NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

MARINE CORPS OFFICER MODELING: RETENTION ANALYSIS BASED ON SOURCE OF ACCESSION

by

Kyle P. McCarley

June 2021

Thesis Advisor:	Samuel E. Buttrey
Second Reader:	Ross J. Schuchard

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 2021	**3. REPORT TYPE AND DATES COVERED** Master's thesis

4. TITLE AND SUBTITLE MARINE CORPS OFFICER MODELING: RETENTION ANALYSIS BASED ON SOURCE OF ACCESSION	**5. FUNDING NUMBERS**
6. AUTHOR(S) Kyle P. McCarley	
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)

The annual Marine officer accession mission is achieved through five primary commissioning sources: U.S. Naval Academy (USNA), Naval Reserve Officer Training Corps (NROTC), Platoon Leaders Class (PLC), Officer Candidate Class (OCC), and Enlisted to Officer programs. Marine Corps Recruiting Command (MCRC) is tasked with finding, selecting, and commissioning qualified applicants that meet quality, diversity, and gender goals. Due to the extensive length of the officer accession pipeline, force generation decisions can take years to manifest in the form of a newly commissioned officer cohort. As Force Design 2030 identifies changes to the structure and capability of the Marine Corps, we seek to equip MCRC with a study of the different sources of accession. This thesis uses survival analysis and logistic regression to model officer retention and selection board performance using data from MCRC and Manpower and Reserve Affairs (M&RA) about active-duty Marine officers commissioned between fiscal years 2006 and 2016. We find that officer survival patterns are different across the commissioning sources, even when controlling for contract type. Additionally, we demonstrate that early indicators of performance are useful predictors when modeling early- and mid-career milestone achievement. MCRC can use these tools and results to inform talent management modernization efforts and help achieve Marine officer procurement objectives to support Force Design 2030.

14. SUBJECT TERMS U.S. Naval Academy, USNA, Naval Reserve Officer Training Corps, NROTC, Platoon Leaders Class, PLC, Officer Candidate Class, OCC, USMC, Marine Corps Recruiting Command, MCRC, Manpower and Reserve Affairs, M&RA, officer accessions, performance, retention, data science, survival analysis, regression analysis, talent management	**15. NUMBER OF PAGES** 101
	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.

MARINE CORPS OFFICER MODELING: RETENTION ANALYSIS BASED ON SOURCE OF ACCESSION

Kyle P. McCarley
Major, United States Marine Corps
BSME, University of Idaho, 2009
MS, University of Arizona, 2019

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
June 2021

Approved by: Samuel E. Buttrey
Advisor

Ross J. Schuchard
Second Reader

W. Matthew Carlyle
Chair, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The annual Marine officer accession mission is achieved through five primary commissioning sources: U.S. Naval Academy (USNA), Naval Reserve Officer Training Corps (NROTC), Platoon Leaders Class (PLC), Officer Candidate Class (OCC), and Enlisted to Officer programs. Marine Corps Recruiting Command (MCRC) is tasked with finding, selecting, and commissioning qualified applicants that meet quality, diversity, and gender goals. Due to the extensive length of the officer accession pipeline, force generation decisions can take years to manifest in the form of a newly commissioned officer cohort. As Force Design 2030 identifies changes to the structure and capability of the Marine Corps, we seek to equip MCRC with a study of the different sources of accession. This thesis uses survival analysis and logistic regression to model officer retention and selection board performance using data from MCRC and Manpower and Reserve Affairs (M&RA) about active-duty Marine officers commissioned between fiscal years 2006 and 2016. We find that officer survival patterns are different across the commissioning sources, even when controlling for contract type. Additionally, we demonstrate that early indicators of performance are useful predictors when modeling early- and mid-career milestone achievement. MCRC can use these tools and results to inform talent management modernization efforts and help achieve Marine officer procurement objectives to support Force Design 2030.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

ACMC	Assistant Commandant of the Marine Corps
AGPA	academic grade point average
AUC	area under the ROC curve
BOC	Basic Officers Course
CD	career designation
DA	Department of the Army
DOD	Department of Defense
DON	Department of the Navy
EAS	end of active service
ECP	enlisted commissioning program
E-O	enlisted to officer
FITREP	fitness report
FY	fiscal year
GCT	general classification test
HQMC	Headquarters United Stated Marine Corps
KM	Kaplan-Meier
M&RA	Manpower and Reserve Affairs
MCO	Marine Corps Order
MCRC	Marine Corps Recruiting Command
MCRISS	Marine Corps Recruiting Information Support System
MECEP	Marine Enlisted Commissioning Education Program
MGPA	military grade point average
MOI	Marine Officer Instructor
MOS	Marine Occupational Specialty
NROTC	Naval Reserve Officer Training Corps
NSTC	Naval Service Training Command
O4	major
O5	lieutenant colonel
OCC	Officer Candidates Class
OCS	Officer Candidate School

ORB	Officer Retention Board
OSO	Officer Selection Officer
PFT	physical fitness test
PLC	Platoon Leaders Class
PRIOR	prior enlisted commissioning program
ROC	receiver operating characteristic
SATTOT	scholastic aptitude test score
TBS	The Basic School
TFDW	Total Force Data Warehouse
USNA	United States Naval Academy
USMC	United States Marine Corps

EXECUTIVE SUMMARY

In his 2019 planning guidance, the 38th Commandant of the Marine Corps identified the need to modernize Marine Corps manpower management modeling to incorporate talent and performance (Berger 2019). This thesis focuses on modeling the Marine officer corps to explore the impact that source of accession has on retention and the achievement of early- and mid-career milestones. The five accession sources considered were the U.S. Naval Academy (USNA), Naval Reserve Officer Training Corps (NROTC), Platoon Leaders Class (PLC), Officer Candidates Class (OCC), and Enlisted to Officer (E-O) programs. We use survival analysis to estimate officer retention and regression analysis to model the relationship between officer commissioning sources and their performance on three different selection boards. The Marine Corps can use this research to help improve talent management practices.

The successful employment of data science methods relies on accurate and complete data. We gather demographic and performance data from three credible sources, each providing information about active-duty Marine officers commissioned between fiscal year (FY) 2006 and FY 2016. We clean and prepare Total Force Data Warehouse (TFDW), The Basic School (TBS), and Marine Corps Recruiting Information Support System (MCRISS) data to produce a sufficient dataset containing 16,311 observations and 20 covariates.

We use survival analysis to model and compare the time in years until officers from each accession source exit the Marine Corps while controlling for contract type, early performance indicators, and gender type. We determine the survival variable using an officer's end of active service (EAS) date and formulate early indicators of performance using Physical Fitness Test (PFT), General Classification Test (GCT), and TBS Leadership performance. We also build three logistic regression models to predict Career Designation (CD), major (O4), and lieutenant colonel (O5) selection board results using a variety of early career variables, to include source of accession and contract type. The primary results from this thesis are:

- The incongruency of variables across the data sources and number of missing values underscore deficiencies in Marine Corps data enterprise systems. Although we employ missing data handling techniques to overcome this challenge, our research could have generated more conclusive findings with the incorporation of more extensive and reliable datasets.

- Officers from the OCC accession source have the lowest estimated median survival time across accession sources and contract types. Ground officers from the E-O accession source and naval aviators from the USNA accession source have the highest estimated median survival time.

- The early indication of leadership model produced the most significant differences in estimated survival across the accession sources and contract types. This result may suggest that leadership ability distinguishes officer retention better than physical fitness or intellectual ability.

- The differences in estimated survival between male and female officers from the OCC, PLC, and E-O accession sources are statistically insignificant. Within the female officer population, officers originating from the NROTC accession source have the lowest estimated median survival time.

- Source of accession is associated with CD and O4 selection but not selection to O5. This result may suggest that source of accession has less of an impact on the achievement of career milestones that occur later in an officer's career. Additionally, we discover that the TBS Leadership covariate is statistically significant in all three regression models, suggesting that it is an important variable when predicting early- and mid-career performance, even though it represents introductory officer aptitude.

This thesis seeks to equip the Marine Corps with a quantitative analysis of officer accession and retention to support talent management modernization efforts. As the Marine Corps strives to leverage its data to improve decision making, this research illustrates the type of impact that data science can have on manpower management modeling.

References

Berger DH (2019) 38th commandant's planning guidance 2019. Washington, DC, https://www.hqmc.marines.mil/Portals/142/Docs/%2038th%20Commandant%27s%20Planning%20Guidance_2019.pdf?ver=2019-07-16-200152-700.

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

This thesis would not have been possible without the guidance and support of Professor Samuel Buttrey and Lieutenant Colonel Ross Schuchard. I thank them for believing in me and providing the encouragement I needed to cross the finish line. I was challenged by the entire thesis process, and it felt like I was "flying blind" most of time. I am indebted to these two mentors who kept me focused on the objective and taught me how to employ and write about operations research along the path to completion.

I am grateful to those who shaped my perspectives on Marine officer talent management and helped get this thesis research off the ground: Lieutenant Colonel David Forbell, Lieutenant Colonel Mark Sykes, Major Jeff Parker, Major Jason Oldenkamp, Major Brian Pugh, Major Alex White, Major Bruce Manuel, Capt Paul McLaughlin, Captain Adam Bicknell, Captain Adam Jatho, Professor Chad Seagren, and Daniel Lehnherr. I cannot thank Tim Johnson, Major Kevin Doherty, and Major Daniel Sanchez enough for providing me with the data that was used in this analysis and Major Robert Jankowski for being a trusted advisor and sounding board during the life cycle of this thesis. I am thankful for the Marine officers in my cohort for tolerating my enthusiasm and politely listening to me pontificate about the Marine Corps; your friendship and camaraderie while completing this degree will always be cherished. To the Marines, midshipmen, and officer candidates who belonged to the University of Arizona NROTC unit from 2016 to 2019, thank you for allowing me to lead you as your Marine Officer Instructor. Those years had a big impact on my personal growth and professional development.

I would be remiss if I did not highlight the amazing feats of my wife, Allison. Thank you for your patience and sacrifice during our two years in Monterey. Without your fortitude, we could not have moved our family to Virginia early, which gave me the time needed to finish this thesis. You are an incredible woman, and I love you, Clayton, Wyatt, and Arlo very much. Lastly, I would like to thank God for being my source of strength and giving me a joyful spirit even in times of adversity.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

> Our manpower model is based primarily on time and experience, not talent or performance or potential for future performance.
>
> —General Berger, 38th Commandant, USMC, 2019

A. BACKGROUND AND MOTIVATION

After two decades of focusing on counterinsurgency operations, the U.S. Marine Corps (USMC) has proposed transformational changes needed to maintain a competitive edge in the age of great power competition. As outlined by General Berger in the 38th Commandant's Planning Guidance, the USMC's unique contribution to the joint force is that of a naval expeditionary force, capable of persisting inside an adversary's weapon engagement zone (Berger 2019). The integration of Navy and Marine Corps capabilities is necessary to ensure the success of operational concepts like Expeditionary Advanced Base Operations (Department of the Navy [DON] 2021). Legacy systems designed to support counterinsurgency operations are no longer relevant and must be replaced with new capabilities that increase the survivability and lethality of the force so that Marine units can support the larger naval campaign (Berger 2019). *Force Design 2030*, a document written by General Berger in 2020, identifies the near- and long-term changes to come. Proposed force structure changes will provide budgetary and capacity opportunities necessary to bring about the commandant's vision for the USMC (Berger 2020). While the core modifications to force structure focus on the reduction of infantry and support battalion strength, the successful integration of emerging technologies and novel operating concepts will require a skilled and properly managed officer corps.

To make informed policy recommendations on officer talent management, the USMC must first assess the impacts of its existing manpower model and the retention patterns of top-performing individuals. As pointed out by Szoldra (2014), the Department of Defense (DOD) is not good at predicting the future, so talent management upgrades should exploit quantitative analysis and not subjective viewpoints. The Army asserts that the complexity of the future security environment justifies the need for new manpower management techniques to attain and retain more sophisticated and diverse knowledge, skills, and abilities. (Department of the Army [DA] 2015). Colarusso et al. (2009) argues that industrial age talent management practices will no longer suffice in the era of an all-volunteer force due to the competitive labor market. The USMC recognizes this problem and is laying the groundwork to evaluate and change existing policies and programs. In 2018, the Assistant Commandant of the Marine Corps (ACMC), serving as the Talent Management Officer, established the Talent Management Executive Council, an organization that seeks to develop guidance and initiatives regarding the attraction, retention, and development of Marines and Sailors (USMC 2021). Despite the progress made in recent years, continued research will present decision makers with relevant data to guide talent management modernization.

There are many ways for someone to become an officer in the Marine Corps. We call these different ways sources of accession. The five primary sources of accession are (Marine Corps Recruiting Command [MCRC] 2016):

- U.S. Naval Academy (USNA),
- Naval Reserve Officer Training Corps (NROTC),
- Platoon Leaders Class (PLC),
- Officer Candidate Class (OCC), and
- Enlisted to Officer (E-O) programs.

A newly commissioned officer cohort represents the talent pool that future field-grade and general officers come from. The USMC, an organization that does not have the luxury of hiring senior executives from the outside, must grow its own future leaders. Lin et al. suggests that the "early prediction of high-potential talent enables organizations to marshal scarce developmental resources and opportunities to those who are best positioned to show distinction in elevated roles" (2020, p. 1). The use of data science can help the USMC identify its top performers early in their careers and incentivize their retention, encouraging them to remain in uniform past their initial contract.

B. THESIS OBJECTIVE

The objective of this thesis is twofold. The first goal is to better understand the retention of officers from each commissioning source. The second is to examine early career performance metrics and determine their effectiveness at predicting future performance. We will create and examine statistical models to complete this research. The statistical models created can be used by Marine Corps Recruiting Command (MCRC), the topic sponsor, to improve officer recruiting practices. Early predictors of long-term performance can potentially inform revisions to the officer accession application and selection process; performance metrics that have historically been used to select officer candidates may no longer indicate the potential for future performance. Additionally, the findings presented in this research will inform the broader talent and manpower management discussion. It is important to highlight that officer accession decisions made in 2021 will not impact the field-grade officer ranks until 2036. The quantitative analysis performed in this thesis will provide the USMC with current officer retention trends to hopefully improve and reinforce talent management policy decisions.

C. SCOPE

This research uses demographic and performance data on active-duty Marine Officers commissioned between fiscal year (FY) 2006 and FY 2016. We acquired data from three different repositories to build an extensive set of covariates for each officer in the sample. Examples of these covariates include commissioning source, gender, Military Occupational Specialty (MOS), and Physical Fitness Test (PFT) score. Survival analysis was used to model retention and regression analysis was used to predict three career milestones: Career Designation (CD), selection to major (O4), and selection to lieutenant colonel (O5).

D. LIMITATIONS

1. Data Availability and Comprehensiveness

Data quality is an essential element of any data science project. Bad data produces model output that is not very useful. The USMC collects copious amounts of data during a

Marine's life cycle, but unfortunately much of this information is neither in a digital format nor amenable for analysis. In some cases, the data used in this research was incomplete, so we used missing data handling techniques to overcome this deficiency. Additionally, isolated data enterprise systems with access restrictions made it challenging to gather and link officer performance data over time. This is the primary reason for the involuntary exclusion of officer fitness report (FITREP) data in this research. Considering that FITREPs contain detailed officer performance ratings and narratives, this high-value data source has the potential to improve the analysis conducted in this research. Specific data limitations are discussed in Chapter III and recommendations for data improvements are presented in Chapter VI.

2. Sample Size and Timeframe

This research selected data based on availability and the existence of promotion board results. Within the year groups selected, 16,311 officers were considered for CD, 4,609 were considered for promotion to O4, and 457 were considered for promotion to O5. Despite the small sample size, enough information is available to make inferences on mid-career retention patterns. By nature of the timeframe, there was Marine Corps involvement in Operation Iraqi Freedom, Operation Enduring Freedom, and Operation Inherent Resolve. The operational tempo of the Marine Corps during this period may have influenced officer retention patterns but combat experience is not a variable captured in this research.

3. Separation Condition

Knowing why an officer left the Marine Corps is an important consideration when analyzing retention. For officers who only completed their initial service obligation, there is no information pertaining to why they decided to exit the Marine Corps. Some career-designated officers decided to refuse CD but exit interview data was not included in this research. Chapter VI recommends further research using exit interview data.

4. Retirement System

The Blended Retirement System (BRS), first introduced in 2017, replaced the legacy retirement system and made significant changes to military retirement benefits (Department of Defense [DOD] 2017). These changes will undoubtedly impact retention decisions. While service members had the ability to opt-in to the BRS until 31 December 2018, this research did not consider an individual's retirement system. This limitation should be considered when evaluating the recommendations presented in Chapter VI as future generations of Marine officers will only have the BRS.

E. ASSUMPTIONS

1. Officers Are a Representative Sample

When comparing officer retention and performance based on source of accession, we assume that officers from a given commissioning source form a representative sample. In other words, the performance of officers from a given source of accession in the dataset adequately describes the future population of officers from that source of accession.

2. Career Designation and Promotion Boards

In the absence of FITREP data, the dependent variables selected were CD and promotion board results. Our assumption is that board selection is an indication of high-performance and non-selection is an indication of low-performance. We also assume that these boards are fair, unbiased, and good at identifying successful Marine officers.

3. Early Indicators of Performance

We used The Basic School (TBS), PFT, and General Classification Test (GCT) performance as early indicators of leadership, fitness, and intellect, respectfully. Our assumption is that high scores on these evaluations are a good indication of high-performance in these traits.

F. THESIS ORGANIZATION

This thesis is organized into six chapters. Chapter II provides a summary of the officer accession and retention strategy, the characteristics of each source of accession, and background information on CD and Marine Corps promotions. Chapter III provides a review of previous studies on officer performance, retention, and talent management. Chapter IV explains the methodology, describes the data, and shows preliminary analysis. Chapter V presents the results from the survival and regression analyses. Finally, Chapter VI summarizes the study, explains the findings, and offers recommendations for future work.

II. OVERVIEW OF THE MARINE CORPS ACCESSION AND RETENTION STRATEGY

> The Officer Corps embodies a unique profession whose culture and core warfighting abilities take years to develop. This means that each new officer cohort represents far more than the Army's latest crop of junior leaders; they are the feedstock for its future field grade and general officers.
>
> —Colarusso et al., 2009

The Marine Corps accession and retention strategy outlines recruiting and retention guidance for Marine officers over the next five fiscal years. Manpower and Reserve Affairs (M&RA) and MCRC are the key stakeholders in this process, each contributing to the total force procurement efforts needed to meet the manpower needs for the entire Marine Corps. While M&RA focuses on the total force structure and end-strength requirements, MCRC has the task of finding, selecting, and commissioning applicants to produce a diverse and skilled officer corps (Headquarters United States Marine Corps [HQMC] 2017). The active component officer accession strategy is primarily achieved through the five sources previously listed in Chapter I. While officer selection criteria are standard across these five programs, each program's requirements, timelines, and experiences differ. A newly commissioned officer cohort possesses a mixture of talents that, when correctly developed and managed, can increase the warfighting readiness of the Marine Corps. This chapter describes each of the officer accession programs and the retention methods used to maintain officer inventory requirements.

