NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

CONTINUATION STUDY FOR THE SELECTED MARINE CORPS RESERVE

by

Lily M. Brose

June 2018

Thesis Advisor: Andrew T. Anglemyer
Second Reader: Thomas W. Lucas

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*

1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** June 2018	**3. REPORT TYPE AND DATES COVERED** Master's thesis

4. TITLE AND SUBTITLE CONTINUATION STUDY FOR THE SELECTED MARINE CORPS RESERVE	**5. FUNDING NUMBERS**
6. AUTHOR(S) Lily M. Brose	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A	**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release. Distribution is unlimited.	**12b. DISTRIBUTION CODE** A

13. ABSTRACT (maximum 200 words)

This continuation study follows a cohort of Marines who enlisted in the Selected Marine Corps Reserve (SMCR) in FY09. For the purposes of this thesis, continuation is defined as members lost to other reserve or active Marine Corps components. These options include transferring to the Individual Ready Reserve (IRR), Active Reserve (AR), Active Component (AC), Individual Mobilization Augmentee (IMA), or leaving the Marine Corps Reserve altogether. Every person in this population was on a 6x2 contract (six-year commitment to the SMCR and the remaining two served in the SMCR or IRR). This study implements two logistic regressions and classification trees at different career milestones. The main research goal is to identify significant impacts of mobilization, prior service history, and demographics on the continuation rates of Reserve enlisted.

At the six-year mark, a combat deployment history and rank (relative to E-1) have a positive impact on transferring to another component. Furthermore, being a minority or living in the Western U.S. (relative to the Northeast) has a negative impact on continuation. The classification tree found rank and deployment rate to be the best classifiers of SMCR retention.

At the eight-year mark, average conduct score has a positive impact on continuation, while number of dependents and commissioning status have negative impacts on continuation. The classification tree found average conduct score and deployment rate to be the best classifiers.

14. SUBJECT TERMS retention, continuation, Marine Corps Reserve, Selected Marine Corps Reserve, logistic regression, classification tree	**15. NUMBER OF PAGES** 77
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
Unclassified	Unclassified	Unclassified	UU

NSN 7540-01-280-5500 Standard Form 298 (Rev. 2-89) Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release. Distribution is unlimited.

CONTINUATION STUDY FOR THE SELECTED MARINE CORPS RESERVE

Lily M. Brose
Ensign, United States Navy
BS, U.S. Naval Academy, 2017

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN APPLIED SCIENCE (OPERATIONS RESEARCH)

from the

NAVAL POSTGRADUATE SCHOOL
June 2018

Approved by: Andrew T. Anglemyer
Advisor

Thomas W. Lucas
Second Reader

Patricia A. Jacobs
Chair, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

This continuation study follows a cohort of Marines who enlisted in the Selected Marine Corps Reserve (SMCR) in FY09. For the purposes of this thesis, continuation is defined as members lost to other reserve or active Marine Corps components. These options include transferring to the Individual Ready Reserve (IRR), Active Reserve (AR), Active Component (AC), Individual Mobilization Augmentee (IMA), or leaving the Marine Corps Reserve altogether. Every person in this population was on a 6x2 contract (six-year commitment to the SMCR and the remaining two served in the SMCR or IRR). This study implements two logistic regressions and classification trees at different career milestones. The main research goal is to identify significant impacts of mobilization, prior service history, and demographics on the continuation rates of Reserve enlisted.

At the six-year mark, a combat deployment history and rank (relative to E-1) have a positive impact on transferring to another component. Furthermore, being a minority or living in the Western U.S. (relative to the Northeast) has a negative impact on continuation. The classification tree found rank and deployment rate to be the best classifiers of SMCR retention.

At the eight-year mark, average conduct score has a positive impact on continuation, while number of dependents and commissioning status have negative impacts on continuation. The classification tree found average conduct score and deployment rate to be the best classifiers.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

AC	Active Component
AR	Active Reserve
ASL	Active Status List
AFQT	Armed Forces Qualification Test
ASVAB	Armed Services Vocational Aptitude Battery
CFT	Combat Fitness Test
CONUS	Continental United States
CSV	Comma-Separated Value
DMDC	Defense Manpower Data Center
DoD	Department of Defense
E/S	End Strength
EDIPI	Electronic Data Interchange Personal Identifier
FMCR	Fleet Marine Corps Reserve
FY	Fiscal Year
GCT	General Classification Test
HQMC	Headquarters, United States Marine Corps
I&I	Inspector and Instructor
IADT	Initial Active Duty for Training
IDT	Inactive Duty Training
IMA	Individual Mobilization Augmentee
IRR	Individual Ready Reserve
ISL	Inactive Status List
MEF	Marine Expeditionary Force
MCR	Marine Corps Reserve
MCRAMM	Marine Corps Reserve Administrative Management Manual
M&RA	Manpower and Reserve Affairs
MOS	Military Occupational Specialty
MSO	Military Service Obligation
NCO	Noncommissioned Officers
NPS	Non-prior Service

OCONUS	Outside the Continental United States
PFT	Physical Fitness Test
PMOS	Primary Military Specialty
PS	Prior Service
RA	Reserve Affairs
RASL	Reserve Active Status List
RC	Reserve Component
REC	Reenlistment Eligibility Code
ROCP	Reserve Officer Commissioning Program
SelRes	Selected Reserve
SMCR	Selected Marine Corps Reserve
SNCO	Staff Noncommissioned Officers
TFDW	Total Force Data Warehouse
USMC	United States Marine Corps

EXECUTIVE SUMMARY

Due to numerous career options available to Marines currently serving in the Selected Marine Corps Reserve (SMCR), it can be difficult to accurately and efficiently track the attrition, retention, and continuation rates of these Marines. Options include transitioning to the Individual Ready Reserve (IRR), Active Reserve (AR), Active Component (AC), serving as an Individual Mobilization Augmentee (IMA), or choosing to end their time in the Marine Corps altogether (Department of the Navy, 2015).

This thesis utilizes multiple logistic regressions (one at the six-year mark and one at the eight-year mark) to analyze predictor variables that are related to Marines leaving the SMCR to another sector of the Marine Corps Reserve (MCR), defined as continuation. This research follows the careers of every non-prior service (NPS) enlisted Marine in the SMCR that enlisted on a 6x2 from FY09 through FY17 (through initial commitment). The models and subsequent analysis of explanatory variables of those who decide to leave the SMCR before their contract is complete, but not the Marine Corps at large, has been generated for the Reserve Affairs (RA) Division, Manpower & Reserve Affairs (M&RA), Headquarters, United States Marine Corps.

This empirical continuation study explored two logistic regressions and two classification trees at different career milestones. The population is a cohort of non-prior service Marines that enlisted in the SMCR in FY09 on a 6x2 contract. With this contract the service member is required to complete the first six years in the SMCR—the remaining years can be served in the SMCR or IRR. Continuation in this thesis is defined as a loss (from the SMCR) to another MCR component.

The response variable for the first logistic regression is binary with value 0 if the Marine is in IRR, IMA, or AR at the six-year mark and 1 if he/she is in the SMCR at the six-year mark. The model had little predictive power (pseudo R^2=11%), but impacts of the predictors can still be ascertained. A non-combat deployment history increases odds of transfer in the first six-years. This concurs with previous research that found that more deployments increased the likelihood of attrition (Price, 2010). Also, White Marines are more likely to transfer as well. This supports previous research by Randall in which NPS

Marines of a minority race had much larger retention rate than White NPS Marines (Randall, 1989). Finally, home region—specifically those living in the Western US—states has a large positive influence on retention in the SMCR. The six-year classification tree found rank and deployment rate at the six-year mark are the best classifiers of the outcome variable.

The sample for the second logistic regression model is everyone who was in the SMCR at the six-year mark. The outcome variable is binary, defined as 0 for IRR, IMA, or AR at the eight-year mark or 1 for SMCR at the eight-year mark. This model found a large positive impact of SMCR retention with commissioning status. If the Marine was commissioned, he/she was over 18 times more likely to be retained in the SMCR on the scale of the odds ratio. An increase in the number of dependents also has a positive impact on retention. Both of these impacts are intuitive: enlisted to officer programs usually impose a longer contract on the service member. Also, when a Marine has more dependents they are more likely to want to keep their military job security and extra income. An increase in average conduct service record, however, increases the chances of transferring to another component between the six- and eight-year mark. For the classification tree, average conduct service record and deployment rate are the best classifiers of SMCR retention.

The poor predictive power of these models is largely due to the inability to capture the elements that are predictive of the outcome in this population. These elements may include why a Marine chooses to remain in a drilling status eligible to deploy as a unit (SMCR) versus deploying as an individual (IMA), fulfilling an active duty billet (AR), or moving to a non-drilling status (IRR). Data on civilian wages and unemployment rate during this time (FY09-FY17) might be able to explain this movement. Also, predicting future human behavior without capturing all of the important elements, as I did here with the likelihood of continuing in SMCR or not, is challenging. Everyone in this population was still retained in the MCR as a whole. Every person used in the predictive analysis was still in the military. Therefore, the differences between the populations were not easily distinguishable.

References

Department of the Navy. (2015). Marine Corps Reserve administrative management manual. (MCO 1001R.1L). Washington, DC: Author. Retrieved from http://www.marines.mil/ Portals/59/MCO%201001R.1L.pdf

Price, J. D. (2010). *Effects of activation on selected Marine Corps Reserve prior enlisted continuation rate in the post-9/11 era* (Master's thesis). Retrieved from http://hdl.handle.net/10945/5437

Randall, J. S. (1989). *Factors influencing the retention of noncommissioned and staff noncommissioned officers in the Selected Marine Corps Reserve* (Master's thesis). Retrieved from http://hdl.handle.net/10945/27049

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

I would like to thank a number of people for their help and support that made the completion of this thesis possible. First, I would like to thank my advisor, Dr. Andrew Anglemyer, for his constant expertise and confidence in me. Next, I would like to extend my gratitude to Dr. Thomas Lucas for being the second pair of eyes everyone needs.

I am also grateful for Major Sean Norton at RA Division, M&RA, HQMC, for his subject-matter knowledge. Also, for his promptness and patience with my many questions.

Thank you to Allison, Amanda, and Audrey for being my constant sounding boards through the high, low, and procrastination points during this past year. I would also like to thank my familial support team: Mom, Dad, my sister, Daisy, and brother, Marshal. Thank you for the love and encouragement throughout this process. I love you guys very much.

Finally, I give thanks and praise God for giving me the strength, wisdom, patience, and determination to finish this thesis.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION AND LITERATURE REVIEW

A. PURPOSE

This thesis utilizes multiple logistic regressions to analyze predictor variables that relate to Marines leaving the Selected Marine Corps Reserve (SMCR) to another sector of the Marine Corps Reserve (MCR), defined as continuation. This research follows the careers of every non-prior service (NPS) enlisted Marine in the SMCR that enlisted on a 6x2 from FY09 through FY17 (through initial commitment). The models and subsequent analysis of explanatory variables of those who decide to leave the SMCR to another sector of the MCR before their contract is complete, but not the Marine Corps at large, has been generated for the Reserve Affairs (RA) Division, Manpower & Reserve Affairs (M&RA), Headquarters, United States Marine Corps.

B. BACKGROUND

Due to numerous career options available to Marines currently serving in the SMCR, it can be difficult to accurately and efficiently track the attrition, retention, and continuation rates of these Marines. Options include transitioning to the Individual Ready Reserve (IRR), Active Reserve (AR), Active Component (AC), serving as an Individual Mobilization Augmentee (IMA), or choosing to end their time in the Marine Corps altogether.

Marine Corps Reserve attrition and retention is easily misunderstood. Herschelmen (2012) identifies the two forms of attrition: losses to another RC or AC (which is how we will define a positive unit attrition), and losses to the civilian world (which is how we will define a negative unit attrition). In identifying key factors that lead to continuation in the service as a whole, Marine Corps recruiters can more effectively identify potential recruits most likely to be retained and Inspector and Instructor (I&I) Staff will have a better understanding of how they can better incentivize SMCR Marines to continue with their service.

The mission of the Reserve Component (RC) of the Marine Corps is "to augment, reinforce, and sustain the AC with trained units and qualified individuals in times of war

or national emergency and at other such times as national security may require" (Department of the Navy, 2015). The MCR consists of an infantry division, air wing and logistics group (acting as a 4th Marine Expeditionary Force (MEF)). In addition, the RC has been created in such a way as to complement AC capabilities and organization, a status Licari (2013) describes as mirror imaging (Licari, 2013). The RC is indistinguishable from the AC in regards to the range of missions they are capable of performing.

C. MARINE CORPS RESERVE COMPOSITION

Depicted in Figure 1 is the breakdown of how the Marine Corps Reserve Component is structured. The MCR is comprised of the Ready Reserve, Standby Reserve, and the Retired Reserve.

