race,Deployments:Prior_E))
#Combine both files
CYOS_20to22_2019_catData= rbind(YG5_catData_2019_2,YG5_catData_2019_3)


#Combine all data frames containing CYOS 20-22 now-------------------
CYOS_20to22_catCombined = rbind(CYOS_20to22_2009_catData,
CYOS_20to22_2010_catData,CYOS_20to22_2011_catData,
```

```
CYOS_20to22_2012_catData,CYOS_20to22_2013_catData,
CYOS_20to22_2014_catData,CYOS_20to22_2015_catData,
CYOS_20to22_2016_catData,CYOS_20to22_2017_catData,
CYOS_20to22_2018_catData,CYOS_20to22_2019_catData)

#Reverse order of data to ensure the most recent record is
#obtained when removing duplicates
revOrd_CYOS_20to22_catComb = CYOS_20to22_catCombined[
order(CYOS_20to22_catCombined$SSAN,CYOS_20to22_catCombined$
CYOS_EFY,decreasing=TRUE),]

#Now remove the duplicated information (records containing older
#information) by identified number:
CYOS_20to22_catNoDups =
revOrd_CYOS_20to22_catComb
[!duplicated(revOrd_CYOS_20to22_catComb$SSAN),]

#Add a new column called 'Retained' with values = 0
CYOS_20to22_catNoDups$Retained = 0

#Change all retained values to 1 if CYOS >= 8. Some
#currently have 0 values.
CYOS_20to22_catFinal = CYOS_20to22_catNoDups %>%
mutate(Retained=replace(Retained,CYOS_EFY >= 22, 1))
%>% as.data.frame()

#Now remove the 'Retain' column since 'Retained' column now has
#the updated information..
CYOS_20to22_catFinal$Retain = NULL
transform(CYOS_20to22_catFinal, Retained = as.numeric(Retained))
#transform to numeric value if necessary

#subset the data so it does not contain SSAN:
CYOS_20to22_catLogRegData = CYOS_20to22_catFinal[c(4,5,7:10,12,13)]
write.csv(CYOS_20to22_catLogRegData,"C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/Developmental
Categories/CYOS_20to22_LogRegData_Categorical.csv",row.names = TRUE)

#Perform logistic regression on the Final dataframe containing all
#fiscal years for CYOS 12 to 19:
catLogReg_CYOS_20to22 = glm(Retained ~ ., data =
CYOS_20to22_catLogRegData,family = "binomial")
```

```
summary(catLogReg_CYOS_20to22)

#Perform Logistic Regression on the Developmental Categories
based on CYOS 20 to 22--------------

#Air Operations and Special Warfare
combAOSWData_CYOS_20to22 = subset(CYOS_20to22_catFinal,Dev_Cat =="1",
select = c(SSAN,DAFSCNR,MARITLST,DEPENT,DG,Spouse_Stat,Race,
Deployments,CYOS_EFY,Prior_E,Retained))
catAOSWData_CYOS_20to22 = subset(combAOSWData_CYOS_20to22,
select = c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,
Prior_E,Retained))
catAOSWLR_CYOS_20to22 = glm(Retained ~ ., data =
catAOSWData_CYOS_20to22,family = "binomial")
summary(catAOSWLR_CYOS_20to22)
write.csv(catAOSWData_CYOS_20to22,"C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/Developmental
Categories/AOSWData_CYOS_20to22_Cat.csv",row.names = TRUE)
confint(catAOSWLR_CYOS_20to22)

#Info Warfare
combIWData_CYOS_20to22 = subset(CYOS_20to22_catFinal,Dev_Cat =="2",
select = c(SSAN,DAFSCNR,MARITLST,DEPENT,DG,Spouse_Stat,Race,
Deployments,CYOS_EFY,Prior_E,Retained))
catIWData_CYOS_20to22 = subset(combIWData_CYOS_20to22, select =
c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,Prior_E,Retained))
catIWLR_CYOS_20to22 = glm(Retained ~ ., data = catIWData_CYOS_20to22,
family = "binomial")
summary(catIWLR_CYOS_20to22)
write.csv(catIWData_CYOS_20to22,"C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/Developmental
Categories/IWData_CYOS_20to22_Cat.csv",row.names = TRUE)
confint(catIWLR_CYOS_20to22)

#Combat Support
combCSData_CYOS_20to22 = subset(CYOS_20to22_catFinal,Dev_Cat =="3",
select = c(SSAN,DAFSCNR,MARITLST,DEPENT,DG,Spouse_Stat,Race,
Deployments,CYOS_EFY,Prior_E,Retained))
catCSData_CYOS_20to22 = subset(combCSData_CYOS_20to22, select =
c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,Prior_E,Retained))
catCSLR_CYOS_20to22 = glm(Retained ~ ., data = catCSData_CYOS_20to22,
```

```
family = "binomial")
summary(catCSLR_CYOS_20to22)
write.csv(catCSData_CYOS_20to22,"C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/Developmental
Categories/CSData_CYOS_20to22_Cat.csv",row.names = TRUE)
confint(catCSLR_CYOS_20to22)

#Force Modernization
combFMData_CYOS_20to22 = subset(CYOS_20to22_catFinal,Dev_Cat =="4",
select = c(SSAN,DAFSCNR,MARITLST,DEPENT,DG,Spouse_Stat,Race,
Deployments,CYOS_EFY,Prior_E,Retained))
catFMData_CYOS_20to22 = subset(combFMData_CYOS_20to22, select =
c(MARITLST,DEPENT,DG,Spouse_Stat,Race,Deployments,Prior_E,Retained))
catFMLR_CYOS_20to22 = glm(Retained ~ ., data = catFMData_CYOS_20to22,
family = "binomial")
summary(catFMLR_CYOS_20to22)
write.csv(catFMData_CYOS_20to22,"C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/Female Only/
Fiscal Year/Final Thesis Data/Developmental
Categories/FMData_CYOS_20to22_Cat.csv",row.names = TRUE)
confint(catFMLR_CYOS_20to22)


#This section contains computational code for odds ratios
#on specific significant characteristics gathered from the
#Logisitic Regression Models for the Developmental Career
#interation. All values are categorical, with Retain = 1 as
#the positive value.

#Air Operations and SpecialWarfare--------------------------------
#CYOS 0 to 6------------------------------------------------------

#Dependent status----------------------------------------------
#Female officer without dependents is the baseline case

#odds ratio for w/ dependents
AOSW_CYOS_0to6_depend_odds =
exp(catAOSWLR_CYOS_0to6$coefficients[4])
AOSW_CYOS_0to6_depend_odds

#Deployments---------------------------------------------------
#No deployments is the baseline case
```

```
#odds ratio for 1 deployment
AOSW_CYOS_0to6_1deploy_odds =
exp(catAOSWLR_CYOS_0to6$coefficients[15])
AOSW_CYOS_0to6_1deploy_odds

#odds ratio for 2+ deployments
AOSW_CYOS_0to6_2deploys_odds =
exp(catAOSWLR_CYOS_0to6$coefficients[16])
AOSW_CYOS_0to6_2deploys_odds


#CYOS 4 to 8-------------------------------------------------------
#Dependent status--------------------------------------------
#Female officer without dependents is the baseline case

#odds ratio for w/ dependents
AOSW_CYOS_4to8_depend_odds =
exp(catAOSWLR_CYOS_4to8$coefficients[4])
AOSW_CYOS_4to8_depend_odds

#Deployments-------------------------------------------------
#No deployments is the baseline case

#odds ratio for 1 deployment
AOSW_CYOS_4to8_1deploy_odds =
exp(catAOSWLR_CYOS_4to8$coefficients[15])
AOSW_CYOS_4to8_1deploy_odds

#odds ratio for 2+ deployments
AOSW_CYOS_4to8_2deploys_odds =
exp(catAOSWLR_CYOS_4to8$coefficients[16])
AOSW_CYOS_4to8_2deploys_odds