A. MARINE CORPS ACCESSION OVERVIEW

Each year, M&RA publishes MEMO-01, a document that tasks MCRC with the total number of officer accessions for the next two fiscal years. To achieve Marine officer end-strength requirements, accession quotas are divided into four contract types: naval aviators, judge advocates, cyber officers, and ground officers. Minimum obligated service length varies depending on the contract type and commissioning source. Contracts last eight years from "winging" date for naval aviators. USNA ground officer contracts last five years while ground officers from other sources last four years (HQMC 2020). A memorandum of agreement between the Navy and Marine Corps establishes the annual allocation of Marine officers to be commissioned through NROTC and USNA (Office of the Chief of Naval Operations 2021). OCC, PLC, and E-O programs comprise the remaining accession quotas, but no more than 10% of the officers commissioned each year can be prior enlisted (Coppage 2021, HQMC 2017). Marine Corps Recruiting Command Order 1100.2A outlines recruiting polices, program requirements, and overall criteria for officer procurement. (MCRC 2016). Table 1 displays the total officer accession requirements by contract type for FY 2021 and FY 2022.

Table 1. FY 2021 and FY 2022 Marine Officer Accession Requirements. Source: HQMC (2020).

	FY 2021	**FY 2022**
Naval Aviators	380	380
Judge Advocates	65	70
Cyber Officers	16	16
Ground Officers	1,156	1,126
Total Commissioned	1,617	1,592

Each accession program targets different groups within the eligible population, specifies unique eligibility prerequisites, and stipulates varying timelines to complete. Regardless of the commissioning program, applicants must (MCRC 2016):

- Possess U.S. citizenship;
- Earn a post-secondary degree from an accredited institution;
- Satisfy certain medical requirements;
- Record a score of at least 1,000 on the Scholastic Aptitude Test (Math and English), a score of at least 22 on the American College Training, or a score of at least 74 on the Armed Forces Qualification Test;
- Satisfy height and weight standards; and
- Achieve a first-class score on the PFT.

Additionally, the USMC seeks to commission individuals of unquestionable moral integrity.

A significant challenge in officer procurement is projecting and reconciling attrition. Injury, poor academic performance, or failure to complete Officer Candidate School (OCS) can lead to disenrollment from an officer program. Since the timetable to commission can range from one to five years, depending on the program, MCRC must continually manage the commissioning quotas for the given fiscal year and following ones. Longer commissioning programs like NROTC and PLC are riskier since there are more time and opportunities for a contracted applicant to dropout. When pressure builds to meet the annual mission, the OCC program is used as the release valve since applicants can be contracted, sent to OCS, and commissioned relatively quickly, in comparison to other sources. USNA is the most reliable program because of the agreed upon percentage of graduating seniors assured to commission as Marine officers. The element of attrition causes some fluctuation in the annual allocation of commissioning quotas to each program. Figure 1 displays the FY 2021 accession mission by commissioning source.

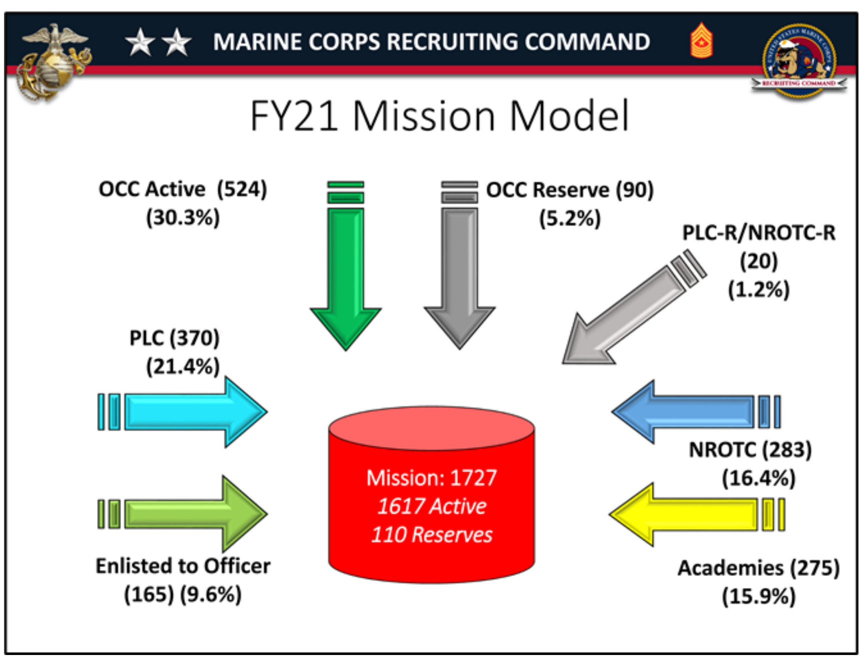

Figure 1. FY 2021 Officer Accession Mission. Source: MCRC (2021).

B. MARINE CORPS ACCESSION PROGRAMS

1. U.S. Naval Academy

The U.S. Naval Academy (USNA) is a four-year undergraduate university that prepares young men and women to become professional officers in the U.S. naval service. Students admitted to USNA become active-duty midshipmen in the Navy, and after graduating with a Bachelor of Science degree, commission as ensigns or second lieutenants.

The application process for USNA is more extensive than that for a typical college. Not only must prospective students have strong academic and standardized test performance, but they also must demonstrate their moral, mental, and physical toughness. Applicants must receive a nomination from an official source, typically a U.S. representative or senator, complete a Candidate Fitness Test, and interview with a Blue and Gold Officer (USNA 2021a). While the core of the applicant pool comes from recent high school graduates, junior enlisted Sailors and Marines are also eligible to apply. USNA attracts many highly competitive applicants, but only 9% of applicants received appointment offers to the most recent freshman class (USNA 2020). Table 2 displays USNA class of 2024 selection statistics.

Table 2. USNA Class of 2024 Snapshot. Source: USNA (2020).

Applications	**15,699**
Offers of Appointment	1,426
Class Size	1,194
SAT Verbal Middle 50th Percentile*	630-760
SAT Math Middle 50th Percentile*	620-750

*50% of the class achieved SAT scores within the range between the 25th and 75th percentile

A major difference between USNA and other accession sources is the degree of military acculturation offered. The academic program at USNA integrates core requirements and major academic courses with the intent of creating an environment that fosters critical thinking and focused education. Core courses in naval engineering, mathematics, history, leadership, and navigation equip midshipmen with the knowledge needed to succeed as commissioned officers (USNA 2021b). Each summer, midshipmen attend training events that are designed and sequenced to provide hands-on experience and professional development in naval operations (Office of the Superintendent 2021).

Despite the extensive officer development opportunities afforded to midshipmen during their time at USNA, midshipmen do not know if they will be commissioned into the Navy or Marine Corps until their senior year. Students can express interest in one service or specialized community early in their tenure at USNA, but the overarching training and education curriculum seeks to instill core competencies for general naval service, not Marine Corps-specific ideologies. Marine officers accessed through USNA do not attend OCS but are required to complete Leatherneck. Leatherneck is a four-week training evolution designed to "train, evaluate, advise, and mentor 1st Class Midshipmen considering a career in the United States Marine Corps" (USNA 2021b). Currently, 25 percent of a USNA graduating class is commissioned as a second lieutenant in the Marine Corps, equating to roughly 275 officers annually (Office of the Chief of Naval Operations 2021).

2. Naval Reserve Officer Training Corps

The NROTC program was designed to provide officer accessions for the Navy and Marine Corps by educating and training qualified college students in the professional, physical, and moral standards needed to serve as a commissioned officer. Throughout the country, there are 63 NROTC units hosted at 77 civilian universities, attracting a diverse group of applicants, and increasing public awareness of opportunities for naval service. (Naval Service Training Command [NSTC] 2021).

There are three primary types of midshipmen in a NROTC unit: Navy Option scholarship midshipmen, Marine Option scholarship midshipmen, and College Program midshipmen. Scholarship students are selected out of high school through a highly competitive national selection board and are designated as Navy or Marine Option during the application process. College Program students participate in all NROTC activities but are not on scholarship nor contracted to commission. While completing their degree, College Program midshipmen apply to earn competitive commissioning slots. MCRC holds a competitive scholarship selection board twice per year, targeting top performing College Program freshmen and sophomores who want to serve as Marine officers. Although MCRC manages the number of Marine officers accessed through the NROTC program, the NROTC program is overseen by NSTC. Tables 3 and 4 display FY 20 Marine Option scholarship selection statistics.

Table 3. FY 2020 NROTC 4-Year Marine Option Scholarship Statistics. Source: Coppage (2021).

Applications	**1,648**
Scholarship Offers	370
Average SAT	1,239
Average PFT	257

Table 4. FY 2020 Marine Corps Competitive Scholarship Statistics. Source: Coppage (2021).

Applications	**289**
Scholarship Offers	116
Average SAT	1,225
Average PFT	278
Average College GPA	3.32

NROTC midshipmen are required to take NROTC-specific classes in addition to classes required for their academic major. Physical training, professional development, and naval science instruction are used to indoctrinate midshipmen into the military lifestyle and teach the fundamentals of officership. Scholarship midshipmen are also required to attend annual summer training sessions that serve as opportunities to gain a better understanding of the operational forces and conduct practical application of core skills in a realistic environment (NSTC 2019). Each NROTC unit has a small active-duty staff that teaches naval science classes and mentors midshipmen as they prepare to commission.

In contrast to USNA midshipmen, NROTC Marine Option midshipmen begin to adopt their Marine identity early in the program. Marine Option midshipmen wear Marine

Corps insignia on their service and dress uniform, conduct field training in the Marine Corps Combat Utility Uniform, and must complete the six-week OCS Commissioning Course between their junior and senior year (NSTC 2019, MCRC 2016). An NROTC unit's Marine Officer Instructor (MOI) is responsible for preparing midshipmen for OCS and typically exposes them to Marine Corps-specific training and education in the years leading up to OCS.

3. Platoon Leaders Class

The PLC program was designed to attract college students attending regionally or nationally accredited colleges or universities. Upon completion of all requirements, these students are commissioned as second lieutenants in the Marine Corps. Freshmen or sophomores enrolled in the PLC program must attend two six-week sessions of OCS. The Pre-Commissioning Course is completed the first summer after enrollment and the Commissioning Course is completed the summer before college graduation. Students enrolled in the PLC program after their junior year attend the 10-week Combined Course of OCS (MCRC 2016). The Officer Selection Officer (OSO) is responsible for meeting assigned accession numbers by canvassing college campuses, finding qualified applicants, and completing all administrative requirements so that selected officer candidates achieve their commissioned grade (MCRC 2016).

Officer candidates in the PLC program attend monthly pool events which serve as opportunities to foster professional growth and generate enthusiasm for service in the Marine Corps. The pool program is managed by the OSO and used to prepare candidates for OCS, generate referrals, and provide mentorship on becoming a Marine officer (MCRC 2015). Pool functions can include pre-OCS training, picnics, trips to military exhibits, and physical training sessions. Since OSOs have limited interaction time with their candidates, the level of military instruction PLC officers receive prior to commissioning is not as extensive as for USNA and NROTC officers. It is important to note that OSOs are heavily focused on the recruiting and administrative aspects of officer accessions while MOIs and USNA Marine staff are focused more on teaching, coaching, and counseling prospective Marine officers.

4. Officer Candidate Class

The target population for the OCC program is college seniors at, or recent graduates of, an accredited college or university. Once selected and after earning their degree, applicants attend the 10-week Combined Course of OCS and are commissioned to the rank of second lieutenant after completion (MCRC 2016). The OCC program and PLC program have the same criteria for selection, but OCC candidates typically spend less time waiting to attend and complete OCS. Since the timeframe from contracting to commissioning is the shortest in the OCC program, this program is used to help achieve the accession mission when other sources cannot deliver their allocated quota. For FY 2021, the OCC program has the "dominant" quota (roughly 30 percent); however, officers commissioned from the OCC program have the least exposure to military standards and the Marine Corps ethos. The 10-week session of OCS is used as the primary mechanism to educate and train OCC officers in the knowledge and skills required to serve as a Marine officer.

5. Enlisted to Officer Programs

The two primary Enlisted to Officer (E-O) programs are the Marine Enlisted Commissioning Education Program (MECEP) and the Enlisted Commissioning Program (ECP). While each program results in the commissioning of a Marine officer, there are key differences in each program. Applicants to the MECEP must have a minimum of three years of active service and must have not yet attained their bachelor's degree. After selection to the program and graduation from the 10-week Combined Course of OCS, MECEP Marines complete their bachelor's degree as an active member of an NROTC unit (HQMC 2015). During this period, MECEP Marines retain their enlisted grade while completing university- and NROTC-specific requirements, including regular professional development, physical training, and naval science classes. Although MECEP Marines are integrated into the NROTC midshipman chain of command, they typically possess multiple years of operational experience which is leveraged to advance the training and preparedness of Marine Option midshipmen. Upon successful completion of their undergraduate degrees, MECEP Marines are appointed to the rank of second lieutenant. Applicants to the ECP must have earned their bachelor's degree, completed at least one year of active service, and attained the rank of lance corporal. After successful completion of the 10-week Combined Course of OCS, ECP Marines are commissioned to the rank of second lieutenant (HQMC 2015). Officers commissioned through E-O programs will be referred to as PRIOR officers for the duration of this thesis.

C. THE BASIC SCHOOL

All Marine officers, regardless of accession source and contract type, attend the Basic Officers Course (BOC) at The Basic School (TBS) immediately following their appointment as a second lieutenant. TBS is in Quantico, Virginia, and the intense 6-month BOC is used to equip all newly commissioned officers with the foundational leadership and tactical skills needed to be an effective Marine officer. TBS is the great equalizer as it attempts to make up for the differences in each source of accession by providing a common curriculum to all officers.

Although the BOC has been modified over the years to reflect updated tactics, techniques, and procedures, the core evaluation categories have remained unchanged. Academic, military skills, and leadership evaluations are used to ensure student officers are proficient in the military knowledge and warfighting skills needed to lead Marines in the operating forces. Performance in these three areas determine a student officer's class standing that is used during MOS assignment to evenly distribute the quality of officers (TBS 2018). Table 5 shows the grading breakdown for each BOC evaluation category.

Table 5. BOC Evaluation Categories

Category	Percent of Total Grade
Academic	30
Leadership	40
Military Skills	30
Total	100

D. OFFICER RETENTION OVERVIEW

The Marine Corps exists as a force in readiness that is manned, trained, and equipped to fight and win its nation's battles. To achieve its full potential, "the Marine Corps must provide the Fleet Marine Force and Supporting Establishment the right Marine, at the right time, with the right skills to support unit operational requirements and mission accomplishment" (HQMC, 2021a, p. 1-4). Although individual desires, career aspirations, and unique circumstances are taken into consideration, the manpower system must balance these factors against the needs of the Marine Corps. Institutional needs will always take priority over individual Marine needs. M&RA carefully manages retention plans to ensure that end-strength requirements and rank composition constraints are achieved within the officer corps. Career Designation (CD) and officer promotion are both used to manage the available inventory of officers to meet staffing requirements (HQMC 2021a).

1. Career Designation

Career Designation (CD) is a competitive process used to select company-grade officers and offer them the opportunity to remain on active duty past their initial active service obligation. The purpose of CD is to identify and retain the most capable officers to achieve inventory requirements. There are four primary methods for qualified officers to be offered CD. The first and primary means for CD is to be selected on the Officer Retention Board (ORB). To be eligible for the ORB, an officer must have a minimum of 540 days of observed time during which a reporting senior evaluates their performance. A description of the ORB is provided in Figure 2. Second, the top five percent of each BOC class is eligible for CD. Third, company-grade officers who accept orders to special duty programs will be automatically career designated. Finally, a set number of CD quotas are available to specific commanding generals who can nominate eligible Marine officers who were not previously selected for CD (HQMC 2021b). Marine Corps Order (MCO) 1001.65A outlines the policies and guidelines for the implementation of the CD program.

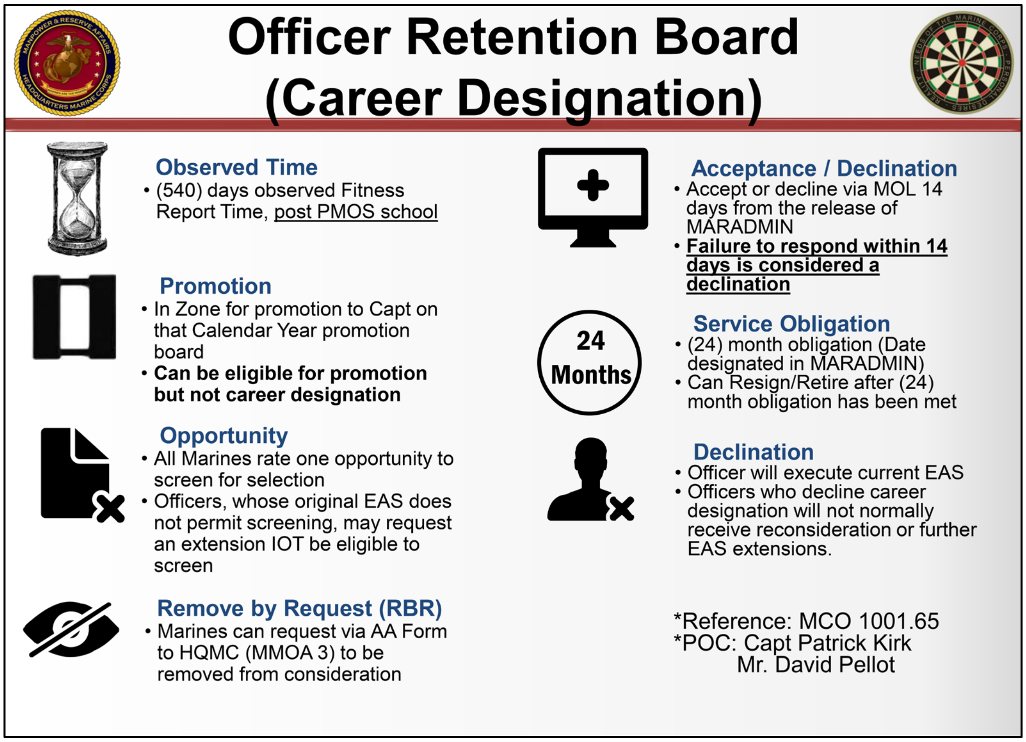

Figure 2. Officer Retention Board Description. Source: Manpower & Reserve Affairs [M&RA] (2021)

2. Officer Promotion

"Officers are selected for promotion for their potential to carry out the duties and responsibilities of the next higher grade" (HQMC 2006, p. 2). Promotion boards take into consideration an officer's past performance as described in their official military personnel file (HQMC 2006). According to Headquarters Marine Corps (2021a), "promotion is not a reward for past performance, rather an indication of the expectation for future performance" (p. 1-2). The intent of the Marine Corps promotion model is to continually identify top performers who have demonstrated the potential to succeed in more senior positions and assume more responsibility. MCO P1400.31C provides the instructions relating to the administration of officer promotions for the Marine Corps (HQMC 2006).