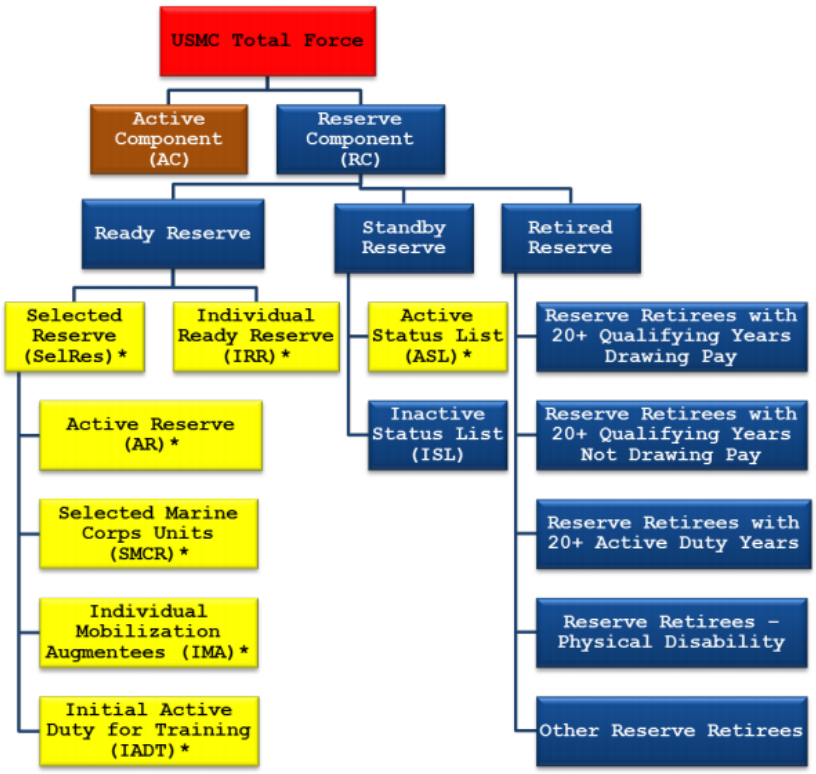

Figure 1. USMC Force Composition Breakdown. Source: MCO 1001R.1L (2015).

1. **<u>Ready Reserve</u>:** Members of the Ready Reserve can be called to Active Duty (AD) during a time of war or national emergency.
    - **(i) Selected Reserve (SelRes):** As of November 2017, the SelRes comprised of 38,721 Marines. The SelRes is made up of the Active Reserve, Selected Marine Corps Reserve, Individual Mobilization Augmentees, and Initial Active Duty for Training.
        - i. **Active Reserve (AR):** Serve on active duty.
        - ii. **Selected Marine Corps Reserve (SMCR):** Members are assigned to a unit within a 100-mile radius of his/her home of record. Waivers must be granted if the member desires to drill at a unit further than 100 miles. Scheduled duty takes place one weekend a month and two weeks a year.
        - iii. **Individual Mobilization Augmentees (IMA):** The IMA provides personnel during the early stages of a crisis. IMA mobilization occurs before IRR mobilization. Marines in the IMA are required to attend 12 days of annual training and usually 48 hours of Inactive Duty Training (depending on billet).
        - iv. **Initial Active Duty for Training (IADT):** IADT status is given to those just entering the SelRes who are undergoing the initial training. IADT Marines are not authorized to be mobilized until the completion of their training.
    - **(ii) Individual Ready Reserve (IRR):** The majority of IRR Marines came from either the AC or SelRes. Marines transfer to the IRR to complete their Military Service Obligation (MSO) or desire to reenlist without being attached to the SelRes.
2. **<u>Standby Reserve:</u>** Members of the Standby Reserve have no unit or training requirements; however, specific skill sets will be mobilized as needed. The Standby Reserve is the smallest RC at 794 Marines.
    - **(i) Active Status List (ASL):** ASL Marines are eligible for promotion and train for retirement point credit only. They are not paid. ASL members

are mobilized for AD by the SECNAV during times or war or national emergencies.

(ii) Inactive Status List (ISL): Members of the ISL are not paid, eligible for promotion, or retirement point credit. Marines consist of those that transferred to the ISL by request, have HIV, or officers with insufficient retirement points.

3. **Retired Reserve:** Retired Reservists fall under various categories of service years and retirement pay. The Retired Reserve is the largest component within the RC, with over 100,000 members (MCRAMM, 2015).

For the purposes of this thesis, our main effort focuses on the Ready Reserve; specifically, the movement within the SMCR and SelRes/RC at large.

D. SMCR

1. Enlisted Accessions

The RC consists of Prior Service (PS) and Non-Prior Service (NPS) Marines. PS Marines have completed their active duty service. NPS Marines choose one of three contract options—4 years active status, 4 years inactive status (4x4); 5 years active status, 3 years inactive status (5x3); and 6 years active status, 2 years inactive status (6x2). All active years are spent in the SMCR, and the inactive years are spent in the IRR by default; however, the service member can choose to serve out those years in any branch of the SelRes. Also, a continuing Marine reenlists into the Reserve Component as a whole, not a specific branch. Therefore, transfer between the different reserve components is not uncommon. In addition, Marines may be eligible for enlistment/reenlistment bonuses with contracts of four to seven years in the SelRes—usually, split between the SMCR and the IRR.

2. Officer Accessions

Officers who have completed or have any remaining AC obligation make up most of the Reserve Officers. If the Marine is still fulfilling his/her initial contract, he/she serves within the Ready Reserve. The other officer accession source is the Reserve Officer

Commissioning Program (ROCP). The ROCP trains and sends officers to SMCR units. When ROCP officers commission, they owe eight years to the MCR—four years in the SMCR, and four years within any branch in the SelRes (this is similar to the first enlisted contract option).

3. Transfer To/From/Within SMCR

(i) Enlisted

Enlisted Marines in the IRR can request a transfer to the SMCR pending satisfactory physical condition and not having a disqualifying Reenlistment Eligibility Code (REC). Marines that have completed their mandatory service in the SMCR, but have remaining time on their contract will be eligible to transfer to the IRR. However, usually they will remain in the SMCR unit. Another common reason of transfer to the IRR is if the commuting distance to the Marine's SMCR unit is beyond the 100-mile radius. Also, if the SMCR does not have an Inactive Drill Training requirement they leave for the IRR upon request.

(ii) Officers

Physically qualified officers who are junior to the SMCR unit Commanding Officer can transfer from the IRR to that SMCR unit upon request. Similar to enlisted, SMCR officers may request and be transferred to the IRR, assuming they do not have a current IDT obligation. They may also be transferred due to unsatisfactory performance or failure to become qualified in MOS (MCRAMM, 2015).

E. SMCR CURRENT FORCE STRENGTH

As of November 2017, according to data provided by Marine Corps Total Force Data Warehouse (TFDW), the SelRes consisted of 38,721 Marines. The SMCR accounts for the largest majority of the SelRes—30,790 Marines: 2,318 officers and 28,472 enlisted members. The complete force total breakdown within the SelRes is shown in Figure 2.

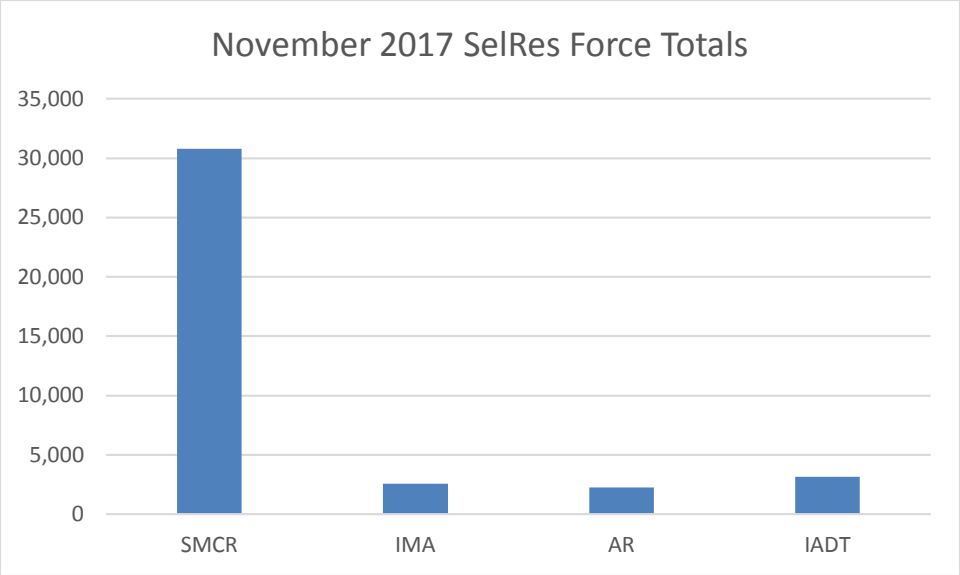

Figure 2. Composition of Entire SelRes in November 2017.

From May to September 2017, the SMCR experienced between 5,197 and 7,748 losses each month, with the number of losses in each month greater than the month before. These numbers are much greater than the losses incurred from in the first half of 2017—ranging from 2,520 to 4,605 losses per month. The average losses per month from FY99–FY17 was 4,466 Marines.

The current FY18 Manpower Plan includes an End Strength (E/S) target as well as specific plans for maintaining enlisted and officer inventory, especially in the needed ranks. The FY18 authorized SelRes E/S is 38,500 personnel. Manpower & Reserves Affairs plans on incentivizing the retention of Senior Staff Non-Commissioned Officers, Company Grade Officers, and Warrant Officers through accession and retention bonuses.

F. OBJECTIVE

The main objective of this thesis is to quantify and determine where and why these NPS enlisted Marines are leaving the SMCR. Since movement within the SelRes is not unusual we might see that many of the SMCR losses are IRR/other RC gains. In other words, these Marines are not actually attriting and are not included in total Marine Corps Reserve losses. Obviously, some have completed their MSO and receive a medical or bad

conduct discharge—or returned to the Active Component. However, by identifying those who continue in the USMC at large, we can give the SMCR a more accurate picture of where their Marines are going and why. Furthermore, we may be equipping recruiters with a better picture of the characteristics/predictors of continuing Marines. In conclusion, this will give them a better understanding of how to incentivize and retain more members of the SMCR.

G. LITERATURE REVIEW

Attrition and retention in the military, including the Marine Corps Reserve, has been researched extensively. Exploring specifically continuation rather than general retention, however, has not. This literature review covers past studies on enlisted, officer, and civilian retention and continuation factors.

Schumacher (2005) utilized a logistic regression to examine how a Marine's mobilization/activation and civilian unemployment affects his or her retention decision. Retention was defined as completing the initial contract and reenlisting (Schumacher, 2005). This study encompasses both enlisted and officers in all three branches of the Marine Corps Reserve. Schumacher found that that the number of activations the service member incurred had a positive impact on retention. An increase in home of record unemployment also had a positive impact on retention behavior. Negative predictors included continued mobilization (longer periods of time) and being recalled from the IRR and retired status. Schumacher discovered that the positive impact of frequency of activation is greater than the adverse impact of length of activation. Furthermore, he recommends more frequent, shorter activation periods (Schumacher, 2005).

Price (2010) focused on 12-month prior service continuation rates of E-3 to E-5s in the SMCR. Price's population was those serving between August 2001 and October 2009. Price used probit regression to analyze continuation behavior at 4-, 12-, and 24- months (Price, 2010). The predictors in his model included activation, demographics, bonuses, and economic conditions. Price, like Schumacher, found that the number of activations have a positive impact on continuation. He warns, however, that this might be caused partially by self-selection bias as Reserve Marines who desire to activate can volunteer to do so (Price,

2010). Other positive predictors are bonuses and unemployment rate. These two together may be a result of not having the same wage opportunities in the civilian sector. Additional significant predictors include prior RC experience, tour length, and multiple tours. Negative explanatory variables discovered were activation length and deploying outside of the continental United States (OCONUS) (Price, 2010). This makes sense because if they wanted to do standard deployments they would have joined the Active Component. Furthermore, poor conduct record, being female, and being older also negatively impact continuation. These results are also expected because those who are not performing well are less likely to receive a promotion and will desire to leave and find a job they are better at. In addition, females who are more traditionally family-oriented are more likely to leave (Price, 2010).

Hansen and MacLeod (2004) studied enlisted retention and continuum of service in reserve and guard components across all branches of the military from FY00-FY03. The authors implemented a logistic regression and found an increase in continuation across the time series (Hansen and MacLeod, 2004). During this three-year period, mobilization rates were increasing, especially post-9/11. However, the data used did not have activation, mobilization, or deployment history. Therefore, a direct positive relationship between the two cannot be inferred. These results are more accurately explained by high civilian unemployment rates and increased patriotism following September 11, 2001. One of the most notable significant predictors of retention was education. Retention increases with education level until the service member attains a college degree, at which point retention drops significantly (Hansen and MacLeod, 2004). This points to a struggle for military wages to compete with those of the civilian sector. In turn, additional incentives are recommended for college graduates to encourage higher rates of retention from these members.

Herschelman (2012) evaluated predictors of attrition of NPS Marines in the Marine Corps Reserve who enlisted between 1995 and 2005. The population was split into three subsections: pre-9/11, overlap-9/11, and post-9/11 among enlistees with a 6x2 contract. Herschelman utilized a probit regression to examine the change in attrition behavior, if any, after 9/11. There was a six percent decrease in attrition between the pre- and post-9/11

cohorts (Herschelman, 2012). This decrease is partially attributed to deployments to Afghanistan and Iraq (those deployed OCONUS elsewhere are more likely to attrite, as seen in previous literature). This research defined attrition as any member who did not fulfill his/her initial commitment. It does not take into consideration those who were lost to other reserve of active components.

Randall (1989) analyzed influential retention factors for Noncommissioned (NCOs) and Staff Noncommissioned Officers (SNCOs) in the SMCR. Randall divided his population into four groups: PS single, PS married, NPS single, and NPS married. Each of the four groups had unique significant predictors. Married NPS Marines were greatly positively influenced by retirement compensation. Also, minorities had a much larger retention rate. PS married members were more likely to stay in if they liked their job and if the opportunity cost of a similar civilian job was less than what they were receiving through SMCR. Single NPS Marines saw the lowest retention rates across the four groups. Finally, NPS single members were mostly influenced by age and potential retirement benefits (Randall, 1989). The main limitation with this study is that it has a small sample size so an accurate magnitude of these impacts cannot be determined.

Buddin (1984) did an analysis of first term attrition of enlisted males who entered active duty in 1979. Buddin found that work history and civilian unemployment rate the year before the service member enlisted impacts the stay or go decision. Recruits who have a high job turnover history are more likely to attrite before their contract expires. Age was also a significant predictor of early attrition. With all other variables held constant, early attrition was expected to raise one percentage point per year after age 17 at initial enlistment (Buddin, 1984).

