2018-06

MODELING FIRST-TERM ENLISTMENT COMPLETION FOR THE SELECTED MARINE CORPS RESERVE

Fletcher, Christopher K.

Monterey, CA; Naval Postgraduate School

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

MODELING FIRST-TERM ENLISTMENT COMPLETION FOR THE SELECTED MARINE CORPS RESERVE

by

Christopher K. Fletcher

June 2018

Thesis Advisor: Andrew T. Anglemyer
Second Reader: Lyn R. Whitaker

Approved for public release. Distribution is unlimited.

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 MODELING FIRST-TERM ENLISTMENT COMPLETION FOR THE SELECTED MARINE CORPS RESERVE	**5. FUNDING NUMBERS**
6. AUTHOR(S) Christopher K. Fletcher	
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)

Military retention studies seek to better explain the factors affecting a service member's decision to complete contractual service agreements. For the purposes of meeting the national security needs of the nation, these studies inform decision makers about recruiting and sustaining a ready and capable force. This thesis explores the significant factors that predict first-term enlistment completion for non-prior service (NPS) enlisted Marines in the Selected Marine Corps Reserve (SMCR).

The data available from the Total Force Data Warehouse (TFDW) and provided by Manpower and Reserve Affairs (M&RA), Quantico, Virginia, contains information about first-term NPS enlisted Marines as they progress in the SMCR. A cohort of Marines who enlisted in the SMCR in fiscal year 2009 is selected and tracked for a period of eight years to determine a binary outcome of successful completion.

Utilizing binary logistic and random forest regression, two distinct models are developed to determine the significant features indicative of successful completion of first-term enlistments. The first model examines information available prior to entering the SMCR. The second model examines data available prior to enlistment and data obtained as the individual completes obligated service. The end state is to select the best model fit for each distinct model to inform decision makers about first-term enlistment completion.

14. SUBJECT TERMS retention, first-term, enlisted, Selected Marine Corps Reserve	**15. NUMBER OF PAGES** 89
	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.

MODELING FIRST-TERM ENLISTMENT COMPLETION FOR THE SELECTED MARINE CORPS RESERVE

Christopher K. Fletcher
Captain, United States Marine Corps
BS, United States Naval Academy, 2010

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
June 2018

Approved by: Andrew T. Anglemyer
Advisor

Lyn R. Whitaker
Second Reader

Patricia A. Jacobs
Chair, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Military retention studies seek to better explain the factors affecting a service member's decision to complete contractual service agreements. For the purposes of meeting the national security needs of the nation, these studies inform decision makers about recruiting and sustaining a ready and capable force. This thesis explores the significant factors that predict first-term enlistment completion for non-prior service (NPS) enlisted Marines in the Selected Marine Corps Reserve (SMCR).

The data available from the Total Force Data Warehouse (TFDW) and provided by Manpower and Reserve Affairs (M&RA), Quantico, Virginia, contains information about first-term NPS enlisted Marines as they progress in the SMCR. A cohort of Marines who enlisted in the SMCR in fiscal year 2009 is selected and tracked for a period of eight years to determine a binary outcome of successful completion.

Utilizing binary logistic and random forest regression, two distinct models are developed to determine the significant features indicative of successful completion of first-term enlistments. The first model examines information available prior to entering the SMCR. The second model examines data available prior to enlistment and data obtained as the individual completes obligated service. The end state is to select the best model fit for each distinct model to inform decision makers about first-term enlistment completion.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS

AC	Active Component
AFQT	Armed Forces Qualification Test
AR	Active Reserve
ASL	Active Status List
CFT	Combat Fitness test
CMC	Commandant of the Marine Corps
CON	Conduct
CONUS	Continental United States
CSS	Combat Service Support
EM	Enlisted Model
GT	General Technical
HMH	Marine Heavy Helicopter Squadron
HMLA	Marine Light Attack Helicopter Squadron
HOR	Home of Record
HQ	Headquarters
IADT	Initial Active Duty Training
IDT	Inactive Duty Training
IIADT	Incremental Initial Active Duty Training
IMA	Individual Mobilization Augment
IRR	Inactive Ready Reserve
LR	Logistic Regression
MACS	Marine Air Control Squadron
MACG	Marine Air Control Group
MAG	Marine Aircraft Group
MALS	Marine Aviation Logistics Squadron
MarDiv	Marine Division
MARFORRES	Marine Forces Reserve
MASS	Marine Air Support Squadron
MATSG	Marine Air Training Support Group
MAW	Marine Air Wing

MCO	Marine Corps Order
MCT	Marine Combat Training
MEPS	Military Entrance Processing Stations
MLG	Marine Logistics Group
MOS	Military Occupational Specialty
M&RA	Manpower and Reserve Affairs
MSO	Military Service Obligation
MTACS	Marine Tactical Air Command Squadron
MWCS	Marine Wing Communications Squadron
MWSS	Marine Wing Support Squadron
NJP	Non-Judicial Punishment
NPS	Non-Prior Service
OccFld	Occupational Field
P-EM	Prior-Enlistment Model
PFT	Physical Fitness Test
PMOS	Primary Military Occupational Specialty
PRO	Proficiency
PS	Prior Service
RC	Reserve Component
RF	Random Forest
ROC	Receiver Operating Characteristic
RTC	Reserve Training Center
SelRes	Selected Reserve
SMCR	Selected Marine Corps Reserve
TA	Tuition Assistance
TFDW	Total Force Data Warehouse
USMC	United States Marine Corps
VMFA	Marine Fighter Attack Squadron
VMFT	Marine Fighter Training Squadron
VMGR	Marine Aerial Refueler Transport Squadron
VMM	Marine Medium Tiltrotor Squadron
VMR	Marine Transport Squadron

VMU	Marine Unmanned Aerial Vehicle Squadron
YOS	Years of Service

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

This thesis identifies the factors associated with successful completion of a non-prior service (NPS) Marine's first-term enlistment in the Selected Marine Corps Reserve (SMCR). Two distinct models were fit to the data obtained from the Total Force Data Warehouse (TFDW) using logistic regression and random forest techniques, tailoring the models via prescriptive selection.

The pre-enlistment model (P-EM), consisting of nine predictor variables and fit to the available data for a Marine prior to the commencement of their SMCR service, failed to produce an accurate model fit based on correct test set prediction rates and receiver operating characteristic (ROC) curves.

The enlisted model (EM) uses data available for a Marine prior to and during SMCR service. The final logistic regression model identified MOS group (combat arms, combat service support, and aviation), MOS change during service, entry marital status, sex, maximum number of dependents, max CFT score, max PFT score, proficiency score at the rank of E4, and conduct scores at the ranks of E3 and E4 as significant predictors of first-term enlistment completion. The final EM has a correct test set prediction rate of 83.3 percent and shows improvement from a full model.

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

I would like to thank Dr. Anglemyer and Dr. Whitaker for their support during the thesis writing process.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. PURPOSE

This thesis identifies the factors associated with successful completion of a non-prior service (NPS) Marine's first-term enlistment in the Selected Marine Corps Reserve (SMCR). Employing techniques based in probability and statistics, primarily logistic regression, this study builds a retention model to estimate the probability of successful first-term enlistment completion. A study of this kind was requested by Manpower and Reserve Affairs (M&RA) to better understand and improve SMCR retention of NPS first-term enlisted Marines.

B. BACKGROUND

According to Title 10 U.S. Code, "The purpose of each reserve component is to provide trained units and qualified persons available for active duty in the armed forces, in time of war or national emergency, and at such other times as the national security may require, to fill the needs of the armed forces whenever more units and persons are needed than are in the regular components" (Purpose of the Reserve 1994). The Marine Corps' defined mission of its Reserve Component (RC) is to "augment, reinforce, and sustain the Active Component (AC) with trained units and qualified individuals in times of war or national emergency and at other such times as national security may require" (Commandant of the Marine Corps [CMC] 2015). To maintain the reserve force necessary to meet the statutory requirements, the United States Marine Corps (USMC) recruits NPS individuals and prior service (PS) Marines who have completed service time on active duty.

1. Reserve Force Structure

The RC, or the Marine Forces Reserve (MARFORRES), is designed to augment the AC and contains an infantry division, an air wing, a logistics group, and a headquarters unit stationed in New Orleans, Louisiana. Figure 1 shows the current command structure

of the MARFORRES, with the 4th Marine Division (4th MarDiv), 4th Marine Air Wing (4th MAW), 4th Marine Logistics Group (4th MLG), the Force Headquarters Group (FHG), and the MARFORRES Headquarters Battalion (HQBN) (CMC 2015). Every Marine that serves in the RC is attached to one of three subgroups: Ready Reserve, Standby Reserve, or Retired Reserve.

Figure 1. MARFORRES force structure. Source: CMC (2015).

Since this thesis focuses primarily on a subcomponent of the Ready Reserve, a brief explanation of the Ready Reserve is provided. For more information on the structure and purpose of the Standby Reserve and the Retired Reserve, please see Chapter 1 of MCO 1001R.1L (CMC 2015). Figure 2 provides an organizational breakdown of the RC.

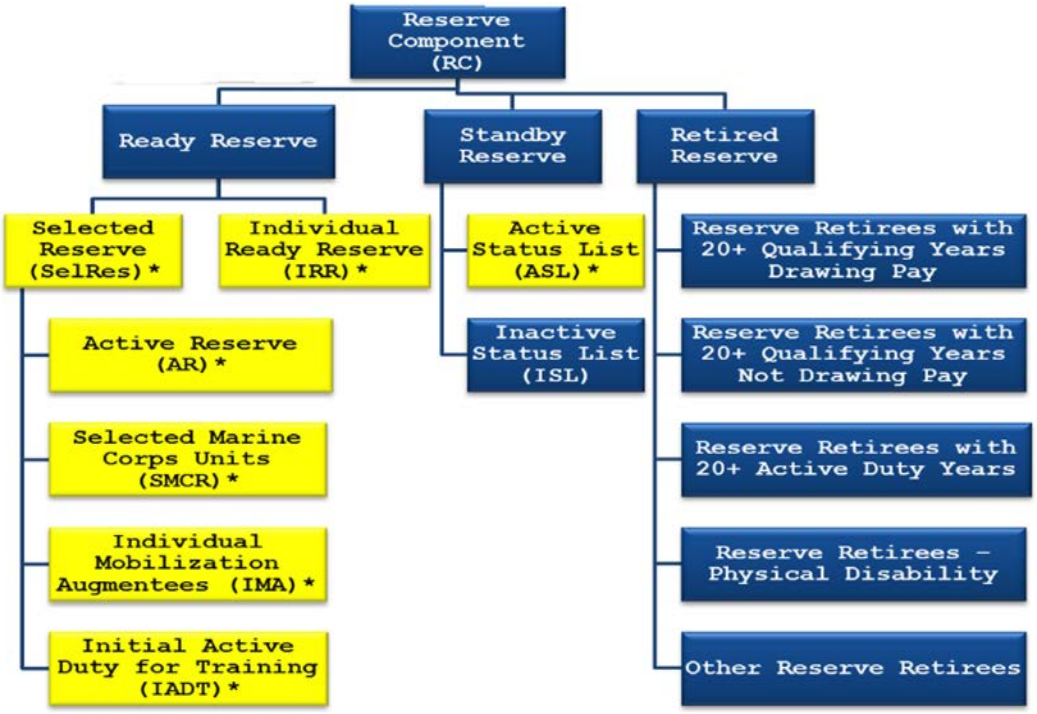

Figure 2. Organization of the Reserve Component. Source: CMC (2015).

2. Ready Reserve

The ready reserve consists of the Individual Ready Reserve (IRR) and the Selected Reserve (SelRes). These Marines can be recalled to the AC in times of war, national emergency, or when authorized by statute (Purpose of the Reserve 1994).

a. IRR

The IRR acts as a "service manpower pool," and is composed of individuals who have completed training, have either completed an obligated service in the AC or the SelRes, and are ready for mobilization (CMC 2015).

b. SelRes

The SelRes is made up of Individual Mobilization Augmentees (IMA), Reserve Marines serving on the Active Reserve (AR) program, Reserve Marines who are completing their initial active duty training (IADT), and SMCR units (CMC 2015).

(1) IMA

The IMA is designed to "facilitate the rapid expansion of the AC" in case of manpower shortages (CMC 2015). IMA Marines are assigned on an as-needed basis to AC units, meaning their billet structure falls outside of the RC. IMA Marines must perform a minimum of 12 days of Annual Training and up to 48 days of Inactive for Duty Training (IDT) per year (CMC 2015).

(2) AR

Marines in the AR program fill full-time, active duty billets that support MARFORRES (CMC 2015).

(3) IADT

Marines assigned to IADT are in the processes of completing their initial training associated with accession to the USMC. Every reservist's career begins within the IADT (CMC 2015).

(4) SMCR

The SMCR is under administrative control of the Commander, MARFORRES, unless otherwise assigned to AC units for mobilization or training (CMC 2015).

3. Selected Marine Corps Reserve

The SMCR consists of approximately 30,700 reservists. These Marines are assigned to SMCR units of the five major subordinate commands which are distributed amongst 47 states, the District of Columbia, and Puerto Rico. As an example, Figure 3 depicts the subordinate units assigned to 4th MAW.

Recruiting for the SMCR is different than for the AC. In the AC, a Marine is recruited to fill a Military Occupational Specialty (MOS), or as an open contract, meaning the Marine will be assigned an MOS based on recruit training performance, test scores, and the needs of the USMC. These Marines are then assigned to an AC unit around the world, and are subject to move to another unit every two to three years. For the SMCR, a Marine is recruited to fill a billet—with a specific MOS—at a specific location. This is an important

recruiting tool because it allows an individual to remain in one geographic area while still maintaining their affiliation with the USMC. If there is no available billet at a certain location, no additional individuals are recruited. If the individual does not meet the MOS requirements for the billet they have been recruited to fill, they are trained and placed in another billet a suitable distance from the original unit's location.

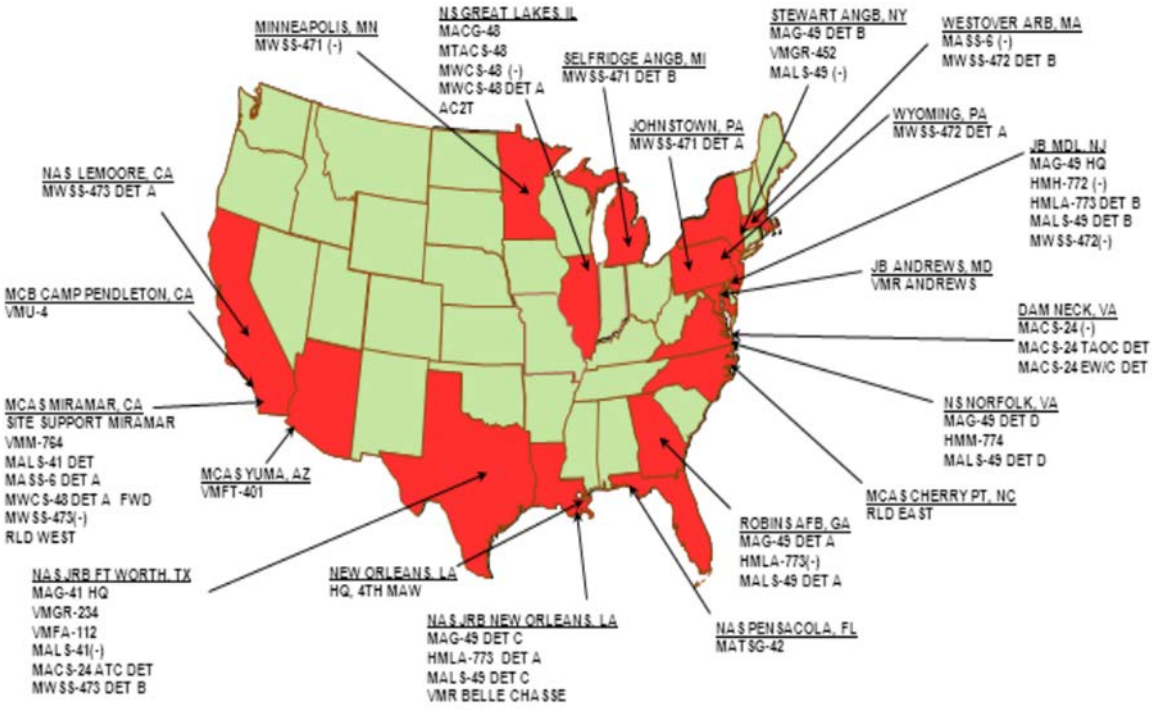

Figure 3. Subordinate units assigned to 4th MAW, SMCR. Source: M&RA (2017).