THIS PAGE INTENTIONALLY LEFT BLANK

III. LITERATURE REVIEW

> Where we have individual leaders and organizations that are trying to adopt the best practices in data science and data analytics, it is often accomplished through the heroic efforts of a few individuals rather than the organized and sustained effort required to transform how we sense, make sense, and act.
>
> —General David H. Berger, USMC, 2019

A. OVERVIEW

This research focuses on determining whether source of accession impacts retention and the value of using early career performance metrics to predict future performance benchmarks. We use TBS, PFT, and GCT scores as early career performance metrics while career designation, selection to O4, and selection to O5 serve as future performance benchmarks. Each commissioning source offers unique opportunities that shape an officer candidate's skill set. Understanding how to leverage this diversity in skill will help improve officer talent management. Many previous studies focused on officer performance and retention. This chapter reviews related scholarly literature on officer performance, retention, statistical modeling, and talent management.

B. SURVIVAL ANALYSIS

Korkmaz (2005) examined the service length of Navy officers based on their source of accession for the purpose of improving officer retention policymaking. Three survival models were built using data from the seven officer cohorts commissioned between 1983 and 1990. This analysis determined that the differences in survival for officers across the commissioning sources were statistically significant; officers from different commissioning sources follow different survival patterns, especially after initial service obligation. According to the research, prior enlisted officers had a lower estimated hazard rate than non-prior enlisted officers, NROTC contract and OCS officers had the highest estimated survival times after 14 years of service, and NROTC scholarship officers had the lowest estimated survival times during all periods. Korkmaz (2005) found that officers who graduated from highly selective colleges had higher estimated hazard rates than those not from highly selective colleges and the author speculated that this was attributed to higher competitiveness in the civilian labor market. While logistic and linear regression are often used in officer performance studies, Korkmaz (2005) showed the benefit of using survival analysis to compare the service lengths of Navy officers from different commissioning sources.

Although Korkmaz (2005) presented valuable insights regarding the survival patterns of Navy officers, it relied upon data that stopped in the year 2000. The results of this thesis presented the survival patterns of officers in a peacetime Navy. It is reasonable to assume that the war on terrorism, which started soon after the attacks on September 11, 2001, may have impacted officers' decisions to stay or leave the Navy. This highlights the need for an updated analysis of officer survival patterns using current data.

Urech (2019) used survival analysis to explain the factors that influence DOD civilian employee attrition. This study focused on data of DOD civilian employees hired in 2009 that total 9,279 blue-collar and 53,478 white-collar employees. Although the study only contained eight years of data, his approach leveraged survival analysis to make the most of the right-censored data. Urech (2019) demonstrated the effectiveness of using the R package "survminer" to determine Kaplan-Meier estimators and survival trees to describe the association a covariate has with employee attrition. Advancements in computing power provided Urech (2019) with updated modeling tools that were not available to Korkmaz (2005). The models produced in this thesis had high explanatory power and made it easy to understand the general impacts of hiring decisions.

C. REGRESSION ANALYSIS

Lin et al. (2020) used logistic regression to evaluate early predictors of performance and their ability to predict Army career outcomes of West Point graduates. Specifically, the authors analyzed the predictive power of a cadet's cognitive ability, academic performance, and military performance to predict early promotion to major, early promotion to lieutenant colonel, and selection for battalion command. The metrics considered as predictors of performance were total Scholastic Aptitude Test score (SATTOT), total academic grade point average (AGPA) and military grade point average (MGPA). MGPA is comprised of job performance evaluations, grades for military science courses that emphasize small unit tactics and military leadership, and combat order delivery (Lin et al. 2020). This research found that MPGA was a better indicator of early promotion and battalion command than either APGA or SATTOT. While AGPA was positively correlated with performance, the relationship only lasted 7 to 9 years (Lin et al. 2020). The authors conclude that leadership skill of West Point officers is the best predictor for long-term military performance.

Ergun (2003) examined the factors that affect the development and performance of Marine officers. The primary hypothesis of this research was that source of accession impacts performance due to the varying amounts of military acculturation provided by each program. USNA, NROTC, and MECEP have more exposure to officership than other programs and the expectation is that job performance will reflect these differences (Ergun 2003).

The underlying dataset used in Ergun (2003) consisted of 20 officer cohorts commissioned between 1980 and 1999. Ergun (2003) built 18 regression models and predicted TBS performance, fitness report scores, and promotion board results based on source of accession and other demographic covariates. This research found that source of accession was statistically significant when predicting officer performance but that many predicted outcomes changed over time. While PLC and OCC officers had lower fitness report scores early in their career, they promoted to O4 at higher rates than USNA and prior enlisted officers.

The research performed by Ergun (2003) provides a wealth of knowledge about officer accession and retention, but like Korkmaz (2005), relied upon outdated data. The economy, military landscape, and civilian job market have changed over the past 20 years.

Hurndon and Wiler (2008) examined the relationship between source of accession, TBS performance, and performance in the operating forces, as measured by fitness report scores. This particular study provided recommendations on modifications to the TBS MOS assignment process. Using the results of six different regression models, Hurndon and Wiler (2008) concluded that TBS leadership performance was the best predictor for company grade officer performance. These findings agree with Lin et al. (2020) as TBS leadership performance represents attributes similar to the West Point MGPA.

Hurndon and Wiler (2008) suffered from a limited scope. The independent variable considered only accounted for performance to the rank of captain. As Ergun (2003) indicated, early career performance for officers from some commissioning sources has little effect on promotion to O4. In comparison, Lin et al. (2020) broadened the analytical scope by examining long-term performance benchmarks like selection to O5 and battalion command; however, officers from accession sources other than West Point are not considered.

D. TALENT MANAGEMENT

In the first article of the *Towards a U.S. Army Officer Corps Strategy* series, Colarusso et al. (2009) presented a new concept for producing and sustaining a highly competent Army officer corps – focus on talent. Mid-career officer deficits, especially those matriculating from West Point and ROTC, expose Army officer talent management problems. West Point and ROTC have high standards for acceptance and seek to attract talented individuals into the officer corps. Mid-career workforce gaps negatively impact the development of junior officers and diminish the talent pool from which more senior officers are selected (Colarusso et al. 2009).

Poor retention from the two accession sources in which the Army invests most creates a persistent manpower management problem. The Army accessed additional officers to account for mid-career gaps without increasing the number of developmental opportunities for these officers. High turnover rates in these experience-building jobs create less capable junior officers, which leads to dissatisfaction and poor retention (Colarusso et al. 2009). Another root cause of the retention problem is the high demand for skilled labor in the civilian sector. Employers seek workers who can lead and manage teams, quicky synthesize information, and solve complex problems. The dynamic national security environment equips Army officers with valuable leadership experience that is desirable and highly profitable outside the Army (Colarusso et al. 2009). These challenges led Colarusso et al. to conclude that "an effective Officer Corps strategy recognizes the interdependency of accessing, developing, retaining, and employing officer talent" (2009, p. 17). Figure 3 depicts the retention rates of Army officers commissioned in 1996 and shows the poor retention of ROTC and West Point officers when compared to OCS officers; after 96 months of service, roughly 75 percent of OCS officers are still serving while only roughly 40 percent of West Point officers are still serving.

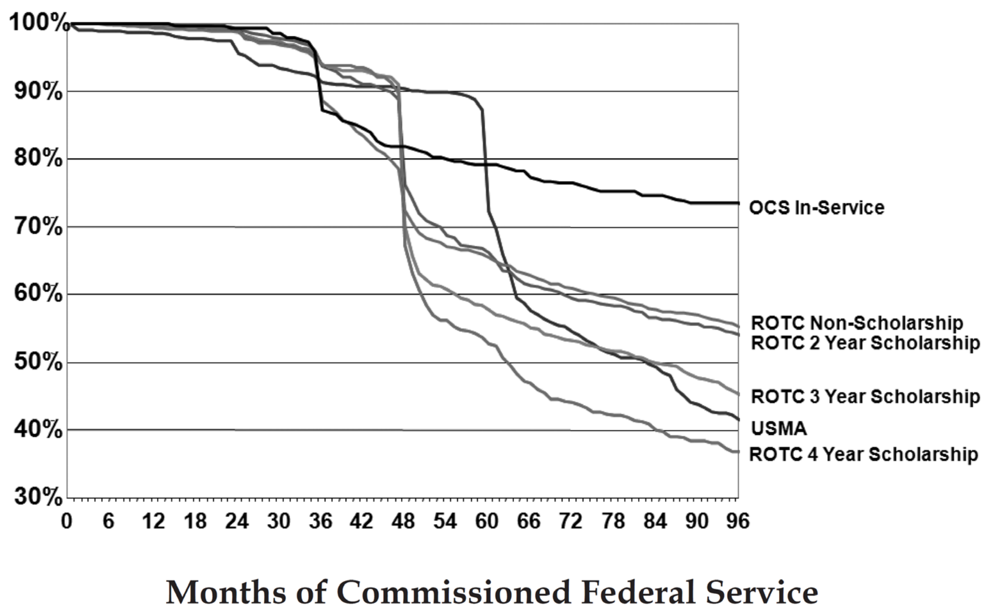

Figure 3. Percent of Year Group 1996 Army Officers Remaining on Active Duty through 96 Months of Service. Source: Colarusso et al. (2009)

Colarusso et al. (2009) offered a solution to the talent management problem in the form of an officer human capital model, seen in Figure 4. The crux of the dilemma is the Army's lack of institutional flexibility and failure to properly leverage officer talent. To overcome this challenge, the authors recognize the potential symbiotic relationship across four key areas: accessing, developing, retaining, and employing talent. In the authors' view, assigning officers to positions to avoid manpower gaps is not talent management. The needs of the officer corps are ever-changing and the ecosystem in which officers exist must be designed to achieve maximum performance. The adjustment of the control levers for these four key areas must happen in unison. Policy decisions regarding one key area must take into consideration the impacts on the adjacent key areas. Some specific talent management improvement actions mentioned in Colarusso et al. (2009) are accessing specific talents, distributing talent across the career fields, offering specialized retention programs to high-performing officers, and the aligning skills with career fields.

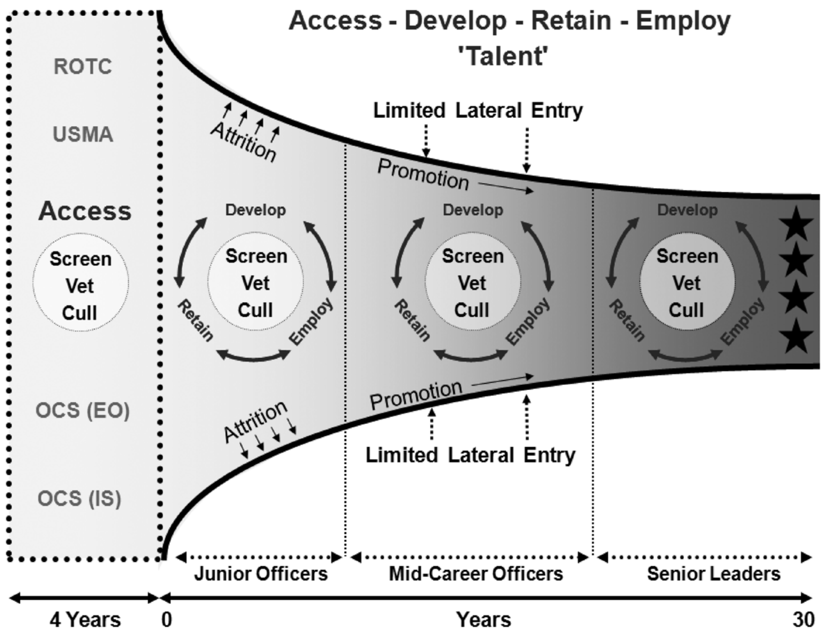

Figure 4. Proposed Army Officer Human Capital Model. Source: Colarusso et al. (2009)

The talent management research conducted by the Army and presented above can help the Marine Corps shape its own officer talent management strategy. This thesis seeks to equip the Marine Corps with a quantitative analysis of officer accession and retention to support that cause. Our research will build upon the survival and regression analysis presented in this chapter. In the next chapter, we will present on overview of the data and methodology used as well as a preliminary analysis of the data.

IV. DATA, METHODOLOGY, AND PRELIMINARY ANALYSIS

> We do not currently collect the data we need systematically, we lack the processes and technology to make sense of the data we do collect, and we do not leverage the data we have to identify the decision space in manning, training, and equipping the force.
>
> —General Berger, 38th Commandant, USMC, 2019

In this chapter, we describe the data and methodology used to conduct our research and present the results of our exploratory data analysis. We followed the data science project workflow as discussed by Rogel-Salazar (2017) and shown in Figure 5. Although we used a general sequence of activities and distinct phases marked progress of our research, we employed an iterative approach to complete the project. This chapter focuses on the first two phases of the workflow: data acquisition and data munging. It is important to note that this project did not start with a well-organized data source. Consequently, we found that the conclusions of Buttrey and Whitaker (2018) hold true; 80% of the project time was spent comprehending and preparing the data for analysis.

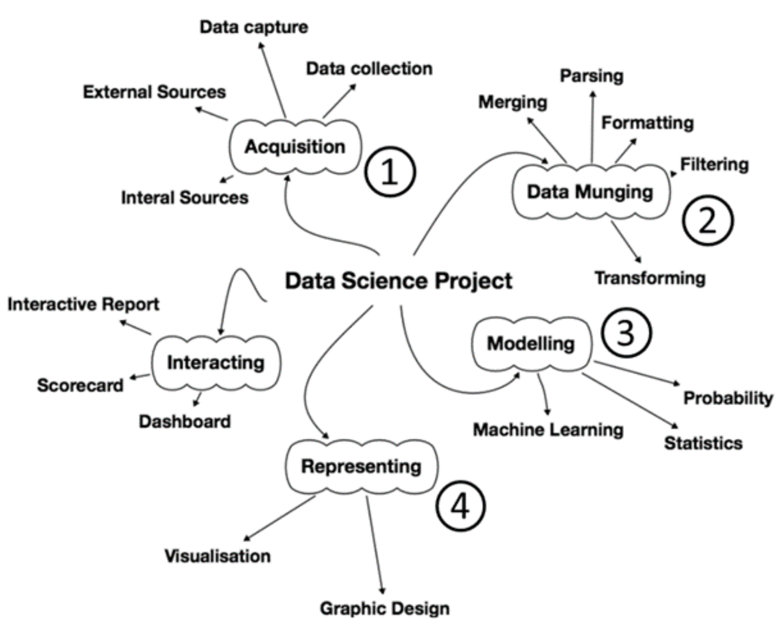

Figure 5. The Data Science Workflow. Adapted from Rogel-Salazar (2017)

A. DATA ACQUISITION

Due to the nature of USMC data management, different organizations collect and store pre-commissioning, early career, and long-term performance variables. To build a set of variables from these distinct periods, we gathered data from three different sources, each containing demographic and performance information about active-duty Marine officers commissioned between FY 2006 and FY 2016. Models were built using cleaned and merged data to describe officer retention and performance based on source of accession.

1. Total Force Data Warehouse Dataset

The core dataset used in this analysis came from the Total Force Data Warehouse (TFDW). TFDW houses more than 30 years of Marine Corps manpower data and is considered the official system used for U.S. Code Title 10 end-strength reporting (Total Force Data Warehouse 2021). We requested and received five comma-separated value files that contained many of the variables needed to complete this research. Table 6 displays the dimensions of each file. A unique identifier distinguished each officer and his or her associated qualitative and quantitative covariates. Some of the variables included in this data were commissioning date, source of accession, date promoted to next rank, GCT score, and end of active service (EAS) date. Since TFDW is comprised of official personnel records, it was treated as the authoritative data source for this research. In other words, when other datasets contained missing or conflicting information, TFDW was considered the correct data.

Table 6. Dimensions of TFDW Data

File	Dataset	Observations	Variables
COMM_CAREER_SCORES	TFDW	16,545	10
DATA	TFDW	101,781	30
GENDER	TFDW	16,545	2
RANK	TFDW	81,075	4
SCHOOLS	TFDW	99,735	6

2. M&RA TBS Dataset

Manpower and Reserve Affairs (M&RA) maintains a dataset of TBS performance records that include final academic, leadership, and military skills grades as well as class standing. Additionally, demographic, collegiate, MOS preference, and MOS assignment data are contained within. The TBS dataset came in the form of one comma-separated values file with 23,613 observations and 83 variables. Poor data quality in observations before 2010 forced us to drop many of the categorical variables from this analysis and assign mean values to continuous variables for 1,270 observations. The final dataset used relied on TBS performance data that estimated 13 percent of its observations, further highlighting data shortfalls faced in this thesis.

3. Marine Corps Recruiting Information Support System Dataset

The Marine Corps Recruiting Information Support System (MCRISS) is a web-based application that supports the collection and management of personnel information and recruiting activities (MCRC 2011). MCRC uses MCRISS to document mental, moral, and physical quality indicators and track an officer candidate's progress during the accession process (MCRC 2016). To support this research, MCRC queried MCRISS and provided a dataset containing demographic and performance information on all officers commissioned between FY 2006 and FY 2016. The data existed in the form of five comma-separated value files and Table 7 displays the dimensions of each file.

Table 7. Dimensions of MCRISS Data

File	Dataset	Observations	Variables
MCRISS_CUM	MCRISS	15,851	8
MCRISS_EDU	MCRISS	12,149	2
MCRISS_PFT	MCRISS	190,201	22
MCRISS_TEST	MCRISS	81,823	5
MCRISS_WAIV	MCRISS	9,710	19

Although the MCRISS dataset contained valuable pre-commissioning performance variables, numerous inconsistencies called into question the quality of the data. USNA officers are not in MCRISS and most performance records for NROTC officers were missing. A comparison of the MCRISS and TFDW datasets revealed that MCRISS was missing 2,812 officers commissioned during this period. This disparity is most likely explained by the fact that NROTC and USNA Marine staff are not in the MCRC chain of command and do not have access to MCRISS. Additionally, many SAT, ACT, and ASVAB scores were mislabeled. MCRISS, being a multi-user system, could be beneficial for information sharing but is only useful if accurate information is entered into the system. As a result of these deficiencies, MCRISS data was cleaned and explored but not considered in the survival and regression analysis.

B. DATA MUNGING

Data munging, also known as data wrangling, is the process of preparing data for the mining of valuable insights. The steps taken to clean raw data are necessary precursors to the modeling phase and help to gain a better understanding of the information contained in the data.

1. Data Preparation

Our research leveraged R (R Core Team 2013), a statistical programming language, to read, clean, merge, and analyze the data gathered for this research. Data was structured in an R data frame so that each observation represented a single Marine officer and each column represented a unique covariate. The process for preparing the data was complex because the raw TFDW data arrived in a long format; data describing a single officer was spread across multiple rows. For example, the TFDW rank data was structured so that each row represented a distinct promotion date and rank for an individual officer. After reshaping long-format data to wide-format data, binary categorical variables were created to signify whether an officer achieved a variety of career milestones.

During data preparation, many anomalies were detected and addressed. Although the TFDW data was the most reliable source, there were still incongruous entries and excessive factor levels. For example, 946 officers had multiple commissioning dates and 150 officers had multiple EAS dates. Categorical variables such as separation narrative, home of record, career designation, and education had too many factor levels to be of any use. General knowledge of the subject matter and research into the data guided the recoding of these categorical variables. The final R data frame contained 16,311 observations and 23 variables. Appendix A contains a complete list of all variables prepared during data munging.