THIS PAGE INTENTIONALLY LEFT BLANK

II. DATA AND METHODOLOGY

A. DATA

The data for this study was pulled from Marine Corps Total Force Data Warehouse (TFDW) by Headquarters Marine Corps Reserve Affairs Division. The data was pulled and collated into monthly sequences. The data for this thesis contains 96 sequences, monthly snapshots from fiscal year (October 1, 2008) 2009 through fiscal year 2017. The sequences contain every Reserve Marine in the SelRes during that monthly pull, and are stored in separate comma-separated values (CSV) files. Each Marine is identified by their Electronic Data Interchange Personal Identifier (EDIPI).

B. COHORT INTRODUCTION

For this thesis, the population of interest is all NPS enlisted who enlisted into the SMCR in FY09 on a 6x2 contract. These individuals and their characteristic data are subset and collated into its own CSV file. Upon identifying the EDIPIs of this cohort, another CSV file was created with each row being one month of data for each individual. If an EDIPI is found that matches one in our defined cohort, the corresponding data for that month is pulled and subset into the CSV file. This is done for every unique EDIPI (individual) of the cohort. The resulting data set contains as many rows for each individual as months they appear in the data—corresponding to how long they stayed in the MCR.

Next, we identified which Marines failed to complete the first six years in the SMCR, but continued elsewhere in the MCR. These individuals comprise my initial population of interest and the final dataset used in the analysis was collapsed to one row per individual (EDIPI) with their corresponding final data fields defined below. The objective of this study is to identify predictors of continuation of service in different reserve branches upon leaving the SMCR. Predictors of continuation of service include demographics, military conduct record, deployment record, and personnel records. Finally, a second analysis is done at the eight-year mark with those who completed their first six years in the SMCR, but finish their remaining two years in a reserve branch other than the SMCR.

C. CONSTRUCTED VARIABLES

Below is a list and description of all of the variables adapted for the logistic regression models. Each variable has been extrapolated explicitly from the original dataset and manipulated so that it can be used as a potential predictor of the dependent variable.

1. Dependent Variable

The binary variables called "DODTCPG_6YEARS" and "DODTCPG_8YEARS" were created for what component the Marine was in at the six- and eight-year point: 0-for IRR, IMA, or AR for six-year case (IRR, IMA, AR, Left MCR for eight-year regression) and 1-for SMCR. These variables act as the dependent variables in each of the two models. With the regression results, we hope to be able to predict the probability of continuation into each of these categories.

2. Demographic Variables

(i) Hometown and Unit State

The variable "HOME" is a categorical variable, with a unique level for each state. "UNIT" is also a categorical variable, and the unit state that the Marine is assigned after he/she finished initial training. Finally, the variable "SAME_STATE" is a binary variable that takes value 1 if the Marine's hometown and unit state are the same, and 0 otherwise.

(ii) Hometown and Unit State Regions

Furthermore, the hometown and unit states are grouped into categorical variables with five levels each: 1-Northeast, 2-Midwest, 3-South, 4-West, and 5-Location Unknown.

(iii) Age at enlistment

"ENTRY_AGE" is age at enlistment and a continuous variable, taken from the very first entry for each individual.

(iv) Race

"RACE" is reported in their personnel files. 84.9% of our enlisted cohort is white, while the remaining four races represent only 15.1%. Because of the large disparity between the categories, the four minority groups are grouped together. Therefore, the "RACE" variable is defined as 0-for non-white and 1-for white. Race upon enlistment is used.

(v) Marriage Status

The Marine's entry, six-year, and eight-year marital status are used. These are categorical variables defined as: 0-Single, 1-Married, and 2-Divorced, 3-Other, referenced as "MARITAL_6" and "MARITAL_8".

(vi) Gender

The variable "GENDER" is defined as: 0-for Female and 1-for Male. No other gender identifications are present in this cohort.

3. Military Record

(i) Disciplinary Action

If the Marine faced any disciplinary action within the first six years (does not matter when), the variable "DISCIPLINED_6YEARS" takes a value of 1; if he/she does not, the variable takes the value 0. The same methodology is used for the eight-year benchmark, creating the variable "DISCILPINED_8YEARS".

(ii) PFT/CFT Scores

The max PFT and CFT scores at the six and eight-year mark are recorded and used in each of the logistic regressions, defined as "PFT_6YEARS", "CFT_6YEARS", "PFT_8YEARS" and "CFT_8YEARS".

(iii) MOS Group

The MOS Group at the six and eight-year mark for each individual is utilized for the regressions. These variables are categorical, and have been split into the following levels: 1-Ground, 2-Support, and 3-Aviation, named "MOS_6YEARS" and "MOS_8YEARS".

4. Deployment Record

(i) Combat Deployments

History of combat deployments, defined as "NO_ DEP_COMBAT_6YEARS" and "NO_ DEP_COMBAT_6YEARS", the Marine completes during his/her service time at six and eight years. These variables are binary with 0-for no combat deployments and 1-for 1+ combat deployments.

(ii) Non-combat Deployments

Number of non-combat deployments (no combat operations conducted) the Marine completes at the six and eight-year point, named "DEP_NON _COMBAT_6YEARS" and "DEP_NON_COMBAT_8YEARS". These variables are binary with 0-for no non-combat deployments and 1-for 1+ non-combat deployments.

(iii) Combat Deployment Days

Directly related to number of combat deployments, "NO_DAYS_DEP_COMBAT_6YEARS" and "NO_DAYS_DEP _COMBAT_8YEARS" corresponds to the number of total combat deployment days the Marine completed during service at those respective years.

(iv) Non-combat Deployment Days

Similarly, "NO_DAYS_DEP_NON_COMBAT_6YEARS" and "NO_DAYS_DEP_NON_COMBAT_8YEARS" is defined as the total number of noncombat deployment days the service member completes during each time period.

(v) Deployment Rate

"DEPLOY_RATE_6YEARS" and "DEPLOY_RATE_8YEARS" are the number of total deployment days per year, divided by the number of months the Marine serves at the six and eight-year mark.

5. Education Level

(i) Years

"EDUCATION_6YEARS" and "EDUCATION_8YEARS" are measured in years. We are interested in education status at the six and eight-year mark. Of particular interest is the difference, if any, between high school (9-12 years) and post-high school (>12 years) education on continuation.

(ii) AFQT Score

The AFQT score is a subset score from a portion of the ASVAB, specifically, Word Knowledge, Paragraph Comprehension, Mathematics Knowledge, and Arithmetic Reasoning. In order to enlist in the Marine Corps, one must achieve at least a 32. The variable "AFQT_SCORE" was created using each Marine's individual score from the ASVAB.

(iii) GCT Score

General Classification Test (GCT) is taken by every Marine upon enlistment or commissioning. The "GCT_SCORE" variable that was recorded for each Marine is used in the analysis.

D. STATISTICAL METHODS

1. LOGISTIC REGRESSION

Separate regressions were performed for the six and eight-year mark. The response variables are "COMPONENT_6YEARS" and "COMPONENT_8YEARS", respectively. The regression for the first analysis is a logistic regression, with 1 representing a Marine who is in the SMCR at the six-year mark, and 0 if he/she transferred to another MCR component (IRR, IMA, AR) sometime within the first six years, ignoring those that left the MCR. This model combines a logit link function with a linear predictor. This logit model uses a probability link function for the predictor variables. The predictive analysis estimates the probability of staying in the SMCR in the first six years, based on the independent variables on the scale of the log-odds. The general model is defined as:

$$\log\left(\frac{p}{1-p}\right) = \beta_o + \beta_1 * x_{i1} + \ldots + \beta_m * x_{im} \; \forall \; i,$$

where *m* is the number of explanatory variables, which in this case is 19. Furthermore, the fitted probabilities of staying in the SMCR are:

$$\hat{p} = \frac{e^{\beta_o + \beta_1 * x_{i1} + \ldots + \beta_m * x_{im}}}{1 + e^{\beta_o + \beta_1 * x_{i1} + \ldots + \beta_m * x_{im}}} \; \forall \; i$$

For example, $\hat{p}$ is the fitted probability that the Marine left the SMCR within the first six years and $1-\hat{p}$ is the probability the Marine transferred to another MCR component given their predictors $x_1, \ldots, x_m$. The x's represent the values of each independent variable (taken directly from the data). The $\beta's$ are the estimated coefficients of the explanatory variables from the regression results (Faraway, 2016, pp. 24-28).

The second logistic regression for the eight-year mark is formed in a similar fashion, with a few exceptions. First, the population is now all the responses that took value 1 for the first regression (everyone in SMCR at six-year mark). This binary response variable takes value 0 includes Marines that transferred to another MCR component and those who separated altogether within the six to eight-year period. The response takes value 1 for those in the SMCR at the eight-year mark. All analyses were performed using the statistical package R (see https://www.r-project.org/).

For both logistic regression models, we implement purposeful selection of covariates. First, a univariate analysis is created for each potential independent variable. All predictors with a significance level less than 0.10 are included in the initial model. Then, using backward stepwise regression, the most insignificant variables are dropped one-by-one until we have an acceptable model. However, if the dropping of a variable changes any of the remaining parameter estimates by more than 20%, it is identified as a confounder and returned back to the model. Confounders are variables that while themselves are not significant, their combined effect with another variable on the outcome is significant. Ultimately, we finish with a final model with all significant covariates and confounders (Bursac, 2008).

2. CLASSIFICATION TREE

A classification tree is another statistical method used to analyze a binary response variable. The node splits "divide the observations within a node so that the class types within a split are mostly of one kind" (Faraway, 2016). In other words, the nodes are split with predictors that best classify the outcome (whether or not a Marine is in the SMCR). The partitions are based on the minimized cross-validated error (Faraway, 2016).

THIS PAGE INTENTIONALLY LEFT BLANK

III. DESCRIPTIVE STATISTICS

A. COHORT PROFILE

The enlisted cohort of NPS Marines on a 6x2 contract who enlisted in FY09 is 4,413 Marines. Figure 3 depicts the totals for each reserve component at the six-year mark, the last year of obligation to the SMCR. 76.1% of the cohort completed the full six years in the SMCR.

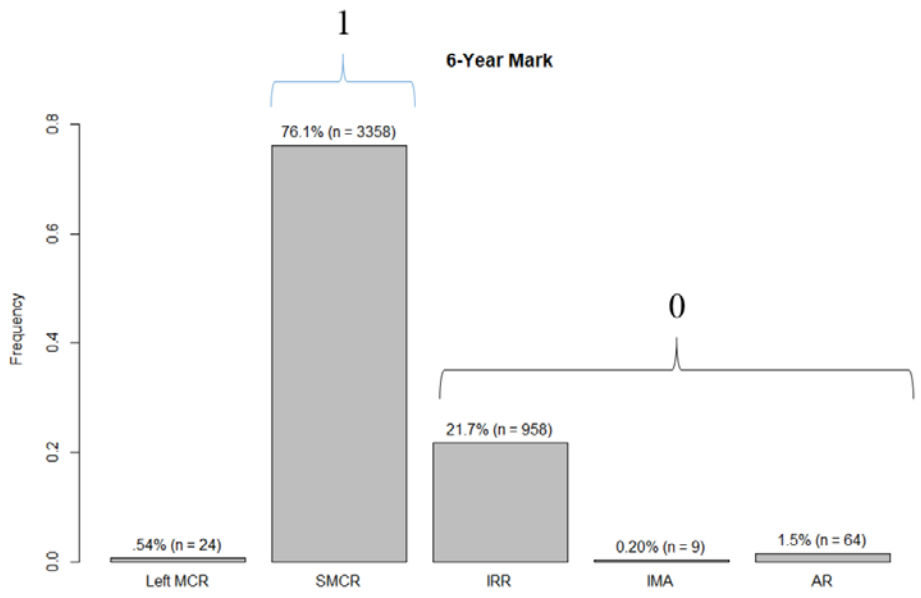

Figure 3. Six-Year Mark Component Totals.

For the purposes of this analysis I am categorizing those who transferred into the IRR, IMA, or AR into one group (i.e., other components within MCR) and those who remain in the SMCR as the comparison group. This leaves 1,031 Marines who did not complete their first six years in the SMCR, but did not leave the MCR as a whole. Over 90% of these transfers went to the IRR, shown in Figure 4. For the first logistic regression, these three components are grouped together in the response variable.

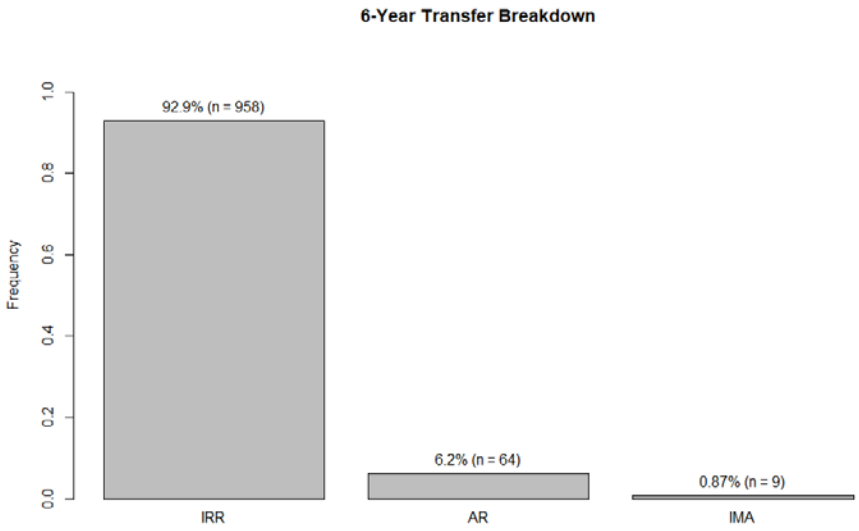

Figure 4. Six-Year Continuation Breakdown.