Every person who enters the armed forces "incurs an eight-year period of obligated service" called a Military Service Obligation (MSO) (CMC 2016). For SMCR Marines, the MSO is typically denoted as a 6x2 contract, meaning the Marine is obligated to serve at least six years in the drilling reserves (SMCR) and two years in the IRR. However, the Marine is not required to transfer to the IRR, and can remain in the SMCR for the full eight-year commitment, if they so choose.

The 6x2 contract is the most common enlistment because portions of a Marine's MSO are spent in training to complete IADT, which includes recruit training, Marine Combat Training (MCT), and Primary Military Occupational Specialty (PMOS) training, an average total of 20 weeks of initial training. The Incremental Initial Active Duty Training (IIADT) program can also affect the training timeline of a NPS Marine enlisted with this contract.

The IIADT program was instituted to attract highly qualified NPS individuals into MARFORRES by incrementing their initial training. The individual would attend recruit training the first summer after their enlistment, MCT the next, and their MOS school the third. This allows the Marine to remain affiliated with the SMCR while attending an educational institution during the fall and winter months (CMC 2016). Under the IIADT program, a Marine is not deployable for three years, until all IADT requirements are complete, even though the Marine is still charged against a unit's end strength. Because of this reason, "The number of applicants enlisting through the IIADT program option may not exceed five percent of the entire Reserve NPS manpower mission and ten percent of each SMCR unit mission" (CMC 2016).

An SMCR Marine will not submit for reenlistment until their final year of the eight-year obligated service. In rare cases, a Marine can request reenlistment early if they have accepted a bonus or a lateral move to another MOS and need obligated time. Regardless of the circumstance, the Marine must meet the qualifications for reenlistment before they are accepted to continue in the SMCR (see CMC 2015 reference for reenlistment requirements).

Aside from the SMCR commitments, each Marine has their own specific circumstances which may include college, a trade school, a 40-hour-a-week job, and a family; all of which may affect the decision of a Marine to remain in the SMCR during their first enlistment.

C. LITERATURE REVIEW

Multiple retention studies have been conducted since the onset of the all-volunteer force in 1973. This section provides a brief description of the literature reviewed regarding retention in the U.S. military, and highlights several studies focused primarily on first-term retention as it relates to the Marine Corps and the RC.

1. Military Studies

Sponsored by the Human Resources Research Office of the Defense Advanced Research Projects Agency (DARPA), the RAND Corporation endeavored to develop a better understanding of the current and future manpower issues faced by the Department of Defense (DoD), commissioning a study in 1977 regarding enlistment bonuses and first-term retention (Enns 1977). The research explored three questions: What was the overall impact of bonuses on first-term reenlistment? Do the bonuses have different impacts depending on the occupational group to which they are offered? And do different bonus levels (or multiple bonuses) produce different results in enlistment rates.

The author concluded that for the three services explored (Army, Air Force, and Navy), the effects of a bonus to first-term enlisted members is positive and statistically significant. They did not find any significant differences between the defined occupational groups and receipt of the bonus. Even though some differences were found between the different services, the magnitude of the difference is not consistent over time and therefore the authors concluded no separate models were needed for each branch. Since the study found no significant differences between the services, even though the study was conducted on active duty service members, the findings of this report may have some application to the USMC reserve component. Though not exploring reenlistment specifically, if a Marine has the option for a bonus contingent upon completing their obligated service, it is expected to affect retention positively (Enns 1977).

Buddin, Levy, Hanley, and Waldman (1992) also associated with RAND, conducted research focused primarily on the rate in which individuals are promoted and how that rate affects retention. The authors examined several retention models used by the DoD and found that none considered promotion tempo. They concluded that retention

models are in fact “sensitive to the specification of individual promotion opportunities at the end of the first-term” enlistment. The expected time required to be promoted to E-5 had a “significant effect on first-term retention.” The analysis showed that a “10 percent promotion slowdown is associated with 14 and 18 percent reductions in Army and Air Force retention rates, respectively” (Buddin et al. 1992).

The authors also used an annualized cost of leaving (ACOL) formulation for the retention model developed. The ACOL describes costs to individuals in their decision to leave the military, both pecuniary and non-pecuniary. Interestingly, the authors found that the ACOL adjustments made to the model indicates there is strong evidence to suggest that non-pecuniary factors play a substantial role, “over and above the monetary value of the promotion itself.” In short, the typical incentives given to first-term enlistees in the form of monetary inducements does not account for the reasons why an individual would depart the military. On-time promotions, naturally associated with satisfactory performance, may provide a significant psychological effect that outweighs any monetary benefit received by the individual (Buddin et al. 1992).

In a third study conducted by RAND, authors Buddin and Kapur (2002) studied the effects of pecuniary incentives on first-term retention, primarily the tuition assistance (TA) program. According to Buddin and Kapur, TA “is a military-sponsored program that reimburses military members for the cost of college classes while on active duty.” The program was designed to increase the quality-of-life of the individual, and attract those also interested in attending college into the military. The authors used data obtained from the Navy and Marine Corps to examine the use of TA and retention behavior of first-term enlistees.

In terms of descriptive statistics, Buddin and Kapur (2002) found that 8 percent in the Navy and 13 percent in the Marine Corps used TA while on active duty. The authors found that women were more likely to use TA than men, with 22 percent of female sailors and 27 percent of female marines participating in the program. The authors found that age was not a factor, but family responsibilities “make a difference for both marines and sailors, but only a small one.” Also, the occupation of the individual made a significant difference in whether they used the TA program or not. Those Marines and sailors in technical,

support, or administrative assignments "participate in the program more frequently than do those in other types of assignments" (Buddin and Kapur 2002). The authors posit that this may be explained by a more predictable work schedule, thus facilitating the time required to take college classes, or that these individuals may be more interested in advancing their education.

The results of the model developed by the authors finds that TA does affect retention but in a negative way. For the Marines, "a typical program participant is about 4 percentage points less likely to reenlist than a comparable marine," while in the Navy, "the participant is 9 percentage points less likely to reenlist" (Buddin and Kapur 2002). These results conflict with prior studies conducted on the same topic, but the authors point out flaws in the previous studies in regards to the amount of time an individual is eligible for TA. When the time an individual is eligible for TA is normalized, contrary to the previous studies, Buddin and Kapur found that TA users were "consistently less likely to remain in the military." The authors believe this is due to the availability of the G.I Bill to those participants at the completion of their active service. In short, those who participate in the TA program are more likely to advance their education and career opportunities, and chose to move to the civilian sector for school after their first enlistment, using the funds provided by the G.I. Bill to be a full-time student (Buddin and Kapur 2002). The TA program is also available to the Reserve component of the USMC. If the RC Marine is activated for a period of 120 days or more, they are eligible for TA funds.

Dolfini-Reed, Parcell, and Horne (2005) evaluated the impact of activation, mobilization, and deployments on reservist attrition for all branches of the military, from September 2001, to January 2005, and compared the results to the attrition statistics for reservists in FY 2000. The study reviewed NPS and PS reservists who were not currently deployed at the time of the study.

Across the services, the post FY 2000 group had higher losses than the FY 2000 group, especially for those who were activated and deployed for longer periods of time. Losses for reservists who were activated and deployed were fewer than those who were activated and not deployed, suggesting that deployments overseas positively impact retention (Dolfini-Reed et al. 2005).

2. Marine Corps Reserve Studies

O'Donohue (1988) produced a thesis study focusing on male, first-term, reserve Marines, and their decision to remain in the SMCR. The data obtained for this study is a combination of survey answers obtained from the 1986 Reserve Component Surveys, and personnel records obtained from the Reserve Component Personnel Data System (pecuniary and non-pecuniary factors on job retention). Using logistic regression, the author developed two models: one for PS reservists, and one for NPS reservists. The author found that the individual's income from the RC has a significant, positive impact on SMCR retention, while civilian income was not found to be statistically significant. Educational benefits, civilian job-related training, and retirement benefits were found to be significant factors in retaining prior service reservists (O'Donohue 1988).

For NPS Marines specifically, Donahue found that age, pay-grade, reserve income, and first-enlistment bonus receipt significantly, positively impacts retention, with reserve income having the largest impact. The age, pay-grade, and reserve income are all counted as "proxies for length of service," and indicates that tenure, even for NPS Marines with fewer than six years of service, plays an important role in retention (O'Donohue 1988).

Hattiangadi and Parcell (2006) explored attrition in the SelRes of the MC and focused primarily on the monetary incentives offered to NPS and PS reserve Marines. The study found that incentives had little effect of mitigating attrition for the NPS Marines, and a far greater impact on PS Marines, contradicting some of the findings from O'Donohue (1988). The authors believe this is due to the six-year initial obligation of the NPS Marines compared to the PS Marines who typically incur a four-year obligation for reenlistment. Also, PS Marines have a higher probability of eligibility for monetary incentives, as most incentives are developed to retain individual after their initial enlistment (Hattiangadi and Parcell 2006).

Herschelman (2012) explored the retention characteristics of NPS Marines in the RC before and after September 11, 2001. The author divided the data into three cohorts, exploring enlisted Marines with varying expectations of deployments, depending on their time of service. For example, the control cohort consisted of Marines who enlisted in FY

1994–1995, who the author stipulates had less of an expectation to deploy than the group who enlisted after FY 2001.

The author found that post-9/11 attrition rates decreased by as much as 6.1 percent, most likely explained by the increase in deployments overseas and rising unemployment in the U.S., which for reservists increased from 4.5 to 7.0 percent between the control cohort (1994–1995) to the post-9/11 cohort. Those reservists who deployed to a combat zone (Afghanistan or Iraq) were consistently less likely to attrite than those who were stationed in the U.S., though it is not clear as to why this may be the case. Herschelman (2012) posits that it could be the needs of the MC, using incentives and mandates to retain Marines during war time, or it could be the individual who develops a greater "sense of purpose" in supporting the war effort abroad, and thus decides to remain in the RC for the entire commitment.

Findings in this study, such as decreased attrition for those with dependents and a decrease in attrition for women, differed from previous literature. Herschelman (2012) did not find education level, aptitude, and regional variables, such as the home of record of the Marine, to be significant. However, he notes the "strongest finding was an increased proficiency/conduct score predict decreased attrition." It is expected to be the same result in this study.

Alstat (2017) examined the effects of commuting distance on SMCR retention for NPS Marines who enlisted between January 1, 2001 and July 31, 2011, and had at least six satisfactory years of service. Commuting distance is the distance traveled from a Marine's home of record (HOR), and the Marine's assigned Reserve Training Center (RTC). The data in this study is divided into four commuting distance groups: 25 miles or less, 25–50 miles, 51–100 miles, and over 100 miles. The regional aspects of the commute are also taken into account, as the Marine's HOR states are aggregated into regions of the U.S. (West, Midwest, South, and Northeast).

The results of Alstat's study indicate that the commute with the highest odds of attrition is the 51–100-mile group, and that a commuting distance of over 100 miles indicates a lower attrition rate. The Mountain region, as compared the Pacific region, has

the highest odds of attrition. Alstat also found that the rank of the Marine played a significant role in attrition. Although a Marine living between 51–100 miles of an RTC is more likely to attrite, if the Marine is a Corporal instead of a Lance Corporal, the likelihood of attrition decreases, suggesting that rank, promotion, and promotion timing (tempo) play an important role in retention (Alstat 2017).

D. SUMMARY

This thesis addresses many of the same factors explored by previous studies, including performance, rank, PMOS, marital status, dependents, deployments, and regional variables. By addressing the impact of these variables on first-term, NPS, enlisted, SMCR Marines, this thesis endeavors to better inform leadership on the characteristics of retention, and improve first-term enlistment completion probability, thus saving time and money, and increasing the overall readiness of the force.

II. DATA AND METHODOLOGY

This thesis focuses on NPS, enlisted Marines under their first-term contract. Though many paths to completing obligated service exist, an SMCR Marine typically signs a 6x2 contract for their first enlistment. The 6x2 contract stipulates that the Marine complete six years of SMCR duty, and two years assigned to the IRR. If they so choose, they can complete the full eight years in the SMCR. The other contract service timelines are similar. Table 1 shows the varied contact options, and historical contract statistics.

Table 1. Number enlisted by contact type and fiscal year for the RC optional enlistment program from 2000–2010. Adapted from M&RA (2017).

Contract Type	*30 Sept 2000*	*30 Sept 2010*
3X5	54 (0.20%)	13 (0.04%)
4X4	510 (1.90%)	493 (1.84%)
5X3	20 (0.07%)	7 (0.03%)
6X2	25,414 (97.75%)	26,785 (98.08%)

A. DEFINING SUCCESS

Success in this context is defined by completing obligated service, as signed by contract. The manner in which this service is completed is irrelevant for the purposes of this study, meaning that an SMCR Marine may complete their initial enlistment in the typical agreement of 6x2, or they can complete their contract by any other means.

Included in the definition of success are those individuals who enter the SMCR as an enlisted Marine and receive a commission as an officer, either in the RC or AC. Even though these Marines may not be filling a billet in the SMCR as originally contracted, the fact that they are retained in the military as an officer and incur additional years of service after commissioning meets the definition of success.

Transferring to the AC is also included in the definition of success. Similar to receiving a commission, Marines who transfer to the AC from the SMCR are considered

to have successfully completed their obligated service, even though they are not filling a billet in the SMCR.

If the data for the following instances is available, SMCR Marines who were activated for combat operations, wounded, and no longer physically capable to complete their service due to service-related injuries, are considered a successful completion. Likewise, those Marines killed in action constitute successful completion.

A wide variety of scenarios do not meet the definition of successful completion. Therefore, if a Marine does not meet any of the above definitions of success, they are considered not successful, regardless of the situation. The varied reasons for unsuccessful completion of first-term enlistments are not explored in this thesis, and are recommended for further study.

B. PROVIDED DATA

USMC M&RA Division, Quantico, VA, extracted data used in this thesis from the Total Force Data Warehouse (TFDW). The data includes records for Marines attached to the RC, including NPS, PS, enlisted Marines and officers.

M&RA maintains information on the RC in the TFDW as far back as the 1970s. The data is collected as monthly snapshots to ensure a Marine can be tracked from month to month as they complete their obligated service. Monthly records provide enhanced insight into the career path of each individual Marine, tracking the changes as they occur more frequently than quarterly or annually. The data provided from M&RA (2017) consists of data sets separated by the month. Each data set contains 48 different variables for each Marine:

- SEQ_NUM: Sequence number that corresponds to the month and year the data is saved to the system. For example, sequence number 236 corresponds to October, 2008.

- COMPONENT_CODE: Component codes identify the military component the Marine is affiliated (AC, RC, etc.).

- RCOMPCODE: Reserve component codes identify the reserve component a Marine is assigned (i.e., SMCR, IRR, etc.).
- CIVILIAN_EDUC_LEVEL_CODE: Civilian education codes indicate the level of civilian education the Marine attained. For example, a code of 12 indicates the Marine completed high school.
- REENLIST_RECOMMENDATION_CODE: The reenlistment recommendation code for each Marine as assigned by their senior leadership.
- RETIREMENT_DATE: A date the Marine is scheduled to retire from military service.
- DATE_OF_BIRTH: The Marine's date of birth.
- DODTCPG: Department of Defense Training Category Pay Group codes indicate the category the Marine is currently affiliated. For example, a code of SA indicates the Marine is assigned to the SMCR.
- PRESGRDE: The present grade is the Marine's current pay grade (rank).
- PMOS: Primary Military Occupational Specialty, or job the Marine is assigned upon enlistment and completion of basic training.
- RELIGION: Code indicating the Marine's religious preference.
- MARITAL: Marine's marital status code.
- RACE: Marine's race code.
- SEX: Sex of the Marine.
- ETHGRP: Ethnic group code of the Marine.
- NUMBER_OF_DEPENDENTS: The number of dependents, or family members, on record for the Marine.

- RUC: Reporting Unit Code denotes the unit the Marine is assigned.
- YOS: Years of service counts the number of years the Marine is affiliated with the USMC.
- TOTSATYRS: Total satisfactory years of service counts the number of years a Marine has met the 50 point RC threshold for a satisfactory year of service. A satisfactory year is counted when a Marine has completed the drilling requirements of the year. It is possible for YOS to be greater than TOTSATYRS.
- EDIPI: Electronic Data Interchange Personal Identifier is a 10-digit unique identifier for each Marine.
- EOS: End of obligated service date is the date the Marine is scheduled to complete their current contract.
- DOR: Date of Rank is the date the Marine achieved their current rank.
- AFADBD: Armed Forces Active Duty Base Date is the date the Marine entered active service.
- PEBD: Pay Entry Base Date is the date a Marine began initial training (boot camp).
- HOR_ZIP: Zip code of the Marine's home of record.
- HOR_CITY: City name of the Marine's home of record.
- HOR_STATE: State name of the Marine's home of record.
- AFQT_SCORE: Armed Force Qualification test score.
- GCT_SCORE: General Technical (GT) score derived from the Armed Services Vocational Aptitude Battery (ASVAB).