#CYOS 8 to 14-----------------------------------------------------
#Dependent status--------------------------------------------
#Female officer without dependents is the baseline case

#odds ratio for w/ dependents
AOSW_CYOS_8to14_depend_odds =
exp(catAOSWLR_CYOS_8to14$coefficients[4])
AOSW_CYOS_8to14_depend_odds
```

```
#Deployments-------------------------------------------------
#No deployments is the baseline case

#odds ratio for 1 deployment
AOSW_CYOS_8to14_1deploy_odds =
exp(catAOSWLR_CYOS_8to14$coefficients[15])
AOSW_CYOS_8to14_1deploy_odds

#odds ratio for 2+ deployments
AOSW_CYOS_8to14_2deploys_odds =
exp(catAOSWLR_CYOS_8to14$coefficients[16])
AOSW_CYOS_8to14_2deploys_odds

#CYOS 12 to 19-----------------------------------------------
#Dependent status--------------------------------------------
#Female officer without dependents is the baseline case

#odds ratio for w/ dependents
AOSW_CYOS_12to19_depend_odds =
exp(catAOSWLR_CYOS_12to19$coefficients[4])
AOSW_CYOS_12to19_depend_odds

#Deployments-------------------------------------------------
#No deployments is the baseline case

#odds ratio for 1 deployment
AOSW_CYOS_12to19_1deploy_odds =
exp(catAOSWLR_CYOS_12to19$coefficients[14])
AOSW_CYOS_12to19_1deploy_odds

#odds ratio for 2+ deployments
AOSW_CYOS_12to19_2deploys_odds =
exp(catAOSWLR_CYOS_12to19$coefficients[15])
AOSW_CYOS_12to19_2deploys_odds


#CYOS 20 to 22-----------------------------------------------

##Dependent status-------------------------------------------
#Female officer without dependents is the baseline case

#odds ratio for w/ dependents
AOSW_CYOS_20to22_depend_odds =
```

```
exp(catAOSWLR_CYOS_20to22$coefficients[4])
AOSW_CYOS_20to22_depend_odds

#Deployments-------------------------------------------------
#No deployments is the baseline case

#odds ratio for 1 deployment
AOSW_CYOS_20to22_1deploy_odds =
exp(catAOSWLR_CYOS_20to22$coefficients[13])
AOSW_CYOS_20to22_1deploy_odds

#odds ratio for 2+ deployments
AOSW_CYOS_20to22_2deploys_odds =
exp(catAOSWLR_CYOS_20to22$coefficients[14])
AOSW_CYOS_20to22_2deploys_odds

#Information Warfare-----------------------------------------
#Spouse Career-----------------------------------------------
#Spouse status = active duty is baseline case

#CYOS 0 to 6---------------------------------------------------
#odds ratio for Reserve/NG
IW_CYOS_0to6_reserve_odds =
exp(catIWLR_CYOS_0to6$coefficients[6])
IW_CYOS_0to6_reserve_odds

#odds ratio for Other
IW_CYOS_0to6_other_odds =
exp(catIWLR_CYOS_0to6$coefficients[7])
IW_CYOS_0to6_other_odds


#CYOS 4 to 8----------------------------------------------------
#odds ratio for Reserve/NG
IW_CYOS_4to8_reserve_odds =
exp(catIWLR_CYOS_4to8$coefficients[6])
IW_CYOS_4to8_reserve_odds

#odds ratio for Other
IW_CYOS_4to8_other_odds =
exp(catIWLR_CYOS_4to8$coefficients[7])
IW_CYOS_4to8_other_odds
```

```
#CYOS 8 to 14-----------------------------------------------------
#odds ratio for Reserve/NG
IW_CYOS_8to14_reserve_odds =
exp(catIWLR_CYOS_8to14$coefficients[6])
IW_CYOS_8to14_reserve_odds

#odds ratio for Other
IW_CYOS_8to14_other_odds =
exp(catIWLR_CYOS_8to14$coefficients[7])
IW_CYOS_8to14_other_odds

#CYOS 12 to 19-----------------------------------------------------
#odds ratio for Reserve/NG
IW_CYOS_12to19_reserve_odds =
exp(catIWLR_CYOS_12to19$coefficients[6])
IW_CYOS_12to19_reserve_odds

#odds ratio for Other
IW_CYOS_12to19_other_odds =
exp(catIWLR_CYOS_12to19$coefficients[7])
IW_CYOS_12to19_other_odds


#CYOS 20 to 22-------------------------------------------------
#odds ratio for Reserve/NG
IW_CYOS_20to22_reserve_odds =
exp(catIWLR_CYOS_20to22$coefficients[6])
IW_CYOS_20to22_reserve_odds

#odds ratio for Other
IW_CYOS_20to22_other_odds =
exp(catIWLR_CYOS_20to22$coefficients[7])
IW_CYOS_20to22_other_odds

#Combat Support-------------------------------------------------

#CYOS 0 to 6----------------------------------------------------
#Dependent status-----------------------------------------------
#Female officer without dependents is the baseline case

#odds ratio for w/ dependents
CS_CYOS_0to6_depend_odds =
```

```
exp(catCSLR_CYOS_0to6$coefficients[4])
CS_CYOS_0to6_depend_odds

#Deployments-----------------------------------------------
#No deployments is the baseline case

#odds ratio for 1 deployment
CS_CYOS_0to6_1deploy_odds =
exp(catCSLR_CYOS_0to6$coefficients[15])
CS_CYOS_0to6_1deploy_odds

#odds ratio for 2+ deployments
CS_CYOS_0to6_2deploys_odds =
exp(catCSLR_CYOS_0to6$coefficients[16])
CS_CYOS_0to6_2deploys_odds

#Marital status--------------------------------------------
#Single females is the baseline case

#odds ratio for married
CS_CYOS_0to6_married_odds =
exp(catCSLR_CYOS_0to6$coefficients[2])
CS_CYOS_0to6_married_odds

#odds ratio for divorced/annulled/etc.
CS_CYOS_0to6_divorce_odds =
exp(catCSLR_CYOS_0to6$coefficients[3])
CS_CYOS_0to6_divorce_odds

#Prior Service---------------------------------------------
#Female officers with no prior service is the baseline case

#odds ratio for Prior service
CS_CYOS_0to6_prior_odds =
exp(catCSLR_CYOS_0to6$coefficients[17])
CS_CYOS_0to6_prior_odds

#CYOS 4 to 8-----------------------------------------------
#Dependent status------------------------------------------
#Female officer without dependents is the baseline case

#odds ratio for w/ dependents
CS_CYOS_4to8_depend_odds =
```

```
exp(catCSLR_CYOS_4to8$coefficients[4])
CS_CYOS_4to8_depend_odds

#Deployments-----------------------------------------------
#No deployments is the baseline case

#odds ratio for 1 deployment
CS_CYOS_4to8_1deploy_odds =
exp(catCSLR_CYOS_4to8$coefficients[15])
CS_CYOS_4to8_1deploy_odds

#odds ratio for 2+ deployments
CS_CYOS_4to8_2deploys_odds =
exp(catCSLR_CYOS_4to8$coefficients[16])
CS_CYOS_4to8_2deploys_odds

#Marital status--------------------------------------------
#Single females is the baseline case

#odds ratio for married
CS_CYOS_4to8_married_odds =
exp(catCSLR_CYOS_4to8$coefficients[2])
CS_CYOS_4to8_married_odds

#odds ratio for divorced/annulled/etc.
CS_CYOS_4to8_divorce_odds =
exp(catCSLR_CYOS_4to8$coefficients[3])
CS_CYOS_4to8_divorce_odds

#Prior Service---------------------------------------------
#Female officers with no prior service is the baseline case

#odds ratio for Prior service
CS_CYOS_4to8_prior_odds =
exp(catCSLR_CYOS_4to8$coefficients[17])
CS_CYOS_4to8_prior_odds

#CYOS 8 to 14----------------------------------------------
#Dependent status------------------------------------------
#Female officer without dependents is the baseline case