Figure 5 shows the continuation at the eight-year mark, the end of the cohort's MSO. The 6x2 contract allows the final two years of the enlistment to be spent in the SMCR or IRR. Of the 3,358 Marines in the SMCR at six years, only 16 Marines continued on to finish their MSO in the AR or IMA. The rest finished in the IRR or SMCR, or left the MCR altogether.

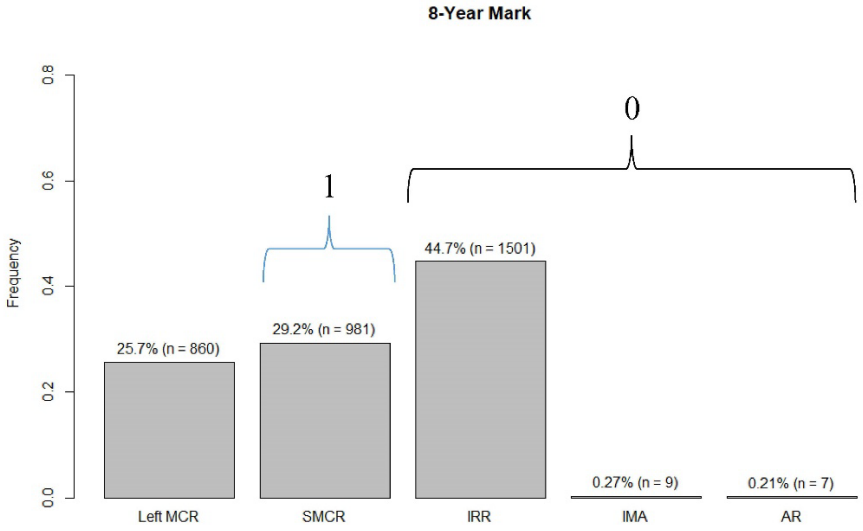

Figure 5. Eight-Year Mark Component Totals.

Overall, 56.2% of the 4,413 Marines that enlisted in FY09 on a 6x2 contract were retained throughout the entirety of their eight-year MSO. Seven transferred to the AR and nine to the IMA. The second logistic regression groups everyone not in the SMCR as 0, and those in the SMCR at the eight-year point as 1.

B. DESCRIPTIVE STATISTICS

For each of the variables used in each of the regressions, I summarize their descriptive statistics in the following tables. Table 1 describes the 4,389 Marines that are in any of the four components of the MCR at the six-year mark. Note, I initially started with 4,413 personnel, but 24 attrited out of the MCR within the first six years. All percentages represented are column percentages and may not add up to 100 due to rounding. A simple logistic regression model was run for each individual independent variable against the outcome. For each, the *p-value* is recorded in the last column of the table. Those *p-values* less than 0.10 are annotated by an asterisk as they will be used in the initial full model, utilizing purposeful covariance selection (Bursac, 2008).

Table 1. Six-Year Descriptive Statistics.

	IRR n=958 (%)	IMA n=9 (%)	AR n=64 (%)	SMCR n=3,358 (%)	Total n=4,389 (%)	P-value
	0			1		
Characteristic						
Gender:						
Female	26 (2.7)	2 (22.2)	4 (6.2)	91 (2.7)	123 (2.8)	Reference
Male	932 (97.3)	7 (77.8)	60 (93.8)	3267 (97.3)	4266 (97.2)	0.577
Race:						
Non-White	104 (10.9)	1 (11.1)	19 (29.7)	535 (15.9)	659 (15.0)	Reference
White	854 (89.1)	8 (88.9)	45 (70.3)	2823 (84.1)	3730 (85.0)	0.003*
Age at enlistment:						
Mean	19.5	20.2	19.9	19.7	19.7	
SD	2.2	3.3	2.6	2.2	2.2	0.185

	IRR n=958 (%)	IMA n=9 (%)	AR n=64 (%)	SMCR n=3,358 (%)	Total n=4,389 (%)	P-value
	0			1		
Characteristic						
Marital Status:						
Single	732 (76.4)	4 (44.4)	28 (43.8)	2638 (78.6)	3402 (77.5)	Reference
Married	212 (22.1)	5 (55.6)	34 (53.1)	666 (19.8)	917 (20.9)	0.003*
Divorced	14 (1.5)	0 (0.0)	2 (3.1)	52 (1.5)	68 (1.5)	0.848
Other	0 (0.0)	0 (0.0)	0 (0.0)	2 (0.1)	2 (0.1)	0.961
Number of Dependents:						
Mean	0.34	0.8	1.2	0.4	0.4	
SD	0.8	1.3	1.3	0.8	0.8	0.517
Unit Region						
Northeast	176 (18.4)	3 (33.3)	12 (18.7)	631 (18.8)	822 (18.7)	Reference
Midwest	224 (23.4)	1 (11.1)	16 (25.0)	893 (26.6)	1134 (25.8)	0.263
South	381 (39.8)	3 (33.3)	25 (39.1)	1124 (33.5)	1533 (34.9)	0.061*
West	177 (18.5)	2 (22.2)	11 (17.2)	699 (20.8)	889 (20.3)	0.308
Unknown	0 (0.0)	0 (0.0)	0 (0.0)	11 (0.3)	11 (0.3)	0.956
Hometown Region:						
Northeast	174 (18.2)	3 (33.3)	12 (18.8)	605 (18.0)	794 (18.1)	Reference
Midwest	185 (19.3)	1 (11.1)	6 (9.4)	658 (19.6)	850 (19.4)	0.513
South	367 (38.3)	3 (33.3)	29 (45.3)	1102 (32.8)	1501 (34.2)	0.159
West	182 (19.0)	2 (22.2)	13 (20.3)	758 (22.6)	955 (21.8)	0.085*
Unknown	50 (5.2)	0 (0.0)	4 (6.3)	235 (7.0)	289 (6.6)	0.082*
Same State (Unit & Hometown)						
Yes	250 (26.1)	1 (11.1)	19 (29.7)	917 (27.3)	1182 (27.0)	Reference
No	708 (73.9)	8 (88.9)	45 (70.3)	2441 (72.7)	3202 (73.0)	0.627
Education Level:						
High School	818 (85.4)	8 (88.9)	54 (84.4)	3066 (91.3)	3946 (89.9)	Reference
At least some college	140 (14.6)	1 (11.1)	10 (15.6)	292 (8.7)	443 (10.1)	<0.001*
AQFT Score						
Mean	67.9	77.1	64.8	63.3	68.3	<0.001*
SD	19.0	11.8	16.8	19.2	16.7	
GCT Score:						
Mean	112.6	117.8	108.4	109.9	112.2	
SD	12.2	9.7	11.8	13.0	11.7	<0.001*

	IRR n=958 (%)	IMA n=9 (%)	AR n=64 (%)	SMCR n=3,358 (%)	Total n=4,389 (%)	P-value
	0			1		
Characteristic						
MOS Group:						
1	116 (12.1)	1 (11.1)	0 (0.0)	437 (13.0)	554 (12.6)	Reference
2	640 (66.8)	6 (66.7)	48 (75.0)	2213 (65.9)	2907 (66.2)	0.147
3	115 (12.0)	1 (11.1)	15 (23.4)	466 (13.9)	597 (13.6)	0.752
Other	87 (9.1)	1 (11.1)	1 (1.6)	242 (7.2)	331 (7.5)	0.043*
CFT:						
Mean	285.3	290.8	294.5	278.3	287.2	
SD	16.1	6.8	6.9	37.1	12.4	<0.001*
PFT:						
Mean	255.8	269.4	273.6	252.5	262.8	
SD	29.2	24.9	21.4	29.0	26.1	<0.001*
Combat Deployment History:						
No	646 (67.4)	3 (33.3)	48 (75.0)	2527 (75.3)	3224 (73.4)	Reference
Yes	312 (32.6)	6 (66.7)	16 (25.0)	831 (24.7)	1165 (26.5)	<0.001*
Non-combat Deployment History:						
No	592 (61.8)	1 (11.1)	46 (71.9)	2374 (70.7)	3013 (68.6)	Reference
Yes	366 (38.2)	8 (88.9)	18 (28.1)	984 (29.3)	1376 (31.4)	<0.001*
Days Deployed Combat:						
Mean	62.4	123.3	45.6	46.8	69.5	
SD	97.4	101.7	87.8	88.9	94.0	<0.001*
Non-combat:						
Mean	131.7	310.8	100.5	103.8	161.7	
SD	174.2	131.6	164.5	169.1	159.9	<0.001*
Deployment Rate:						
Mean	0.14	0.30	0.11	0.11	0.17	
SD	0.21	0.23	0.20	0.20	0.21	<0.001*
Disciplinary Action:						
No	926 (96.7)	0 (0.0)	59 (92.2)	3156 (94.0)	4150 (94.6)	Reference
Yes	32 (3.3)	9 (100.0)	5 (7.8)	202 (6.0)	239 (5.4)	0.0029*

	IRR n=958 (%)	IMA n=9 (%)	AR n=64 (%)	SMCR n=3,358 (%)	Total n=4,389 (%)	P-value
	0			1		
Characteristic						
Commissioned						
No	927 (96.8)	9 (100.0)	62 (96.9)	3335 (99.3)	4333 (98.7)	Reference
Yes	31 (3.2)	0 (0.0)	2 (3.1)	23 (0.7)	56 (1.3)	<0.001*
Proficiency Average Service:						
Mean	41.6	42.2	42.5	41.1	41.9	
SD	1.6	0.7	0.9	1.7	1.2	<0.001*
Conduct Average Service:						
Mean	41.6	42.2	42.5	41.0	41.8	
SD	1.6	0.8	0.9	1.7	1.3	<0.001*
Rank:						
E-1	1 (0.1)	0 (0.0)	0 (0.0)	72 (2.1)	73 (1.7)	Reference
E-2	1 (0.1)	0 (0.0)	0 (0.0)	207 (6.2)	208 (4.7)	0.433
E-3	130 (13.6)	0 (0.0)	4 (6.3)	1206 (35.9)	1340 (30.5)	0.059*
E-4	614 (64.1)	1 (11.1)	52 (81.3)	1224 (36.5)	1891 (43.1)	<0.001*
E-5	185 (19.3)	6 (66.7)	8 (12.5)	631 (18.8)	830 (18.9)	0.004*
E-6	0 (0.0)	2 (22.2)	0 (0.0)	3 (0.1)	5 (0.1)	0.007*
O-1	14 (1.5)	0 (0.0)	0 (0.0)	8 (0.2)	22 (0.5)	<0.001*
O-2	12 (1.3)	0 (0.0)	0 (0.0)	6 (0.2)	18 (0.4)	<0.001*
O-3	1 (0.1)	0 (0.0)	0 (0.0)	1 (0.02)	2 (0.05)	0.019*
Note: Variables considered candidates for multivariate models have *p-values* less than 0.10 and are denoted with *						

There are some notable differences among the components, specifically in terms of race, marital status, and PFT scores. For race, among those who transfer to the AR almost 30% are minority, non-White. This is double the overall percentage, 15%, of non-White Marines in the sample. The minority ethnicity (non-White) represented in the IRR and SMCR at the six-year mark is 11% and 16%, respectively. For marital status at six years, overall 77.5% of the population is single and 20.9% married. Those who remained in the SMCR or transferred to the IRR saw similar proportions—78.6% and 76.4% single and 19.8% and 22.1% married, respectively. The AR and IMA's numbers are much more balanced—43.8% of the Marines who moved to the AR were single at the time and 53.1%

married. In the IMA, the breakdown was 44.4% single, 55.6% married. Finally, there are some significant differences among the mean PFT scores of the components (*p-value* < 0.001). The average score for the Marines who remained in the SMCR or transferred to the IRR was 252.5 and 255.8, respectively. Those who moved to the AR had a max average score almost 20 points higher of 273.6. Similarly, Marines in the IMA at the six-year mark had a max average PFT score of 269.4.

In Table 2, I describe the characteristics of 3,358 Marines (who were in the SMCR at the six-year mark) at the eight-year point. Between six and eight years, 860 Marines left the MCR altogether, leaving 2,498 Marines in the MCR at the end of the eight-year MSO.

Table 2. Eight-Year Descriptive Statistics Table.