- MAND_DRILL_STOP_DATE: The date the Marine is scheduled to complete obligated service, and is dependent on contract.
- PRIOR_SERVICE: Indicator of prior service in the AC or in another branch of the military.
- AGE: Age of the Marine.
- REASON_TRANSFERRED_IRR: Reason transferred to the IRR code.
- REASON_TRASFERRED_SMCR: Reason transferred to the SMCR code.
- UNIT_NAME: Name of the unit the Marine is assigned.
- UNIT_ZIP: Zip code of the Marines current unit.
- UNIT_CITY: City of the Marine's current unit.
- UNIT_STATE: State of the Marine's current unit.
- PERSONAL_AWARD: Indicator if a Marine receives an award during their service.
- DISCIPLINARY_ACTION: Indicator if a Marine has received disciplinary action during their service.
- CBT_SCORE: Combat Fitness Test score.
- PFT_SCORE: Physical Fitness Test score.
- CONDUCT: Marine's current Conduct grade as assigned by senior leadership.
- PROFICIENCY: Marine's current Proficiency score as assigned by senior leadership.

- NO_DAYS_DEPLOYED_COMBAT: The number of days a Marine has been deployed to a combat zone.
- NO_DEPLOYMENTS_COMBAT: The number of combat zone deployments a Marine has completed.
- NO_DAYS_DEPLOYED_NON_COMBAT: The number of days a Marine has been deployed in a non-combat zone.
- NO_DEPLOYMENTS_NON_COMBAT: The number of no-combat deployments the Marine has completed.

C. COHORT DATA COLLECTION

A contemporary cohort of SMCR Marines is selected for this study, based on the constraints of time. In order to track successful completion of obligated service, at least eight years of data is required for each Marine, since this is the minimum MSO. For this reason, the Marines who enlisted in the SMCR in FY 09 are the cohort of study in this thesis. To ensure the data is sufficient to track Marines for their entire service obligation, the FY 09 cohort is the most contemporary set of data available at the commencement of this thesis.

Since the U.S. military, including the USMC, increased overall end strength due to combat operations in Afghanistan and Iraq, the use of the FY 09 cohort for this study does warrant discussion. In order to increase the end strength of the force, recruiting and retention quotas were expanded for the AC and RC. Part of that expansion occurred in FY 09. It is necessary then to view the change in RC end strength and evaluate the impact of increased retention on the FY 09 cohort. There are turbulent changes in end strength (see Figure 4), with the maximum yearly increase in end strength between the period of FY 09 and FY 16 of 7.71 percent (M&RA 2017). This flux in end strength due to the national military needs of the time may affect the long-term applicability of models, based on the FY 09 cohort.

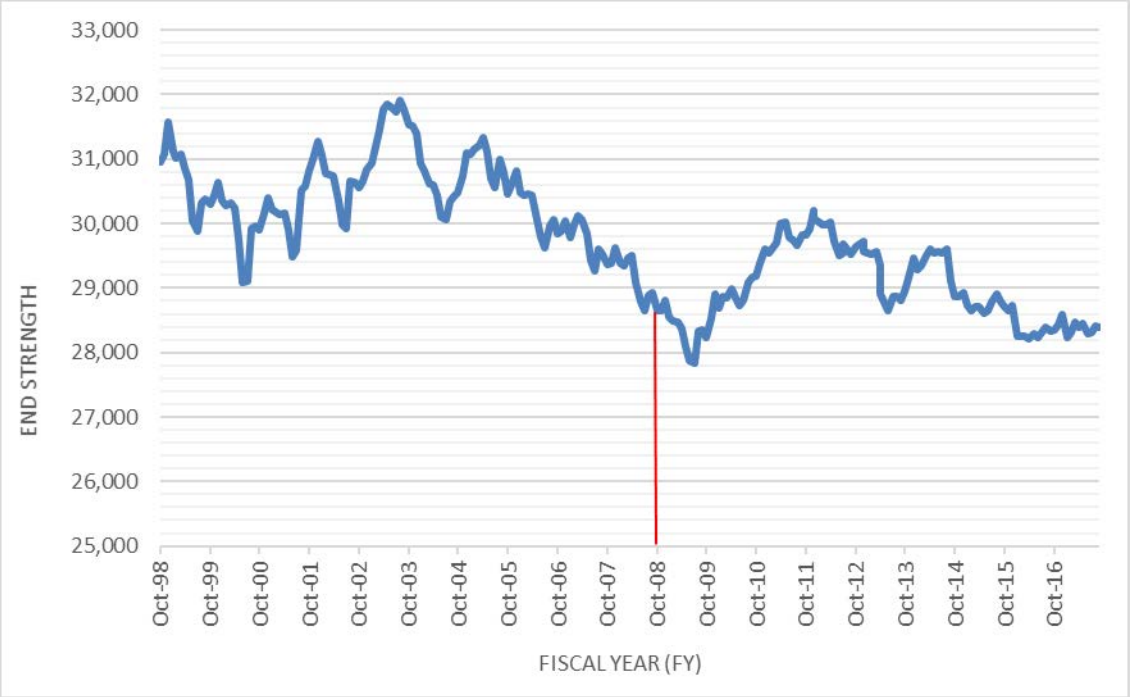

Figure 4. SMCR end strength by year. Adapted from M&RA (2017).

1. Isolating the Cohort

The data received from M&RA (2017) included all of the SelRes, tracking each individual from October 2008, to October 2017, in monthly data retrievals. For this study, the enlisted, NPS, first-term Marines are filtered from the data to establish the final study cohort.

The first filter used in the data is based on YOS. In order for the Marine to be in their first-term, the YOS must equal zero. This indicates that the individual is within their first year of service in the SelRes.

Secondly, the PRESGRDE variable is used to remove any who have the rank of an officer. Only those individuals with an enlisted rank are retained in the data.

Records with PRIOR_SERVICE variable, indicating "Y" for prior service, are removed.

The DODTCPG variable contains a series of codes for each individual Marine. Codes PJ and RE refer to the IRR. Individuals containing any of these two codes are filtered out of the data.

The RCOMPCODE variable also contains a series of codes for each individual Marine. The KP code refers to the Platoon Leader's Course (PLC), which is an officer's training program. These individuals are counted toward the SelRes end strength, but do not fill a specific SMCR billet, and are therefore removed from the data.

The PEBD variable is used for the final, major filter. PEBD is the date in which a Marine entered service, or the date a Marine first arrives to boot camp for initial training. To isolate those Marines who enlisted in FY 09, the individuals selected have a PEBD between 1 October 2008 and 30 September 2009.

This method of extracting the FY 09 cohort is used for each of the first twelve data sets, which correspond to the twelve months in the FY. The total number of Marines who meet the qualifications above is 5,085. From interactions with M&RA, the total number of Marines who enlist in any given year into the SMCR averages between 4,500 to 5,500. Therefore, 5,085 Marines is a reasonable number of enlistees to expect in the data.

2. Collecting Cohort Data

Once the study cohort is established and known, it is necessary to track those individuals throughout the various data sets available, allowing for an accurate understanding of changes in each variable as time and service progresses. To track these changes for each Marine, a transactional data structure is developed. This structure follows only those assigned in the cohort and tracks their monthly records in each data set. Depending on how long a Marine remained in the RC, each individual will have a varying number of records. For example, if a Marine completed six years of service, he or she should have at least 72 observations (12 months multiplied by 6 years).

The transactional data set not only isolates information related solely to our identified cohort, it also allows for a two-tiered approach to identifying service time. The years of service and total satisfactory years of service are tracked throughout the data as

changes occur. This is the primary means of determining if the Marine actually served the amount of time agreed to by contract, signifying successful completion. In addition, creating the transactional data set permits the enumeration of records, which corresponds directly to the number of months in which a person appears in the data. The number of records for each Marine is then compared to the years of service and total satisfactory years of service to ensure the data collected is correct.

Once verification is complete, the cohort is established. Of the 5,085 Marines currently included in the cohort, 519 are removed for having incorrect or missing data. These consist primarily of observations with missing values. For example, 515 of those removed showed zero years of service and zero total satisfactory years but were included in the provided data and more than 50 observations had blank or missing values in most fields. The remaining excluded observations have similar disparities that, if left in the data, would serve only to bias results. The cohort sample size is thus reduced to 4,566.

3. Variable Development

The cohort data is reshaped from the transactional data structure to create a data set with one row (observation) per individual. Some information compressed, including the present grade, marital status, combat and non-combat deployments for instance, change as the Marine is followed throughout their career. To adjust for the changes, new variables are constructed from the original data.

a. Dependent Variable

The dependent variable is successful completion. Successful completion is identified by the number of times a Marine appears in the data. For example, if a Marine appears in the data 96 times or more, meaning the Marine was present in 96 months of data, then the Marine has served eight years in the RC. This is verified using the years of service and total satisfactory years of service data. If there are discrepancies in the number of observations and years of service, the data is further explored to determine the cause. Using the definition of success in Section A of this chapter arbitrates the final determination of success.

b. Independent Variables (Features)

The data compiled consists of 27 independent variables:

- CIV_EDU: Level of civilian education.
- MOS_ENTRY_GROUP: MOS group at entry into the SMCR.
- MOS_CHANGE: Indicated if the Marine changed MOSs during their SMCR service.
- ENTRY_MARITAL: Marital status at entry into the SMCR.
- LAST_MARITAL: Marital status at completion of service, or the last known marital status before separating from the RC.
- RACE: Race of the Marine.
- SEX: Sex of the Marine.
- ETHNIC_GROUP: Ethnic group of the Marine.
- MAX_NUMBER_DEPENDENTS: The maximum number of dependents claimed by the Marine during SMCR service.
- ENTRY_AGE: Age of the Marine at entry into the SMCR.
- AFQT_SCORE: Marine's AFQT score.
- GT_SCORE: Marine's GT score.
- SAME_STATE: Indicated if a Marine's assigned unit is the same as their home of record state. The Marine's home of record zip code (HOR_ZIP) and assigned unit's zip code (UNIT_ZIP) are better measures of commuting distance; however, a significant percentage of the Marines' data did not contain a HOR_ZIP. Therefore, this variable is used as a proxy for commuting distance.

- DISC_ACTION: Indicated if a Marine has received a disciplinary action during their service.
- MAX_CFT: Maximum CFT score the Marine achieved during their service.
- MAX_PFT: Maximum PFT score the Marine achieved during their service.
- PRO_ABOVE_E1: Indicated if a Marine's average Proficiency score is above the cohort's average Proficiency score for the rank of E1.
- PRO_ABOVE_E2: Indicated if a Marine's average Proficiency score is above the cohort's average Proficiency score for the rank of E2.
- PRO_ABOVE_E3: Indicated if a Marine's average Proficiency score is above the cohort's average Proficiency score for the rank of E3.
- PRO_ABOVE_E4: Indicated if a Marine's average Proficiency score is above the cohort's average Proficiency score for the rank of E4.
- CON_ABOVE_E1: Indicated if a Marine's average Conduct score is above the cohort's average Conduct score for the rank of E1.
- CON_ABOVE_E2: Indicated if a Marine's average Conduct score is above the cohort's average Conduct score for the rank of E2.
- CON_ABOVE_E3: Indicated if a Marine's average Conduct score is above the cohort's average Conduct score for the rank of E3.
- CON_ABOVE_E4: Indicated if a Marine's average Conduct score is above the cohort's average Conduct score for the rank of E4.
- DEPLOY_RATE: Deployment rate is determined by adding the number of days the Marine deployed on non-combat and combat deployments, and dividing by the total number of days served. The deployment rate provides

insight into the amount of time a Marine spent deployed compared to the total amount of time served.

- ENTRY_RANK: Rank of the Marine when they entered SMCR service.
- REGION: Marine's HOR region.

D. METHODOLOGY

The goal of this thesis is to identify factors relating to successful completion of first-term enlistments. In order to accomplish this goal, two models are developed. The first, called the pre-enlistment model (P-EM), is developed using information available for a Marine prior to the commencement of military service. The second, referred to as the enlisted model (EM), combines the data used in the pre-enlistment model with data obtained as the Marine progresses through their contract.

Univariate analysis is performed for each of the features for an initial analysis of feature importance. For the P-EM and the EM, two separate models are developed using logistic regression and random forest statistical methods. These methods are used to identify predictors of successful completion.

1. Univariate Analysis

The univariate approach is used for each of the independent variables or features. This is a statistical approach sometimes used for initial exploration and screening of the independent variable. To screen the independent variables, a conservative significance level, α, of less than 0.1 is used.

2. Binary Logistic Regression

Logistic regression, an extension of linear regression, is used to model the probability that an observation falls into one of two categories of the dependent variable, as a function of the independent variables. In this study, the dependent variable, successful completion, is the binary response variable.

A logistic regression model is fit for the P-EM and the EM. Purposeful selection of predictors is used to further reduce the number of predictor variables (Zhang 2016). The predictive quality for each model is determined using a combination of pseudo-R^2 values, misclassification rates, and ROC curves.

3. Random Forest

The random forest method uses bootstrap aggregation, called bagging, to draw samples without replacement from available data. A classification tree is fit to each bootstrapped sample where the candidate variables considered for each split are themselves a random sample from the list of independent variables. The "forest" of trees is then aggregated by averaging an observation's estimated probabilities over the trees in the forest (Breiman 2001).

Random forests can be used to measure feature importance by using the observations not used in a tree's formulation (known as the out-of-bag sample) to count the number times the correct prediction is made. Next, a single variable's values are permuted for each out-of-bag observation, leaving other variables' values the same, again counting the number of correct classifications. The difference between the correct predictions in the original out-of-bag sample and the permuted sample are averaged over the trees in the forest. This difference measures the feature importance, with higher values indicating greater importance (Breiman 2001). Each independent variable is measured in this manner to determine its importance in the presence of other variables.

Though the random forest method has shown robust performance in prediction on a wide range of data, they are not traditionally used to determine relationships between a response variable and predictors. This method is used in this thesis as a secondary means to identify feature importance and to compare the results found using the univariate and binary logistic regression methods (Faraway 2016).

THIS PAGE LEFT INTENTIONALLY BLANK

III. DESCRIPTIVE STATISTICS

As discussed in Chapter II, the study cohort consists of 4,566 NPS Marines who enlisted in FY 09. The data provided by M&RA (2017) is used to identify those individuals who successfully completed their first-term enlistment in the SMCR and those who did not, which becomes the dependent variable (successful completion). The data is also used to develop 27 independent variables that are used as predictors. This chapter describes the data compiled for this study. Each variable created has information describing the cohort and how the variable is associated with the response.

(1) Successful Completion

The data shows that approximately 80 percent of the cohort completed their first-term enlistment successfully, while the remaining 20 percent did not. Figure 5 shows this information, along with the number of observations for each group. This is consistent with other studies regarding SMCR retention.

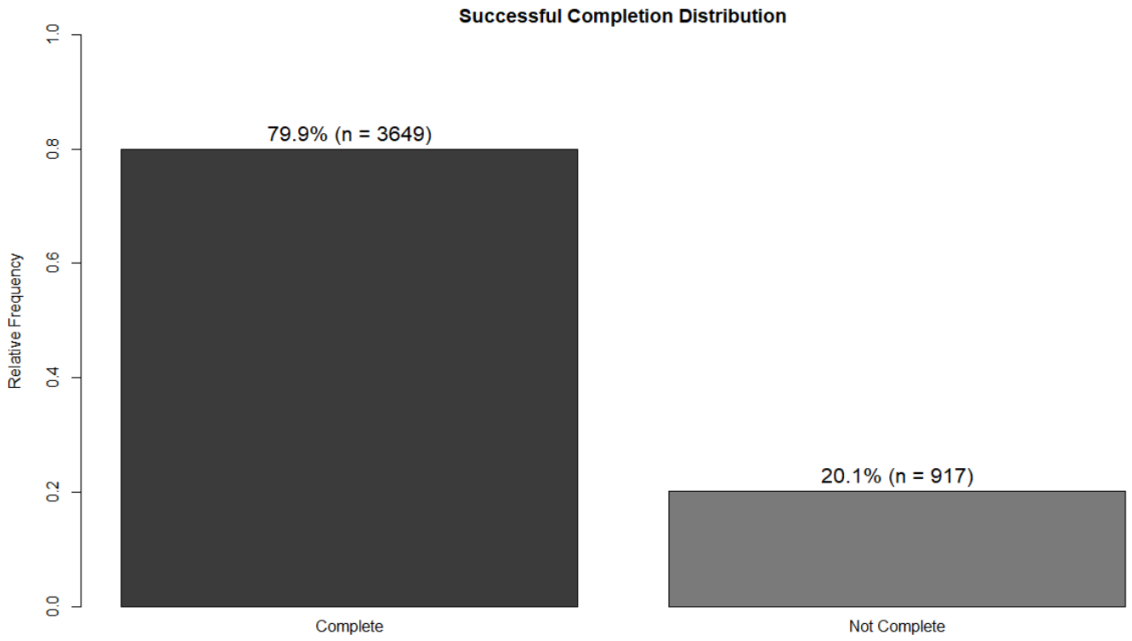

Figure 5. Successful completion distribution for the FY 09 cohort.