#odds ratio for w/ dependents
CS_CYOS_8to14_depend_odds =
```

```
exp(catCSLR_CYOS_8to14$coefficients[4])
CS_CYOS_8to14_depend_odds

#Deployments------------------------------------------------
#No deployments is the baseline case

#odds ratio for 1 deployment
CS_CYOS_8to14_1deploy_odds =
exp(catCSLR_CYOS_8to14$coefficients[15])
CS_CYOS_8to14_1deploy_odds

#odds ratio for 2+ deployments
CS_CYOS_8to14_2deploys_odds =
exp(catCSLR_CYOS_8to14$coefficients[16])
CS_CYOS_8to14_2deploys_odds

#Marital status---------------------------------------------
#Single females is the baseline case

#odds ratio for married
CS_CYOS_8to14_married_odds =
exp(catCSLR_CYOS_8to14$coefficients[2])
CS_CYOS_8to14_married_odds

#odds ratio for divorced/annulled/etc.
CS_CYOS_8to14_divorce_odds =
exp(catCSLR_CYOS_8to14$coefficients[3])
CS_CYOS_8to14_divorce_odds

#Prior Service----------------------------------------------
#Female officers with no prior service is the baseline case

#odds ratio for Prior service
CS_CYOS_8to14_prior_odds =
exp(catCSLR_CYOS_8to14$coefficients[17])
CS_CYOS_8to14_prior_odds

#CYOS 12 to 19---------------------------------------------
#Dependent status------------------------------------------
#Female officer without dependents is the baseline case

#odds ratio for w/ dependents
CS_CYOS_12to19_depend_odds =
```

```
exp(catCSLR_CYOS_12to19$coefficients[4])
CS_CYOS_12to19_depend_odds

#Deployments------------------------------------------------
#No deployments is the baseline case

#odds ratio for 1 deployment
CS_CYOS_12to19_1deploy_odds =
exp(catCSLR_CYOS_12to19$coefficients[15])
CS_CYOS_12to19_1deploy_odds

#odds ratio for 2+ deployments
CS_CYOS_12to19_2deploys_odds =
exp(catCSLR_CYOS_12to19$coefficients[16])
CS_CYOS_12to19_2deploys_odds

#Marital status---------------------------------------------
#Single females is the baseline case

#odds ratio for married
CS_CYOS_12to19_married_odds =
exp(catCSLR_CYOS_12to19$coefficients[2])
CS_CYOS_12to19_married_odds

#odds ratio for divorced/annulled/etc.
CS_CYOS_12to19_divorce_odds =
exp(catCSLR_CYOS_12to19$coefficients[3])
CS_CYOS_12to19_divorce_odds

#Prior Service----------------------------------------------
#Female officers with no prior service is the baseline case

#odds ratio for Prior service
CS_CYOS_12to19_prior_odds =
exp(catCSLR_CYOS_12to19$coefficients[17])
CS_CYOS_12to19_prior_odds


#CYOS 20 to 22----------------------------------------------
#Dependent status-------------------------------------------
#Female officer without dependents is the baseline case

#odds ratio for w/ dependents
```

```
CS_CYOS_20to22_depend_odds =
exp(catCSLR_CYOS_20to22$coefficients[4])
CS_CYOS_20to22_depend_odds

#Deployments-----------------------------------------------
#No deployments is the baseline case

#odds ratio for 1 deployment
CS_CYOS_20to22_1deploy_odds =
exp(catCSLR_CYOS_20to22$coefficients[15])
CS_CYOS_20to22_1deploy_odds

#odds ratio for 2+ deployments
CS_CYOS_20to22_2deploys_odds =
exp(catCSLR_CYOS_20to22$coefficients[16])
CS_CYOS_20to22_2deploys_odds

#Marital status---------------------------------------------
#Single females is the baseline case

#odds ratio for married
CS_CYOS_20to22_married_odds =
exp(catCSLR_CYOS_20to22$coefficients[2])
CS_CYOS_20to22_married_odds

#odds ratio for divorced/annulled/etc.
CS_CYOS_20to22_divorce_odds =
exp(catCSLR_CYOS_20to22$coefficients[3])
CS_CYOS_20to22_divorce_odds

#Prior Service----------------------------------------------
#Female officers with no prior service is the baseline case

#odds ratio for Prior service
CS_CYOS_20to22_prior_odds =
exp(catCSLR_CYOS_20to22$coefficients[17])
CS_CYOS_20to22_prior_odds

#Force Modernization----------------------------------------

#CYOS 0 to 6------------------------------------------------
#Marital status---------------------------------------------
#Single females is the baseline case
```

```
#odds ratio for married
FM_CYOS_0to6_married_odds =
exp(catFMLR_CYOS_0to6$coefficients[2])
FM_CYOS_0to6_married_odds

#odds ratio for divorced/annulled/etc.
FM_CYOS_0to6_divorce_odds =
exp(catFMLR_CYOS_0to6$coefficients[3])
FM_CYOS_0to6_divorce_odds

#CYOS 4 to 8--------------------------------------------------
#Marital status----------------------------------------------
#Single females is the baseline case

#odds ratio for married
FM_CYOS_4to8_married_odds =
exp(catFMLR_CYOS_4to8$coefficients[2])
FM_CYOS_4to8_married_odds

#odds ratio for divorced/annulled/etc.
FM_CYOS_4to8_divorce_odds =
exp(catFMLR_CYOS_4to8$coefficients[3])
FM_CYOS_4to8_divorce_odds

#CYOS 8 to 14------------------------------------------------
#Marital status----------------------------------------------
#Single females is the baseline case

#odds ratio for married
FM_CYOS_8to14_married_odds =
exp(catFMLR_CYOS_8to14$coefficients[2])
FM_CYOS_8to14_married_odds

#odds ratio for divorced/annulled/etc.
FM_CYOS_8to14_divorce_odds =
exp(catFMLR_CYOS_8to14$coefficients[3])
FM_CYOS_8to14_divorce_odds

#CYOS 12 to 19-----------------------------------------------
#Marital status----------------------------------------------
#Single females is the baseline case
```

```
#odds ratio for married
FM_CYOS_12to19_married_odds =
exp(catFMLR_CYOS_12to19$coefficients[2])
FM_CYOS_12to19_married_odds

#odds ratio for divorced/annulled/etc.
FM_CYOS_12to19_divorce_odds =
exp(catFMLR_CYOS_12to19$coefficients[3])
FM_CYOS_12to19_divorce_odds

#CYOS 20 to 22-----------------------------------------------
#Marital status----------------------------------------------
#Single females is the baseline case

#odds ratio for married
FM_CYOS_20to22_married_odds =
exp(catFMLR_CYOS_20to22$coefficients[2])
FM_CYOS_20to22_married_odds

#odds ratio for divorced/annulled/etc.
FM_CYOS_20to22_divorce_odds =
exp(catFMLR_CYOS_20to22$coefficients[3])
FM_CYOS_20to22_divorce_odds
```

Appendix B: R Code for Survival Analysis

```
#This section contains Survival Analysis code for
#covariates considered significant in CYOS logistic regression
#model of the study. Dataframes created call from logistic
#regression code for ease of use

#Call necessary packages for analysis:
library(ggplot2)
library(survival)
library(survminer)
library(dplyr)
library(ggfortify)
library(ggsci)

#Perform Proportional Hazard

#Create the dataframe to look at survival analysis for the
entire group (COHORT LEVEL) of female officers:
survivalCatDat = rbind(CYOS_0to6_catFinal,CYOS_4to8_catFinal
CYOS_8to14_catFinal,CYOS_12to19_catFinal,CYOS_20to22_catFinal)
#survival_Dat_NoDups = survivalDat
[!duplicated(survivalDat$SSAN),]