	IRR n=1,501 (%)	**IMA n=9 (%)**	**AR n=7 (%)**	**SMCR n=981 (%)**	**Total n=2,498 (%)**	**P-value**
	0			**1**		
Characteristic						
Gender:						
Female	31 (2.1)	1 (11.1)	1 (14.3)	25 (2.5)	58 (2.3)	Reference
Male	1470 (97.9)	8 (88.9)	6 (85.7)	956 (97.5)	2440 (97.7)	0.465
Race:						
Non-White	224 (14.9)	2 (22.2)	0 (0.0)	182 (18.6)	408 (16.3)	Reference
White	1277 (85.1)	7 (77.8)	7 (100.0)	799 (81.4)	2090 (83.7)	0.002*
Age at enlistment:						
Mean	19.7	19.8	19	19.8	19.8	
SD	2.3	2.1	1.1	2.3	2.3	0.160
Marital Status:						
Single	1103 (73.5)	5 (55.6)	6 (85.7)	718 (73.2)	1832 (73.3)	Reference
Married	363 (24.2)	4 (44.4)	1 (14.3)	240 (24.5)	608 (24.3)	0.833
Divorced	35 (2.3)	0 (0.0)	0 (0.0)	22 (2.2)	57 (2.3)	0.968
Other	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.1)	1 (0.1)	0.951
Number of Dependents:						
Mean	0.5	0.9	1.0	0.6	0.5	
SD	0.9	1.3	0.8	1.0	0.9	0.032*

	IRR n=1,501 (%)	IMA n=9 (%)	AR n=7 (%)	SMCR n=981 (%)	Total n=2,498 (%)	P-value
	0			1		
Characteristic						
Unit Region						
Northeast	270 (18.0)	1 (11.1)	2 (28.6)	186 (19.0)	459 (18.4)	Reference
Midwest	430 (28.6)	4 (44.4)	1 (14.3)	239 (24.4)	674 (27.0)	0.101
South	462 (30.8)	3 (33.3)	4 (57.1)	374 (38.1)	843 (33.7)	0.170
West	331 (22.1)	1 (11.1)	0 (0.0)	180 (18.3)	512 (20.5)	0.095*
Unknown	8 (0.5)	0 (0.0)	0 (0.0)	2 (0.2)	10 (0.4)	0.226
Hometown Region:						
Northeast	253 (16.9)	1 (11.1)	2 (28.6)	186 (19.0)	442 (17.7)	Reference
Midwest	317 (21.1)	1 (11.1)	0 (0.0)	173 (17.6)	491 (19.7)	0.044*
South	452 (30.1)	5 (55.6)	4 (57.1)	357 (36.4)	818 (32.7)	0.411
West	368 (24.5)	2 (22.2)	0 (0.0)	196 (20.0)	566 (22.7)	0.025*
Unknown	111 (7.4)	0 (0.0)	1 (14.3)	69 (7.0)	181 (7.2)	0.266
Same State (Unit & Hometown)						
No	406 (27.0)	5 (55.6)	2 (28.6)	298 (30.4)	711 (28.5)	Reference
Yes	1095 (73.0)	4 (44.4)	5 (71.4)	683 (69.6)	1787 (71.5)	0.164
Education Level:						
High School (0-12)	1338 (89.1)	6 (66.7)	5 (71.4)	883 (90.0)	2232 (89.4)	Reference
College (>12)	163 (10.9)	3 (33.3)	2 (28.6)	98 (10.0)	266 (10.6)	0.506
AQFT Score						
Mean	65.3	79.9	55.6	62.1	65.7	
SD	19.3	18.1	17.3	19.5	18.6	<0.001*
GCT Score:						
Mean	111.2	121.2	105.4	109.2	111.8	
SD	12.9	16.9	10.5	13.3	13.4	<0.001*
MOS Group:						
1	188 (12.5)	0 (0.0)	0 (0.0)	123 (12.5)	311 (12.4)	Reference
2	968 (64.5)	7 (77.8)	6 (85.7)	635 (64.7)	1616 (64.7)	0.889
3	226 (15.1)	2 (22.2)	0 (0.0)	133 (13.6)	361 (14.5)	0.578
Other	119 (7.9)	0 (0.0)	1 (14.3)	90 (9.2)	210 (8.4)	0.421
CFT:						
Mean	285.8	292	295.6	278.7	288.0	
SD	16.1	6.9	6.9	40.0	14.4	<0.001*
PFT:						
Mean	256.3	263.6	272.7	252.8	261.4	
SD	27.1	23.6	16.4	28.7	24.0	0.042*

	IRR n=1,501 (%)	IMA n=9 (%)	AR n=7 (%)	SMCR n=981 (%)	Total n=2,498 (%)	P-value
	0			1		
Characteristic						
Combat Deployment History:						
No	1057 (70.4)	6 (66.7)	5 (71.4)	750 (76.5)	1818 (72.8)	Reference
Yes	444 (29.6)	3 (33.3)	2 (28.6)	231 (23.5)	680 (27.2)	0.005*
Non-combat Deployment History:						
No	984 (65.6)	5 (55.6)	5 (71.4)	703 (71.7)	1697 (67.9)	Reference
Yes	517 (34.4)	4 (44.4)	2 (28.6)	278 (28.3)	801 (32.1)	0.008*
# Days Deployed Combat:						
Mean	58.2	77.1	26.4	43.6	51.3	
SD	96.3	115.8	48.4	87.6	87.0	<0.001*
Non-combat:						
Mean	122.1	169.8	106.1	101.9	125.0	
SD	173.5	201.4	182.0	172.6	182.4	0.02*
Deployment Rate:						
Mean	0.5	0.5	0.4	0.5	0.4	
SD	0.1	0.02	0.04	0.1	0.1	0.008*
Disciplinary Action:						
No	1428 (95.1)	9 (100.0)	7 (100.0)	903 (92.0)	2347 (94.0)	Reference
Yes	73 (4.9)	0 (0.0)	0 (0.0)	78 (8.0)	151 (6.0)	<0.001*
Commissioned						
No	1499 (99.9)	9 (100.0)	6 (85.7)	961 (98.0)	2475 (99.1)	Reference
Yes	2 (0.1)	0 (0.0)	1 (14.3)	20 (2.0)	23 (0.9)	<0.001*
Proficiency Average Service:						
Mean	42.2	42.7	42.7	41.3	42.2	
SD	0.9	0.8	0.7	2.1	1.1	<0.001*
Conduct Average Service:						
Mean	42.1	42.7	42.6	41.2	42.2	
SD	1.0	0.8	0.8	2.1	1.2	<0.001*

	IRR n=1,501 (%)	IMA n=9 (%)	AR n=7 (%)	SMCR n=981 (%)	Total n=2,498 (%)	P-value
	0			1		
Characteristic						
Rank:						
E-1	0 (0.0)	0 (0.0)	0 (0.0)	34 (3.5)	34 (1.4)	Reference
E-2	2 (0.1)	0 (0.0)	0 (0.0)	80 (8.2)	82 (3.3)	0.966
E-3	138 (9.2)	0 (0.0)	0 (0.0)	409 (41.7)	547 (21.9)	0.959
E-4	726 (48.4)	1 (11.1)	6 (85.7)	100 (10.2)	833 (33.3)	0.951
E-5	614 (40.9)	7 (77.8)	1 (14.3)	326 (33.2)	948 (40.0)	0.955
E-6	19 (1.3)	1 (11.1)	0 (0.0)	12 (1.2)	32 (1.3)	0.955
O-1	2 (0.1)	0 (0.0)	0 (0.0)	7 (0.7)	9 (0.4)	0.981
O-2	0 (0.0)	0 (0.0)	0 (0.0)	10 (1.0)	10 (0.4)	1.0
O-3	0 (0.0)	0 (0.0)	0 (0.0)	3 (0.3)	3 (0.1)	1.0

The variables described in the above two tables are utilized in each of the two logistic regression models, respectively. Due to small sample sizes, evaluations of differences between the components using a multinomial regression could not be performed. Furthermore, there are no obvious distinguishable characteristics between the different components in terms of the descriptive statistics.

IV. ANALYSIS AND MAIN RESULTS

A. SIX-YEAR LOGISTIC REGRESSION RESULTS

The first logistic regression predicts the probability of finishing the first six years of the MSO in the SMCR, on the log-odds scale. There are 75 observations that are dropped due to missing values in the data, all regarding the max CFT score. This is because the CFT was not implemented until mid-2008, so not everyone had been required to take it yet. Before constructing a model, the correlations between the numerical predictors were calculated (Figure 6). Correlations between a variable and itself (represented on the diagonal) are always 1, with the darker the color, the higher the correlation.

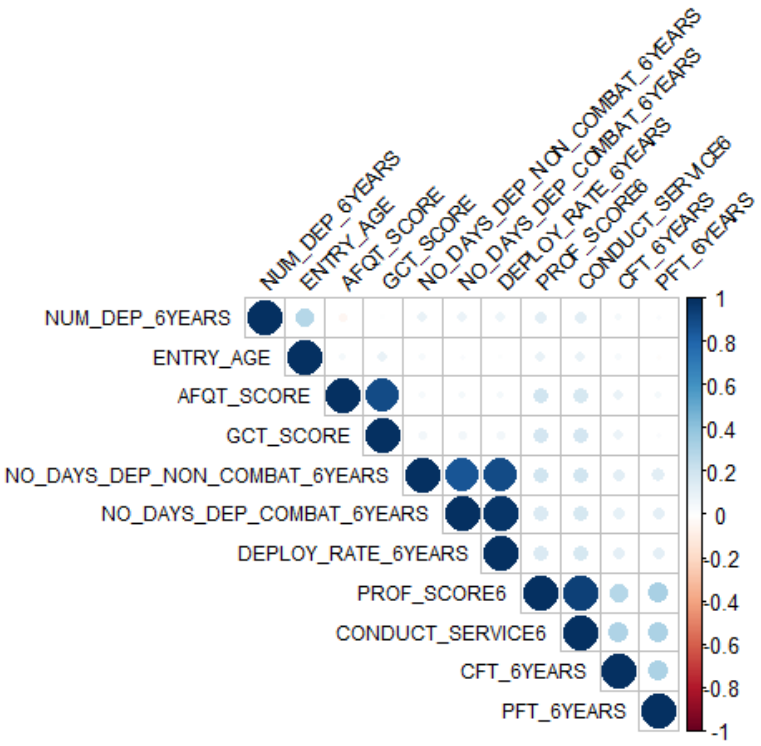

Figure 6. Six-Year Mark Correlation Diagram for Numerical Variables.

There is a high correlation (>0.8) between GCT and AFQT scores, number of days deployed combat and non-combat, number of days deployed non-combat and deployment rate, number of days deployed combat and deployment rate, and proficiency and conduct scores. All other correlations are minor. This indicates dependency between the variables.

Therefore, the AFQT Score, number of days deployed combat, deployment rate, and average proficiency score variables are excluded from the model.

Using Table 1, the initial full model is constructed with all variables with a *p-value* less than 0.1 (without the correlated variables aforementioned) when regressed individually against the response variable (Hosmer and Lemeshow, 2000, pp. 91-128). Then, employing backward stepwise regression, the most statistically nonsignificant results are dropped one by one until an acceptable model is found. If dropping a nonsignificant variable subsequently changes the coefficients of the remaining variables by at least 20% (Zhang, 2016a), said variable is identified as a confounder. Confounders are kept in the model because while they themselves might be insignificant, they are interacting with other independent variables in such a way that affects the outcome. Table 3 depicts the final model results, including coefficient estimates, 95% confidence intervals (CIs), and whether or not they are confounding. The coefficients shown below are exponentiated; thus, representing the odds ratios and their respective 95% CI.

Table 3. Six-Year Logistic Regression Results.

Variable	Odds Ratio (95% CI)	P-value	Confounder
Ethnicity:			
Non-White	Reference	Reference	
White	0.8 (0.6,1.0)	0.039	
Hometown Region:			
Northeast	Reference	Reference	
Midwest	1.1 (0.7,1.8)	0.765	
South	1.1 (0.7,1.8)	0.627	
West	2.0 (1.1,3.7)	0.026	
Unknown	1.5 (1.0,2.3)	0.071	
Unit Region:			
Northeast	Reference	Reference	
Midwest	1.1 (0.7,1.8)	0.556	Yes
South	0.7 (0.5,1.2)	0.180	
West	0.6 (0.3,1.2)	0.145	
Location Unknown	638,539.2 (0.0, NA)	0.959	
GCT Score	1.0 (1.0,1.0)	0.042	
Combat Deployment History:	Reference	Reference 0.012	

Variable	Odds Ratio (95% CI)	P-value	Confounder
No Yes	0.5 (0.3,0.9)		
# Days Deployed: Non-Combat	1.0 (1.0,1.0)	0.011	
Commissioned: No Yes	Reference 0.4 (0.1,1.3)	Reference 0.123	Yes
Conduct Average Service	0.9 (0.8,0.9)	<0.001	
Rank: E-1 E-2 E-3 E-4 E-5 E-6 O-1 O-2 O-3	 Reference 4.6 (0.2,118.9) 0.3 (0.0,1.2) 0.1 (0.0,0.3) 0.1 (0.0,0.6) 0.1 (0.0,1.0) 0.1 (0.0,0.5) 0.1 (0.0,0.4) 0.1 (0.0,6.1)	 Reference 0.282 0.192 0.007 0.043 0.062 0.020 0.017 0.288	

1. Coefficient Interpretations

Two of the variables are confounders: unit region and commissioning status. Removing them from the regression changes the coefficients of at least one of the other variables by at least 20%. Unit region is confounded with home region. Commissioning status is confounded with rank at the six-year mark. This makes sense because if the Marine is commissioned he/she will be an O-1, O-2, or O-3. The other variables are all statistically significant. The confounders themselves have no significance on the outcome.

The influence of each of the predictor variables on the outcome (retained in the SMCR through six years, rather than transferring to another component) is explained in this section.

A White Marine has 0.8 times the odds of being retained in the SMCR through six years (instead of transferring to another component) relative to non-White Marines (*p-value* = 0.039). Service members whose home of record is in the West US, have two times

the odds of retention, relative to those who live in the Northeast US (95% CI = 1.1-3.7). Living in any other regions are not statistically different than the Northeast.

The increase of each additional point in GCT score results in equal odds of retention, with a *p-value* of 0.042. This means that there is evidence supporting that performance on the GCT does not affect retention. Having a non-combat deployment history decreases the odds of retention in half (0.5; 95% CI of (0.3, 0.9)). However, the amount of days spent in non-combat deployments over the first six years neither increases nor decreases the odds of retention (*p-value* = 0.011). Next, each point increase in the Marine's Conduct Service Record decreases odds of retention by 10% (95% CI of (0.8, 0.9)). Finally, compared to E-1s, the only ranks that are held at the six-year mark that affect retention are E-4, E-5, E-6, O-1, and O-2. A Marine whose rank is any of these has 0.1 times the odds of being in the SMCR at six years, relative to E-1, with *p-values* of 0.007, 0.043, 0.062, 0.020, and 0.017, respectively.

2. Diagnostics and Checking Model Assumptions

(1) Linearity

Scatter plots were created for the continuous (numerical) data to check the linearity assumption of the model. The plots are shown in Figure 7.