(2) Civilian Education Level

The civilian education level variable is tracked for each Marine as they enter the SMCR. This numeric variable is transformed into a categorical variable with three categories: high school (12 years), some college (less than four years of college), and Bachelor's Degree or higher (four years of college or more). Figure 6 shows the distribution of education levels for the cohort.

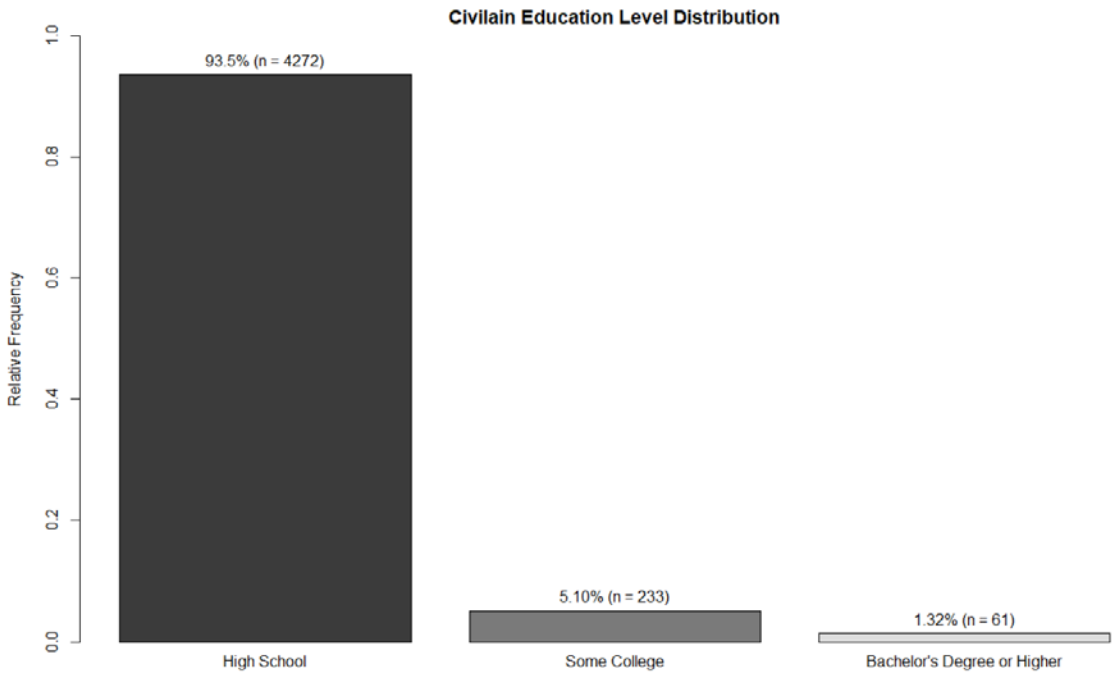

Figure 6. Distribution of civilian education level for the FY 09 cohort.

As expected for first-term enlistees, the overwhelming majority (~ 93.5 percent) of those who entered the SMCR in FY 09 have a high school diploma only. A higher percentage of those with a Bachelor's Degree or higher and those with at least some college completed their first-term enlistment compared to those with only a high school diploma (see Table 2). The table shows the sample size and percentage for each group compared to successful completion.

Table 2. Successful completion by civilian education level.

	CIVILIAN EDUCATION LEVEL		
COMPLETION	**High School**	**Some College**	**Bachelor's or higher**
Complete	3394 (79.4%)	203 (87.1%)	52 (85.2%)
Not Complete	878 (20.6%)	30 (12.9%)	9 (14.8%)

(3) MOS Entry Group

This variable describes the Marine's MOS at entry into the SMCR. Typically, a Marine is assigned an MOS after basic training, called a Primary MOS (PMOS). An MOS is assigned using a four-digit code. The first two digits indicate the occupational field (OccFld), which are groupings of related MOSs. The last two-digits identify the skill-knowledge set. Since dozens of MOSs and OccFlds exist, and to avoid granularity, the MOSs are divided into three distinct groups depending on the OccFld: Combat Arms, Combat Service Support, and Aviation. Table 3 defines the MOS categories divided by OccFld.

Table 3. Entry MOS categories.

GROUP	MOS DESIGNATOR (FIRST TWO DIGITS)
Combat Arms	03, 08, 13, 18
CSS	01, 02, 04, 05, 06, 09, 11, 21, 23, 26, 27, 28, 30, 31, 33, 34, 35, 41, 43, 44, 46, 55, 57, 58, 59
Aviation	60, 61, 62, 63, 64, 65, 68, 70, 72, 80

Figure 7 shows the distribution of MOS categories and Table 4 displays the sample size and percentage of those who completed their first-term enlistment by category. For Combat Arms and CSS, completion rates are near 80 percent, whereas those assigned to an Aviation MOS have a completion rate of 85 percent, though the number of Marines in the Aviation category is much smaller than the others.

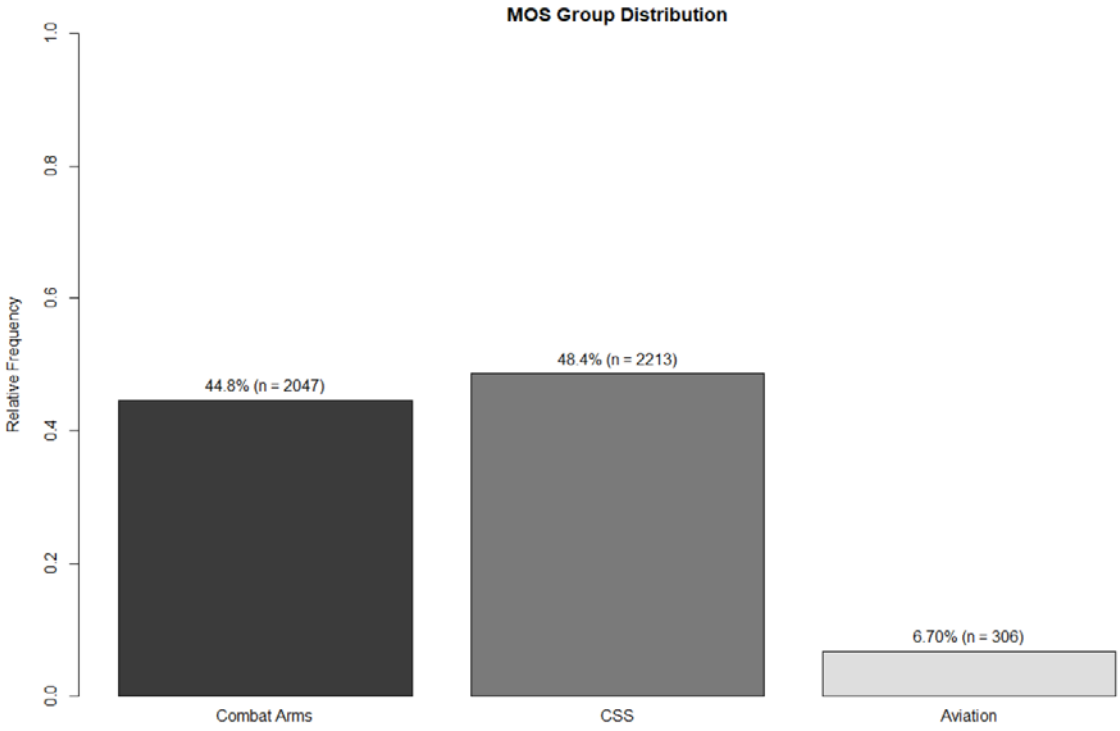

Figure 7. Entry MOS group distribution for the FY 09 cohort.

Table 4. Successful completion by MOS group.

	MOS GROUP		
COMPLETION	***Combat Arms***	***CSS***	***Aviation***
Complete	*1580 (77.2%)*	*1809 (81.7%)*	*260 (84.9%)*
Not Complete	*467 (22.8%)*	*404 (18.3%)*	*46 (15.1%)*

(4) MOS Change

The MOS change variable is a binary variable that identifies those Marines that changed their PMOS during their service in the SMCR. Changing an MOS occurs for a number of reasons, but typically a PMOS change indicates that the Marine received an incentive or was allowed to change specialty based on personal preference. Figure 8 shows the MOS change distribution. The majority (~81.4 percent) of Marines did not change their MOS during their service. A higher percentage of those who changed their PMOS completed their first-term enlistment compared to those who did not change their MOS (see Table 5).

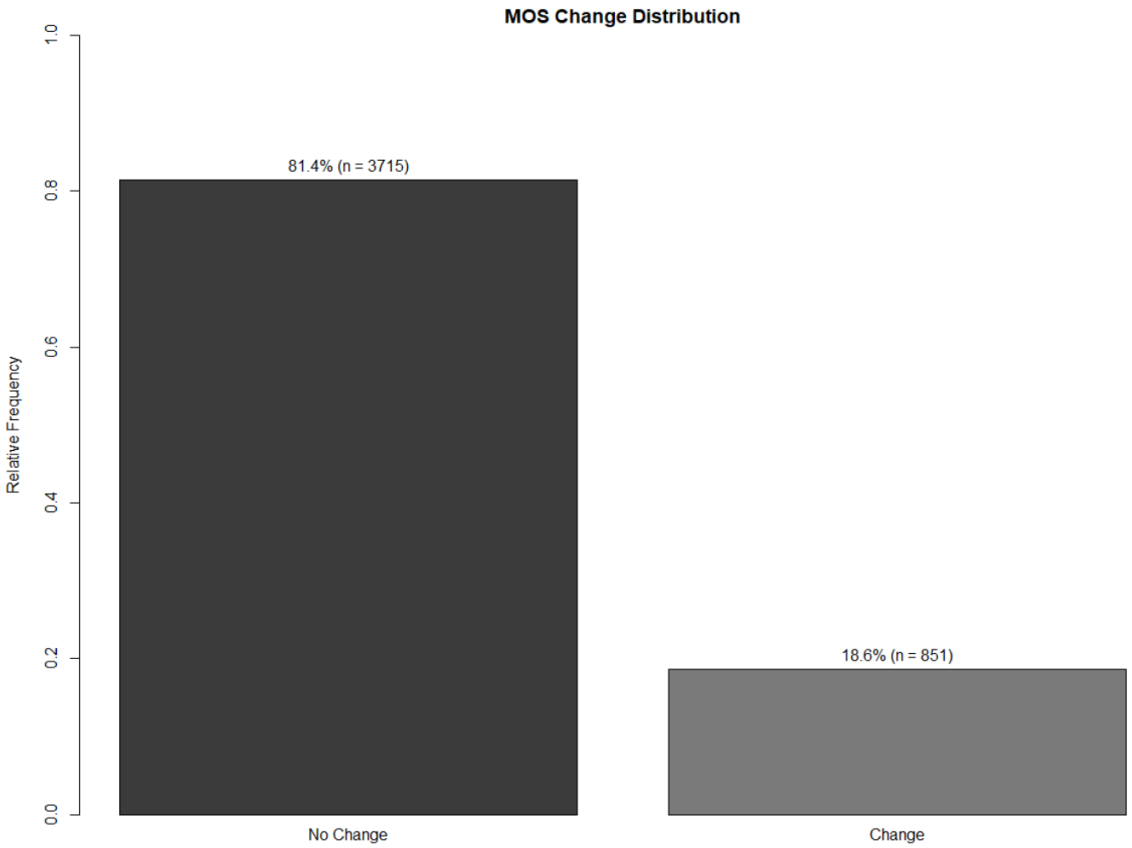

Figure 8. MOS change distribution for the FY 09 cohort.

Table 5. Successful completion by MOS change.

	MOS CHANGE	
COMPLETION	**No Change**	**Change**
Complete	*2895 (77.9%)*	*754 (88.6%)*
Not Complete	*820 (22.1%)*	*97 (11.4%)*

(5) Marital Status

Marital status is divided into two independent variables: entry and last. Entry refers to the marital status as the Marine entered SMCR service. Last refers to the marital status of the Marine at the end of their SMCR service, regardless of how their service ended. Both variables are categorical and divided into three groups. Most Marines entered and remained single during their service. The data shows increased completion percentages for Marines in the Married or Other categories compared to the Single category (see Table 6).

Table 6. Successful completion by marital status.

	MARITAL STATUS					
	At Entry			At End of Service		
COMPLETION	**Single**	**Married**	**Other**	**Single**	**Married**	**Other**
Complete	*3503 (79.8%)*	*140 (83.3%)*	*6 (60%)*	*2634 (76.8%)*	*934 (89.4%)*	*81 (90%)*
Not Complete	*885 (20.2%)*	*28 (16.7%)*	*4 (40%)*	*797 (23.2%)*	*111 (10.6%)*	*9 (10%)*

(6) Race

Race is a categorical variable with six categories: White, American Indian, Asian, Black or African American, Pacific Islander, and a last category that represents those who declined to respond when asked for their race at entry into the SMCR. Table 7 shows the successful completion percentages for each category. Most categories had completion rates of near 80 percent.

Table 7. Successful completion by race.

	RACE					
COMPLETION	**White**	**American Indian**	**Asian**	**Black or African American**	**Pacific Islander**	**Declined to Respond**
Complete	3098 (79.9%)	14 (82.3%)	129 (85.4%)	329 (77.2%)	26 (74.3%)	53 (86.9%)
Not Complete	778 (20.1%)	3 (17.6%)	22 (14.6%)	97 (22.8%)	9 (25.7%)	8 (13.1%)

Most who enlisted in the SMCR in FY 09 indicated they were white (~84.9 percent). The second highest category consists of those who indicated Black or African American (~9.32 percent). Figure 9 shows the distribution of race.

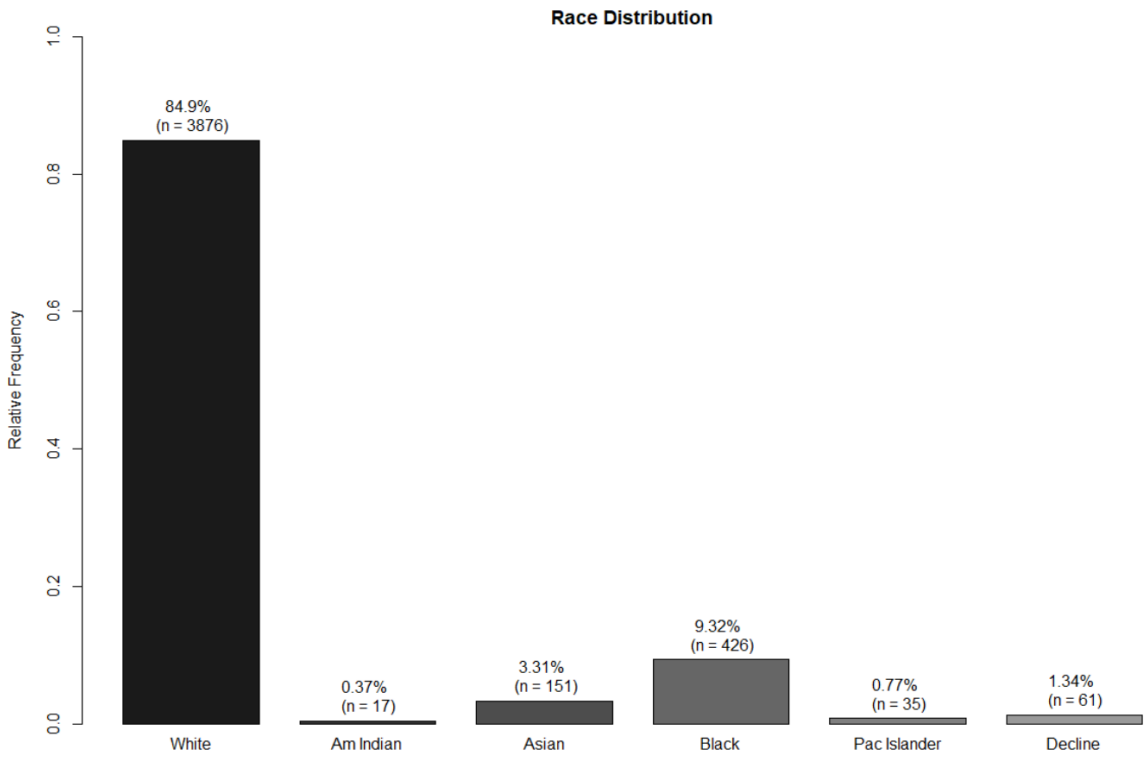

Figure 9. Race distribution for the FY 09 cohort.