#Reverse order of data to reflect most recent record for
#each female officer:
revOrd_catSurvivalDat = survivalCatDat
[order(survivalCatDat$SSAN,
survivalCatDat$CYOS_EFY,decreasing=TRUE),]
#Now remove the duplicated information (records containing
#older information) by identified number:
survival_catDat_NoDups2 = revOrd_catSurvivalDat
[!duplicated(revOrd_catSurvivalDat$SSAN),]

cohortSurvVariables = subset(survival_catDat_NoDups2,
select =c(DG,DEPENT,Deployments,MARITLST,Prior_E,
Race,Spouse_Stat,
CYOS_EFY,Retained))

#Fit survival data using Kaplan-Meier method
km_Cohort = with(cohortSurvVariables,Surv(CYOS_EFY,Retained))
head(km_Cohort,100) #check to see if censored data is included
```

```
#This section creates the model that shows the survival rate of
female officers over the course of 30 yrs
#use the formula surv(time = x, event = y) and
#survfit function to produce the Kaplan-Meier estimates
km_Cohort_fit = survfit(Surv(CYOS_EFY,Retained)~1,data =
cohortSurvVariables)
#now show the summary of the km estimates (times parameter
#of summary() gives control over which times to print)
summary(km_Cohort_fit,times=c(10,15,20,25*(1:5)))
#summary(km_Cohort_fit,times=c(6,7,8,9,10,15,20,25,30*(1:5)))

#Now display the 'Kaplan-Meir' plot
cohortGGSurvPlot = ggsurvplot(km_Cohort_fit, palette = "blue",
xlab = "CYOS", ylab = "Probability of Retention",
legend.title = "", legend.labs = "Cohort Level",
conf.int = TRUE, conf.int.style = "step",
conf.int.alpha = 0.5, risk.table = TRUE,
risk.table.y.text = FALSE)
print(cohortGGSurvPlot)

#Now perform survival analysis on the developmental
#categories together
km_devcat_fit = survfit(Surv(CYOS_EFY,Retained)~ Dev_Cat,
data = cohortSurvVariables)
summary(km_devcat_fit,times=c(10,15,20,25*(1:5)))
cohortDevCatSurvPlot = ggsurvplot(km_devcat_fit,
palette = c("blue","dark green","dark orange","red"),
xlab = "CYOS", ylab = "Survival Probability",
legend.title = "", legend.labs = c("AOSW","IW","CS","FM"),
conf.int = FALSE, conf.int.style = "ribbon",
conf.int.alpha = 0.5, risk.table = TRUE,
risk.table.y.text = TRUE,
tables.theme = theme_cleantable(base_size = 10),
submain = "Cohort Level Survival Estimates
for Developmental Categories")
print(cohortDevCatSurvPlot)


#Now we look at survival curves by significant
#factors---------------------------------------------

#Distinguished graduate:
```

```
km_DG_fit = survfit(Surv(CYOS_EFY,Retained) ~ DG,
data= cohortSurvVariables)
dgSurvPlot = ggsurvplot(km_DG_fit,
palette = c("blue","dark green"),
xlab = "CYOS", ylab = "Survival Probability",
legend.title = "",
legend.labs = c("Regular ", "Distinguished"),
conf.int = TRUE, conf.int.style = "ribbon",
conf.int.alpha = 0.5, risk.table = TRUE,
risk.table.y.text = TRUE,
tables.theme = theme_cleantable(),
submain = "Cohort Level Survival Estimates
for Distinguished Graduate")
print(dgSurvPlot)

#Number of dependents:
km_DEPENT_fit = survfit(Surv(CYOS_EFY,Retained) ~
DEPENT, data = cohortSurvVariables)
dependSurvPlot = ggsurvplot(km_DEPENT_fit,
palette = c("blue","dark green"), xlab = "CYOS",
ylab = "Survival Probability", legend.title = "",
legend.labs = c("No Dependents ", "Dependents"),
conf.int = TRUE, conf.int.style = "ribbon",
conf.int.alpha = 0.5, risk.table = TRUE,
risk.table.y.text = TRUE,
tables.theme = theme_cleantable(),
submain = "Cohort Level Survival
Estimates for Dependents")
print(dependSurvPlot)

#Marital Status:
km_MaritStat_fit = survfit(Surv(CYOS_EFY,Retained) ~
MARITLST, data = cohortSurvVariables)
maritStatSurvPlot = ggsurvplot(km_MaritStat_fit,
palette = c("blue","dark green", "dark orange"),
xlab = "CYOS", ylab = "Survival Probability",
legend.title = "",
legend.labs = c("Single", "Married", "Divorced/etc."),
conf.int = TRUE,  conf.int.alpha = 0.5,
risk.table = TRUE, risk.table.y.text = TRUE,
tables.theme = theme_cleantable(),
submain = "Cohort Level Survival Estimates
for Marital Status")
```

```
print(maritStatSurvPlot)

#Number of Deployments:
km_DeployStat_fit = survfit(Surv(CYOS_EFY,Retained) ~
Deployments, data = cohortSurvVariables)
deploySurvPlot = ggsurvplot(km_DeployStat_fit,
palette = c("blue","dark green", "dark orange"),
xlab = "CYOS", ylab = "Survival Probability",
legend.title = "", legend.labs = c("0", "1","2+"),
conf.int = TRUE, conf.int.alpha = 0.5,
risk.table = TRUE, risk.table.y.text = TRUE,
tables.theme =  theme_cleantable(),
submain = "Cohort Level Survival Estimates
for Deployments")
print(deploySurvPlot)

#Spouse Career:
km_SpouseStat_fit = survfit(Surv(CYOS_EFY,Retained) ~
Spouse_Stat, data = cohortSurvVariables)
spouseSurvPlot = ggsurvplot(km_SpouseStat_fit,
palette = c("blue","dark green","dark orange", "red"),
xlab = "CYOS", ylab = "Survival Probability",
legend.title = "",
legend.labs = c("Active","RANG","Other","No Spouse"),
conf.int = TRUE,  conf.int.alpha = 0.5,
risk.table = TRUE, risk.table.y.text = TRUE,
tables.theme =  theme_cleantable(),
submain = "Cohort Level Survival Estimates for
Spouse's Career Stataus")
print(spouseSurvPlot)

#Prior Service:
km_PriorServ_fit = survfit(Surv(CYOS_EFY,Retained) ~
Prior_E, data = cohortSurvVariables)
priorSurvPlot = ggsurvplot(km_PriorServ_fit,
palette = c("blue","dark green"),
xlab = "CYOS", ylab = "Survival Probability",
legend.title = "", legend.labs = c("Not Prior", "Prior"),
conf.int = TRUE, conf.int.alpha = 0.5,
risk.table = TRUE, risk.table.y.text = TRUE,
tables.theme =  theme_cleantable(),
submain = "Cohort Level Survival Estimates
for Prior Service")
```

```
print(priorSurvPlot)

#Race:
km_Race_fit = survfit(Surv(CYOS_EFY,Retained) ~
Race, data = cohortSurvVariables)
raceSurvPlot = ggsurvplot(km_Race_fit,
palette = c("blue","dark green","dark orange",
"red", "maroon","purple", "dark grey"),
xlab = "CYOS", ylab = "Survival Probability",
legend.title = "", legend.labs = c("Am Ind/AK Nat",
"Asian", "Black", "Declined", "Multiple",
"Nat Haw/Oth Pac","White"),
conf.int = FALSE, risk.table = TRUE,
risk.table.y.text = TRUE, tables.
theme = theme_cleantable(),
submain = "Cohort Level Survival Estimates for Race")
print(raceSurvPlot)

#------------------------------------------------------------
#------------------------------------------------------------
#------------------------------------------------------------

#This section contains the Survival Analysis code for
#covariates considered significant in each developmental
#category

##-----------------------------------------------------------
#------------------------------------------------------------
#Air Operations and Special Warfare-------------------------