2. Predictors

Predictors, also known as independent or explanatory variables, are input variables that describe the attributes or characteristics of an observation. The predictors used in this research describe qualitative and quantitative aspects of Marine officers commissioned between FY 2006 and FY 2016. Table 8 lists the predictors that were included during modeling.

Table 8. Predictors

Name	Source	Type	Description	Factor Level
FY	TFDW	Categorical	Fiscal year commissioned	11
Gender	TFDW	Categorical	Male or female	2
Source	TFDW	Categorical	Source of accession	5
Coast	TFDW	Categorical	Home of record coast	4
Education	TFDW	Categorical	Title of college degree	200
STEM	TFDW	Categorical	Degree classification	3
Type	TFDW	Categorical	Contract Type[a]	2
PFT	TFDW	Numeric	First PFT score	-
GCT	TFDW	Numeric	GCT score	-
TBS Academic	TBS	Numeric	TBS Academic GPA	-
TBS Leadership	TBS	Numeric	TBS Leadership GPA	-
TBS Military Skills	TBS	Numeric	TBS Military Skills GPA	-
Early Leadership	TBS	Categorical	90th TBS Leadership	2
Early Fitness	TBS	Categorical	First PFT >= 285	2
Early Intellect	TBS	Categorical	90th GCT	2

[a] All non-naval aviator contract types were grouped into one group called “ground.”

3. Responses

Responses, also known as dependent variables, are output variables that are measured or investigated in a scientific experiment. Statistical models are used to explain the relationship between predictors and responses. The responses used in this research describe an officer's length of active service and performance on three different selection boards. Length of active service was measured in years and performance on selection boards was measured using a binary response variable. Table 9 lists the responses that were included during modeling.

Table 9. Responses

Name	Type	Description	Factor Level
Survival	Numeric	Years of commissioned service	-
Failed	Logical	Censored data indicator	2
O4 Select	Categorical	Selected to O4	2
O5 Select	Categorical	Selected to O5	2
Career Designated	Categorical	Career Designation results	2

All dependent variables came from the TFDW data set.

C. EXPLORATORY DATA ANALYSIS

Exploratory data analysis was used to gain understanding about the data, visualize important relationships, and enhance model fitting. The dataset contained both categorical and continuous variables that can best be understood using summary statistics, plotting techniques, and statistical hypothesis testing.

1. Summary Statistics

Although most variables used in this analysis were categorical, boxplots and summary statistics were used to represent the continuous variables and illustrate the distribution of performance. Figure 6 shows officer performance during TBS training based on source of accession. When comparing TBS training success across the accession sources, we see that certain populations tend to achieve higher scores than others. PRIOR officer median performance on academic and leadership evaluations are two and three percentage points higher than the lowest scoring accession source, respectfully. This may be attributed to PRIOR officer military experience and familiarization with Marine Corps leadership traits and principles. USNA officer median performance on military skills evaluations are 0.7 percentage points higher than the lowest scoring accession source, possibly because of the four years of military experience gained while in college.

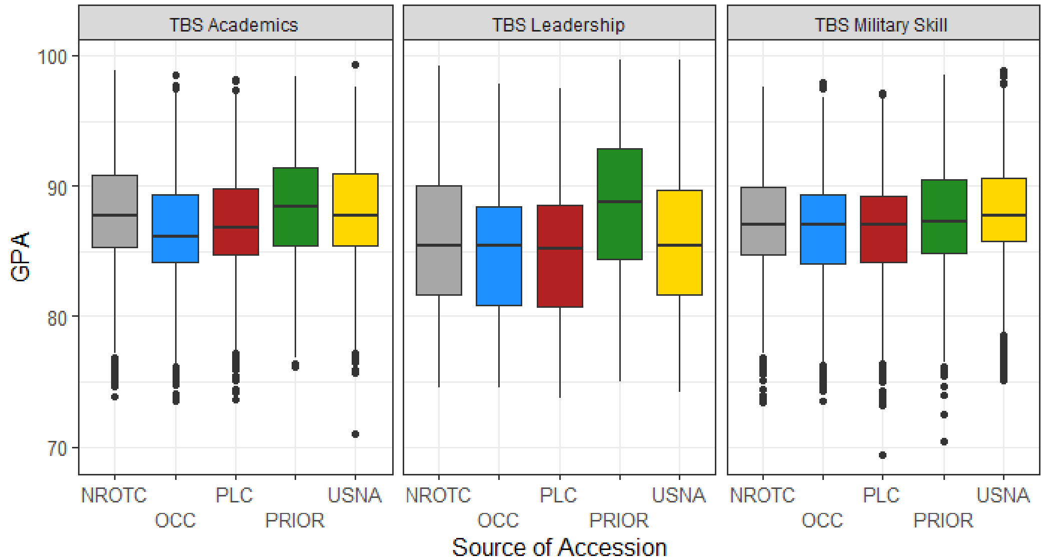

Figure 6. TBS Performance by Accession Source

Figure 7 shows performance during an officer's first PFT. NROTC and USNA officer median PFT performance is four points higher than the lowest scoring accession source.

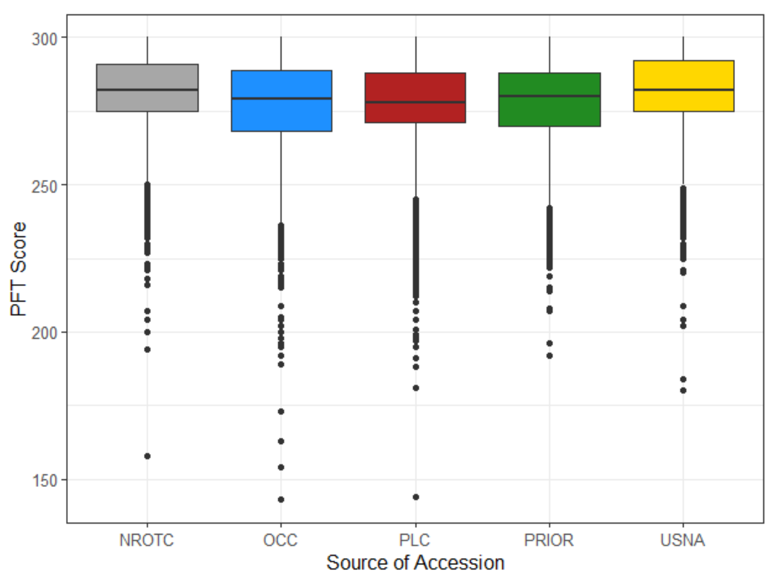

Figure 7. PFT Performance by Accession Source

Figure 8 shows officer performance on the GCT. USNA officer median GCT performance is five points higher than the lowest scoring accession source.

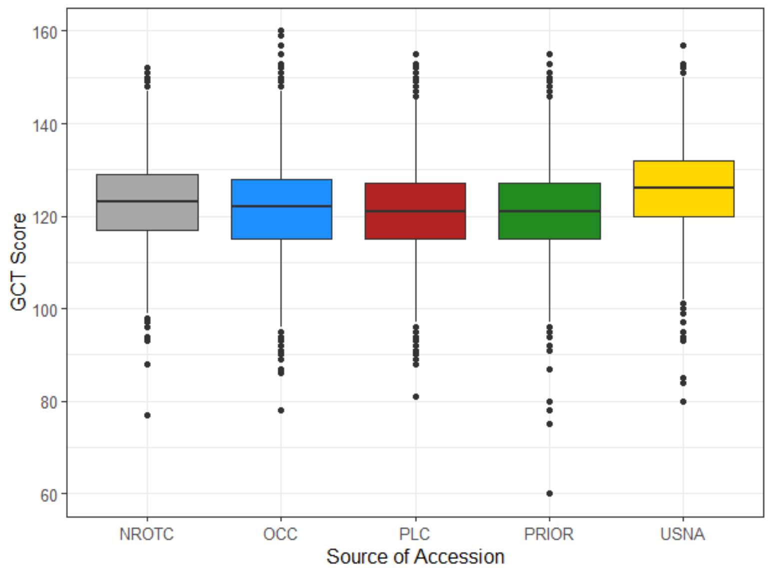

Figure 8. GCT Performance by Accession Source

Appendix B presents detailed summary statistics for each source of accession.

2. Source of Accession

Figure 9 displays the number of officers commissioned each year, categorized by source of accession. The proportion of the total accession mission dedicated to each source of accession is also depicted. To meet end-strength objectives, officer accession numbers from each accession source fluctuate year-to-year. When officer corps requirements increase or decrease, Figure 9 shows that the OCC program serves as the major tuning dial for officer accessions. OCC numbers go up when workforce requirements increase and down when workforce requirements decrease. For the most part, NROTC and USNA accession numbers have not changed while PRIOR and PLC accession numbers have decreased.

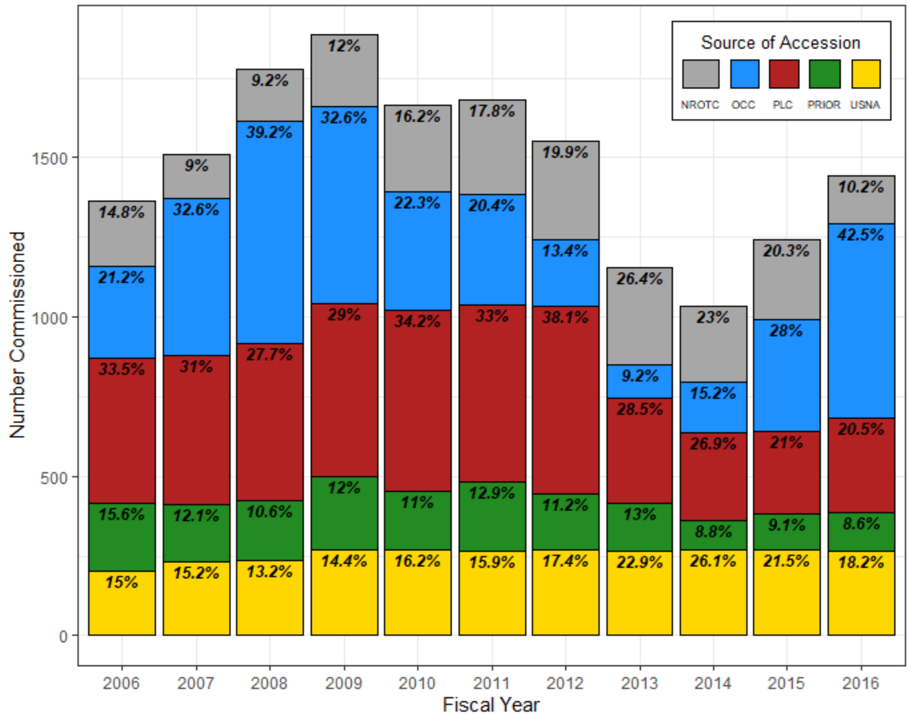

Figure 9. Accessions by Source Over Time

3. Contract Type

Figure 10 shows the distribution of officers by contract type from each source of accession. The preponderance of naval aviators originate from the USNA and PLC programs while very few originate from the PRIOR program. In fact, both USNA and PLC commit one-third of their commissioning quotas to naval aviation contract types.

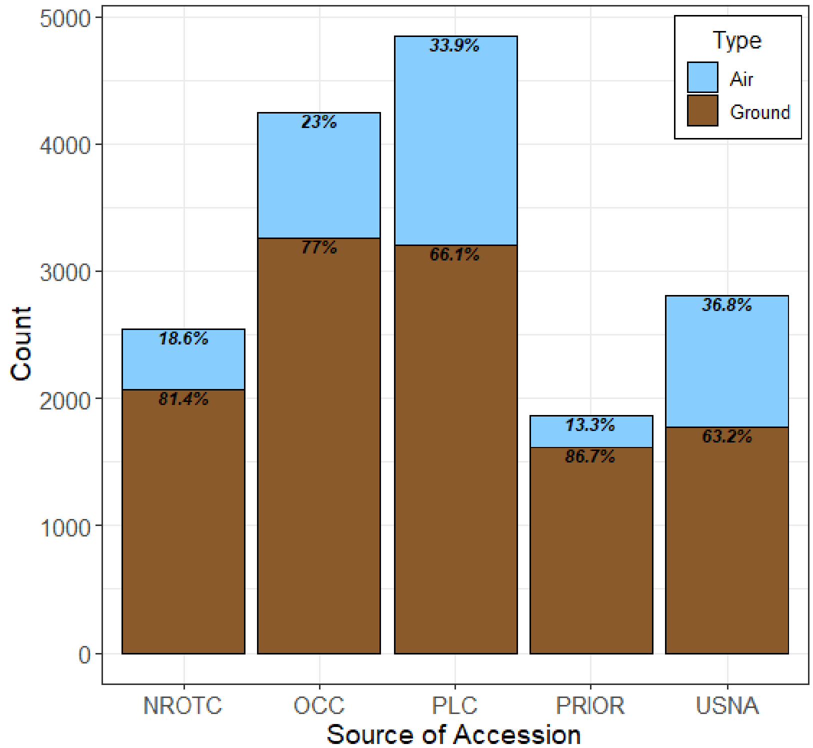

Figure 10. Contract Type by Accession Source

4. Gender

A 2020 report produced by the DOD (2020) on diversity and inclusion identified that "greater demographic diversity in accessions is essential to improving demographic diversity among senior military leaders in the future" (p. 20). To better understand female Marine officer retention patterns, we decided to investigate the relationship between source of accession and retention. It is possible that greater demographic diversity in accessions from certain sources will have a greater impact on achieving a more diverse senior officer corps. Figure 11 shows female officer accession numbers and proportions. There is a noticeable negative correlation between the total number of officers commissioned and the number of female officers commissioned. When larger numbers of officers are commissioned, the percentage of females commissioned is smaller.

Figure 12 shows female accession proportions by commissioning source over time. OCC and USNA programs consistently produce the largest portion of female officers.

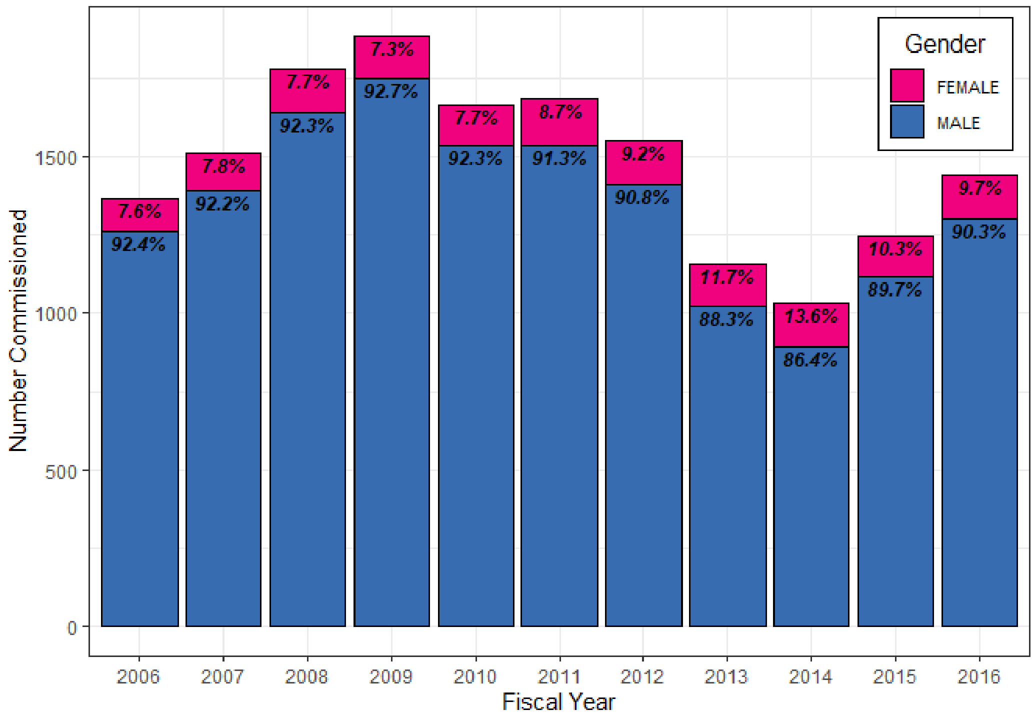

Figure 11. Accessions by Gender Over Time

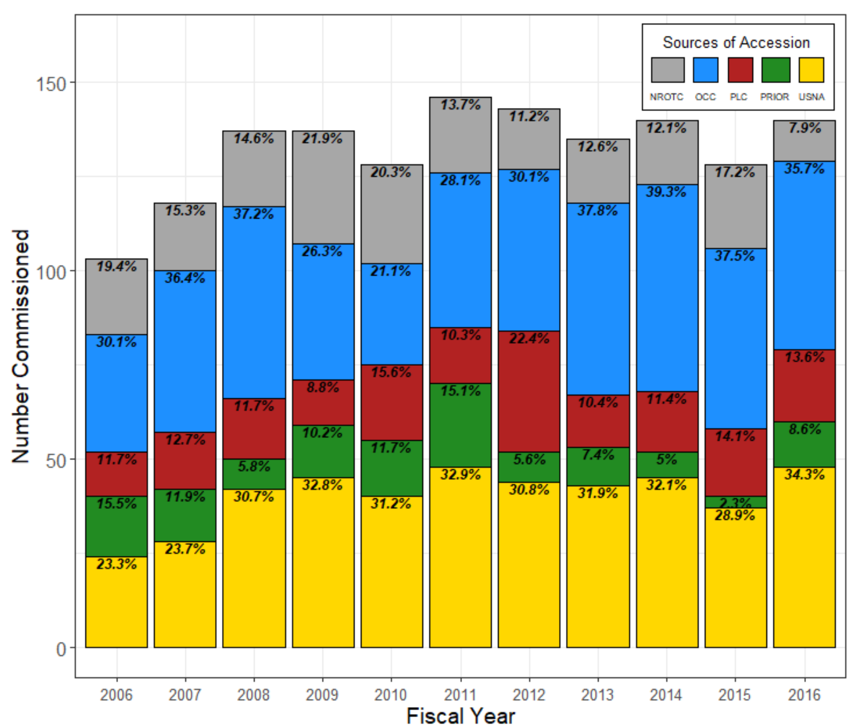

Figure 12. Female Accessions by Source Over Time

5. Career Designation

Since CD is the first major career milestone in an officer's career, it is important to understand. Officers who are not selected for CD exit the Marine Corps after the expiration of their initial contract. OCC and PLC officers have the lowest selection rates for CD while PRIOR and USNA officers have the highest. An interesting finding that was discovered during exploratory data analysis was the rate at which officers refuse CD. NROTC, a source that the Marine Corps invests in more than PLC and OCC, commissions officers who have the highest rate of CD refusal. Understanding CD trends can help increase the retention of talented individuals. Figure 13 shows consolidated CD results by source of accession and Figure 14 shows the rate at which CD was declined for each source of accession over time. NROTC and USNA both have a general increasing trend of refusing CD.