(7) Ethnic Group

Ethnic group is a categorical variable divided into three groups: Hispanic, Non-Hispanic, and a third category for those who either declined to respond or did not respond when asked to indicate their ethnic group at entry into the SMCR. Figure 10 shows the distribution of ethnic groups, and Table 8 shows the completion rate between each group is nearly 80 percent. Since the majority of the FY 09 cohort declined to respond or did not respond (~73.8 percent), the ethnic group variable is not expected to be a significant predictor or an influential variable in interpreting the results.

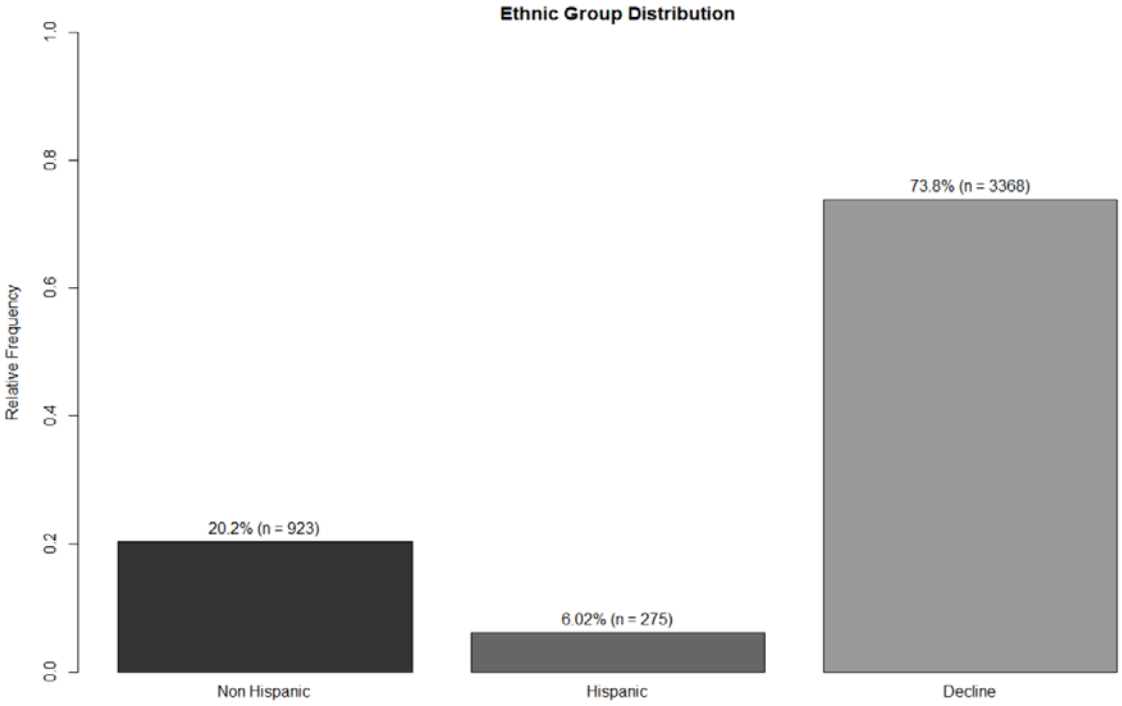

Figure 10. Ethnic group distribution for the FY 09 cohort.

Table 8. Successful completion by ethnic group.

	ETHNIC GROUP		
COMPLETION	**Non-Hispanic**	**Hispanic**	**Declined to Respond**
Complete	*743 (80.5%)*	*222 (80.7%)*	*2684 (79.7%)*
Not Complete	*180 (19.5%)*	*53 (19.3%)*	*684 (20.3%)*

(8) Sex

The sex variable is divided into male or female. The majority of the FY 09 cohort (~97.1 percent) are males and the remaining (~2.9 percent) are female. Males had a slightly higher completion rate than females, though the sample size for females is relatively small (see Table 9).

Table 9. Successful completion by sex.

	SEX	
COMPLETION	***Male***	***Female***
Complete	*3554 (80.2%)*	*95 (72%)*
Not Complete	*880 (19.8%)*	*37 (28%)*

(9) Number of Dependents

This variable is calculated by tracking the maximum number of dependents a service member had during their time affiliated with the SMCR. Figure 11 shows the distribution of dependents.

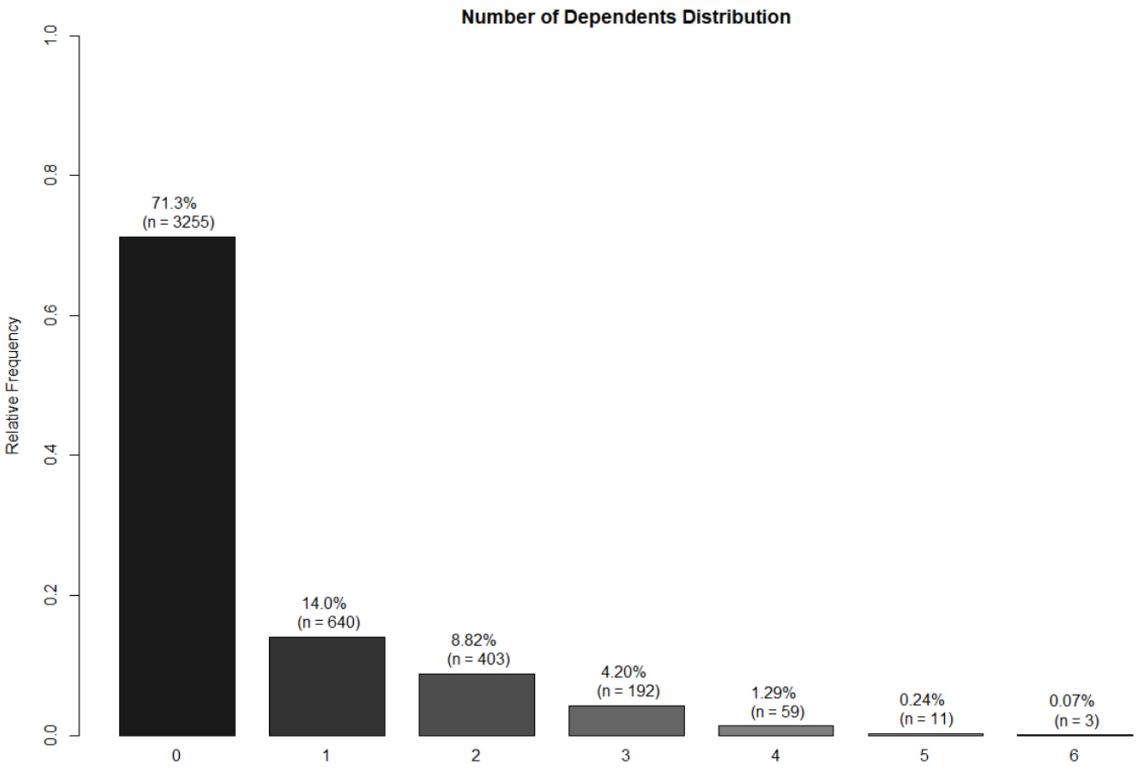

Figure 11. Maximum number of dependents distribution for the FY 09 cohort.

Because this variable can change throughout a Marine's service in the SMCR, the maximum number of dependents reported by the Marine is chosen to indicate the number of dependents the Marine is supporting financially. It is expected that those with a higher number of dependents are more likely to complete SMCR service than those with fewer dependents. Table 10 shows that Marines who reported having two dependents or more completed their first-term at higher rates than those with fewer dependents.

Table 10. Successful completion by the maximum number of dependents.

	NUMBER OF DEPENDENTS						
COMPLETION	***0***	***1***	***2***	***3***	***4***	***5***	***6***
Complete	*2489 (76.5%)*	*551 (86.1%)*	*367 (91.1%)*	*174 (91%)*	*52 (88.1%)*	*11 (100%)*	*3 (100%)*
Not Complete	*766 (23.5%)*	*89 (13.9%)*	*36 (8.90%)*	*18 (9%)*	*7 (11.9%)*	*0 (0%)*	*0 (0%)*

(10) Entry Age

This numeric variable is the age of the Marine at the beginning of their SMCR service. The average age of the FY 09 cohort is 19.68 years. Figure 12 shows the entry age distribution.

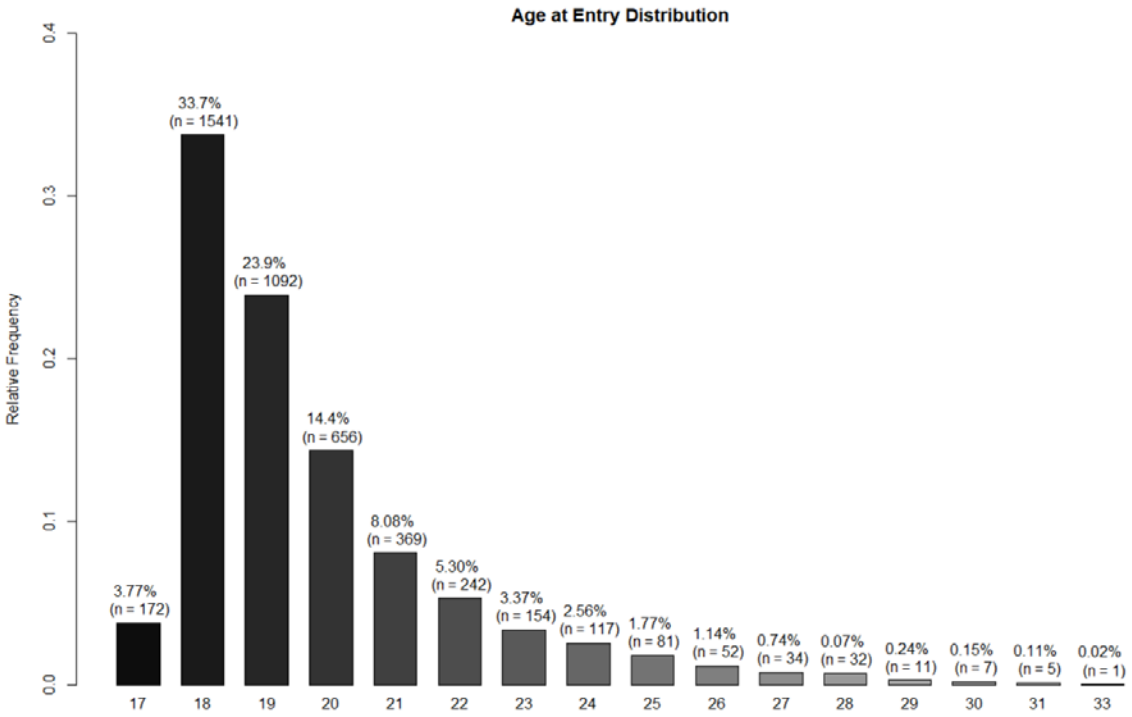

Figure 12. Entry age distribution for the FY 09 cohort.

Table 11 shows completion information for the FY 09 cohort. The data in this table is divided into three age groups, with the majority consisting of those from age 17–22. The data shows a relatively equal successful completion rate across the three age groups.

Table 11. Successful completion by entry age.

	ENTRY AGE		
COMPLETION	***17 - 22***	***23 - 28***	***29 - 33***
Complete	*3233 (79.4%)*	*396 (84.3%)*	*20 (83.3%)*
Not Complete	*839 (20.6%)*	*74 (15.7%)*	*4 (16.7%)*

(11) AFQT Score

The AFQT score is a qualification test score received by every individual who enters the U.S. Military, including the SMCR. The maximum achievable score on the AFQT is a 99. The mean score for the FY 09 cohort is 64.22, with a median score of 64. Approximately 50.4 percent (n = 2302) scored below average, with the remaining 49.6 percent (n = 2264) scoring at or above average. A larger percentage of those with an average or above average score did complete their first-term enlistment compared to those who scored below average (see Table 12).

Table 12. Successful completion by AFQT score.

	AFQT SCORE	
COMPLETION	***Below Average***	***At or Above Average***
Complete	*1762 (76.5%)*	*1887 (83.3%)*
Not Complete	*540 (23.5%)*	*377 (16.7%)*

(12) GT Score

The GT is the general technical score that each Marine receives after completing qualification testing. A GT score is used for MOS placement by the USMC once a Marine completes the basic entry requirements. The mean GT score for the FY 09 cohort is 110.39, with a median score of 110. Approximately 50.7 percent scored below the average GT score, with the remaining 49.3 percent scoring at or above the mean. These above and below average score percentages are similar to the AFQT scores described above.

Approximately 45.3 percent (n = 2069) of the cohort scored at or above average in both the AFQT and GT scores. Approximately 42.2 percent (n = 1925) scored below average in both the AFQT and GT scores. Table 13 shows a slightly higher percentage of completion for those with an average or higher GT score.

Table 13. Successful completion by GT score.

	GT SCORE	
COMPLETION	Below Average	At or Above Average
Complete	1789 (77.2%)	1860 (82.7%)
Not Complete	528 (22.8%)	389 (17.3%)

(13) HOR and Unit State

This binary variable indicates whether a Marine is assigned an SMCR billet in a unit located in their home state, or state listed as their HOR. As discussed in Chapter II, this variable is used as a proxy for commuting distance. If a Marine is assigned a billet in a unit located in the same state as their HOR, it is assumed the commuting distance is less than if a Marine is assigned a billet in a unit not located in their HOR state.

The standard practice of billet assignments is to place a Marine in a unit as near to their HOR as possible, dependent on the Marine's MOS and the unit or billet availability. Approximately 95.2 percent (n = 4345) of the FY 09 cohort were assigned billets in the same state as their HOR. The remaining 4.8 percent (n = 216) were not assigned to a unit in their HOR state. Table 14 shows similar completion rates for each category.

Table 14. Successful completion by HOR state and unit assignment.

	HOR & UNIT STATE	
COMPLETION	Different State	Same State
Complete	171 (79.2%)	3477 (80%)
Not Complete	45 (20.8%)	868 (20%)

(14) Disciplinary Action

Disciplinary action is a binary variable indicating whether a Marine received a disciplinary action of some kind during their SMCR service. A disciplinary action includes actions taken that remain in a Marine's record of service, including Non-Judicial Punishment (NJP) and courts martial. The provided data only tracks if a Marine received a disciplinary action but not the level of action or if the Marine received more than one during their service.

Approximately 94.1 percent (n = 4295) of the FY 09 cohort did not receive a disciplinary action during their SMCR service. 5.87 percent (n = 268) did receive at least one disciplinary action at some point during their service. Table 15 shows slightly higher completion rates for those who did not receive a disciplinary action.

Table 15. Successful completion by disciplinary action.

	DISCIPLINARY ACTION	
COMPLETION	***No Action***	***Action***
Complete	*3445 (80.2%)*	*202 (75.4%)*
Not Complete	*850 (19.8%)*	*66 (24.6%)*

(15) CFT Scores

CFT score is a categorical variable that indicates a Marine's maximum score on the CFT during their service in the SMCR in terms of class. Fitness test scores are divided into classes depending on the Marines' numeric score during the tests. A score of 235 and above, with a maximum score of 300, is a first class; a 185 to 235 score is a second class; and a score of 184 and below is a third class.

Approximately 96 percent (n = 4380) of the FY 09 cohort fall into the first class score category. Approximately 4 percent (n = 186) had either a second class, third class, or no recorded score. Table 16 shows a much higher completion rate for those who scored in the first-class category.

Table 16. Successful completion by CFT score.