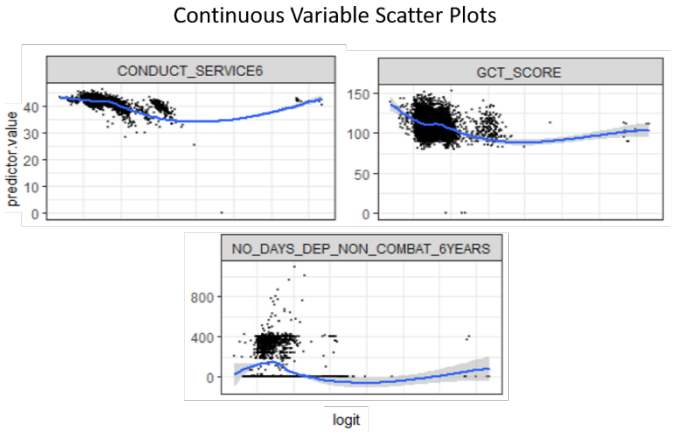

Figure 7. Six-Year Scatterplots.

Average conduct service record score and GCT score look to have a quadratic relationship with the response variable. Number of days of non-combat deployment days has a linear relationship, but is pulled away because of a few influential observations. The linearity assumptions are variable. Therefore, a quadratic term was put in the model for conduct service and GCT score and tested for statistical significance. The ANOVA produced a *p-value* < 0.001, meaning there is a statistical difference between the models. However, the addition of these terms made every other term in the model statistically nonsignificant and decreased the pseudo R^2 value (explained further in the Model Fit section). Therefore, the added terms are subsequently dropped, returning to the original model, the lower order terms retained.

(2) Checking for Interactions

Interaction effects were tested with home and unit region as well as commissioned and rank at six years. An ANOVA test found that the first interaction is not statistically significant at the 0.05 level (*p-value* = 0.1399. The interaction term of commissioned and rank was, however, statistically significant (*p-value* = 0.027). However, six of the interaction terms cannot be defined because of singularities. This means that these variables can be expressed as a linear combinations of other predictor variables. Therefore, the interaction is not included.

(3) Outliers, Leverage, and Influence Points

Next, the model was checked for outlier, leverage, and influence points. Outliers have a response value "unusual conditional on covariate pattern" (Zhang, 2016b). Leverage points have a predictor pattern very different than the average, and an influential observation is the product of outlier and leverage. Each of these types of observations can have large effects on model fitting (Zhang, 2016b).

Figure 8 displays the plot testing for each of these observations for the six-year model. Observation 1130 is a significant outlier—over four standard deviations away from the mean. This is because this observation seems to be a data entry error. There are no

proficiency, conduct, CFT, or PFT scores recorded throughout the entire dataset for this individual. Also, rank never exceeds E-1. That is, he/she did not get promoted. This Marine likely dropped but was not removed from the system. Leverage points are measured by their hat-values, displayed on the *x-axis*. The size of the circle corresponds to that observation's Cook distance (measure of influence). Observations 1894 and 1882 have high influences and high leverage. These observations are flagged because rank at the six-year mark for these two service-members is O-3. These are the only two Marines that have commissioned and earned O-3 by the six-year mark. These two observations are also removed, and the model is rerun to check the impact on the coefficients and overall model.

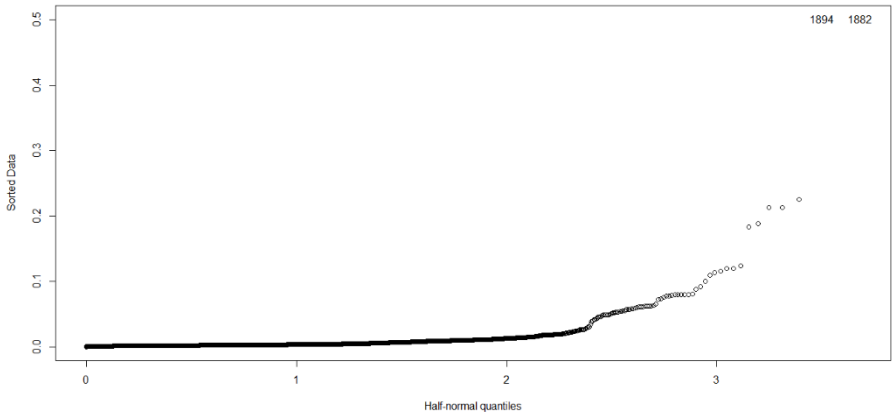

Figure 8. Leverage and Influence Test: Six-Year Model.

The model is re-examined for the impact of these potential outliers and influential points. The exponentiated coefficients are compared in Table 4. None of the coefficients change at a statistically significant level. E-1's estimate has a large change; however, the variable remains insignificant (*p-value* > 0.05). Therefore, the outlier and potential influential variables do not greatly influence the model fit. The coefficient for O-3 is NA because there are no more O-3 observations in the data.

Table 4. Six-Year Model Coefficient Comparison.

	Coefficients	
	Original	**New**
Ethnicity:		
No	Reference	Reference
Yes	0.8	0.8
Hometown Region:		
Northeast	Reference	Reference
Midwest	1.1	1.1
South	1.1	1.1
West	2.0	2.0
Location Unknown	1.5	1.5
Unit Region:		
Northeast	Reference	Reference
Midwest	1.1	1.1
South	0.7	0.7
West	0.6	0.6
Location Unknown	638539.2	4830483
GCT Score	1.0	1.0
Non-combat Deployment History:		
No	Reference	Reference
Yes	0.5	0.5
Days Deployed Non-combat	1.0	1.0
Commissioned:		
No	Reference	Reference
Yes	0.4	0.4
Conduct Service	0.9	0.8
Rank:		
E-1	Reference	Reference
E-2	4.6	0.0
E-3	0.3	0.0
E-4	0.1	0.0
E-5	0.1	0.0
E-6	0.1	0.0
O-1	0.1	0.0
O-2	0.0	0.0
O-3	0.1	NA

3. Model Fit

(1) ROC Curve

Finally, a receiving operating characteristic curve is constructed. The ROC curve assesses the overall discrimination power of the model (Zhang, 2016a). The ROC curve for the six-year model is displayed in Figure 9. An area under the curve of 0.9-1.0 is considered exceptional. 0.8-0.9 excellent, 0.7-0.8 acceptable, 0.6-0.7 poor and 0.5 means no discrimination (Mandrekar, 2010). The area under the curve for this model is 0.72.

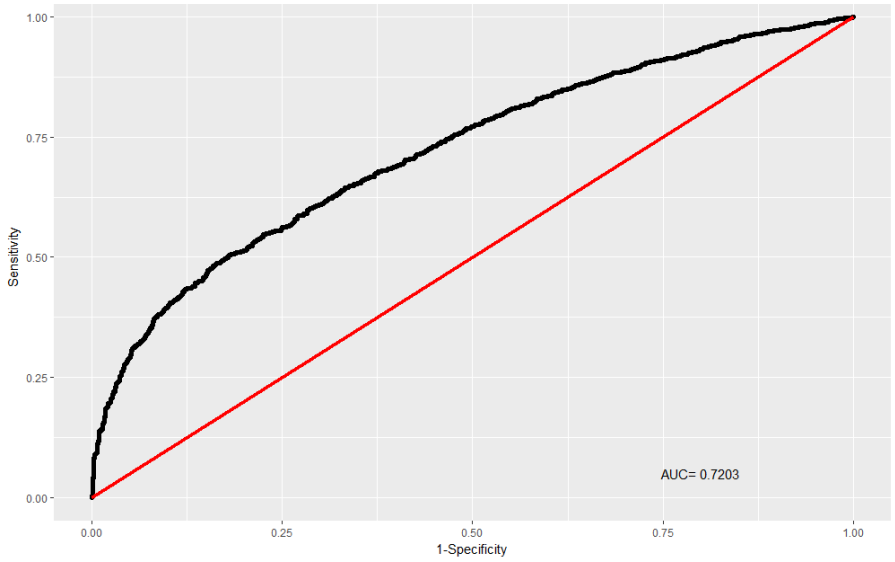

Figure 9. Six-Year ROC Curve.

Next, McFadden's pseudo R^2 value was calculated. The amount of variability that can be explained by the model is 11%. While small, there is still significance in this model.

B. SIX-YEAR CLASSIFICATION TREE

Classification trees are another method to model binary response data. The data are split into groups, rather than given a coefficient estimate. Each node (branch) represents the likelihood of a certain outcome given a particular attribute (one of the predictor variables) (Faraway, 2016, pp. 354-358). The initial tree is fit with the same predictor variables as the initial logistic regression. It is then pruned based on the cross-validated

error. Figures 10 and 11 depict the classification tree for Marines in the SMCR versus another MCR component at the six-year mark. The figures are split for readability.

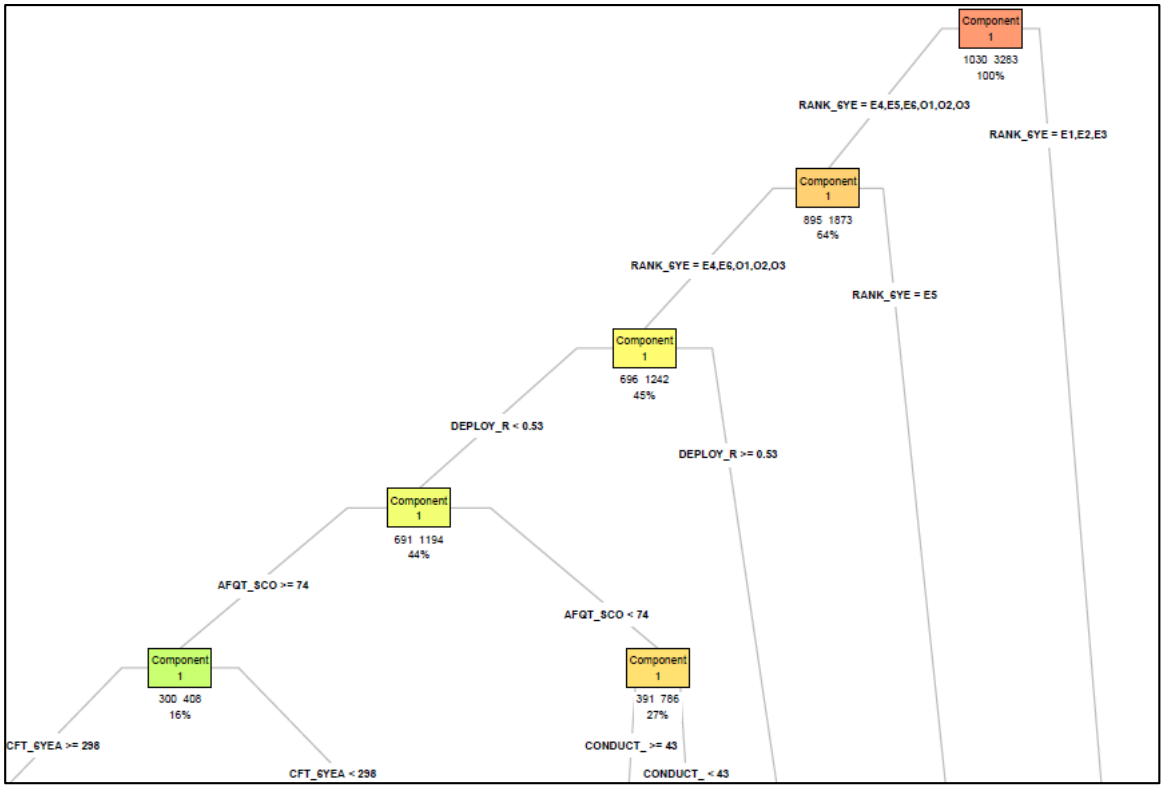

Figure 10. Six-Year Classification Tree: Top Half.

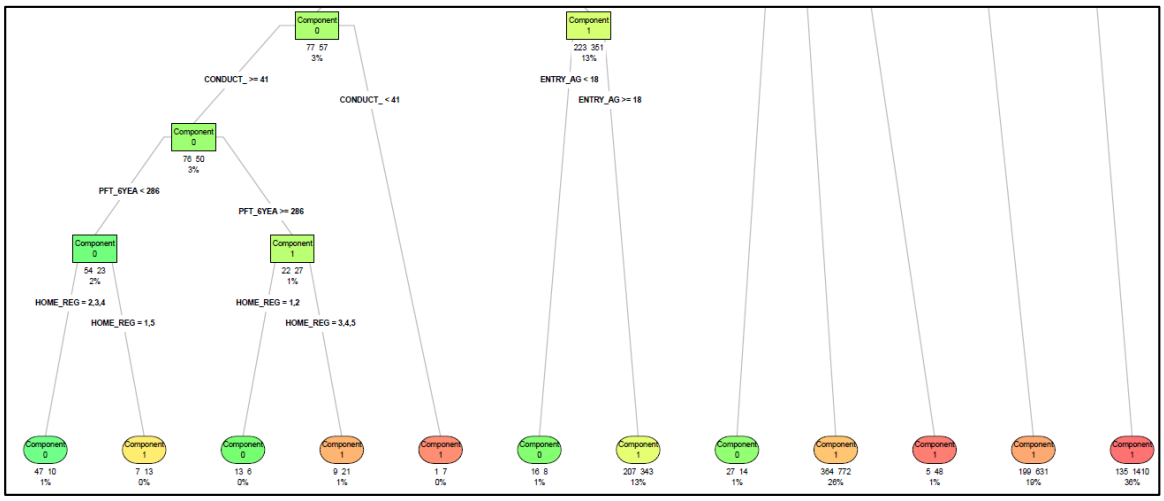

Figure 11. Six-Year Classification Tree: Bottom Half.

The final classification tree has 11 node splits and a cross-validation error of 1.0048. The first predictor (best classifier) to split the tree is the Marine's rank at six years. 64% were E-4, E-5, E-6, O-1, O-2, and O-3s and were still in the SMCR at the six-year mark. Moving down the tree, 44% of the population that were E-4, E-6, O-1, O-2, and O-3s that had a deployment rate less than 0.53 over the first six years were retained in the SMCR. Each node split is read in similar fashion.