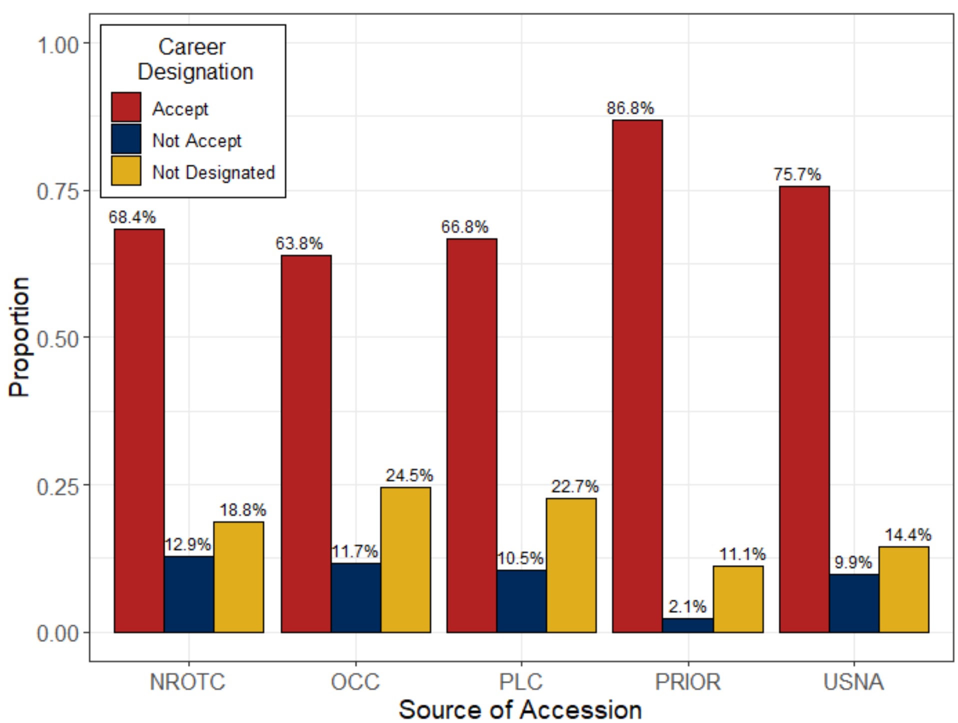

Figure 13. CD by Accession Source

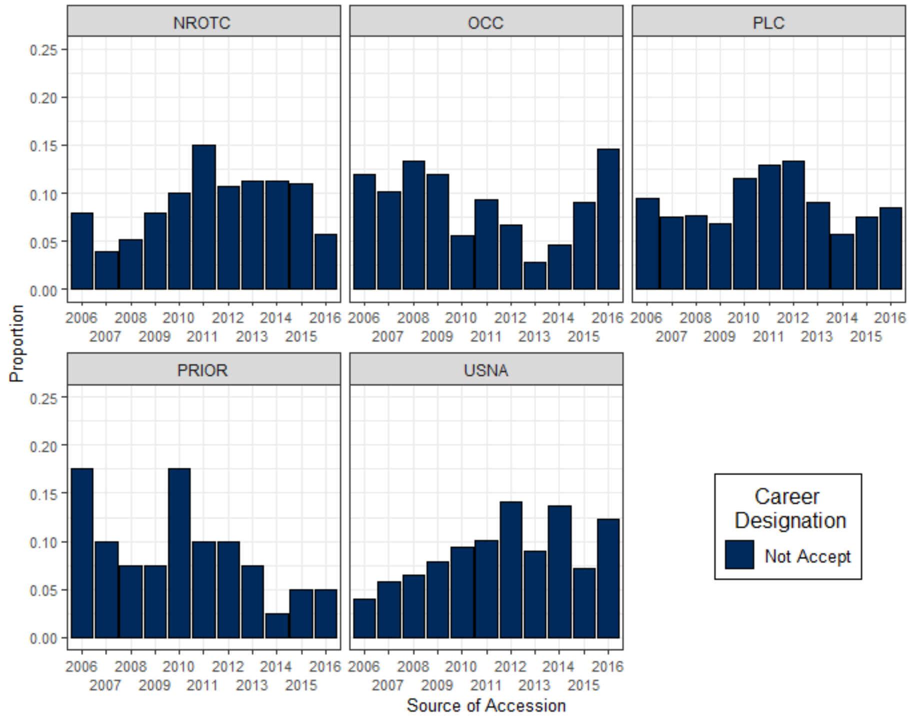

Figure 14. CD Refusal by Accession Source Over Time

6. Promotion

Figure 15 displays the proportion of officers who were selected for promotion to O4 and O5. Only officers who stayed on active service long enough to be considered for promotion were used for this analysis. Elements of six officer cohorts were considered for selection to O4 and elements of one officer cohort were considered for selection to O5. Since selection for promotion is based on performance, it is a good metric for job success. Each source has similar selection rates to O4, but larger differences emerge in O5 selection. Although officers from OCC had the lowest CD rates, their performance on both O4 and O5 selection boards were either the best or second best of the five sources.

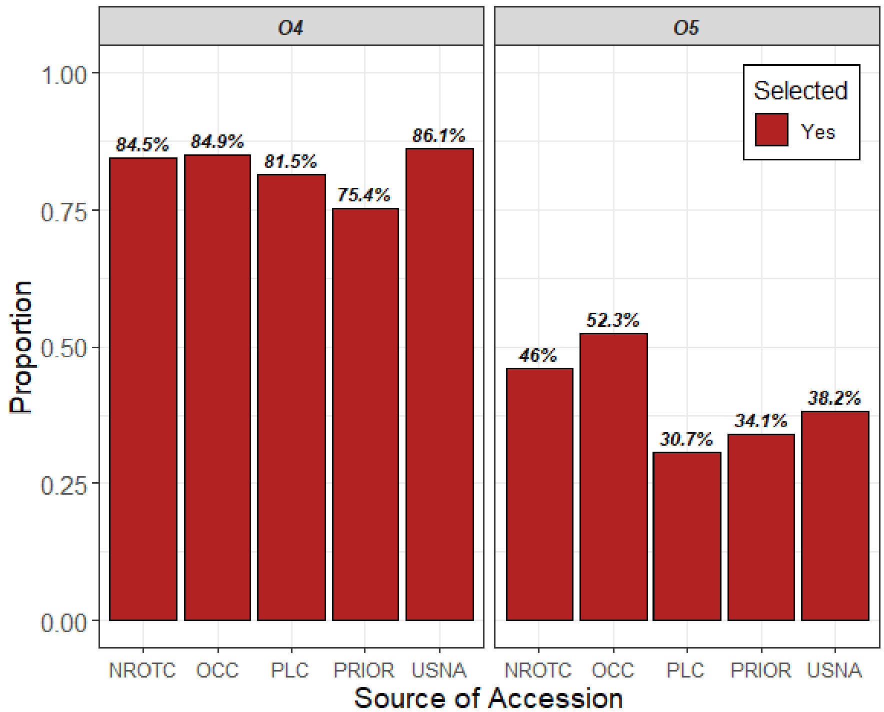

Figure 15. O4 and O5 Selection Results by Accession Source

7. Early Indicators of Performance

In addition to classifying officers by their source of accession, we used early career performance metrics to classify officers based on their physical fitness, intellect, and leadership. We created these early indicators of performance and cutoff values for the scope of this research. Officers who scored at least a 285 on their first PFT were classified as possessing early indication of physical fitness. Officers who scored at least a 136 on the GCT (90th percentile) were classified as possessing early indication of intellect. Officers who achieved at least a 92.83 TBS Leadership GPA (90th percentile) were classified as possessing early indication of leadership. Figure 16 shows the distribution of early indicators possessed by officers from each commissioning source. The results confirm our general intuition about expected officer performance from each source of accession. NROTC and USNA have longer periods of military acculturation, and we see that they have higher proportions of early fitness indication. The academically challenging curriculum at USNA produces a group of officers with higher proportion of early intellect indication. PRIOR officers have the highest percentage of early leadership indication, possibly attributed to their previous experience in the operational forces as active-duty enlisted Marines.

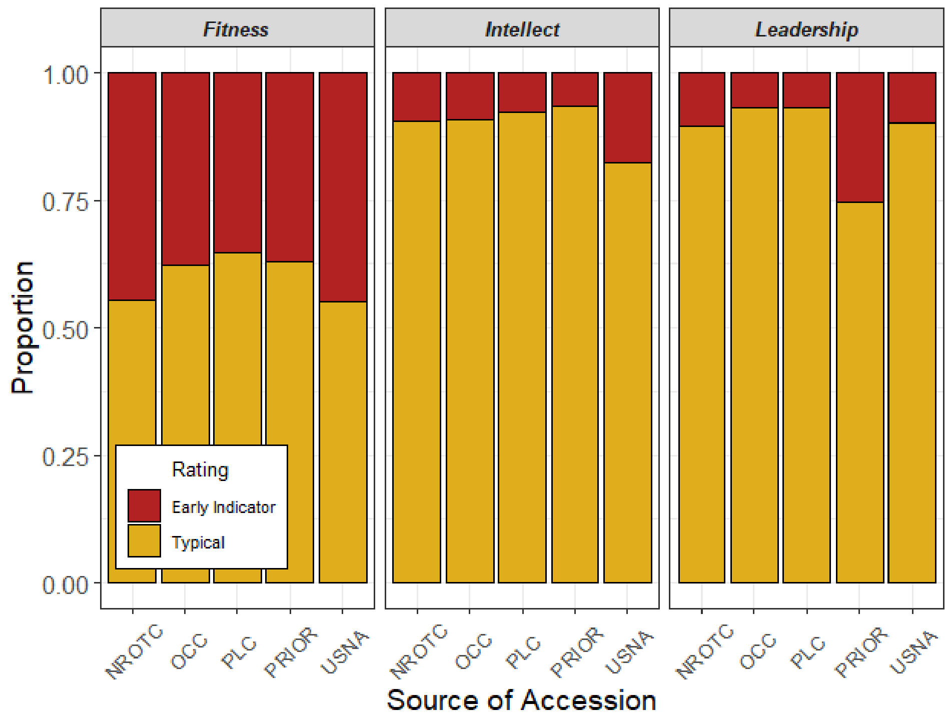

Figure 16. Early Indicators of Performance by Source

8. Pearson Chi-square Test

The Person Chi-square statistic is used to test independence in categorical count data. The null hypothesis states that there is no relationship between the variables. A test result that provides evidence that rejects the null suggests that there is an association between the variables (Smeeton and Sprent 2007). Contingency tables were constructed to compare source of accession and the achievement of each career milestone. Table 10 is an example of a contingency table that compares career designation and source of accession count data.

Table 10. Source versus CD

Source	CD	Not CD
NROTC	2063	393
OCC	3076	738
PLC	3724	952
PRIOR	1645	123
USNA	2045	365

Table 11 displays the results of three Pearson Chi-square tests used to determine if there was any relationship between source of accession and CD, selection to O4, and selection to O5. The low *p*-values for each of these tests indicate that there is evidence against the null hypothesis. In fact, at a level of 0.05, the null hypothesis is rejected for all three tests, and we conclude that there is an association between source of accession and career designation, selection to O4, and selection to O5. Although all *p*-values are significant, we see that *p*-value significance decreases for later career milestone.

Table 11. Source versus Milestone Pearson Chi-square Test Results

Milestone	Statistic	DOF	*p*-value
CD	105.54	4	2.2e-16
Selection to O4	48.27	4	8.2e-10
Selection to O5	12.31	4	0.015

V. ANALYSIS AND RESULTS

> We will make strategic investments in data science, machine learning, and artificial intelligence. Initial investments will be focused on challenges we are confronting in talent management, predictive maintenance, logistics, intelligence, and training.
>
> —General Berger, 38th Commandant, USMC, 2019

In this chapter, we introduce statistical modeling techniques and present key findings. Model building, the next step in the data science project workflow, is used "to describe the world and make predictions about what will happen next" (Buttrey and Whitaker 2018, p. xvii). Appropriate modeling algorithms were selected based upon the attributes of the data and the research objectives. We used survival and logistic regression models and visualizations to describe Marine officer retention and performance.

A. SURVIVAL ANALYSIS MODELING

Survival analysis is a statistical procedure used when the response variable of interest represents the time until an event occurs (Kleinbaum and Klein 2012). Survival analysis is often used when studying the effects of different medical treatments where time is observed from the beginning of treatment follow-up until the event, typically death, occurs. Intuitively, the time variable is referred to as survival time because it represents the time an individual in the study "survived" over a given period (Kleinbaum and Klein 2012). A comparison of survival times across treatment groups can help determine which treatment improves patient survival. In the case of this research, we are interested in modeling the time in years until an officer exits the Marine Corps while controlling for source of accession and early performance indicators.

A benefit to survival analysis is that it accounts for censored data. The exact survival time for every Marine officer in our data is unknown. While many have experienced the event of exiting the Marine Corps, there are some officers from the sample still on active duty and one can only speculate how long they will continue to serve. This type of attribute exists in survival data and is called censoring. More specifically, our analysis involves right-censored data since the ambiguity in survival time is on the right side of the observation period (Kleinbaum and Klein 2012). Figure 17 shows an example of a right-censored observation. Officers who exited the Marine Corps during the study period were classified as “true failures” and officers remaining on active duty at the end of the study period were classified as “censored.” Roughly 51 percent of the observations in our dataset were right-censored. Other statistical models like linear regression can attempt to represent time-to-event data, but biased results are inevitable because these methods cannot properly deal with censored observations.

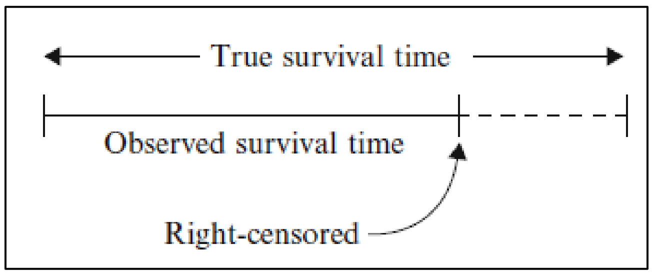

Figure 17. Right-Censored Data. Source: Kleinbaum and Klein (2012)

1. Survival Function

The purpose of survival analysis is to estimate the survival function, *S(t)*, an expression that gives the probability that a person’s survival time exceeds an indicated unit of time. The random variable capital letter *T* represents an individual’s survival time, and the small letter *t* represents a specific time of interest (Kleinbaum and Klein 2012). We use the data referenced in Chapter IV and the survival function to estimate and compare Marine officer survival probabilities. Equation 1 shows the survival function.

$$S(t) = \Pr(T > t) \tag{1}$$

The survival function is non-increasing as t ranges from 0 to infinity. When $t = 0$, the date an officer is commissioned, the survival function is equal to 1, meaning that all Marine officers are serving on active duty the day they are commissioned. As t increases and officers exit the Marine Corps, the survival function decreases toward zero. When data is used to estimate the survival function, a step function emerges as depicted in Figure 18.

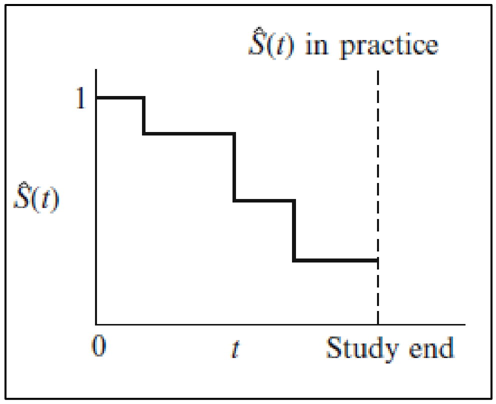

Figure 18. Survival Curve Example. Source: Kleinbaum and Klein (2012)

We used the Kaplan-Meier (KM) method (Kleinbaum and Klein 2012) by means of the “survminer” R package (Biecek et al. 2021) to estimate survival functions and visualize all KM survival curves. The KM method requires two response variables to estimate the survival function: a continuous variable that represents survival time and a logical vector that represents censoring.

2. Survival Models

We used KM survival curves to model and compare Marine officer longevity in active service while controlling for selected covariates. The evaluation of KM curves included visual inspection and the log-rank test. The log-rank test determines if the survival functions of two or more groups are statistically equivalent. At its core, the log-rank test is a large-sample Chi-square test that assesses observed count data and expected count data to calculate a test statistic (Kleinbaum and Klein 2012). Like the Pearson Chi-square test described in Chapter IV, the null hypothesis of the log-rank test states that there is no difference in survival between the groups and *p*-values below a threshold of 0.05 suggest that there is a difference in survival between the two groups. All *p*-values displayed are unadjusted. Since we compare the survival of multiple accession sources, contract types, and early indicators of performance, an adjustment like the Bonferroni one mentioned in Smeeton and Sprent (2007) might be employed for more accurate multiple hypothesis testing.

a. Contract Type

The KM survival curves shown in Figure 19 display officer survival based on source of accession and contract type. As stated in Chapter IV, all non-naval aviators are classified as "ground" officers. The variation in KM curve shape is helpful for distinguishing estimated officer survival variability over time. The KM curves that depict naval aviators are higher than the KM curves depicting ground officers, likely attributed to the dissimilarity in initial contract length for these two groups. Officers who exit the Marine Corps after their initial contract are depicted in the steep vertical lines at the four- and five-year points. The low log-rank test *p*-values displayed in Figure 19 indicate that there are differences in officer survival across accession source and contract type. To help distinguish the differences in survival, each survival curve displays 95 percent confidence intervals. Additionally, vertical "tick" marks signify censored observations, and in some figures, a dashed line signifies the median survival time for a group.