	CFT SCORE			
COMPLETION	**1st Class**	**2nd Class**	**3rd Class**	**NA**
Complete	*3573 (81.6%)*	*20 (40%)*	*22 (39.3%)*	*34 (42.5%)*
Not Complete	*807 (18.4%)*	*30 (60%)*	*34 (60.7%)*	*46 (57.5%)*

(16) PFT Scores

Similar to the CFT score, the PFT score variable is a categorical variable divided by class. The class delineations are the same as for the CFT. Each Marine in the cohort has at least one recorded PFT score. Approximately 75 percent (n = 3440) have a first class PFT score, 20 percent have a second class score, and 5 percent have a third class score. Table 17 shows higher completion rates for those in the first class and second class categories.

Table 17. Successful completion by PFT score.

	PFT SCORE		
COMPLETION	**1st Class**	**2nd Class**	**3rd Class**
Complete	*2876 (83.6%)*	*648 (70.9%)*	*125 (59.8%)*
Not Complete	*564 (16.4%)*	*266 (29.1%)*	*84 (40.2%)*

(17) Deployment Rate

As discussed in Chapter II, deployment rate is calculated by dividing the total number of days a Marine spent deployed by the total number of days the Marine served in the SMCR. Both the non-combat and combat deployment periods are considered.

The USMC policy for the AC on deployment-to-dwell, or the time spent deployed compared to the time spent at the CONUS unit, is 1-to-2 (CMC 2014). For example, if a Marine is deployed for a period of six months, the target "dwell" period is at least 12 months before being deployed again, or 1/3 deployed to 2/3 dwell. If a Marine has a deployment rate greater than or equal to 1/3, they are considered to have a high deployment

rate. Figure 13 shows the distribution of deployment rates for the FY 09 cohort. Table 18 shows completion information as it relates to deployment rate.

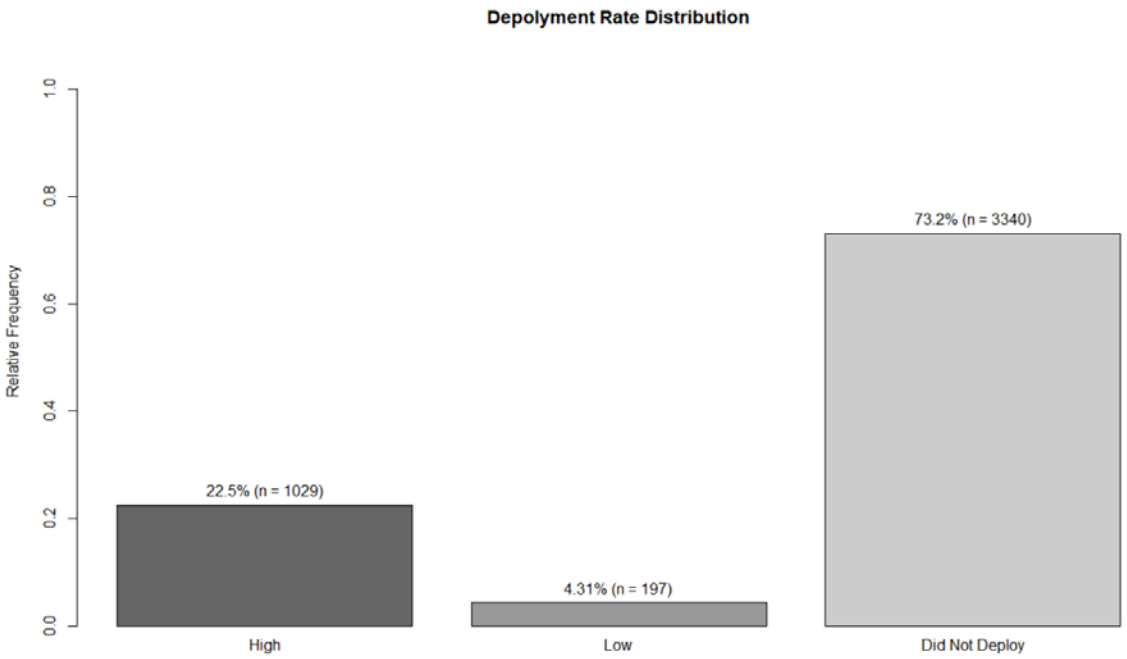

Figure 13. Deployment rate distribution for the FY 09 cohort

Only 26.9 percent (n = 1226) of the FY 09 cohort were deployed for any period of time. Of those who deployed, approximately 84 percent are considered to have a high deployment rate. The high deployment rates remain relatively similar across the MOS groups. Of those who have a Combat Arms MOS, 22.9 percent have high deployment rates; those with a CSS MOS have 23.9 percent with high deployment rates; and those with an Aviation MOS have 12.8 percent with high deployment rates. Table 18 shows those who did deploy have a higher completion percentage compared to those who did not deploy. This variable is expected to be significant in the model.

Table 18. Successful completion by deployment rate.

	DEPLOYMENT RATE		
COMPLETION	**High**	**Low**	**Did Not Deploy**
Complete	*937 (91.1%)*	*179 (90.9%)*	*2533 (75.8%)*
Not Complete	*92 (8.90%)*	*18 (9.10%)*	*807 (24.2%)*

(18) Region

Region is a categorical variable based on the Marine's HOR state. The six categories are determined using the U.S. Census Bureau regional divisions with an added category for those Marines that have no information regarding their HOR state. Table 19 shows relatively equal completion rates across the different regions. Figure 14 depicts the U.S. Census Bureau's regional map, and Figure 15 shows the distribution of the regional variable.

Table 19. Successful completion by region.

	REGION					
COMPLETION	**Pacific**	**West**	**Mid-West**	**South**	**Northeast**	**No Information**
Complete	*16 (80%)*	*714 (79.5%)*	*706 (80.4%)*	*1241 (79.9%)*	*649 (79.3%)*	*323 (80.9%)*
Not Complete	*4 (20%)*	*184 (20.5%)*	*172 (19.6%)*	*312 (20.1%)*	*169 (20.7%)*	*76 (19.1%)*

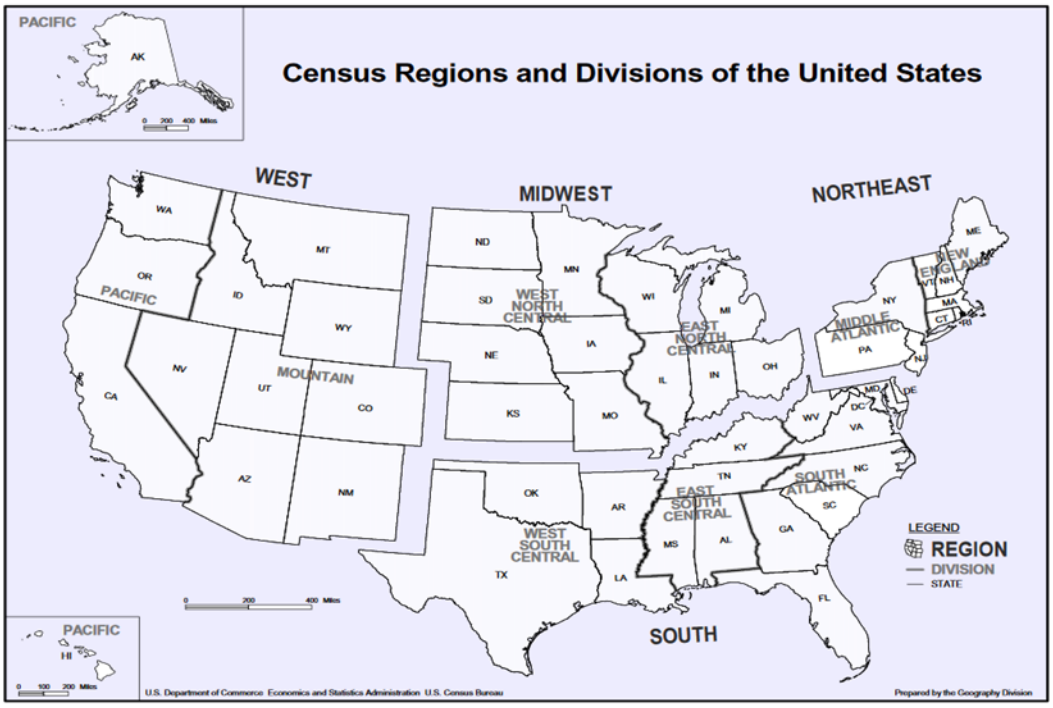

Figure 14. U.S. Census Bureau regional division map. Source: U.S. Census Bureau (2010).

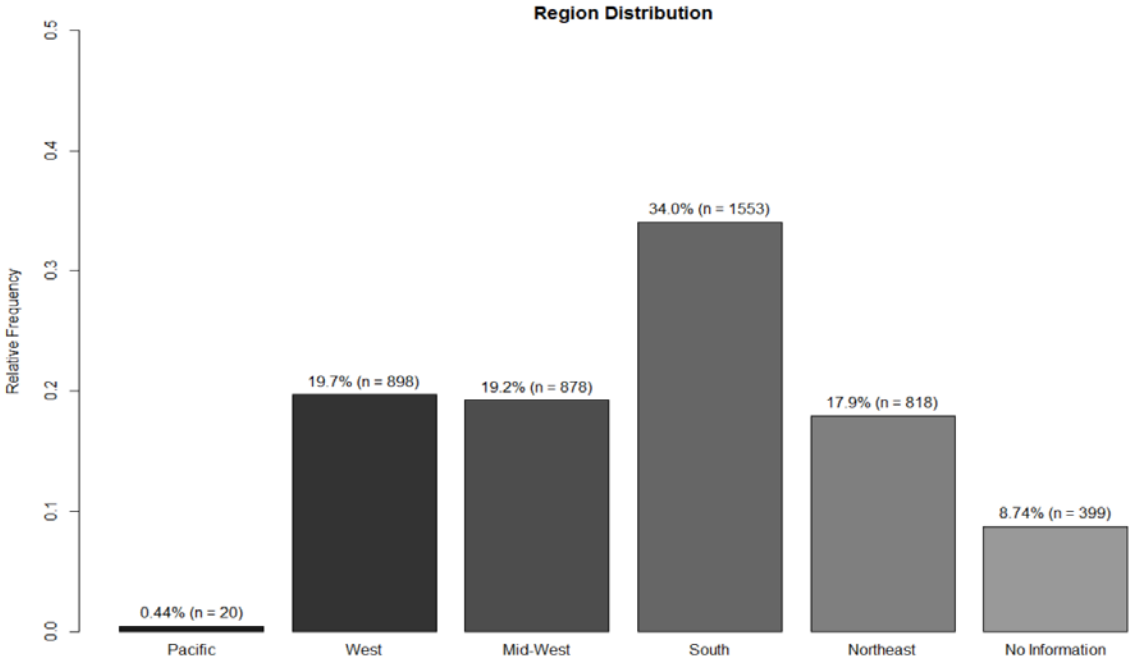

Figure 15. Region distribution for the FY 09 cohort.

(19) Above-Average PRO and CON

Four variables are created for both Proficiency and Conduct scores for each rank from E1 to E4. These variables are categorical by indicating whether a Marine scored above or below the mean PRO or CON score for their respective rank. The mean score is calculated by averaging the PRO or CON score by rank and year. For example, the average PRO score for a Marine when they are an E1 in FY 10 is calculated by averaging all of the PRO scores for all Marines who are at the rank of E1 in FY 10. This method ensures the Marines are not compared to Marines of other ranks and different year groups. Figures 16 and 17 show the distribution of PRO and CON scores for the FY 09 cohort.

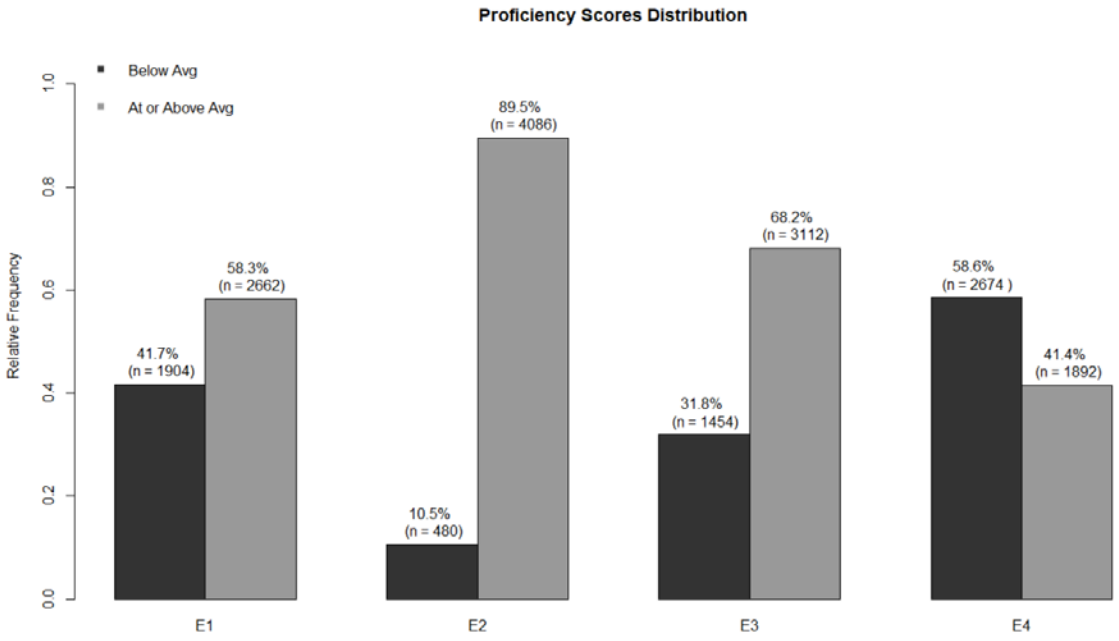

Figure 16. PRO scores by rank for the FY 09 cohort.

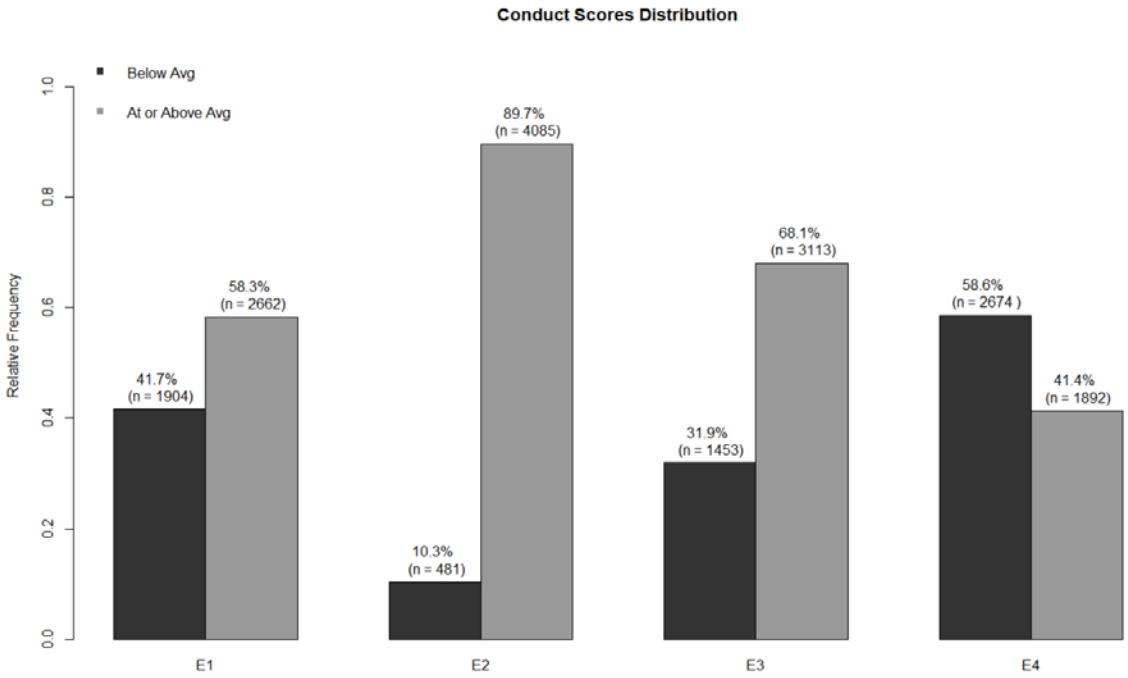

Figure 17. CON scores by rank for the FY 09 cohort.

Approximately 90.1 percent of the time, on average across ranks, if a Marine is rated as at or above average in Proficiency, they are also rated as at or above average for Conduct. Tables 20 and 21 show completion information as it relates to PRO and CON scores. Other than for the rank of E1, the data shows higher completion rates for those who scored above the average.