#Create the dataframe to look at survival analysis for
#Air Ops & Special Warfare female officers:
survAirOpsDat = rbind(combAOSWData_CYOS_0to6,
combAOSWData_CYOS_4to8,combAOSWData_CYOS_8to14,
combAOSWData_CYOS_12to19,combAOSWData_CYOS_20to22)

#Reverse order of data to reflect most recent record
#for each female officer:
revOrd_survAirOps = survAirOpsDat[order(survAirOpsDat$SSAN,
survAirOpsDat$CYOS_EFY,decreasing=TRUE),]
#Now remove the duplicated information (records
#containing older information) by identified number:
survAirOps_NoDups = revOrd_survAirOps
```

```
[!duplicated(revOrd_survAirOps$SSAN),]

#Subset the data to the necessary variables
AirOps_SurvDat = subset(survAirOps_NoDups,
select = c(DG,DEPENT,Deployments,MARITLST,
Prior_E,Race,Spouse_Stat,CYOS_EFY,Retained))

#Fit survival data using Kaplan-Meier method
km_AirOps = with(AirOps_SurvDat2,Surv(CYOS_EFY,
Retained))
head(km_AirOps,100)
#check for censored variables (will have + at end of CYOS)
tail(km_AirOps,100)
#check for censored variables (will have + at end of CYOS)

#This section creates the model that shows the
#survival rate of female officers over the course of 30 yrs
#use the formula surv(time = x, event = y) and survfit
#function to produce the Kaplan-Meier estimates
km_AirOps_fit = survfit(Surv(CYOS_EFY,Retained)~1,
data = AirOps_SurvDat)

#now show the summary of the km estimates (times
#parameter of summary() gives control over which times to print)
summary(km_AirOps_fit,times=c(10,15,20,25*(1:5)))

#Now display the 'Kaplan-Meir' plot
airOpsSurvPlot = ggsurvplot(km_AirOps_fit,
palette = "blue", xlab = "CYOS", ylab =
"Survival Probability", legend.title = "",
legend.labs = "Air Operations & Special Warfare",
conf.int = TRUE, conf.int.style = "step",
conf.int.alpha = 0.5, risk.table = TRUE,
risk.table.y.text = FALSE)
print(airOpsSurvPlot)

#Now we look at survival curves by significant
#factors-------------------------------------------

#Number of dependents:
km_airOpsDEPENT_fit = survfit(Surv(CYOS_EFY,Retained) ~
DEPENT, data = AirOps_SurvDat)
AOSWdependSurvPlot = ggsurvplot(km_airOpsDEPENT_fit,
```

```
palette = c("blue","dark green"), xlab = "CYOS",
ylab = "Survival Probability", legend.title = "",
legend.labs = c("No Dependents ", "Dependents"),
conf.int = TRUE, conf.int.style = "ribbon",
conf.int.alpha = 0.5, risk.table = TRUE,
risk.table.y.text = TRUE,
tables.theme =  theme_cleantable(),
submain = "Air Ops & Special Warfare
Survival Estimates for Dependents")
print(AOSWdependSurvPlot)

#Number of Deployments:
km_airOpsDeploy_fit = survfit(Surv(CYOS_EFY,Retained) ~
Deployments, data= AirOps_SurvDat)
AOSWdeploySurvPlot = ggsurvplot(km_airOpsDeploy_fit,
palette = c("blue","dark green", "dark orange"),
xlab = "CYOS", ylab = "Survival Probability",
legend.title = "", legend.labs = c("0", "1","2+"),
conf.int = TRUE, conf.int.alpha = 0.5,
risk.table = TRUE, risk.table.y.text = TRUE,
tables.theme =  theme_cleantable(),
submain = "Air Ops & Special Warfare
Survival Estimates for Deployments")
print(AOSWdeploySurvPlot)

#Information Warfare----------------------------------------

#Create the dataframe to look at survival analysis for
#Information Warfare female officers:
survInfoWarDat = rbind(combIWData_CYOS_0to6,
combIWData_CYOS_4to8,combIWData_CYOS_8to14,
combIWData_CYOS_12to19,combIWData_CYOS_20to22)

#Reverse order of data to reflect most recent record for
each female officer:
revOrd_survInfoWar = survInfoWarDat[
order(survInfoWarDat$SSAN,survInfoWarDat$CYOS_EFY,
decreasing=TRUE),]
#Now remove the duplicated information
#(records containing older information) by identified number:
survInfoWar_NoDups = revOrd_survInfoWar[
!duplicated(revOrd_survInfoWar$SSAN),]
```

```
#Subset the data to the necessary variables
InfoWar_SurvDat = subset(survInfoWar_NoDups,
select = c(DG,DEPENT,Deployments,MARITLST,Prior_E,
Race,Spouse_Stat,CYOS_EFY,Retained))

#Fit survival data using Kaplan-Meier method
km_InfoWar = with(InfoWar_SurvDat,Surv(CYOS_EFY,
Retained))
head(km_InfoWar,100)
#check for censored variables (will have + at end of CYOS)
tail(km_InfoWar,100)
#check for censored variables (will have + at end of CYOS)

#This section creates the model that shows the survival
#rate of female officers over the course of 30 yrs
#use the formula surv(time = x, event = y) and survfit
#function to produce the Kaplan-Meier estimates
km_InfoWar_fit = survfit(Surv(CYOS_EFY,Retained)~1,
data = InfoWar_SurvDat)

#now show the summary of the km estimates (times
#parameter of summary() gives control over which times to print)
summary(km_InfoWar_fit,times=c(10,15,20,25*(1:5)))

#Now display the 'Kaplan-Meir' plot
infoWarSurvPlot = ggsurvplot(km_InfoWar_fit,
palette = "blue", xlab = "CYOS",
ylab = "Survival Probability", legend.title = "",
legend.labs = "Information Warfare", conf.int = TRUE,
conf.int.style = "step", conf.int.alpha = 0.5,
risk.table = TRUE,risk.table.y.text = FALSE)
print(infoWarSurvPlot)

#Now we look at survival curves by significant
#factors---------------------------------------------

#Number of Deployments:
km_InfoWardeployStat_fit = survfit(Surv(CYOS_EFY,Retained) ~
Deployments, data = InfoWar_SurvDat)
IWdeploySurvPlot = ggsurvplot(km_InfoWardeployStat_fit,
palette = c("blue","dark green", "dark orange"),
xlab = "CYOS", ylab = "Survival Probability",
legend.title = "", legend.labs = c("0", "1","2+"),
```

```
conf.int = TRUE, conf.int.alpha = 0.5,
risk.table = TRUE, risk.table.y.text = TRUE,
tables.theme =  theme_cleantable(),
submain = "Information Warfare Survival Estimates
for Deployments")
print(IWdeploySurvPlot)

#Spouse Career:
km_InfoWarspouseStat_fit = survfit(Surv(CYOS_EFY,Retained)
~ Spouse_Stat, data = InfoWar_SurvDat)
IWspouseSurvPlot = ggsurvplot(km_InfoWarspouseStat_fit,
palette = c("blue","dark green","dark orange", "red"),
xlab = "CYOS", ylab = "Survival Probability",
legend.title = "",
legend.labs = c("Active","RANG","Other","No Spouse"),
conf.int = FALSE,  conf.int.alpha = 0.5,
risk.table = TRUE, risk.table.y.text = TRUE,
tables.theme =  theme_cleantable(),
submain = "Information Warfare Survival
Estimates for Spouse's Career Status")
print(IWspouseSurvPlot)

#Combat Support---------------------------------------------

#Create the dataframe to look at survival analysis for
#Combat Support female officers:
survCombSupDat = rbind(combCSData_CYOS_0to6,
combCSData_CYOS_4to8,combCSData_CYOS_8to14,
combCSData_CYOS_12to19,combCSData_CYOS_20to22)