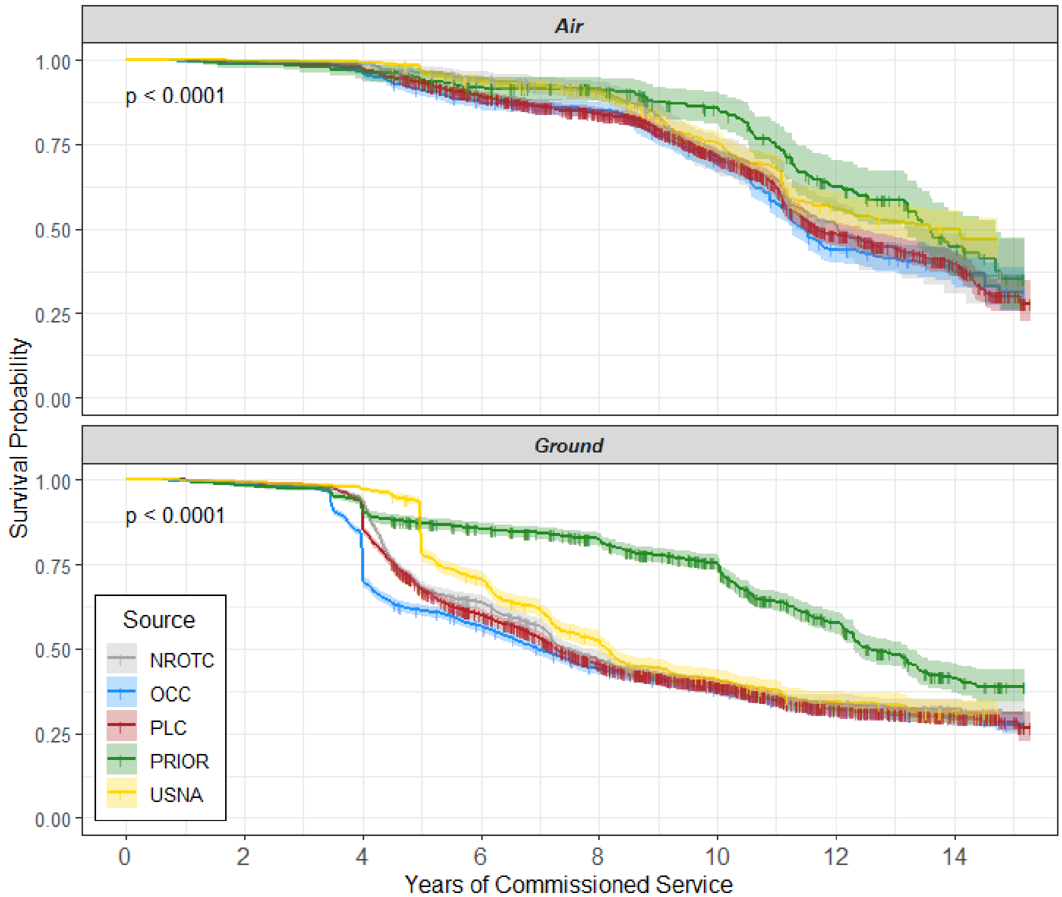

Figure 19. KM Estimated Survival Function by Accession Source and Contract Type

A principal benefit of survival analysis is the ability to estimate survival statistics from the survival function. Table 12 shows a summary of the estimated median survival times. For naval aviators, USNA officers have the highest median survival time and OCC officers have the lowest median survival time. For ground officers, PRIOR officers have the highest median survival time and OCC officers have the lowest median survival time. Table 13 shows a summary of the estimated 14-year survival probabilities for officers from each source of accession. The probabilities listed indicate the likelihood that an officer's survival exceeds 14 years. All estimated survival probabilities are displayed in Appendix C.

Table 12. Summary of KM Estimated Median Survival Times

Source	N	Event	Median	.95 LCL	.95 UCL
NROTC (A)	473	128	12.05	11.32	13.50
OCC (A)	978	376	11.45	11.26	11.75
PLC (A)	1644	609	11.69	11.45	12.43
PRIOR (A)	248	84	13.59	13.17	14.71
USNA (A)	1036	262	13.60	12.27	NA
NROTC (G)	2075	1156	7.41	7.23	7.74
OCC (G)	6265	1943	6.96	6.71	7.26
PLC (G)	3201	1894	7.26	7.12	7.53
PRIOR (G)	1615	548	12.55	12.31	13.23
USNA (G)	1776	886	8.11	8.0	8.27

Table 13. Summary of KM Estimated 14-Year Survival Probabilities

Source	Time	N	Survival Probability	Std Err	.95 LCL	.95 UCL
NROTC (A)	14	17	0.386	0.045	0.308	0.484
OCC (A)	14	53	0.377	0.025	0.332	0.429
PLC (A)	14	98	0.395	0.019	0.358	0.435
PRIOR (A)	14	31	0.445	0.049	0.359	0.551
USNA (A)	14	49	0.499	0.026	0.450	0.553
NROTC (G)	14	57	0.321	0.015	0.294	0.351
OCC (G)	14	102	0.292	0.011	0.271	0.314
PLC (G)	14	105	0.299	0.011	0.278	0.322
PRIOR (G)	14	80	0.412	0.022	0.371	0.458
USNA (G)	14	35	0.310	0.018	0.277	0.347

b. Early Indication of Fitness

The KM survival curves shown in Figure 20 display officer survival based on source of accession, contract type, and early indication of fitness. In most cases, officers possessing early indication of fitness have higher estimated survival probabilities early in their career, indicated by the light-green line being higher than the dark-green line. Still, the results of the log-rank test suggest that there is no difference in survival when controlling for early indication of fitness. At a level of 0.05, the only group with a statistically significant *p*-value is NROTC ground officers.

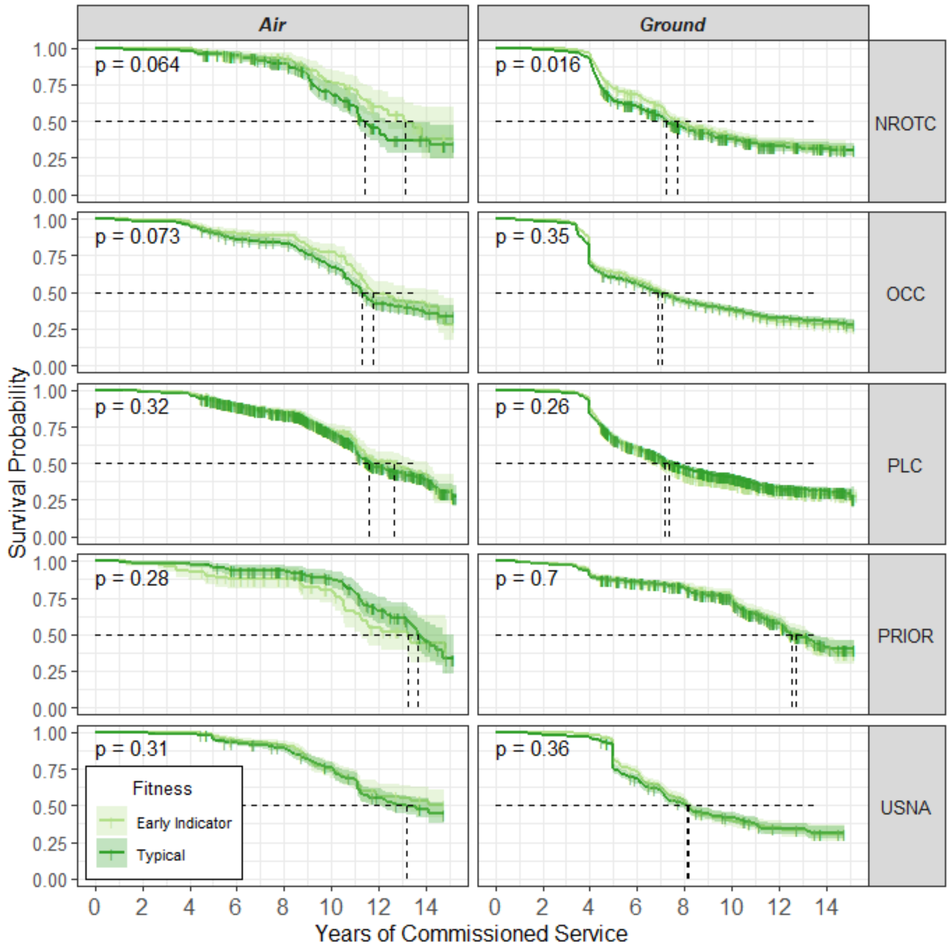

Figure 20. KM Estimated Survival Function by Accession Source, Contract Type, and Early Indication of Fitness

c. *Early Indication of Intellect*

The KM survival curves shown in Figure 21 display officer survival based on source of accession, contract type, and early indication of intellect. Although the only group with statistically significant results is PLC ground officers, an interesting finding emerges when comparing the shape of the KM survival curves based on contract type. PLC ground officers possessing early indication of intellect have lower estimated survival than those lacking while naval aviators possessing early indication of intellect have higher estimated survival than those lacking. In other words, PLC ground officers possessing early indication of intellect have worse retention than those lacking.

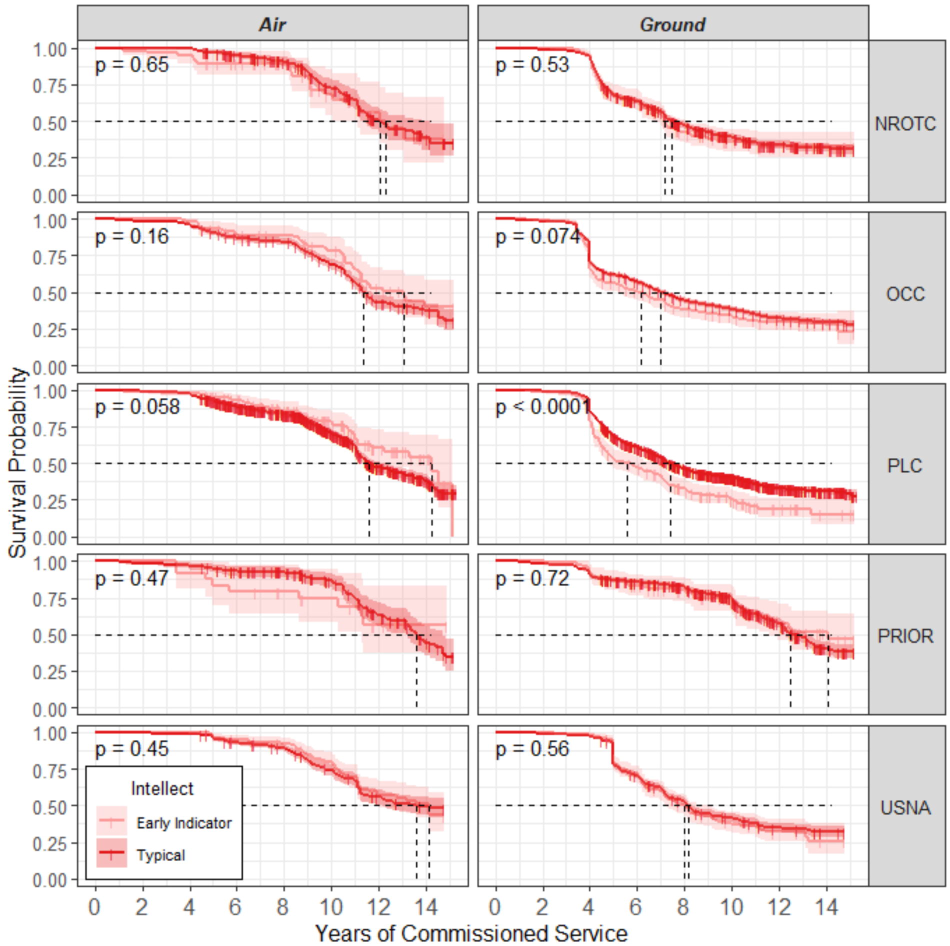

Figure 21. KM Estimated Survival Function by Accession Source, Contract Type, and Early Indication of Intellect

Figure 22 helps visualize the differences in survival based on contract type and early indication of intellect. When officer survival is controlled by these two variables only, it can be clearly seen that the aviation community retains officers possessing early indication of intellect at higher rates than the ground community.

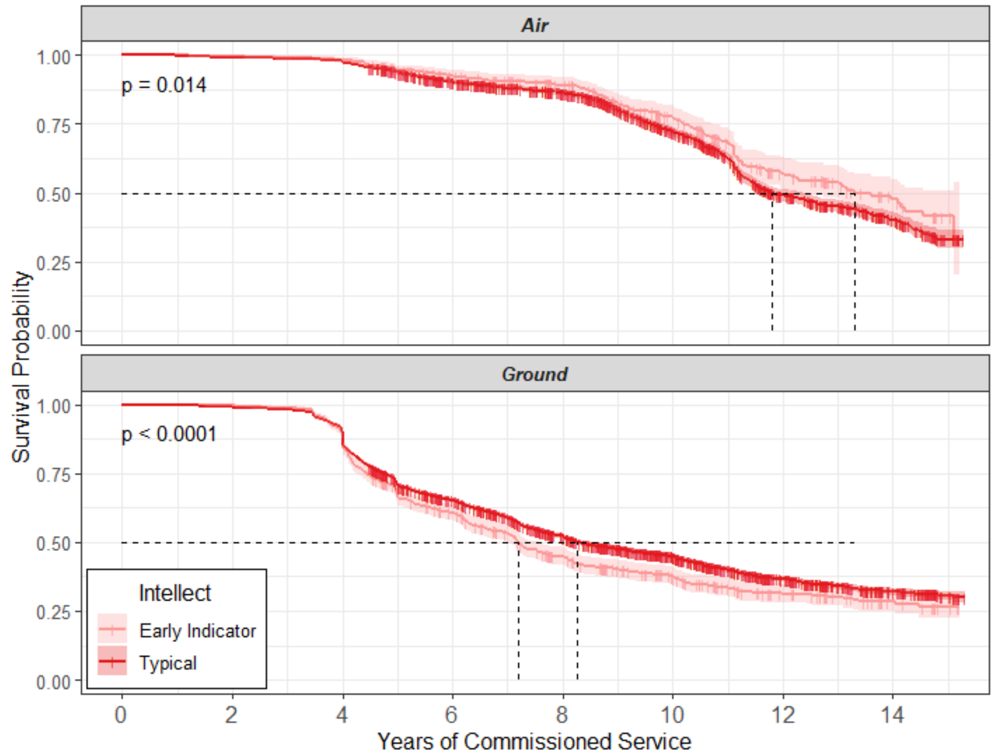

Figure 22. KM Estimated Survival Function by Contract Type and Early Indication of Intellect

d. *Early Indication of Leadership*

The KM survival curves shown in Figure 23 display officer survival based on source of accession, contract type, and early indication of leadership. In every case, officers possessing early indication of leadership have higher estimated survival than those lacking. In some cases, the difference in estimated survival is substantial. For example, the log-rank test *p*-value for OCC ground officer estimated survival is less than 0.0001. These results confirm our intuition that officers who achieve high leadership grades at TBS continue to perform well later in their career and are targeted by selection boards. It also suggests that the Marine Corps places a heavier influence on an officer's leadership abilities than other attributes. In six of ten cases, the difference in estimated survival based on early indication of leadership is statistically significant at a level of 0.05. Notably, the differences in estimated survival of USNA officers based on leadership indication are statistically insignificant. In other words, the estimated survival function of USNA officers with early indication of leadership and without are equivalent.

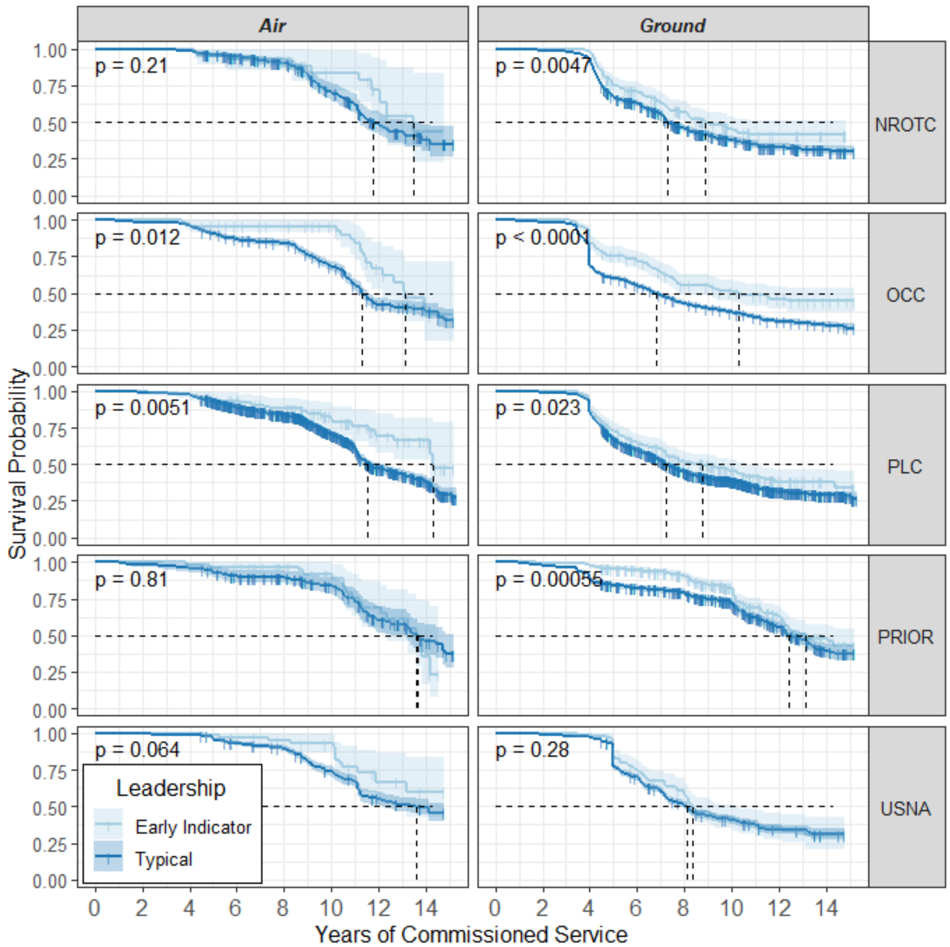

Figure 23. KM Estimated Survival Function by Accession Source, Contract Type, and Early Indication of Leadership

e. Gender

The KM survival curves shown in Figure 24 display officer survival based on source of accession, contract type, and gender. The survival curves for OCC, PLC, and PRIOR officers by gender are statistically equivalent. Figure 24 also shows that female NROTC and USNA air officers have lower estimated survival than males from the same source.

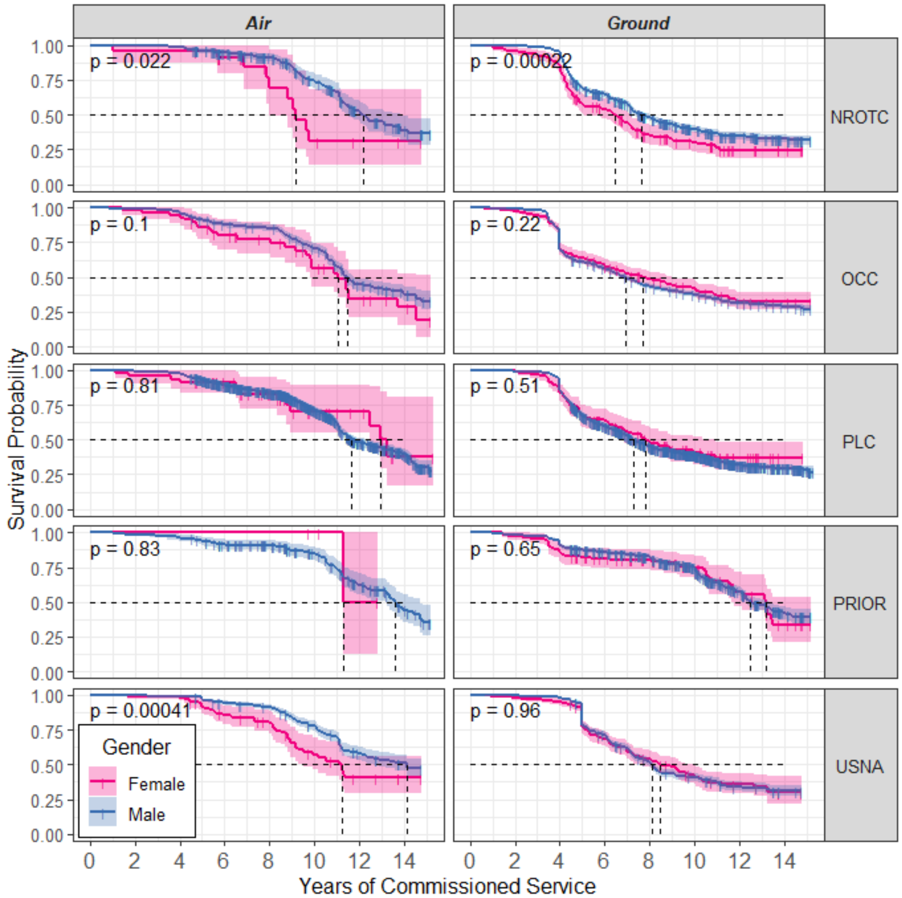

Figure 24. KM Estimated Survival Function by Accession Source, Contract Type, and Gender

Table 14 shows a summary of the estimated median survival times for female and male officers by source of accession and contract type. Within the female population, PLC naval aviators have the highest estimated median survival time and NROTC ground officers have the lowest estimated survival time. In fact, female NROTC ground officers have the lowest estimated median survival of all the groups examined. This is a surprising finding due to the years of investment made to train and commission female NROTC officers. We would expect to see higher estimated survival from officers who received the amount of military acculturation provided in the NROTC program. Within the male population, USNA naval aviators have the highest estimated median survival time and OCC ground officers have the lowest estimated median survival time. Table 14 also shows that female officers have higher estimated survival times than male officers within the OOC ground officer, PLC air and ground officer, PRIOR ground officer, and USNA ground officer communities.