Table 20. Successful completion by PRO scores.

	PROFICIENCY SCORES							
	E1		E2		E3		E4	
COMPLETION	Above	Below	Above	Below	Above	Below	Above	Below
Complete	2097 (78.8%)	1552 (81.5%)	3331 (81.5%)	318 (66.3%)	2703 (86.9%)	846 (62.5%)	1826 (96.5%)	1823 (68.2%)
Not Complete	565 (21.2%)	352 (18.5%)	755 (18.5%)	162 (33.7%)	409 (13.1%)	508 (37.5%)	66 (3.5%)	851 (31.8%)

Table 21. Successful completion by CON scores.

	CONDUCT SCORES							
	E1		E2		E3		E4	
COMPLETION	Above	Below	Above	Below	Above	Below	Above	Below
Complete	2099 (78.8%)	1550 (81.5%)	3342 (81.6%)	307 (65.2%)	2720 (87.5%)	929 (63.8%)	1824 (96.6%)	1825 (68.2%)
Not Complete	565 (21.2%)	352 (18.5%)	753 (18.4%)	164 (34.8%)	390 (12.5%)	527 (36.2%)	65 (3.4%)	852 (31.8%)

THIS PAGE INTENTIONALLY LEFT BLANK

IV. MODELING AND ANALYSIS

Each of the models are fit using a training data set, which includes a random selection of approximately 80 percent (3,652 observations) of the original data. A test set is constructed using the remaining observations in the original data (914 observations). The test set is used after model development to assess the predictive ability of each model.

The P-EM uses data available prior to a Marine's enlistment in the SMCR, which consists of nine independent variables:

- Civilian education level
- Entry marital status
- Race
- Sex
- Ethnic group
- Entry age
- AFQT score
- GT score
- Region

The EM uses the data in the P-EM model and data collected throughout a Marine's service in the SMCR. In addition to the variables in the P-EM model listed above, the EM also includes the following predictors, for a total of 27 independent variables:

- MOS group
- MOS change
- Maximum number of dependents

- Last marital status
- Disciplinary action
- Max PFT score
- Max CFT score
- Same state
- Above average PRO score (rank of E1)
- Above average PRO score (rank of E2)
- Above average PRO score (rank of E3)
- Above average PRO score (rank of E4)
- Above average CON score (rank of E1)
- Above average CON score (rank of E2)
- Above average CON score (rank of E3)
- Above average CON score (rank of E4)
- Deployment rate
- Entry rank

A. UNIVARIATE ANALYSIS

Univariate analysis is performed exploring the relationship between each independent variable and the probability of successful completion. Separate logistic regressions are fit, one for each independent variable, based on the training set. The results of these fits are given in Table 22. For categorical variables such as MOS group, one category, the reference category (denoted by "*Ref*" in Table 22), has coefficient set to zero. For MOS Group the reference category is Combat Arms. The estimates of the

coefficients corresponding the other categories, their standard errors, and p-values for the two-sided test of the null hypothesis that the coefficient is zero are given in Table 22.

Table 22. Results of univariate analysis.

Independent variable	*Estimate*	*Standard Error*	*P-Value*
Civilian Education Level			
High School	*Ref*	*Ref*	*Ref*
Some College	0.69	0.24	*0.0038**
Bachelor's or Higher	0.17	0.34	*0.6326*
MOS Group			
Combat Arms	*Ref*	*Ref*	*Ref*
CSS	0.30	0.08	0.0004*
Aviation	0.59	0.17	0.0022*
MOS Change			
No Change	*Ref*	*Ref*	*Ref*
Change	0.79	0.01	< 0.0001*
Entry Marital Status			
Single	*Ref*	*Ref*	*Ref*
Married	0.19	0.23	0.4115
Other	-1.67	0.12	0.0281*
Race			
White	*Ref*	*Ref*	*Ref*
American Indian	*-0.01*	0.54	*0.9790*
Asian	*0.14*	0.24	*0.5469*
Black or African American	*-0.16*	0.14	*0.2487*
Pacific Islander	*-0.06*	0.50	*0.8923*
Declined to Respond	*0.18*	0.39	*0.6435*
Sex			
Female	*Ref*	*Ref*	*Ref*
Male	0.35	0.05	0.1067
Ethnic Group			
Non-Hispanic	*Ref*	*Ref*	*Ref*
Hispanic	0.04	0.15	0.8062
Declined to Respond	0.04	0.34	0.6435
Max Number of Dependents	0.43	0.05	< 0.0001*
Entry Age	0.06	0.04	0.0011*
AFQT Score	0.01	0.002	< 0.0001*
GT Score	0.01	0.003	< 0.0001*
Same State			
No	*Ref*	*Ref*	*Ref*
Yes	-0.18	0.21	0.3652
Disciplinary Action			
No	*Ref*	*Ref*	*Ref*
Yes	-0.34	0.11	0.0361*
Max CFT			
First Class	*Ref*	*Ref*	*Ref*

Independent variable	*Estimate*	*Standard Error*	*P-Value*
Second Class	-2.10	0.25	< 0.0001*
Third Class	-1.86	0.38	< 0.0001*
Max PFT			
First Class	*Ref*	*Ref*	*Ref*
Second Class	-0.75	0.43	< 0.0001*
Third Class	-1.14	0.12	< 0.0001*
Proficiency Above Average E1			
No	*Ref*	*Ref*	*Ref*
Yes	-0.19	0.14	0.0218*
Proficiency Above Average E2			
No	*Ref*	*Ref*	*Ref*
Yes	0.67	0.34	< 0.0001*
Proficiency Above Average E3			
No	*Ref*	*Ref*	*Ref*
Yes	1.23	0.32	< 0.0001*
Proficiency Above Average E4			
No	*Ref*	*Ref*	*Ref*
Yes	2.47	0.02	< 0.0001*
Conduct Above Average E1			
No	*Ref*	*Ref*	*Ref*
Yes	-0.19	0.23	0.0237*
Conduct Above Average E2			
No	*Ref*	*Ref*	*Ref*
Yes	0.72	0.43	< 0.0001*
Conduct Above Average E3			
No	*Ref*	*Ref*	*Ref*
Yes	1.31	0.21	< 0.0001*
Conduct Above Average E4			
No	*Ref*	*Ref*	*Ref*
Yes	2.46	0.17	< 0.0001*
Deployment Rate	1.23	0.19	< 0.0001*
Entry Rank			
E1	*Ref*	*Ref*	*Ref*
E2	0.39	0.01	< 0.0001*
Region			
Pacific	*Ref*	*Ref*	*Ref*
West	-0.23	0.45	0.7122
Mid-West	-0.16	0.43	0.7929
South	-0.10	0.46	0.8713
Northeast	-0.15	0.48	0.8155
No Information	-0.04	0.45	0.9484

* indicates significance (p-value < 0.1)

The univariate analysis shows that Marines with at least some years of college education (including an Associate's Degree) are more likely to successfully complete their first-term enlistment compared to those with only a high school education.

Marines assigned to CSS and Aviation MOSs are more likely to complete their first-term when compared to those in Combat Arms MOSs. Also, Marines who changed their MOS at some point in their enlistment are more likely to complete their first-term enlistment than those who did not.

Those who had a marital status at entry other than single or married are less likely to complete their first-term than those who were single.

Marines who did not receive a disciplinary action are more likely to complete their first-term enlistment than those who did receive a disciplinary action at some point during their career.

For both CFT and PFT scores, Marines who attained a second or third class score are less likely to complete their first-term enlistment than those who attained a first class score.

For the proficiency and conduct variables, and for each rank level (E1 to E4), Marines who scored at or above average are more likely to complete their first-term enlistment than those who scored below average.

Those Marines who entered the SMCR with the rank of E2 are more likely to complete their first-term enlistment compared to those who entered with the rank of E1.

We note that even treating numeric independent variables as is, with no transformation, all logistic regression fits give p-values much less than 0.01 for the null hypothesis that the coefficient corresponding to the numeric independent variable is zero. This is strong evidence that each one of the independent numeric variables separately contributes to estimating the probability of successful completion.

Overall, 21 independent variables have p-values < 0.1 excluding race, sex, ethnic group, same state and region. These 21 independent variables include four of the nine independent variables in P-EM and five of the 27 independent variables in EM.

B. PRE-ENLISTMENT MODEL (P-EM)

The purpose of the P-EM is to find predictors of successful completion based on the information known about a Marine prior to entering the SMCR. Based on the univariate analysis, five of the nine independent variables in the P-EM are potential predictors of successful completion. Since there are relatively few independent variables, the logistic regression and random forest models use all available independent variables in their respective models.

A logistic regression model is fit using the training set data. This model is refined using a modified backwards elimination procedure called purposeful selection (Zhang 2016). The results of the variable selection show that AFQT score, entry age, sex, and entry marital status are significant (p-value < 0.05). Table 23 summarizes the independent variables, coefficient estimates, standard errors, and p-values for each significant variable in the final model.

Table 23. Logistic regression final model for the P-EM.

Independent variable	***Estimate***	***Standard Error***	***P-Value***
Entry Marital Status			
Single	*Ref*	*Ref*	*Ref*
Married	-0.01	0.32	0.9456
Other	-1.65	0.12	0.0228*
Sex			
Female	*Ref*	*Ref*	*Ref*
Male	0.43	0.21	0.0445*
Entry Age	0.06	0.03	0.0053*
AFQT Score	0.01	0.01	< 0.0001*

* indicates significance (p-value < 0.05)

A likelihood ratio test compared the full model (containing all independent variables) to the final model containing only the independent variables described in Table 23. The test resulted in a non-significant p-value (p-value > 0.05), meaning the two models are not statistically different. The smaller model, containing only the four independent variables (entry marital status, sex, entry age, and AFQT score), is chosen for the principal of parsimony (Zhang 2016).

The model shows that as the age of the Marine at entry increases, the likelihood of successful first-term completion increases. For every unit change in the entry age, the log-odds of successful completion changes by 0.06. A higher AFQT score also signifies an increase in the likelihood of successful completion. For every unit change in the AFQT score, the log-odds of successful completion increases by 0.01. These unit increases in log-odds are assuming all other independent variable values remain constant. A Marine with an entry marital status of "other" is significantly less likely to successfully complete their first-term enlistment when compared to Marines who are listed as "single." The odds for a Marine in the "other" category are about five times lower than the odds for those Marines in the "single" category, in terms of successful completion. Marines listed as male have a higher odds ratio of completion compared to females. The odds for males are about 55 percent higher than the odds for females, in terms of successful completion.

The logistic regression model is compared to a random forest model in terms of variable importance, receiver operating characteristic (ROC) curves, and correct classification percentage. The random forest model determines feature importance as described in Chapter II. The test set of data is used for evaluation of correct classification and ROC curves.

Figure 18 shows the ROC curve comparison between the logistic regression (LR) and the random forest (RF) models. The LR model has an 80.5 percent correct test set prediction rate. The RF has an 80.2 correct test set prediction rate. Both models seem to be predicting correctly at nearly the same rate, and the ROC curve shows little difference between the models. There is a near linear relationship between sensitivity (true positive rate) and 1—specificity (false positive rate), showing that the models are not predicting very accurately.

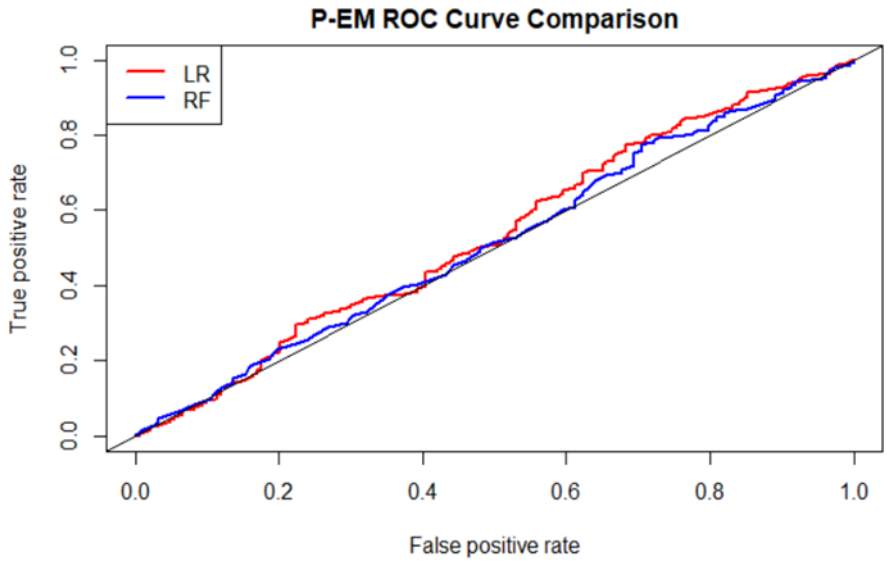

Figure 18. ROC curve comparisons of the P-EM LR and RF models.

Figure 19 shows the feature importance of the RF model. Two of the top five important variables in the RF model match that of the LR model's significant variables (AFQT score and entry age). Marital status at entry and sex are found to be less important to the RF model compared to the LR model.

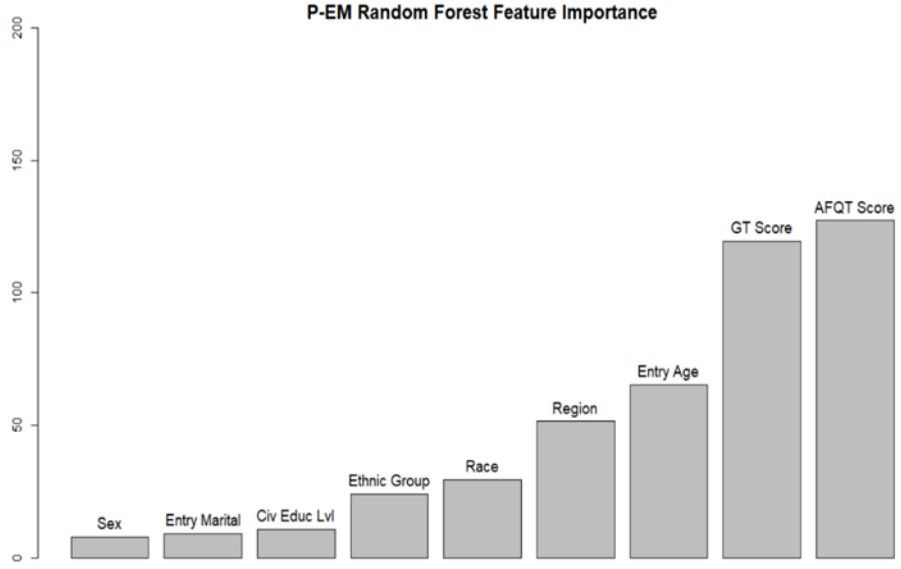

Figure 19. P-EM RF feature importance plot.

C. ENLISTED MODEL (EM)

The purpose of the EM is to identify predictors of first-term completion based on the information available for a Marine during their service in the SMCR. This includes the data used in the P-EM.

Similar to the P-EM, a LR model is fit, using all the independent variables available, is used to develop a model fit of the training data set. The LR model is refined manually with the same criteria used to fit the P-EM LR model. The results of the variable selection show that MOS change, entry marital status, maximum number of dependents, max CFT score, proficiency score above average E1, conduct score above average E2, conduct score above average E4, sex, and MOS group are significant (p-value < 0.05). Table 24 summarizes the independent variables, coefficient estimates, standard errors, and p-values for each significant variable in the final model.

Table 24. Logistic regression final model for the EM.

Independent variable	***Estimate***	***Standard Error***	***P-Value***
MOS Group			
Combat Arms	*Ref*	*Ref*	*Ref*
CSS	0.08	0.02	0.3787
Aviation	0.60	0.23	0.0098*
MOS Change			
No Change	*Ref*	*Ref*	*Ref*
Change	0.46	0.04	0.0019*
Entry Marital Status			
Single	*Ref*	*Ref*	*Ref*
Married	0.23	0.32	0.3125
Other	-1.34	0.12	0.0063*
Sex			
Female	*Ref*	*Ref*	*Ref*
Male	0.73	0.27	0.0041*
Max Number of Dependents	0.43	0.07	< 0.0001*
Max CFT			
First Class	*Ref*	*Ref*	*Ref*
Second Class	-0.87	0.33	0.0056*
Third Class	-0.35	0.34	0.0005*
Proficiency Above Average E4			
No	*Ref*	*Ref*	*Ref*
Yes	1.28	0.22	< 0.0001*

Independent variable	*Estimate*	*Standard Error*	*P-Value*
Conduct Above Average E3			
No	*Ref*	*Ref*	*Ref*
Yes	0.40	0.13	< 0.0001*
Conduct Above Average E4			
No	*Ref*	*Ref*	*Ref*
Yes	1.29	0.22	< 0.0001*

* indicates significant predictor (p-value < 0.05)

A likelihood ratio test is performed on the full model and the model using only the independent variables described in Table 24. A non-significant p-value is obtained, suggesting that the two models are not statistically different.