#Reverse order of data to reflect most recent
#record for each female officer:
revOrd_survCombSup = survCombSupDat[
order(survCombSupDat$SSAN,survCombSupDat$CYOS_EFY,
decreasing=TRUE),]
#Now remove the duplicated information
#(records containing older information) by identified number:
survCombSup_NoDups = revOrd_survCombSup[
!duplicated(revOrd_survCombSup$SSAN),]
write.csv(survCombSup_NoDups,
"C:/Users/User/Desktop/Astudillo_Backup/Data/
Data in Excel/offinv_annual/Female Only/Fiscal Year/
Final Thesis Data/Developmental Categories/
```

```
CS_survivability_data.csv",row.names = TRUE)


#Subset the data to the necessary variables
CombSup_SurvDat = subset(survCombSup_NoDups,
select = c(DG,DEPENT,Deployments,MARITLST,Prior_E,
Race,Spouse_Stat,CYOS_EFY,Retained))


#Fit survival data using Kaplan-Meier method
km_CombSup = with(CombSup_SurvDat,Surv(CYOS_EFY,
Retained))
head(km_CombSup,100)
#check for censored variables (will have + at end of CYOS)
tail(km_CombSup,100)
#check for censored variables (will have + at end of CYOS)

#This section creates the model that shows the survival
#rate of female officers over the course of 30 yrs
#use the formula surv(time = x, event = y) and
#survfit function to produce the Kaplan-Meier estimates
km_CombSup_fit = survfit(Surv(CYOS_EFY,Retained)~1,
data = CombSup_SurvDat)

#now show the summary of the km estimates (times
#parameter of summary() gives control over
#which times to print)
summary(km_CombSup_fit,times=c(10,15,20,25*(1:5)))

#Now display the 'Kaplan-Meir' plot
combSupSurvPlot = ggsurvplot(km_CombSup_fit,
palette = "blue", xlab = "CYOS",
ylab = "Survival Probability",
legend.title = "", legend.labs = "Combat Support",
conf.int = TRUE, conf.int.style = "step",
conf.int.alpha = 0.5, risk.table = TRUE,
risk.table.y.text = FALSE)
print(combSupSurvPlot)

#Graph the range of survival function curves for all
#possible combinations of combat support:

km_CombSup_fit2 = survfit(Surv(CYOS_EFY,Retained)~
```

```
DEPENT+Deployments+MARITLST+Prior_E+Race+Spouse_Stat+DG,
data = CombSup_SurvDat)

#now show the summary of the km estimates (times
#parameter of summary() gives control over which times to print)
summary(km_CombSup_fit2,times=c(10,15,20,25*(1:5)))

#Now display the 'Kaplan-Meir' plot
combSupSurvPlot2 = ggsurvplot(km_CombSup_fit2,
palette(), xlab = "CYOS", ylab = "Survival Probability",
legend.title = "", conf.int = TRUE, conf.int.style = "step",
conf.int.alpha = 0.5, risk.table = TRUE, risk.table.y.text = FALSE)
print(combSupSurvPlot2)

#Now we look at survival curves by significant
#factors------------------------------------------
#Number of dependents:
km_CSdepend_fit = survfit(Surv(CYOS_EFY,Retained) ~
DEPENT, data = CombSup_SurvDat)
CSdependSurvPlot = ggsurvplot(km_CSdepend_fit,
palette = c("blue","dark green"), xlab = "CYOS",
ylab = "Survival Probability", legend.title = "",
legend.labs = c("No Dependents ", "Dependents"),
conf.int = TRUE, conf.int.style = "ribbon",
conf.int.alpha = 0.5, risk.table = TRUE,
risk.table.y.text = TRUE,
tables.theme =  theme_cleantable(),
submain = "Combat Support Survival Estimates
for Dependents")
print(CSdependSurvPlot)

#Marital Status:
km_CSMaritStat_fit = survfit(Surv(CYOS_EFY,Retained)
~ MARITLST, data = CombSup_SurvDat)
CSmaritStatSurvPlot = ggsurvplot(km_CSMaritStat_fit,
palette = c("blue","dark green", "dark orange"),
xlab = "CYOS", ylab = "Survival Probability",
legend.title = "",
legend.labs = c("Single", "Married", "Divorced/etc."),
conf.int = TRUE,  conf.int.alpha = 0.5,
risk.table = TRUE, risk.table.y.text = TRUE,
tables.theme =  theme_cleantable(),
submain = "Combat Support Survival Estimates
```

```
for Marital Status")
print(CSmaritStatSurvPlot)

#Number of Deployments:
km_CSdeployStat_fit = survfit(Surv(CYOS_EFY,Retained) ~
Deployments, data = CombSup_SurvDat)
CSdeploySurvPlot = ggsurvplot(km_CSdeployStat_fit,
palette = c("blue","dark green", "dark orange"),
xlab = "CYOS", ylab = "Survival Probability",
legend.title = "", legend.labs = c("0", "1","2+"),
conf.int = TRUE, conf.int.alpha = 0.5,
risk.table = TRUE, risk.table.y.text = TRUE,
tables.theme =  theme_cleantable(),
submain = "Combat Support Survival Estimates
for Deployments", pval = TRUE)
print(CSdeploySurvPlot)

#Force Modernization------------------------------------

#Create the dataframe to look at survival analysis for
Force Modernization female officers:
survForModDat = rbind(combFMData_CYOS_0to6,
combFMData_CYOS_4to8,combFMData_CYOS_8to14,
combFMData_CYOS_12to19,combFMData_CYOS_20to22)

#Reverse order of data to reflect most recent
#record for each female officer:
revOrd_survForMod = survForModDat[order
(survForModDat$SSAN,survForModDat$CYOS_EFY,
decreasing=TRUE),]
#Now remove the duplicated information
#(records containing older information) by identified number:
survForMod_NoDups = revOrd_survForMod[
!duplicated(revOrd_survForMod$SSAN),]

#Subset the data to the necessary variables
ForMod_SurvDat = subset(survForMod_NoDups,
select = c(DG,DEPENT,Deployments,MARITLST,Prior_E,
Race,Spouse_Stat,CYOS_EFY,Retained))

#Fit survival data using Kaplan-Meier method
km_ForMod = with(ForMod_SurvDat,Surv(CYOS_EFY,Retained))
head(km_ForMod,100)
```

```
#check for censored variables (will have + at end of CYOS)
tail(km_ForMod,100)
#check for censored variables (will have + at end of CYOS)

#This section creates the model that shows
#the survival rate of female officers over the course
#of 30 yrs use the formula surv(time = x, event = y)
#and survfit function to produce the Kaplan-Meier estimates
km_ForMod_fit = survfit(Surv(CYOS_EFY,Retained)~1,
data = ForMod_SurvDat)

#now show the summary of the km estimates
#(times parameter of summary() gives control
#over which times to print)
summary(km_ForMod_fit,times=c(0,10,15,20,25*(1:5)))

#Now display the 'Kaplan-Meir' plot
ForModSurvPlot = ggsurvplot(km_ForMod_fit,
palette = "blue", xlab = "CYOS",
ylab = "Survival Probability",
legend.title = "",
legend.labs = "Force Modernization",
conf.int = TRUE, conf.int.style = "step",
conf.int.alpha = 0.5, risk.table = TRUE,
risk.table.y.text = FALSE)
print(ForModSurvPlot)


#Now we look at survival curves by significant
#factors------------------------------------------

#Number of Deployments:
km_FMdeployStat_fit = survfit(Surv(CYOS_EFY,Retained) ~
Deployments, data = ForMod_SurvDat)
FMdeploySurvPlot = ggsurvplot(km_FMdeployStat_fit,
palette = c("blue","dark green", "dark orange"),
xlab = "CYOS", ylab = "Survival Probability",
legend.title = "",
legend.labs = c("0", "1","2+"),conf.int = TRUE,
conf.int.alpha = 0.5, risk.table = TRUE,
risk.table.y.text = TRUE,
tables.theme =  theme_cleantable(),
submain = "Force Modernization Survival Estimates
```

```
for Deployments")
print(FMdeploySurvPlot)