While some of the same variables are found to be significant or important between the logistic regression and classification tree, many differed. The logistic regression found more significance in demographic (ethnicity, location, and education) and military service history (combat deployment history and days of non-combat deployments) variables. Meanwhile, the partition tree found that performance predictors, such as PFT and CFT scores better classified the outcome, shown in Table 5. The classification tree produced a simpler model-splitting using eight variables. The final logistic regression model had nine.

Table 5. Six-Year Model Comparison.

Variable	Logistic Regression	Classification Tree
Ethnicity	X	
Unit Region	X	
Hometown Region	X	X
Education Level		
AFQT Score		X
GCT Score	X	
Combat Deployment History	X	
Non-combat Days Deployed	X	
Commissioned	X	
Conduct Average Service	X	X
Rank	X	X
Deployment Rate		X
PFT Score		X
CFT Score		X
Age at Enlistment		X

C. EIGHT-YEAR LOGISTIC REGRESSION

I constructed the eight-year logistic regression the same way as the previous model. This model identifies predictors of whether or not a Marine will either be in the SMCR or another MCR component upon the completion of their MSO. 32 observations are dropped due to missing CFT score values. The correlation between the potential predictors were computed to identify any collinearity in the data (Figure 12).

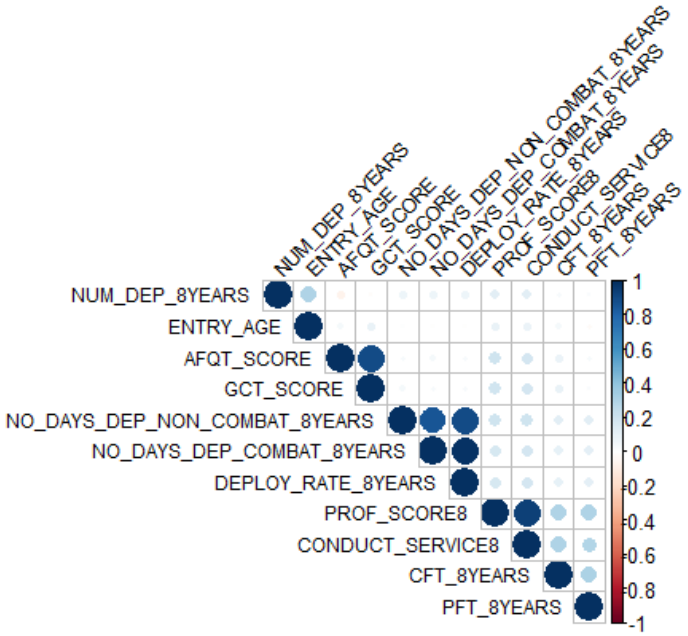

Figure 12. Eight-Year Mark Correlation Diagram for Numerical Variables.

Like the six-year data, there is high correlation (> 0.8) between GCT and AFQT scores, number of days deployed combat and non-combat, number of days deployed non-combat and deployment rate, number of days deployed combat and deployment rate, and proficiency and conduct scores. Similarly, the AFQT Score, number of days deployed combat, deployment rate, and average proficiency score variables are excluded from the model.

As before, independent variables that are significant (*p-value* < 0.10) when regressed individually against the response variable are in the initial full model. After the

backwards stepwise regression, the exponentiated coefficient estimates and corresponding confidence intervals and *p-values* are shown in Table 6.

Table 6. Eight-Year Logistic Regression Results.

Variable	Odds Ratio (95% CI)	P-value	Confounder
Number of Dependents	1.2 (1.1,1.4)	<0.001	
CFT Score	1.0 (1.0,1.0)	<0.001	
PFT Score	1.0 (1.0,1.0)	0.001	
Combat Deployment History: No Yes	 Reference 0.8 (0.5,1.2)	 Reference 0.247	Yes
Non-combat Deployment History: No Yes	 Reference 0.7 (0.3,1.3)	 Reference 0.223	Yes
# Days Deployed: Non-combat	1.0 (1.0,1.0)	0.043	
Commissioned: No Yes	 Reference 18.5 (6.0,80.6)	 Reference <0.001	
Conduct Average Service	0.6 (0.6,0.7)	<0.001	

1. Coefficient Interpretation

The final eight-year model only has two confounding variables, combat and non-combat deployment history. Non-combat deployment history is confounding with combat deployment history and number of days deployed non-combat. Combat deployment history is confounding with non-combat deployment history. All others factors are statistically significant at the 0.05 level and the effect each variable has on the outcome is when all other variables are fixed.

Each additional dependent the Marine has results in 1.2 times the odds of retention for said Marine (*p-value* < 0.001). Next, increases in max PFT and CFT scores as well as days combat deployed result in equal odds of staying in the SMCR (*p-values* < 0.001, < 0.001, and 0.043, respectively). Getting commissioned also has a large positive effect on the decision to continue in the SMCR the last two years—18.5 times the odds of those not commissioned (95% CI = 6.0, 80.6). Finally, similar to the six-year results, a one-point increase in conduct service score decreases odds of retention by 40% (95% CI = 0.6, 0.7).

2. Diagnostics and Checking Model Assumptions

(i) Linearity

Next, I constructed scatterplots for the numerical variables in the final model. Figure 13 shows that each of the five continuous variables are acceptably linearly related with the response variable. No transformations are necessary.

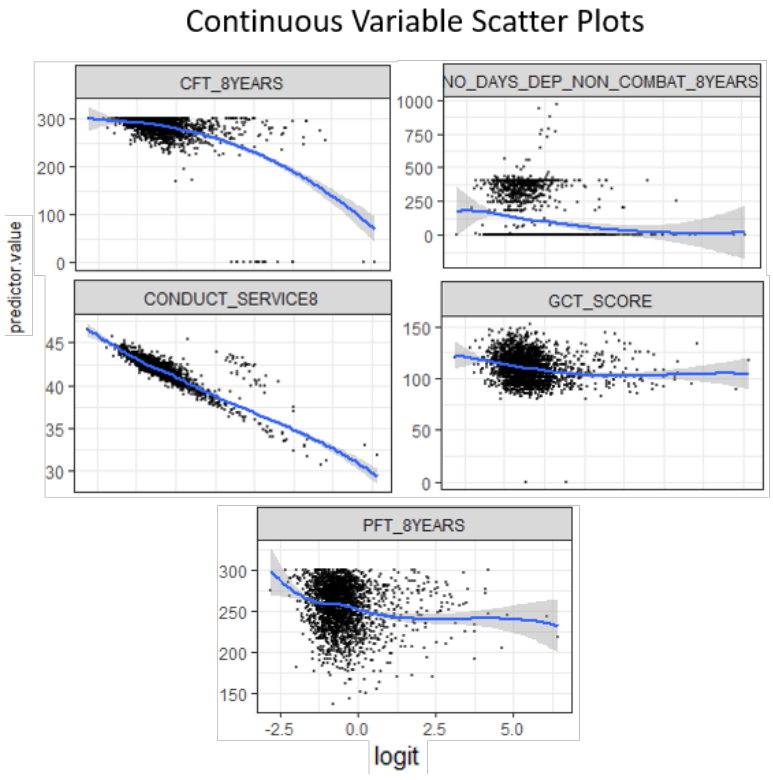

Figure 13. Eight-Year Scatterplots.

(ii) Checking for interactions

Interactions were tested between non-combat and combat deployment history as well as between PFT and CFT scores. In each case, the ANOVA results when comparing each of the new models to the original found that the interaction terms were not needed—with *p-values* of 0.259 and 0.104, respectively.

(iii) Outliers, Leverage, and Influence Points

As in the six-year model, this model was also checked for outliers, leverage and influence points. There is no statistical evidence of outliers. Observations 163 and 1992 are removed for their high influence potential, shown in Figure 14.

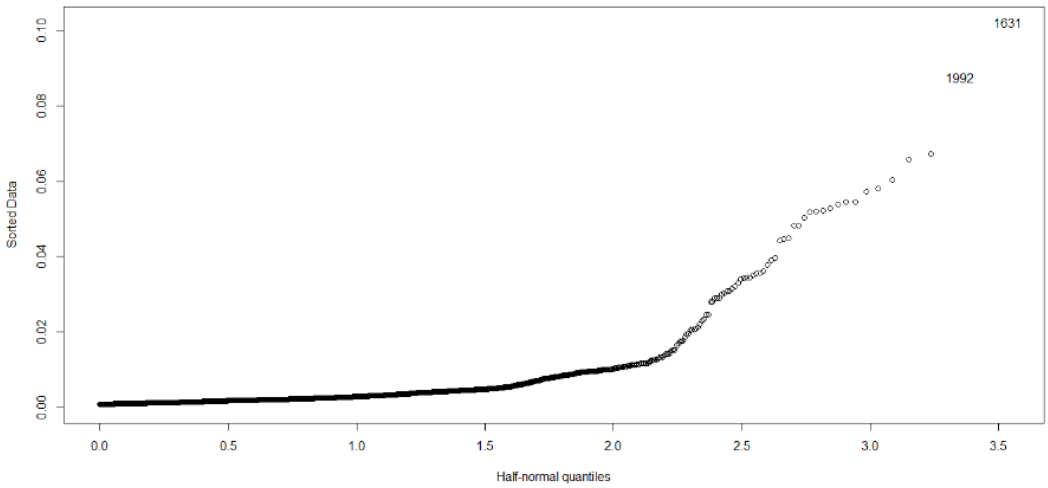

Figure 14. Leverage and Influence Test: Eight-Year Model.

The model is re-run and the new coefficients are compared to the original final model; the coefficients do not change (Table 7). The points removed from the model do not have any impact on the model.

Table 7. Eight-Year Model Coefficient Comparison.

	Coefficients	
	Original	**New**
Number of Dependents	1.2	1.2
CFT Score	1.0	1.0
PFT Score	1.0	1.0
Combat Deployment History:		
No	Reference	Reference
Yes	0.8	0.8
Non-combat Deployment History:		
No	Reference	Reference
Yes	0.7	0.7
Days Deployed Non-combat	1.0	1.0
Commissioned:		
No	Reference	Reference
Yes	18.5	18.3
Conduct Average Service	0.6	0.6

3. Model Fit

(1) ROC Curve

The ROC curve for the eight-year model is shown in Figure 15. The area under the curve is 0.67, just less than the six-year model.

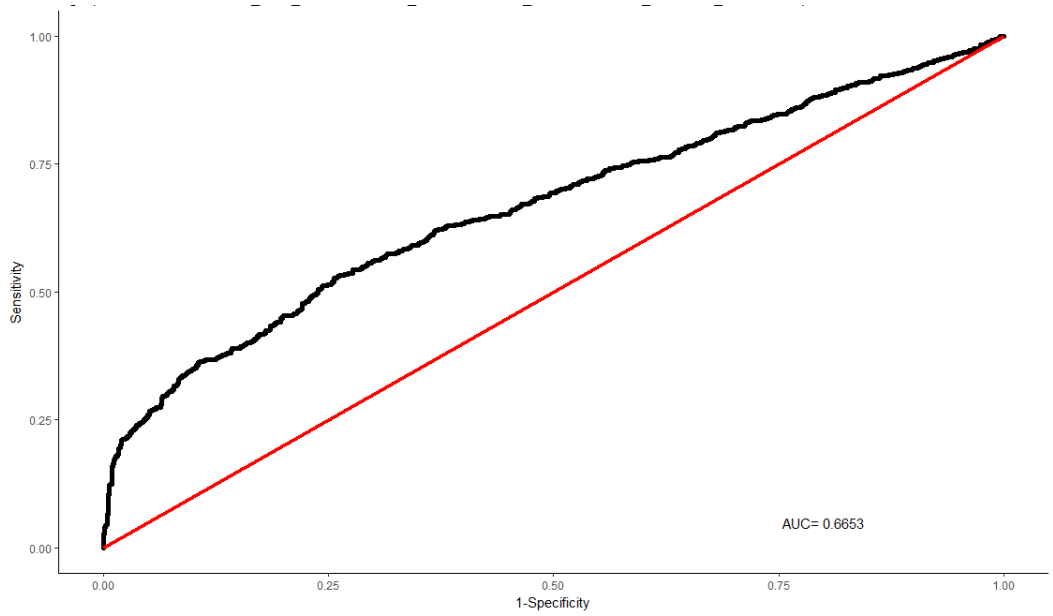

Figure 15. Eight-Year ROC Curve.

Finally, I calculated McFadden's pseudo R^2 value. The amount of variability that can be explained by the model is 8.4%. This model is slightly less predictive than the six-year model. This is attributed to the inability to distinguish any differences between the components. This is not a "stay" or "go" decision. Regardless of the component the Marine is in, he/she is still in the MCR. Thus, indicating little difference in those transferring to another component versus being retained in the SMCR.

D. EIGHT-YEAR CLASSIFICATION TREE

The eight-year classification tree was performed with the same variables as the initial logistic regression model. The tree produced in this model is also a smaller model, shown below in Figure 16.