Table 14. Summary of KM Estimated Median Survival Times by Accession Source, Contract Type, and Gender

Source	Gender	N	Event	Median	.95 LCL	.95 UCL
NROTC (A)	F	23	10	9.18	8.00	NA
NROTC (A)	M	450	118	12.20	11.45	13.84
NROTC (G)	F	194	129	6.48	5.05	7.17
NROTC (G)	M	1881	1027	7.61	7.29	8.07
OCC (A)	F	58	27	11.07	9.81	NA
OCC (A)	M	920	349	11.48	11.26	11.78
OCC (G)	F	418	225	7.71	6.64	8.98
OCC (G)	M	2847	1718	6.92	6.66	7.19
PLC (A)	F	46	13	12.96	12.47	NA
PLC (A)	M	1598	596	11.65	11.42	12.43
PLC (G)	F	143	74	7.81	6.77	10.96
PLC (G)	M	3058	1820	7.26	7.11	7.49
PRIOR (A)	F	4	1	11.27	11.27	NA
PRIOR (A)	M	244	83	13.59	13.17	14.71
PRIOR (G)	F	125	44	13.19	11.79	NA
PRIOR (G)	M	1490	504	12.50	12.31	13.29
USNA (A)	F	115	42	11.20	9.54	NA
USNA (A)	M	921	220	14.10	12.64	NA
USNA (G)	F	329	155	8.44	7.59	9.98
USNA (G)	M	1447	731	8.09	7.85	8.26

B. LOGISTIC REGRESSION MODELING

Logistic regression is a supervised learning technique used to predict the probability that a response variable belongs to specific category (Hastie et al. 2013). Like linear regression, logistic regression models the relationship between the predictors and response. The logistic function, p(X), is an expression that predicts the classification probability of an observation where all predictions lie between 0 and 1. The logistic function can incorporate p predictors by representing each predictor using X, where $X = (X_1,\ldots,X_p)$ for each unique predictor. During model fitting, estimated regression coefficients (β terms) provide insight into a predictor's association with the response. Equation 2 shows the generalized logistic function.

$$p(X) = \frac{e^{\beta_o + \beta_1 X_1 + \cdots + \beta_p X_p}}{1 + e^{\beta_o + \beta_1 X_1 + \cdots + \beta_p X_p}} \tag{2}$$

Hypothesis testing is used to determine which predictors are important enough to include in the model. The null hypothesis states that there is no relationship between a predictor and the response. Tests that yield low p-values provide evidence to reject the null and suggest that there is an association between a predictor and the response (Hastie et al. 2013). This is a necessary step in the modeling process because it helps to eliminate independent variables that are insignificant; keeping unimportant variables in a model could lead to misleading results. The so-called feature selection "procedure" is further explained in Hastie et al. (2013). In the case of this research, we are interested in using logistic regression to predict CD, O4, and O5 selection board results.

Logistic regression model performance can be evaluated in many ways, but we will use the receiver operating characteristic (ROC) curve. Since the output of a logistic regression model generates predicted probabilities, performance is determined by assessing the accuracy of predictions by comparing them to the truth. Considering that data is split into training and test sets, and only training set data is used to fit a model, our discussion of model performance is about the accuracy of test set predictions. There are two types of errors that can arise when making predictions: false positive and false negative. The ROC curve will be used for evaluating model performance because it displays both types of errors for all possible cutoff thresholds (Hastie et al. 2013). An ROC curve that approaches the top left corner of the plot is desirable as it indicates a high true positive rate and a low false positive rate. The area under the ROC curve (AUC) is also used as a numeric indicator of performance where an AUC of 1 describes a model that perfectly classifies test data.

We created three logistic regression models to predict three career milestones using a variety of early career predictors. Each binary response variable was recoded to 0 for officers who were not selected or 1 for those who were selected. For the CD model, officers who did not accept CD were considered selected and recoded to 1. After fitting logistic regression models, we used the "pROC" R package (Doering et al. 2021) to prepare the ROC curves, and the "ggplot2" R package (Chang et al. 2020) for all visualizations. The remainder of this chapter will present an analysis of these three models.

1. Career Designation

First, we examine the CD selection logistic regression model. Table 15 displays the predictors used to fit this model after performing feature selection. We encoded categorical predictors using dummy variables and removed predictors with non-statistically significant *p*-values from the model. The estimate value represents the regression coefficient which helps to illustrate a predictor's leverage on the response. The sign of the OCC and PLC coefficients indicate that officers from these sources were less likely to be selected for CD than the reference source, NROTC. Ground officers were also more likely to be selected for CD than naval aviators. Low *p*-values reveal an association between source of accession, PFT, TBS Leadership, TBS Military Skills and CD outcome.

Table 15. Career Designation Model Predictors

Source	Estimate	Std Err	*z*-stat	*p*-value
(Intercept)	1.266	0.070	18.154	0.000
Source [OCC]	−0.264	0.070	−3.750	0.000
Source [PLC]	−0.143	0.070	−2.048	0.041
Source [Prior]	0.529	0.102	5.191	0.000
Source [USNA]	0.339	0.084	4.031	0.000
Type [Ground]	0.249	0.050	5.013	0.000
PFT	0.116	0.022	5.377	0.000
TBS Leadership	0.248	0.026	9.703	0.000
TBS Military Skills	0.064	0.025	2.529	0.011

To assess CD model performance, several diagnostic plots are presented in Figure 25. Although the model has a less than ideal AUC of 0.64, the boxplot shows that the model is detecting a signal for CD selection. Officers who were truly selected for CD generally have higher predicted probabilities than those who were not. Finally, the scatterplot describes goodness of fit by displaying the average predicted probabilities against the proportion of actual CD selection by decile of predicted probability. Since the correlation between the points is 0.96, we conclude that the ordering of the probabilities is adequate.

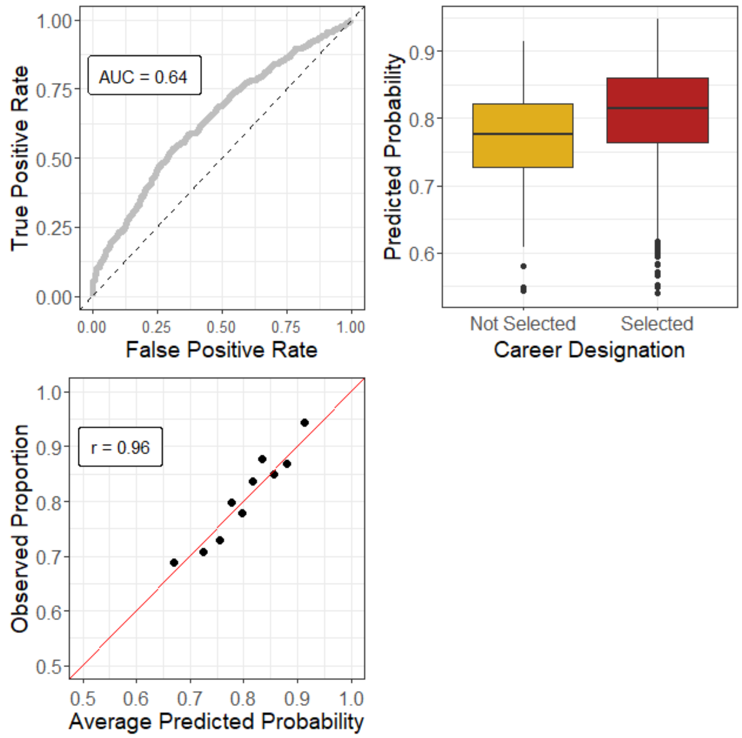

Figure 25. Career Designation Diagnostic Plots

2. Selection to O4

The second logistic regression model we examine is the selection to O4 model. Table 16 displays the predictors used to fit this model. Not surprisingly, the predictors chosen during feature selection for the selection to O4 model were the same as the CD selection model. It is plausible that the types of officers who are selected for CD are also selected for O4 if they remain on active duty long enough. There are, however, key differences from the CD selection model. The negative sign of the PRIOR coefficient shows that officers from this source are less likely to be selected for promotion to O4 than the reference source, NROTC. Relative to the other coefficients, the ground officer coefficient is dominant and indicates that contract type has a big impact on selection to O4. The TBS Military Skills coefficient is larger than the TBS Leadership one in the selection to O4 model but smaller in the CD selection model.

Table 16. Selection to O4 Model Predictors

Source	Estimate	Std Err	*z*-stat	*p*-value
(Intercept)	0.920	0.146	6.299	0.000
Source [OCC]	0.266	0.166	1.604	0.109
Source [PLC]	0.263	0.157	1.677	0.093
Source [Prior]	−0.707	0.166	−4.265	0.000
Source [USNA]	0.412	0.183	2.250	0.024
Type [Ground]	1.022	0.098	10.481	0.000
PFT	0.181	0.042	4.282	0.000
TBS Military Skills	0.330	0.051	6.516	0.000
TBS Leadership	0.224	0.051	4.438	0.000

Figure 26 presents the same diagnostic plots that were discussed earlier but for the selection to O4 model. The AUC for this model is 0.65 and no more appealing than the CD selection model. Fortunately, the average predicted probability versus selection boxplot is encouraging as it indicates that the model does detect a weak signal for selection to O4. The goodness of fit is 0.08 percentage points inferior to the CD selection model, reflected by scatterplot points that have drifted from the red line and a lower correlation.

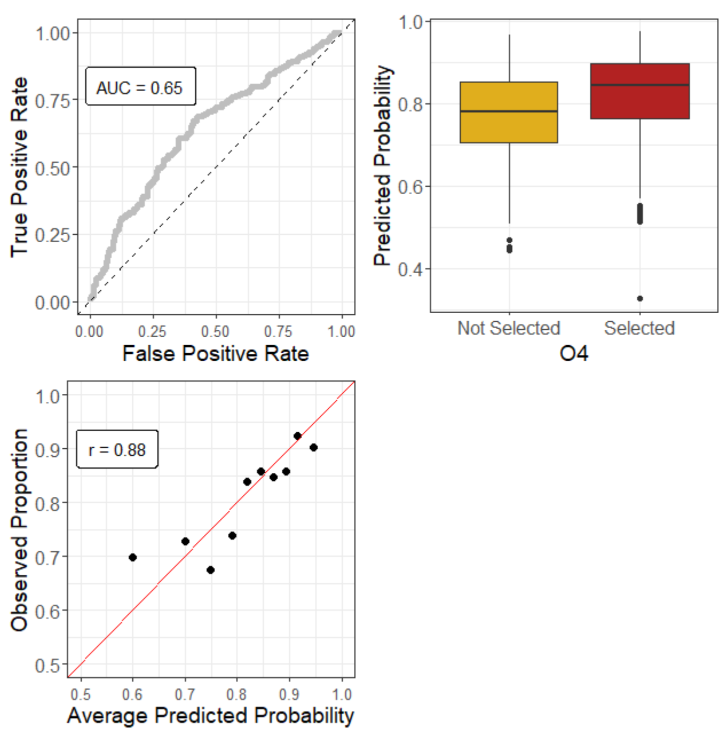

Figure 26. Selection to O4 Diagnostic Plots

3. Selection to O5

The final logistic regression model we examine is the selection to O5 model. Table 17 displays the predictors used to fit this model. At this point in the analysis, we see a noticeable difference in the predictors chosen during feature selection. While source of accession was a statistically significant predictor in previous logistic regression models, it is not for the selection to O5 model. The only predictors that are associated with selection to O5 are PFT, TBS Leadership, and TBS Academics. This result suggests that source of accession has less of an impact on the achievement of career milestones that happen later in an officer's career. A possible explanation for this is that early career performance is impacted by the experience gained in the commissioning program, but mid-career performance is impacted by the experience gained as a junior officer. Furthermore, contract type is not a statistically significant predictor in the selection to O5 model, implying that ground officers and naval aviators are equally likely to be selected for O5.

Table 17. Selection to O5 Model Predictors

Source	**Estimate**	**Std Err**	***z*-stat**	***p*-value**
(Intercept)	−0.546	0.117	−4.669	0.000
PFT	0.362	0.126	2.877	0.004
TBS Leadership	0.380	0.124	3.065	0.002
TBS Academics	0.439	0.130	3.386	0.001

Figure 27 presents the diagnostic plots for the selection to O5 model. The AUC for this model is 0.68 and is better than that of the other two models. The average predicted probability versus selection boxplot also shows suitable signal detection, but there still exists lots of overlap in the predicted probabilities of selected and not selected officers. Goodness of fit is 0.06 percentage points worse than the previous model. A possible explanation for this is the small amount of data available to fit this model. There were only 457 officers in the dataset who remained on active duty long enough to experience an O5 selection board. After the data was divided into train and test sets, 365 observations were used to fit the model. Conversely, the CD selection model was fit using 13,048 observations and the selection to O4 model was fit using 3,687 observations. Generally, larger amounts of training data will result in better model performance (Rogel-Salazar 2017).

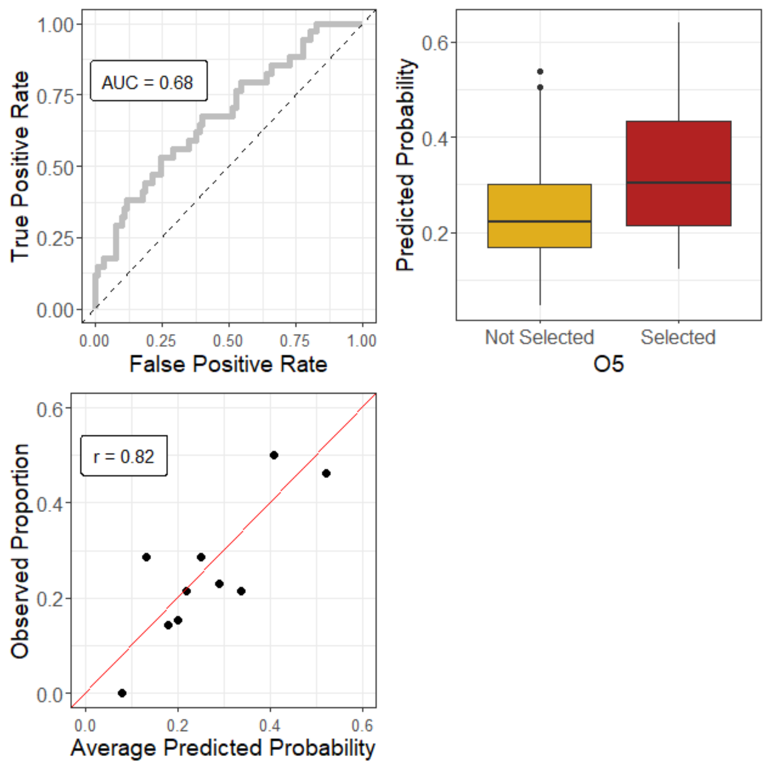

Figure 27. Selection to O5 Diagnostic Plots

VI. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

In this chapter, we will summarize the analysis conducted, explain several key findings, and recommend areas for future work.

A. SUMMARY

The motivation for this research stems from the recognition that the current Marine officer manpower model is lacking an emphasis on talent retention. Our objective was to use data science methodologies to model officer retention and performance based on source of accession and equip MCRC with quantitative results that can help shape possible changes to the officer accession and retention strategy. We followed the data science project workflow which included data acquisition, munging, modeling, and representation. Data was acquired from three different sources, prepared for analysis, and used to build survival and logistic regression models. We found that officer survival patterns were different across the commissioning sources, even when controlling for contract type. Additionally, we demonstrated that early indicators of performance were useful predictors when modeling early- and mid-career milestone achievement.

B. CONCLUSIONS

We identified several key findings during this research that are worth presenting. Challenges that materialized during data acquisition and munging revealed data availability and reliability concerns. The isolated nature of Marine Corps data collection and management practices limited our ability to gather and link officer performance over time. It is possible that non-standardized data entry and storage methods contributed to data accuracy shortfalls. Officer performance data is stored across several repositories, and it was difficult to gain support and receive data from each repository. For example, OCS and FITREP performance scores were requested but not made available to us.

The statistical models used in this thesis expounded on previous research conducted on officer performance and retention based on source of accession. Exploratory data analysis exposed that a higher proportion of NROTC officers deny CD than officers from the other sources of accession. Survival analysis clearly emphasized differences in officer retention based on contract type and source of accession. When controlling for contract type, the difference in officer survival was statistically significant across the sources of accession. We found that both ground and air officers from the OCC source had the lowest estimated median survival time. The early indication of leadership model displayed the most significant difference in estimated survival between those with and without early indication. Officers with early indication of intellect had higher rates of estimated survival in the aviation community but lower rates of estimated survival in the ground community. The difference between the estimated survival of male and female officers from the OCC, PLC, and PRIOR accession sources were not statistically significant. Additionally, we found that female NROTC ground officers had the lowest estimated median survival time of all groups examined.

Logistic regression analysis showed that source of accession was associated with CD and O4 selection but not selection to O5. The TBS Leadership predictor was present in all three logistic regression models, suggesting that it is an important variable when predicting early- and mid-career performance, even though it represents introductory officer aptitude. Although weak, career milestone selection and non-selection signals were detected by each model. The level of model performance achieved may indicate that the inclusion of more important predictors, such as OCS and FITREP scores, could improve model performance.

C. RECOMMENDATIONS

This thesis highlights the type of workforce analysis that can be performed in the future, especially if larger datasets that contain more predictors become available. Many desirable predictors are collected during the officer screening, selection, and training process, but data is not stored or managed in a standard way. This ultimately restricts the kind of analysis that can be performed. As the Marine Corps seeks to leverage its data to improve decision making, advancements in enterprise-level data collection and management should be prioritized. Officer performance data should be collected and managed with the intention of using it to model behavior and make predictions about the future. While TFDW possess an extensive set of predictors, the USMC could channel a broader range of performance data into this repository to make it more encompassing. An all-inclusive repository of officer performance data could have improved our analysis by generating more conclusive findings. The Marine Corps will not be able to leverage the power of machine learning until relevant and expansive datasets are created and used for modeling.