#Spouse Career:
km_ForModSpouseStat_fit = survfit(Surv(CYOS_EFY,Retained) ~
Spouse_Stat, data = ForMod_SurvDat)
FMspouseSurvPlot = ggsurvplot(km_ForModSpouseStat_fit,
palette = c("blue","dark green","dark orange", "red"),
xlab = "CYOS", ylab = "Survival Probability",
legend.title = "",
legend.labs = c("Active","RANG","Other","No Spouse"),
conf.int = FALSE, conf.int.alpha = 0.5,
risk.table = TRUE, risk.table.y.text = TRUE,
tables.theme =  theme_cleantable(),
submain = "Force Modernization Survival
Estimates for Spouse's Career Category")
print(FMspouseSurvPlot)

#-------------------------------------------------------------
#-------------------------------------------------------------
#-------------------------------------------------------------
#This section contains Cox Proportional Hazards code for
#explanatory factors considered significant in
#the Survival Analysis section
#--------------------------------------------------------------

#Read in excel spreadsheet containing start and stop times
Start_End_Data = read_excel("C:/Users/User/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/
Female Only/Fiscal Year/Final Thesis Data/coxPropData2.xlsx")

#Subset the data with necessary variables
coxPropvariables = subset(Start_End_Data,select =
c(DG,DEPENT,Deployments,MARITLST,Prior_E,Race,Spouse_Stat,
CYOS_Start,CYOS_End,Retained))

#Compute a Cox regression model using the cohort dataset
cox.fit = coxph(Surv(CYOS_Start,CYOS_End,Retained)~DG+DEPENT+
Deployments+MARITLST+Spouse_Stat+Race, data = coxPropvariables)

#display the results
summary(cox.fit)
```

```
#Display hazard ratios of each covariate
Cohorthazratios = ggforest(fit, data = coxPropvariables)
print(Cohorthazratios)

#------------------------------------------------------------
#Now collect the regression covariate p-values for all
#developmental categories and plot their hazard ratios

#Read in appropriate file with start/end times
AOSW_CPHData = read_excel("C:/Users/Jess/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/
Female Only/Fiscal Year/Final Thesis Data/
AirOps_SurvDat2.xlsx")

#Subset data to appropriate variables
AOSWcoxprop = subset(AOSW_CPHData, select =
c(DEPENT,Deployments,MARITLST,Prior_E,Race,
Spouse_Stat,CYOS_Start,CYOS_End,Retained))

#Compute a Cox regression model using the AOSW data
AOSW.cox.fit = coxph(Surv(CYOS_Start,CYOS_End,Retained)
~DEPENT+Deployments, data = AOSWcoxprop)

#display the results
summary(AOSW.cox.fit)

#Now read in individual levels of each covariate to
#show their hazard ratios
AOSW.cox.fit2 = coxph(Surv(CYOS_EFY,Retained)~
DEPENT+Deployments, data = AirOps_SurvDat)
summary(AOSW.cox.fit2)
AOSWhazvariables = subset(AirOps_SurvDat, select =
c(DEPENT,Deployments))

#Display hazard ratios of each covariate
AOSWhazratios = ggforest(AOSW.cox.fit2,
data = AOSWhazvariables)
print(AOSWhazratios)


#IW-------------------------------------------------------
#Read in appropriate file with start/end times
IW_CPHData = read_excel("C:/Users/Jess/Desktop/
```

```
Astudillo_Backup/Data/Data in Excel/offinv_annual/
Female Only/Fiscal Year/Final Thesis Data/
InfoWar_SurvDat2.xlsx")

#Subset data to appropriate variables
IWcoxprop = subset(IW_CPHData, select =
c(DEPENT,Deployments,MARITLST,Prior_E,Race,Spouse_Stat,
CYOS_Start,CYOS_End,Retained))

#Compute a Cox regression model using the cohort dataset
IW.cox.fit = coxph(Surv(CYOS_Start,CYOS_End,Retained)
~Spouse_Stat+Deployments, data = IWcoxprop)

#display the results
summary(IW.cox.fit)

#Now read in individual levels of each covariate
#to show their hazard ratios
IW.cox.fit2 = coxph(Surv(CYOS_EFY,Retained)~
DEPENT+Deployments, data = InfoWar_SurvDat)
summary(IW.cox.fit2)
IWhazvariables = subset(InfoWar_SurvDat, select
= c(DEPENT,Deployments))

#Display hazard ratios of each covariate
IWhazratios = ggforest(IW.cox.fit2, data =
IWhazvariables)
print(IWhazratios)

#CS-----------------------------------------------------------
#Read in appropriate file with start/end times
CS_CPHData = read_excel("C:/Users/Jess/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/
Female Only/Fiscal Year/Final Thesis Data/
CombSup_SurvDat2.xlsx")

#Subset data to appropriate variables
CScoxprop = subset(CS_CPHData, select =
c(DEPENT,Deployments,MARITLST,Prior_E,Race,
Spouse_Stat,CYOS_Start,CYOS_End,Retained))

#Compute a Cox regression model using the cohort dataset
CS.cox.fit = coxph(Surv(CYOS_Start,CYOS_End,Retained)
```

```
~Deployments+DEPENT+MARITLST, data = CScoxprop)

#display the results
summary(CS.cox.fit)

#Now read in individual levels of each covariate
#to show their hazard ratios
CS.cox.fit2 = coxph(Surv(CYOS_EFY,Retained)~DEPENT+
Deployments, data = CombSup_SurvDat)
summary(CS.cox.fit2)
CShazvariables = subset(CombSup_SurvDat,
select = c(DEPENT,Deployments))

#Display hazard ratios of each covariate
CShazratios = ggforest(CS.cox.fit2, data =
CShazvariables)
print(CShazratios)


#FM----------------------------------------------------------
#Read in appropriate file with start/end times
FM_CPHData = read_excel("C:/Users/Jess/Desktop/
Astudillo_Backup/Data/Data in Excel/offinv_annual/
Female Only/Fiscal Year/Final Thesis Data/ForMod_SurvDat2.xlsx")

#Subset data to appropriate variables
FMcoxprop = subset(FM_CPHData, select =
c(DEPENT,Deployments,MARITLST,Prior_E,Race,
Spouse_Stat,CYOS_Start,CYOS_End,Retained))

#Compute a Cox regression model using the cohort dataset
FM.cox.fit = coxph(Surv(CYOS_Start,CYOS_End,Retained)
~Spouse_Stat+Deployments+MARITLST, data = FMcoxprop)

#display the results
summary(FM.cox.fit)

#Now read in individual levels of each covariate to
#show their hazard ratios
FM.cox.fit2 = coxph(Surv(CYOS_EFY,Retained)~
Spouse_Stat+Deployments+MARITLST, data = ForMod_SurvDat)
summary(FM.cox.fit2)
FMhazvariables = subset(ForMod_SurvDat, select =
```

```
c(Deployments,Spouse_Stat,MARITLST))