The LR model shows that for every unit change in the maximum number of dependents, or the maximum number of dependents claimed by a Marine during their SMCR service, the log-odds of successful first-term completion increases by 0.43. Marines in the Aviation MOS group have a higher odds ratio when compared to Marines with a Combat Arms MOS. Aviation Marines have approximately 83 percent higher odds of completion than their counterparts in the Combat Arms MOS group. Marines who changed their MOS during their service have 59 percent higher odds of completion than those who did not. Similar to the findings in the P-EM, Marines with a marital status at entry of "other" have approximately four times lower odds of completion than those Marines reported as "single." The EM also shows male Marines have a higher odds-ratio of completing their first-term than female Marines. A Male's odds of completion are two times higher than a female's, a finding akin to that of the P-EM. Marines with a second or third class CFT score are less likely to complete their first-term than those with a first-class score. The odds of completion for a Marine with a second-class score is two times lower than the odds of completion for a Marine with a first-class score. The odds of completion for a Marine with a third-class score are 40 percent lower than those with a first-class score. The odds of completion for a Marine with an average or above average proficiency score at the rank of E4 are nearly three times higher than those who did not score at or above average. Likewise, for the rank of E4, the odds of a Marine completing their first-term if they scored at or above average in conduct is three times higher than those who did not score at or above average. For those with a conduct score at or above average

have 50 percent higher odds of completion than those who did not score at or above average at the rank of E3.

The LR model is compared to a RF model in the same manner as the P-EM. Figure 20 shows the ROC curve comparison between the LR and the RF models. The LR model has an 83.6 percent correct test-set prediction rate. The RF has an 84.4 percent correct test prediction rate. Similar to the P-EM, both the LR and RF models have nearly the same correct prediction rate. The ROC curve also shows similar performance between the LR and RF models and an increase in accurate prediction compared to the P-EM.

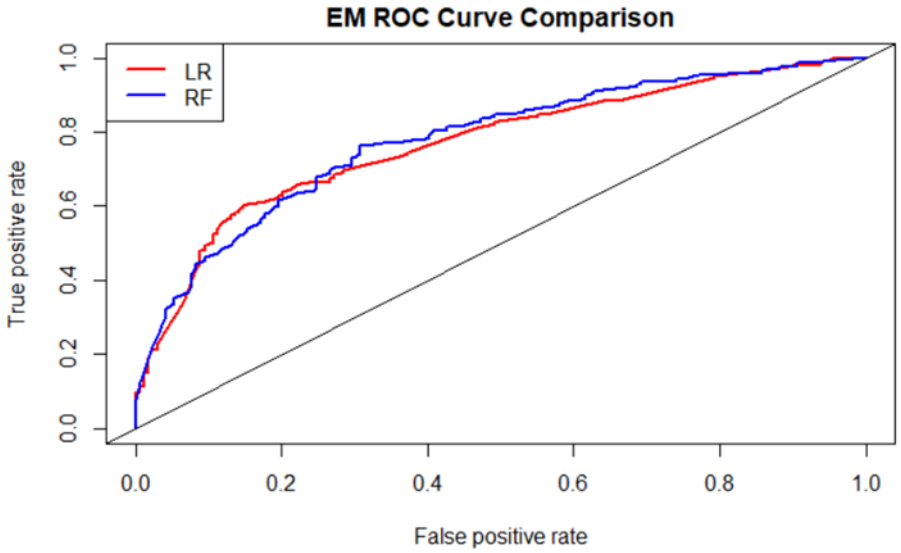

Figure 20. ROC curve comparisons of the EM LR and RF models.

Figure 21 shows the feature importance plot for the RF model. Two of the top five important variables in the RF match the significant variables in the LR model (PRO above average E4 and CON above average E4). The other significant variables found in the LR model were found to be less important in the RF model, with AFQT and GC scores remaining in the top five for the RF model with the EM data.

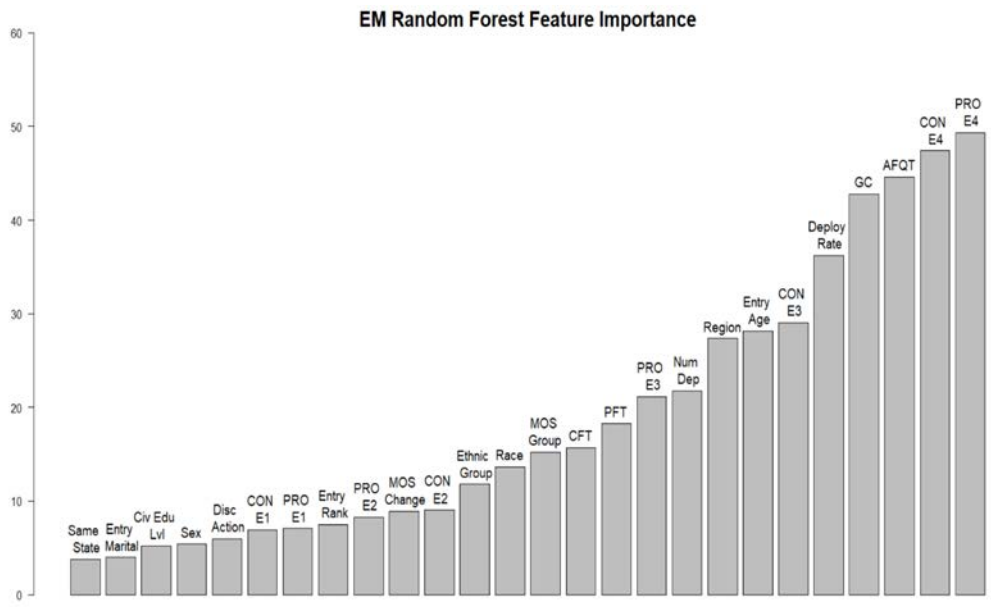

Figure 21. P-EM RF feature importance plot.

D. DISCUSSION

For both the P-EM and the EM, the LR and RF models show a test set predictive ability of above 80 percent and similar performance as displayed in the ROC curves. The ROC curve comparison and correct test set prediction rate, which shows near equal performance between the LR and RF models, indicates that there is no further need to explore adding interactions or transforming numeric variables to improve the LR model, as the RF method of modeling takes this into account.

There were some differences in terms of significant predictors and important features between the two modeling methods for both the P-EM and EM. The top 50 percent of the important features in the RF models did not include all of the independent variables found to be significant in the LR models, a fact that requires further exploration.

The RF models used all available independent variables, whereas the LR models are tailored using purposeful selection. Thus, the RF method has more data and predictors with which to build the final model. In principle, the RF model should perform better than the LR in terms of prediction because it starts with a larger set of available variables and can account for interactions and nonlinear effects of the variables. However, the analysis

shows this is not the case, as both the LR and RF methods produced models that performed nearly the same in each metric. To explore the differences in significant variables and feature importance, only those significant variables found in the LR model were used to build a new RF model. If the performance of the new RF model does not change noticeably, then we know the features the RF method is using to build the model are so closely aligned that the method finds little to no difference between them. This idea is tested using the EM, as there is more data available.

The new RF model, using only the predictors found significant in the LR model, successfully predicts completion on the test set at 83.3 percent. Figure 22 shows the ROC curve of the new RF model. As this figure shows, the new RF model performs almost exactly the same as the previous RF model. This is further evidence that the LR model is effective at identifying significant predictors in the data.

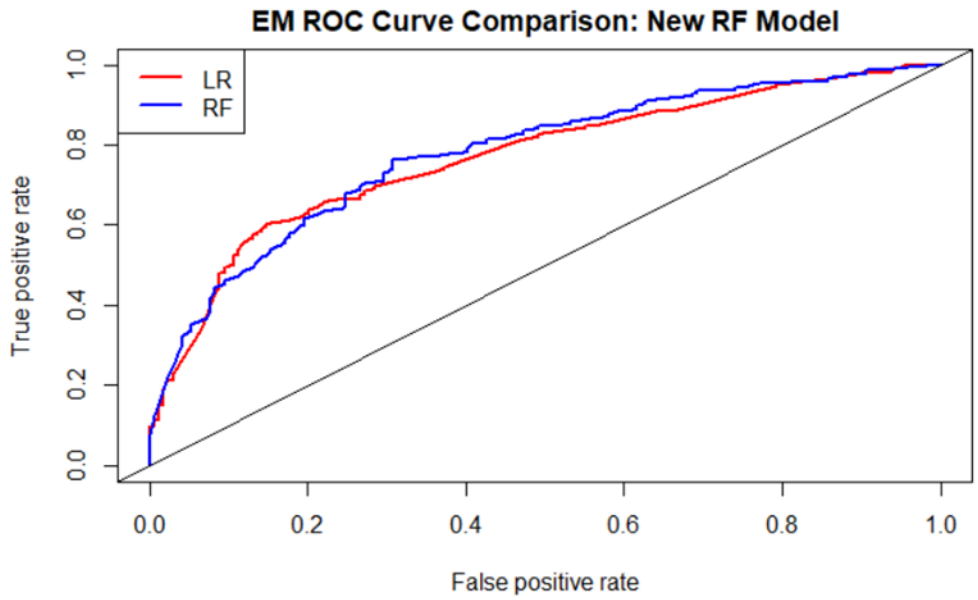

Figure 22. ROC curve comparisons of the LR and new RF models.

A Hosmer-Lemeshow goodness-of-fit test is performed on the P-EM and the EM LR fit (Zhang 2016). The Hosmer-Lemeshow's test statistic is based on the differences between observed and predicted values. A p-value of 0.85 is obtained for the EM suggesting no significant evidence to reject the null hypothesis that the LR model is adequate, with significance determined by a p-value < 0.05. A p-value of 0.15 is obtained for the P-EM, also suggesting there is not enough evidence to conclude that the LR model is a poor fit.

V. CONCLUSIONS

A. RESULTS

This thesis developed a LR and RF model to identify the factors associated with successful completion of a NPS, enlisted Marine's first-term enlistment in the SMCR. The LR method for the EM identified MOS group, MOS change, entry marital status, sex, maximum number of dependents, max CFT, max PFT, PRO scores at the rank of E4, and CON scores at the ranks of E3 and E4 as significant, verified by the feature importance from the RF model. The P-EM, seeking to identify the factors associated with successful completion based on information prior to a Marine's entry into the SMCR, did not show the positive results expected, and therefore is rejected as an informative model.

An analysis of the two distinct models suggests that the P-EM is not an accurate fit to the data. Even though the LR model for the P-EM correctly predicts completion slightly above 80 percent on the test set, the ROC curve shows the LR is not a good model fit to describe the data. Since the RF model for the P-EM performs at nearly the same rate and takes into account interactions and nonlinear effects, it is unlikely that the LR model fit can be improved. The poor fit is most likely due to the number and quality of the variables available in this model, as the P-EM included only the data available for a Marine prior to the commencement of their SMCR service. This assertion is confirmed by the EM which outperforms the P-EM in every metric and includes more available variables. Because the P-EM may not be an accurate model, concluding that entry marital status, entry age, sex, and AFQT scores are significant predictors of completion when combined with other information available for a Marine prior to their SMCR service commencement would be erroneous.

Further exploration of the predictive ability of each model validates the model fit. The training set contained approximately 80 percent (n = 3,652) of the observations, with the test set containing the remaining 20 percent of the observations (n = 914). As is discussed in Chapter III, approximately 80 percent of the data are successful completion observations. Therefore, the training and test sets, randomly selected and verified,

contained approximately 80 percent successful completion observations. Because of the training and test set structure, if the model developed were to predict "successful completion" for each observation regardless of the corresponding data, the model would predict correctly at a rate of at least 80 percent. Although the P-EM is not predicting "successful completion" for each observation regardless of data, the correct prediction rate is still near 80 percent, showing no significant improvement beyond guessing. However, the EM does have an improved correct prediction rate.

Some of the categorical predictors in the final EM, including sex and entry marital status, have relatively small sample sizes amongst the categories. Females comprise only 2.9 percent of the cohort (n = 132). Similarly, those listed in the "other" category in entry marital status contain only 0.2 percent of the observations (n = 10). The significance of these categorical predictors may be less informative due to relatively small sample sizes.

B. RECOMMENDATIONS

Future research on this topic is necessary to improve upon the results discussed above. The cohort from FY 09, which may have been influenced by socio-economic and military manning restrictions, may not be a fully accurate representation of retention characteristics within the SMCR. As new MOSs are created, physical requirements and recruiting practices are altered, and operational tempo changes, the important predictors of first-term completion may also change.

Additional information in terms of TA use, IIADT contracts, accurate HOR information, and moral waivers should also be tracked by the TFDW. As previous studies have shown, this information may be significant in predicting first-term completion. These characteristics were not provided in the data and were not studied in this thesis.

An accurate P-EM would be beneficial for SMCR recruiting and retention goals. To achieve an accurate model of completion for Marines as they are recruited, the author suggests obtaining more information than the provided data includes, if available. Using data from Military Entrance Processing Stations (MEPS), which includes more detailed information on a perspective reservist, may increase the accuracy of developed models, allowing for more precise insight into recruiting and retention policy.

LIST OF REFERENCES

Alstat JA (2017) Effects of commuting distance on participation rates for obligated Selected Marine Corps Reserves. Master's thesis, Graduate School of Business and Public Policy, Naval Postgraduate School, Monterey, CA. https://calhoun.nps.edu/handle/10945/56896.

Breiman L (2001) *Machine Learning* (Kluwer Academic Publishers, Netherlands).

Buddin R, Kapur K (2002) Tuition assistance usage and first term military retention. RAND Corporation, Santa Monica, CA.

Buddin R, Levy DS, Hanley JM, Waldman DM (1992) Promotion Tempo and Enlisted Retention. RAND Corporation, Santa Monica, CA.

Commandant of the Marine Corps (CMC). (2014). Deployment-to-dwell, mobilization-to-dwell policy revision. MARADMIN 346/14, USMC, Washington, DC. http://www.marines.mil/News/Messages/Messages-Display/Article/896510/deployment-to-dwell-mobilization-to-dwell-policy-revision/

Commandant of the Marine Corps (CMC). (2015). Marine Corps Reserve administrative management manual. MCO 1001R.1L, USMC, Washington, DC. http://www.marines.mil/Portals/59/MCO%201001R.1L.pdf

Commandant of the Marine Corps (CMC). (2016). Reserve optional enlistment program and incremental initial active duty training program. MCO 1133R.27, USMC, Washington, DC. http://www.marines.mil/Portals/59/MCO%201133R.27.pdf

Dolfini-Reed MA, Parcell AD, Horne BC (2005) Patterns of Reserve Officer attrition since September 11, 2001. CNA Corporation, Arlington, VA.

Enns JH (1977) Reenlistment bonuses and first-term retention. RAND Corporation, Santa Monica, CA.

Faraway J (2016) *Linear Models with R 1st ed*. (Chapman and Hall/CRC, London).

Hattiangadi AU, Parcell AD (2006) SelRes attrition and selected reserve incentive program in the Marine Corps Reserve. CNA Corporation, Arlington, VA.

Herschelman PR (2012) United States Marine Corps Reserve first term attrition characteristics. Master's thesis, Graduate School of Business and Public Policy, Naval Postgraduate School, Monterey, CA. https://calhoun.nps.edu/handle/10945/6807.

Manpower and Reserve Affairs (M&RA), personal communication, November 12, 2017.

O'Donohue DJ (1988) First-term retention of enlisted Selected Marine Corps (SMCR) Reservists. Master's thesis, Graduate School of Business and Public Policy, Naval Postgraduate School, Monterey, CA. https://calhoun.nps.edu/handle/10945/23233.

Purpose of the reserve components, 10 U.S.C. § 10102 (1994). /www.law.cornell.edu/uscode/text/10/10102?qt-us_code_temp_noupdates=0#qt-us_code_temp_noupdates

U.S. Census Bureau. (2010). Census regions and divisions of the United States. https://www2.census.gov/geo/pdfs.maps-data/maps/referecne/us_regdiv.pdf

Zhang Z (2016) Model building strategy for logistic regression: Purposeful selection. *Ann Transl Med* 4(6):111, http://dx.doi.org/10.21037/atm.2016.02.15.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School