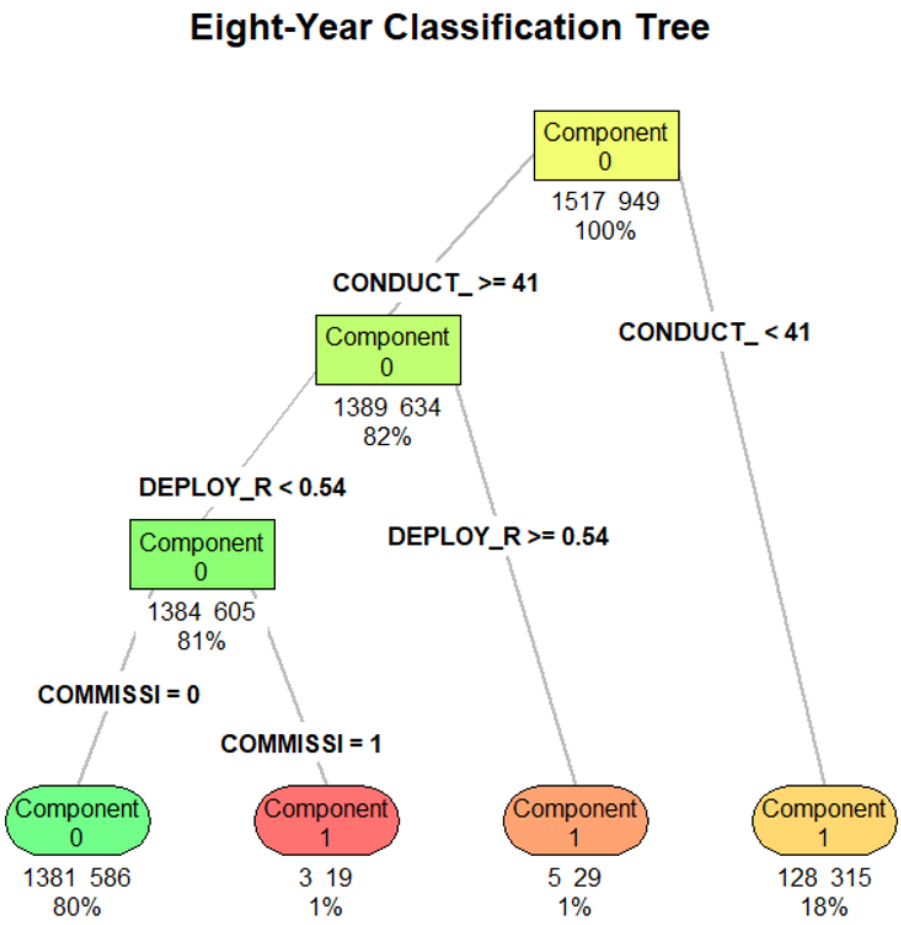

Figure 16. Eight-Year Classification Tree.

This partition tree has three node splits—splitting on average conduct service, deployment rate and commissioning status. 82% with an average conduct service score greater than or equal to 41 transferred to another MCR component within the last two years of their commitment. Furthermore, 80% of the population did not complete their MSO in the SMCR. These Marines had a conduct score greater than or equal to 41, a deployment rate less than 0.54, and did not commission.

The final logistic regression model has eight predictors, while the classification has three. There is overlap—the three classifiers of the outcome in the tree are included in the regression model, depicted in Table 8.

Table 8. Eight-Year Model Comparison.

Variable	Logistic Regression	Classification Tree
Number of Dependents	X	
CFT Score	X	
PFT Score	X	
Combat Deployment History	X	
Non-combat Deployment History	X	
Non-combat Days Deployed	X	
Deployment Rate		X
Commissioned	X	X
Conduct Average Service	X	X

E. SIX- TO EIGHT-YEAR TRENDS

Between the six- and eight-year mark, the totals for each component is too imbalanced to do a statistical analysis for each component. For example, at the eight-year point, there are only nine people in the IMA, and seven in the AR. Everyone else was either discharged, transferred to the IRR, or continued in the SMCR.

In Figure 17, the monthly departures and entry totals for each component are shown over the two-year period.

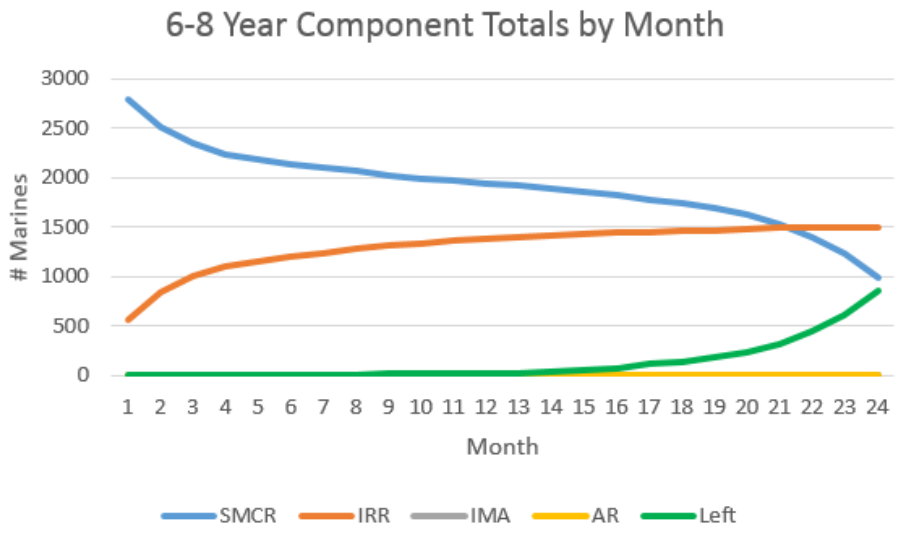

Figure 17. Two-Year Component Totals per Month.

As expected, over time many Marines are leaving the SMCR to join the IRR (as they can choose to stay in the SMCR or go to the IRR upon the six-year mark). For the first year, very few are leaving the MCR as a whole. Almost everyone who is leaving the SMCR moves to the IRR. After one year, seven years into service, the number leaving the MCR rises in an exponential fashion throughout the final 12 months. Meanwhile, the SMCR total continues to fall and the IRR's numbers grow at a much slower rate. This demonstrates that between the seven and eight-year period more of the Marines leaving the SMCR are leaving the MCR, and less are transferring to the IRR.

F. FINDINGS

In this section, the key findings for each of the two models are explained.

1. Six-Year Model

While the logistic regression has eight predictors, only six are statistically significant. Of those, ethnicity and non-combat deployments history have the largest impact on retention in the SMCR. White Marines are less likely (0.8 times odds) to complete their first six years in the SMCR than non-White Marines, holding all other variables constant. Also, a positive non-combat deployment history increases the odds of

transfer by 50%. Finally, if the service member lives in the Western US, he/she has twice the odds to stay in the SMCR the entire six years, compared to living in the Northeast.

The classification tree found different variables to better classify whether or not the service member will be retained. Rank at the six-year mark is the best predictor. Those who do not get promoted to at least E-4 only have a 36% chance to stay in the SMCR for the whole six years, the rest transfer to another component. Next, higher deployment rates over the course of the first six years is a deterrent to SMCR retention.

2. Eight-Year Model

The eight-year logistic regression model found that getting commissioned increases SMCR retention odds significantly, by 18.5. Also, the more dependents the Marine has decreases the odds of transferring to another component by 20%. Finally, a one-point increase in average conduct service score decreases odds of full retention in the SMCR by 40%.

The partition tree shows that average conduct service score best classifies the response variable. If the Marine had a score above 41, they are 82% more likely to transfer to another component somewhere between the six and eight-year period. From there, if the deployment rate over the eight-year period is less than 0.54, there is 81% probability they will transfer to another MCR component. This supports the logistic regression model. Finally, being commissioned also has a large positive influence on the SMCR retention decision—also backing up the regression model.

V. CONCLUSION

A. CONCLUSION

This empirical continuation study explored two logistic regressions at different career milestones. The population was a cohort of non-prior service Marines that enlisted in the SMCR in FY09 on a 6x2 contract. With this contract the service member is required to complete the first six years in the SMCR—the remaining years can be served in the SMCR or IRR. Continuation in this thesis is defined as a loss (from the SMCR) to another MCR component. The possible transfer components are the IRR, IMA, and AR. Due to the number of observations, we were unable to construct multinomial regressions.

The response variable for the first logistic regression is binary, with value 0 if the Marine is in IRR, IMA, or AR at the six-year mark and 1 if he/she is in the SMCR at the six-year mark. The model has little predictive power (pseudo R^2=11%), but impacts of the predictors can still be ascertained. A non-combat deployment history increases odds of transfer in the first six-years. This concurs with previous research that found that more deployments increased the likelihood of attrition (Price, 2010). Also, White Marines are more likely to transfer as well. This supports previous research by Randall which found that NPS Marines of a minority race had much larger retention rate than white NPS Marines (Randall, 1989). Finally, home region—specifically those living in the Western US—states has a large positive influence on retention in the SMCR. The six-year classification tree found rank and deployment rate at the six-year mark are the best classifiers of the outcome variable.

The sample for the second logistic regression model is everyone who was in the SMCR at the six-year mark. The outcome variable is binary defined as 0-for IRR, IMA, AR at eight-year mark or 1-for SMCR at eight-year mark. This model found a large positive impact of SMCR retention with commissioning status. If the Marine was commissioned, he/she was over 18 times more likely to be retained in the SMCR. An increase in the number of dependents also has a positive impact on retention. Both of these impacts are intuitive: enlisted to officer programs usually impose a longer contract on the service

member. Also, when a Marine has more dependents they are more likely to want to keep their military job security and extra income.

An increase in average conduct service record, however, increases the chances of transferring to another component between the six- and eight-year mark. This confirms previous studies that a poor conduct record results in higher levels of attrition (Price, 2010). This model has slightly less predictive ability than the first logistic regression (8.4%). The eight-year classification tree found that average conduct service record, deployment rate, and commissioning status are the best classifiers of the outcome variable.

Finally, Figure 15 depicts the monthly entries/departures of each component in the last two years of the MSO. The figure shows that over the two years more people are leaving the SMCR and leaving the MCR altogether (rather than transferring to the IRR); therefore, not fulfilling their contract.

B. RECOMMENDATION FOR FUTURE WORK

I believe that the poor predictive power of our models is largely due to the inability to capture the elements that are predictive of the outcome in this population. These elements may include why a Marine chooses to remain in a drilling status eligible to deploy as a unit (SMCR) versus deploying as an individual (IMA), fulfilling an active duty billet (AR), or moving to a non-drilling status (IRR). Data on civilian wages and unemployment rate during this time (FY09-FY17) might be able to explain this movement. Also, predicting future human behavior, as I did here with likelihood of continuing in SMCR or not is hard to capture. Everyone in this population was still retained in the MCR as a whole and the differences between the components are very minute. Everyone used in the predictive analysis was still in the military. Therefore, the differences between the population were not easily distinguishable. Next, research in this area, specifically in the psychological field, usually results in smaller pseudo-R^2 (Veall and Zimmermann, 1996). Future research in this area of continuation should be analyzed with a number of cohorts. With more observations, a multinomial approach could be taken to analyze the likelihood of continuing in each component (compared to SMCR versus all others). Also, there were a number of assumptions with the data that needed to be made; however, those assumptions

could be wrong. For example, logistic regression assumes a linear relationship between the predictors and the log-odds. If this condition is not met, it leads to rejection of statistical predictors (Faraway, 2016). This assumption might be in violation with this data. Finally, with more data a deeper look can be taken at the monthly departures in the final two years of the eight-year MSO.

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Buddin, R. J. (1984). *Analysis of Early Military Attrition Behavior*. Santa Monica, CA: Rand.

Bursac, Z., Gauss, C. H., Williams, D. K., & Hosmer, D. W. (2008). Purposeful selection of variables in logistic regression. *Source Code for Biology and Medicine*, 3(17). Retrieved from http://doi.org/10.1186/1751-0473-3-17

Department of the Navy. (2015). Marine Corps Reserve administrative management manual. (MCO 1001R.1L). Washington, DC: Author. Retrieved from http://www.marines.mil/%20Portals/59/MCO%201001R.1L.pdf

Faraway, J, J. (2016). *Extending Linear Models with R* (2nd ed.) (pp. 28–30, pp. 354–358). Chapman & Hall/CRC.

Hansen, M. L. & MacLeod, I, D. (2004). *Retention in the Reserve and Guard Components*. CRM D0009534.A4/1REVl. Alexandria, VA: CNA.

Herschelman, P. R. (2012). *United States Marine Corps Reserve First Term Attrition Characteristics* (Master's thesis). Retrieved from http://hdl.handle.net/10945/6807

Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression, Textbook and Solutions Manual*, 2nd ed. (pp. 91-128). John Wiley & Sons.

Licari, A. D. (2013). *Developing a Markov Model for Forecasting End Strength of Selected Marine Corps Reserve (SMCR) Officers* (Master's thesis). Retrieved from http://hdl.handle.net/10945/32856

Mandrekar, J. N. (2010). Receiver Operating Characteristic Curve in Diagnostic Test Assessment. *Journal of Thoracic Oncology*, 5(9), 1315. Retrieved from http://doi.org/10.1097/JTO.0b013e3181ec173d

Price, J. D. (2010). *Effects of activation on selected Marine Corps Reserve prior enlisted continuation rate in the post-9/11 era* (Master's thesis). Retrieved from http://hdl.handle.net/10945/5437

The R Project for Statistical Computing. Retrieved from https://www.r-project.org/

Randall, J. S. (1989). *Factors influencing the retention of noncommissioned and staff noncommissioned officers in the Selected Marine Corps Reserve* (Master's thesis). Retrieved from http://hdl.handle.net/10945/27049

Schumacher, J. F. (2005). *Forecasting retention in the United States Marine Corps Reserve* (Master's thesis). Retrieved from http://hdl.handle.net/10945/2254

Veall, M. R. & Zimmermann, K. F. (1996). Pseudo-R^2 for Some Common Limited Dependent Variable Models. *Journal of Economic Surveys*, 10(3), 241. Retrieved from https://doi.org/10.1111/j.1467-6419.1996.tb00013.x

Zhang, Z. (2016). Model building strategy for logistic regression: purposeful selection. *Annals of Translational Medicine*, *4*(6), 111. Retrieved from http://doi.org/10.21037/atm.2016.02.15

Zhang, Z. (2016). Residuals and regression diagnostics: focusing on logistic regression. *Annals of Translational Medicine*, *4*(10), 195. Retrieved from http://doi.org/10.21037/atm.2016.03.36

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California