#Display hazard ratios of each covariate
FMhazratios = ggforest(FM.cox.fit2, data = FMhazvariables)
print(FMhazratios)
```

Bibliography

Allen, S. V. (2018), Female officer retention in the marine corps: An exploratory study of early supervisors and mentors, Technical report, Naval Postgraduate School Monterey United States.

Asch, B. J., Mattock, M. G. and Hosek, J. (2017), *The blended retirement system: Retention effects and continuation pay cost estimates for the armed services*, Rand Corporation, Santa Monica CA US.

Asch, B. J., Miller, T. and Malchiodi, A. (2012), A new look at gender and minority differences in officer career progression in the military, Technical report, RAND NATIONAL DEFENSE RESEARCH INST, Santa Monica CA.

Bissonnette, P. (2012), A quantitative analysis of factors affecting retention of female aviators in us naval aviation, Technical report, Naval Postgraduate School, Monterey CA.

Blau, F. D. and Kahn, L. M. (2007), 'The gender pay gap: Have women gone as far as they can?', *Academy of Management Perspectives* **21**(1), 7–23.

Bureau of Labor Statistics (2020), 'Current population survey', `https://www.bls.gov/news.release/pdf/empsit.pdf?mod=article_inline&mod=article_inline`. Retrieved 23 February 2021.

Carter, A. (2016), 'Department of defense press briefing by secretary carter on force of the future reforms in the pentagon press briefing room'.

Chamberlain, A. (2016), 'Demystifying the gender pay gap', *Mill Valley, CA: Glassdoor* .

Delle Donne, A., Hatch, A., Carr, N. R., Aden, J. and Shapiro, J. (2019), 'Extended maternity leave and breastfeeding in active duty mothers', *Pediatrics* **144**(2), e20183795.

Demirel, T. (2002), A statistical analysis of officer retention in the us military, Technical report, Naval Postgraduate School, Monterey CA.

Franzen, C. N. (2017), Survival analysis of us air force officer retention rate, Master's thesis, School of Engineering and Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH. AD-1051631.

Gildea, D. (2013), 'Air force personnel center: Milpds upgrade begins', `https://www.afpc.af.mil/News/Article-Display/Article/422050/milpds-upgrade-begins/`. Retrieved 23 February 2021.

Hosmer Jr, D. W., Lemeshow, S. and Sturdivant, R. X. (2013), *Applied logistic regression*, Vol. 398, John Wiley & Sons, Hoboken, NJ.

Jansen, D. J., Kamarck, K. N., Kapp, L. and Salazar Torreon, B. (2015), 'Fy2016 national defense authorization act: Selected military personnel issues'.

Johnson-Freese, J., Haring, E. and Ulrich, M. (2014), The counterproductive sea of sameness in pme, Technical report, National Defense University, Fort McNair D.C.

Joint Economic Committee (2016), Gender pay inequality: Consequences for women, families and the economy, Technical report, United States Congress, Washington D.C.

Jomana, A. (2020), 'Military woman and the force of the future', *Defense & Peace Economics* **31**(1), 1–3. Retrieved 23 February 2021.

Kanter, R. M. (2008), 'Men and women of the corporation revisited', *Management Review* **76**(3), 14.

Kapp, L. and Salazar Torreon, B. (2020), Reserve component personnel issues: Questions and answers, Technical report, Congressional Research Service, Washington D.C. RL30802.

Karmarck, K. N. (2020), 'Defense primer: Personnel tempo (perstempo)', `https://fas.org/sgp/crs/natsec/IF11007.pdf`. Retrieved 23 February 2021.

Keller, K. M., Hall, K. C., Matthews, M., Payne, L. A., Saum-Manning, L., Yeung, D., Schulker, D., Zavislan, S. and Lim, N. (2018), Addressing barriers to female officer retention in the air force, Technical report, RAND Corporation, Santa Monica United States.

Kleinbaum, D. G. and Klein, M. (2010), *Survival analysis*, Springer.

Kolano, B. M. (2019), Multivariate analysis of diversity and inclusion data, Master's thesis, School of Engineering and Management, Air Force Institute of Technology (AU), Wright-Patterson AFB, OH.

Losey, S. (2018), 'Air force expands parental leave for new parents', `https://www.airforcetimes.com/news/your-air-force/2018/06/08/air-force-expands-parental-leave/`. Retrieved 23 February 2021.

Mahoney, M. (2018), 'The military's new blended retirement system (brs): Will the new brs pension plan achieve the equitable benefits and financial incentives to retain long-term service members?', *Proceedings of the Northeast Business & Economics Association* .

Matthews, M. (2018), 'The us air force is working harder to retain female', https://www.rand.org/blog/2018/08/the-us-air-force-is-working-harder-to-retain-female.html. Retrieved 23 February 2021.

McMahon, C. J. and Bernard, C. J. (2019), 'Storm clouds on the horizon', *Naval War College Review* **72**(3), 84–100.

Meeker, W. Q. and Escobar, L. A. (2014), *Statistical methods for reliability data*, John Wiley & Sons.

Montgomery, D. C., Peck, E. A. and Vining, G. G. (2021), *Introduction to linear regression analysis*, John Wiley & Sons.

Patten, E. (2016), 'Racial, gender wage gaps persist in u.s. despite some progress', https://www.pewresearch.org/fact-tank/2016/07/01/racial-gender-wage-gaps-persist-in-u-s-despite-some-progress/. Retrieved 23 February 2021.

Pourhoseingholi, M. A., Baghestani, A. R. and Vahedi, M. (2012), 'How to control confounding effects by statistical analysis', *Gastroenterology and hepatology from bed to bench* **5**(2), 79.

Roberts, K., Dowell, A. and Nie, J.-B. (2019), 'Attempting rigour and replicability in thematic analysis of qualitative research data; a case study of codebook development', *BMC medical research methodology* **19**(1), 1–8.

Schieder, J. and Gould, E. (2016), '"women's work" and the gender pay gap: how discrimination, societal norms, and other forces affect women's occupational choices-and their pay', *Economic Policy Institute* **20**.

Schofield, J. A. (2015), Non-rated air force line officer attrition rates using survival analysis, Master's thesis, School of Engineering and Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH. AD-A615579.

Secretary of the Air Force Public Affairs (2019), 'Air force announces new officer developmental categories', https://www.jbsa.mil/News/News/Article/1996781/air-force-announces-new-officer-developmental-categories/. Retrieved 23 February 2021.

Smith, D. G. and Rosenstein, J. E. (2017), 'Gender and the military profession: Early career influences, attitudes, and intentions', *Armed Forces & Society* **43**(2), 260–279.

United States Government Accountability Office (2020), Female active-duty personnel: Guidance and plans needed for recruitment and retention efforts, Technical report, Government Accountability Office, Washington D.C.

Zais, M. and Zhang, D. (2016), 'A markov chain model of military personnel dynamics', *International Journal of Production Research* **54**(6), 1863–1885.

Zens, C. L. (2016), Application of non-rated line officer attrition levels and career, Master's thesis, School of Engineering and Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH. AD-1054019.

Zimmermann, J. T. (2017), Application of enlisted force retention levels and career field stability, Master's thesis, School of Engineering and Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH. AD-1051596.

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704–0188*

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
25–03–2021	Master's Thesis	August 2019 — March 2021

4. TITLE AND SUBTITLE

An Examination into Retention Behavior of Air Force Female Officers

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Jessica M. Astudillo, Capt, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENS-MS-21-M-142

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

HAF/A1XD
Douglas A. Boerman
1550 W. Perimeter Rd,, Rm 4710
Joint Base Andrews NAF Washington, MD 20762-5000
Email: douglas.a.boerman.civ@mail.mil

10. SPONSOR/MONITOR'S ACRONYM(S)

HAF/A1XD

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

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A:
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Female retention rates in the US military have been considerably lower than that of their male counterparts for numerous years. In the Air Force, women represent 14 percent of officer ranks from O-5 level and above. Comparatively, the overall rate of women officers in service is 20 percent. Understanding the negative factors associated with the attrition rate of this group can help the Air Force leverage positive change. It may also influence adjustments that will increase the number of women serving, and improve diversity throughout both the officer and enlisted ranks. In this study, logistic regression and survival analysis are applied to model retention and some understanding of how to diversify the Air Force, through increasing our female officer population. Demographic, organizational, and political elements are considered to ensure all affecting issues are measured. Programs that have gone into effect in the past five years, such as the Force of the Future, and Blended Retirement, are also considered to determine their statistical significance. Applying logistic regression determines potential factors affecting retention rates. All elements are include in survival analysis to characterize female officer retention behavior. Implementing and providing such analysis will help generate a prediction model for retention rates amongst female officers, and how to further amplify diversity.

15. SUBJECT TERMS

Retention modeling, personnel analysis, survivability modeling, logistic regression

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dr. Raymond R. Hill, AFIT/ENS
					19b. TELEPHONE NUMBER *(include area code)*
U	U	U	UU	196	(937) 255-3636, x7469; rhill@afit.edu