Our research showed that TBS Leadership performance is important when predicting CD, O4 and O5 selection. Further study should focus on predicting TBS Leadership performance using only pre-commissioning predictors to help improve the officer application process. Currently, standardized tests like the PFT and SAT are used to screen and select officers for commissioning programs, but these evaluations do not assess leadership aptitude. Letters of recommendation and officer interviews, which could be perceived as subjective, are used to access the leadership potential of an applicant. A study that determines which pre-commissioning predictors are associated with high TBS Leadership performance could be useful when allocating limited commissioning slots among many highly competitive applications. Additionally, the Marine Corps may benefit from developing an assessment that would be given to all officer applicants to help evaluate the baseline leadership abilities of those desiring to serve as Marine officers. Commercially available emotional intelligence assessments could be modified by the USMC and adopted to appraise the self-awareness and relationship management abilities of its officer applicants.

We recommend a deeper study into the CD and survival patterns of NROTC officers. NROTC officers had the highest rate of CD refusal and female NROTC ground officers had the lowest estimated median survival. Despite this finding, the data showed that NROTC officers had the second highest selection rate to O5. The fusion of exit interview data and survival analysis may elicit the factors contributing to these findings. Since the Marine Corps invests up to four years of time to commission an officer through the NROTC program, a study that seeks to understand how to improve NROTC officer return on investment would be valuable.

Since we showed that officers from the PRIOR source had heightened estimated survival, a study that determines the optimal window to select an enlisted Marine for a commissioning program is applicable. There is a perception that prior enlisted officers will only serve for a total of 20 years, and if they are commissioned too late, are less desirable for mid-career opportunities. We recommend a study be performed that identifies the prime career timeframe to commission prior enlisted Marines to maximize their longevity in active service and achievement of mid-career performance milestones.

APPENDIX A. COMPLETE VARIABLE SUMMARY

Covariate	Source	Type	Description	Factor Level
FY	TFDW	Categorical	Fiscal year commissioned	11
Gender	TFDW	Categorical	Male or female	2
Source	TFDW	Categorical	Source of accession	5
Enter	TFDW	Date	Date commissioned	-
Separation	TFDW	Categorical	Separation narrative	8
HOR	TFDW	Categorical	Home of record	66
Coast	TFDW	Categorical	Home of record coast	4
OCS	TFDW	Categorical	Attended OCS	2
TBS	TFDW	Categorical	Attended TBS	2
EWS	TFDW	Categorical	Attended resident EWS	2
CSC	TFDW	Categorical	Attended resident CSC	2
Education	TFDW	Categorical	Title of college degree	200
STEM	TFDW	Categorical	STEM degree classification	3
Field	TFDW	Categorical	Career field	7
Type	TFDW	Categorical	Contract Type	2
PFT	TFDW	Numeric	First PFT score	-
GCT	TFDW	Numeric	GCT score	-
TBS Academic	TBS	Numeric	TBS Academic GPA	-
TBS Leadership	TBS	Numeric	TBS Leadership GPA	-
TBS Military Skills	TBS	Numeric	TBS Military Skills GPA	-
Early Leadership	TBS	Categorical	90th Percentile	2
Early Fitness	TBS	Categorical	> 285 PFT	2
Early Intellect	TBS	Categorical	90th Percentile GCT	2
Survival	TFDW	Numeric	Years commissioned service	-
Failed	TFDW	Logical	Censored data indicator	2
O4 Select	TFDW	Categorical	Selected to O4	2
O5 Select	TFDW	Categorical	Selected to O5	2
Career Designated	TFDW	Categorical	Career Designation results	2

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B. SUMMARY STATISTICS

Name	Count	Mean	Median	Min	Max	S.D.
NROTC						
PFT	2548	280.6	282	158	300	14.6
GCT	2548	123.4	123	77	152	9.1
TBS Academic	2548	87.8	87.8	73.9	98.9	4.1
TBS Leadership	2548	85.7	85.4	74.6	99.2	5.3
TBS Military Skills	2548	87.3	87.1	73.4	97.6	3.9
OCC						
PFT	4243	276.4	279	143	300	17.8
GCT	4243	122.1	122	78	160	9.9
TBS Academic	4243	86.6	86.2	73.2	98.5	4.1
TBS Leadership	4243	84.9	85.4	74.5	97.8	5.1
TBS Military Skills	4243	86.7	87.1	73.5	97.9	4
PLC						
PFT	4845	277.4	278	144	300	16.1
GCT	4845	121.6	121	81	155	9.4
TBS Academic	4845	87	86.9	73.7	98.1	3.9
TBS Leadership	4845	84.8	85.2	73.8	97.4	5.1
TBS Military Skills	4845	86.6	87.1	69.4	97.4	4
PRIOR						
PFT	1863	277.3	280	192	300	16.2
GCT	1863	120.9	121	60	155	9.7
TBS Academic	1863	88.2	88.4	76.2	98.4	4.2
TBS Leadership	1863	88	88.7	75	99.7	5.5
TBS Military Skills	1863	87.5	87.3	70.4	98.5	4.2
USNA						
PFT	2812	280.7	282	180	300	15.4
GCT	4243	126.4	126	80	157	9.8
TBS Academic	4243	87.9	87.8	71	99.3	4.1
TBS Leadership	4243	85.5	85.4	74.2	99.6	5.3
TBS Military Skills	4243	87.9	87.7	75.1	98.8	3.7

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX C. SUMMARY OF SURVIVAL PROBABILITIES

NROTC Air					
Time	**N**	**Survival Probability**	**Std Err**	**.95 LCL**	**.95 UCL**
1	473	1.000	0.000	1.000	1.000
2	471	0.996	0.003	0.990	1.000
3	470	0.994	0.004	0.987	1.000
4	467	0.987	0.005	0.977	0.997
5	426	0.962	0.008	0.945	0.979
6	351	0.943	0.012	0.922	0.965
7	302	0.923	0.013	0.898	0.949
8	243	0.904	0.015	0.875	0.933
9	184	0.830	0.022	0.789	0.874
10	140	0.714	0.028	0.663	0.775
11	108	0.648	0.032	0.589	0.713
12	56	0.509	0.036	0.443	0.586
13	37	0.446	0.039	0.376	0.529
14	17	0.386	0.045	0.308	0.484

OCC Air					
Time	**N**	**Survival Probability**	**Std Err**	**.95 LCL**	**.95 UCL**
1	975	0.997	0.002	0.993	1.000
2	967	0.989	0.003	0.982	0.995
3	965	0.987	0.004	0.980	0.994
4	945	0.956	0.007	0.943	0.969
5	808	0.909	0.009	0.891	0.927
6	671	0.875	0.011	0.854	0.896
7	613	0.858	0.012	0.836	0.882
8	580	0.850	0.012	0.826	0.874
9	482	0.775	0.015	0.746	0.805
10	386	0.700	0.017	0.667	0.875
11	298	0.571	0.020	0.533	0.611
12	160	0.439	0.021	0.399	0.483
13	97	0.412	0.022	0.371	0.458
14	53	0.377	0.025	0.332	0.429

PLC Air					
Time	**N**	**Survival Probability**	**Std Err**	**.95 LCL**	**.95 UCL**
1	1642	0.999	0.001	0.997	1.000
2	1632	0.993	0.002	0.989	0.997
3	1619	0.985	0.003	0.979	0.991
4	1604	0.969	0.004	0.961	0.977
5	1480	0.930	0.006	0.918	0.943
6	1316	0.890	0.008	0.875	0.905
7	1173	0.860	0.009	0.843	0.878
8	1061	0.838	0.010	0.820	0.857
9	847	0.792	0.011	0.770	0.813
10	647	0.711	0.013	0.686	0.737
11	464	0.621	0.015	0.592	0.651
12	277	0.484	0.017	0.452	0.518
13	185	0.441	0.018	0.407	0.477
14	98	0.395	0.019	0.358	0.435

PRIOR Air					
Time	**N**	**Survival Probability**	**Std Err**	**.95 LCL**	**.95 UCL**
1	248	1.000	0.000	1.000	1.000
2	245	0.988	0.007	0.974	1.000
3	243	0.980	0.009	0.963	0.997
4	239	0.960	0.012	0.936	0.984
5	230	0.939	0.015	0.910	0.970
6	210	0.919	0.017	0.885	0.953
7	191	0.914	0.018	0.880	0.950
8	171	0.914	0.018	0.880	0.950
9	145	0.875	0.023	0.832	0.920
10	127	0.856	0.025	0.809	0.905
11	106	0.739	0.034	0.675	0.808
12	77	0.622	0.039	0.550	0.704
13	52	0.586	0.041	0.512	0.672
14	31	0.445	0.049	0.359	0.551

USNA Air					
Time	**N**	**Survival Probability**	**Std Err**	**.95 LCL**	**.95 UCL**
1	1036	1.000	0.000	1.000	1.000
2	1033	0.997	0.002	0.994	1.000
3	1030	0.994	0.002	0.990	0.999
4	1028	0.992	0.003	0.987	0.998
5	906	0.961	0.006	0.949	0.973
6	778	0.936	0.008	0.921	0.952
7	671	0.920	0.009	0.903	0.938
8	568	0.900	0.010	0.881	0.921
9	441	0.825	0.014	0.798	0.854
10	339	0.754	0.017	0.722	0.789
11	250	0.681	0.019	0.644	0.721
12	140	0.558	0.024	0.513	0.606
13	97	0.520	0.025	0.474	0.572
14	49	0.499	0.026	0.450	0.553

NROTC Ground					
Time	**N**	**Survival Probability**	**Std Err**	**.95 LCL**	**.95 UCL**
1	2074	1.000	0.000	0.999	1.000
2	2061	0.993	0.002	0.990	0.997
3	2043	0.985	0.003	0.979	0.990
4	1956	0.937	0.005	0.927	0.948
5	1339	0.678	0.010	0.958	0.698
6	1153	0.636	0.011	0.616	0.658
7	911	0.564	0.011	0.543	0.587
8	619	0.467	0.012	0.444	0.490
9	441	0.423	0.012	0.400	0.448
10	295	0.388	0.012	0.365	0.414
11	197	0.355	0.013	0.330	0.382
12	135	0.335	0.014	0.309	0.363
13	84	0.321	0.015	0.294	0.351
14	57	0.321	0.015	0.294	0.351

OCC Ground					
Time	**N**	**Survival Probability**	**Std Err**	**.95 LCL**	**.95 UCL**
1	3249	0.995	0.001	0.993	0.997
2	3226	0.988	0.002	0.984	0.992
3	3200	0.979	0.003	0.974	0.984
4	2600	0.699	0.008	0.684	0.715
5	1736	0.613	0.009	0.597	0.630
6	1427	0.566	0.009	0.549	0.583
7	1190	0.496	0.009	0.478	0.514
8	1010	0.439	0.009	0.421	0.458
9	888	0.406	0.009	0.388	0.425
10	722	0.380	0.010	0.362	0.399
11	572	0.345	0.010	0.326	0.364
12	400	0.318	0.010	0.300	0.338
13	229	0.305	0.010	0.286	0.325
14	102	0.292	0.011	0.271	0.314

PLC Ground					
Time	**N**	**Survival Probability**	**Std Err**	**.95 LCL**	**.95 UCL**
1	3200	1.000	0.000	0.999	1.000
2	3180	0.993	0.001	0.991	0.996
3	3157	0.986	0.002	0.982	0.990
4	2963	0.853	0.006	0.840	0.865
5	2041	0.671	0.008	0.655	0.688
6	1733	0.600	0.009	0.583	0.618
7	1436	0.532	0.009	0.514	0.550
8	1111	0.452	0.009	0.434	0.471
9	862	0.414	0.009	0.396	0.433
10	626	0.390	0.009	0.372	0.409
11	412	0.349	0.010	0.331	0.369
12	272	0.321	0.010	0.301	0.342
13	169	0.309	0.011	0.288	0.331
14	105	0.299	0.011	0.278	0.322

PRIOR Ground					
Time	N	Survival Probability	Std Err	.95 LCL	.95 UCL
1	1611	0.998	0.001	0.995	1.000
2	1589	0.984	0.003	0.978	0.990
3	1575	0.975	0.004	0.968	0.983
4	1479	0.905	0.007	0.890	0.919
5	1330	0.872	0.008	0.856	0.889
6	1221	0.856	0.009	0.839	0.873
7	1135	0.842	0.009	0.825	0.861
8	1018	0.822	0.010	0.803	0.842
9	836	0.774	0.011	0.753	0.796
10	651	0.754	0.012	0.732	0.777
11	431	0.639	0.015	0.610	0.668
12	287	0.575	0.017	0.544	0.609
13	155	0.484	0.019	0.447	0.523
14	80	0.412	0.022	0.371	0.458

USNA Ground					
Time	N	Survival Probability	Std Err	.95 LCL	.95 UCL
1	1774	0.999	0.001	0.997	1.000
2	1762	0.992	0.002	0.988	0.996
3	1752	0.986	0.003	0.981	0.992
4	1727	0.972	0.004	0.965	0.980
5	1351	0.780	0.010	0.760	0.800
6	1012	0.707	0.011	0.685	0.730
7	781	0.617	0.012	0.593	0.642
8	565	0.525	0.013	0.500	0.552
9	399	0.441	0.014	0.415	0.469
10	288	0.410	0.014	0.384	0.439
11	197	0.379	0.015	0.351	0.408
12	126	0.344	0.015	0.315	0.375
13	79	0.336	0.016	0.306	0.368
14	35	0.310	0.018	0.277	0.347

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Asch BJ, Hosek J, Mattock MG (2017) The Blended Retirement System: Retention effects and continuation pay cost estimated for the Armed Services. Report RR1887, RAND Corporation, Santa Monica, CA. https://www.rand.org/content/dam/rand/pubs/research_reports/RR1800/RR1887/RAND_RR1887.pdf.

Berger DH (2019) 38th commandant's planning guidance 2019. Washington, DC, https://www.hqmc.marines.mil/Portals/142/Docs/%2038th%20Commandant%27s%20Planning%20Guidance_2019.pdf?ver=2019-07-16-200152-700.

Berger DH (2020) Force design 2030. Washington, DC, https://www.hqmc.marines.mil/Portals/142/Docs/CMC38%20Force%20Design%202030%20Report%20Phase%20I%20and%20II.pdf?ver=2020-03-26-121328-460.

Biecek P, Fabian S, Kassambara A, Kosinski M (2021) Drawing survival curves using 'ggplot2'. R package version 0.4.9. Accessed April 6, 2021, https://cran.r-project.org/web/packages/survminer/survminer.pdf.

Buttrey SE, Whitaker LR (2018) *A Data Scientist's Guide to Acquiring, Cleaning, and Managing Data in R,* (John Wiley and Sons, Hoboken, NJ).

Chang W, Dunnington D, Henry L, Pedersen T, Takahashi K, Wickham H, Wilke C, Woo K, Yutani H, (2020), Create elegant data visualizations using the grammar of graphics. R package version 3.3.3. Accessed April 6, 2021, https://cran.r-project.org/web/packages/ggplot2/ggplot2.pdf.

Colarusso M, Lyle D, Wardynski C (2009) Towards a U.S. Army officer corps strategy: A proposed human capital model focused upon talent. *Strategic Studies Institute,* https://oema.army.mil/pub/2009_Wardynski_Lyle_Colarusso_v1_Towards_Officer_Corps_Strategy_Human_Capital_Model.pdf.

Coppage AP (2021) Marine officer accession information provided to the author via personal communication, April 5.

Department of the Army (2015) Talent management concepts of operations for force 2025 and beyond. Fort Leavenworth, KS, https://usacac.army.mil/sites/default/files/publications/TM%20CONOP%20Final-Signed%20v2.pdf.

Department of Defense (2017) Introduction to the Blended Retirement System. Presentation. Washington, DC, https://militarypay.defense.gov/Portals/3/Documents/BlendedRetirementDocuments/Intro_BRS_Website_06262017.pdf?ver=2017-06-26-142742-403

Department of Defense (2020) Recommendations to improve racial and ethnic diversity and inclusion in the U.S. Military. Washington, DC. https://media.defense.gov/2020/Dec/18/2002554852/-1/-1/0/DOD-DIVERSITY-AND-INCLUSION-FINAL-BOARD-REPORT.PDF.

Department of the Navy (2021) *Tentative Manual for Expeditionary Advanced Base Operations.* TM EABO, Washington, DC, https://intelshare.intelink.gov/sites/mcwl/TMEABO/_layouts/15/WopiFrame.aspx?sourcedoc=/sites/mcwl/TMEABO/SiteAssets/TM%20EABO%20-%20First%20Edition%20Rev%2020210415.pdf&action=default.

Doering M, Hainard A, Lisacek F, Muller M, Robin X, Sanchez J, Siegert S, Tiberti N, Turck N (2021), Display and analyze ROC curves. R package version 1.17.0.1. Accessed April 6, 2021, https://cran.r-project.org/web/packages/pROC/pROC.pdf.

Ergun E (2003) An analysis of officer accession programs and the career development of U.S. Marine Corps officers. Master's thesis, Graduate School of Business and Public Policy, Naval Postgraduate School, Monterey, CA. http://hdl.handle.net/10945/1118.

Hastie T, James G, Tibshirani R, Witten D (2013) *An Introduction to Statistical Learning with Applications in R*, (Springer, New York, NY).

Headquarters United States Marine Corps (2006) *Marine Corps Promotion Manual, Volume 1, Officer Promotions*. MCO P1400.31C, Washington, DC, https://www.hqmc.marines.mil/Portals/135/MCO%20P1400.31C.pdf.

Headquarters United States Marine Corps (2015) *Enlisted to Officer Commissioning Programs.* MCO 1040.43B, Washington, DC, https://www.mcrc.marines.mil/Portals/95/E-O/Reference%20Documents/MCO%201040.43B%20(E-O%20ORDER).pdf?ver=Yk0Y3BWeHDx4qqsKE20aiw%3d%3d.

Headquarters United States Marine Corps (2017) Fiscal years 2017–2021 Marine Corps manpower accession and retention strategy. Washington, DC.

Headquarters United States Marine Corps (2020) Fiscal year 2021 manpower accession and retention plan. MEMO-01. Quantico, VA.

Headquarters United States Marine Corps (2021a) *Human Resource Development Process (HRDP)*. MCO 5250.1, Washington, DC, https://www.marines.mil/Portals/1/Publications/MCO%205250.1.pdf?ver=6s0NePig2zbEMfLoh82Suw%3d%3d.

Headquarters United States Marine Corps (2021b) *Active Component Officer Retention and Prior Service Accessions.* MCO 1001.65A, Washington, DC, https://www.marines.mil/Portals/1/Publications/MCO%201001.65A.pdf?ver=ayJ8OToRkilm7SIoexDV0Q%3d%3d.

Hurndon N and Wiler D (2008) An analysis of performance at The Basic School as a predictor of officer performance in the operating forces. Master's thesis, Graduate School of Business and Public Policy, Naval Postgraduate School, Monterey, CA. http://hdl.handle.net/10945/4223.

Kleinbaum DG, Klein M (2012) *Survival Analysis: A Self-Learning Text*, 3rd ed. (Springer, New York, NY).

Korkmaz I (2005) Analysis of the survival patterns of United States Naval officers. Master's thesis, Graduate School of Business and Public Policy, Naval Postgraduate School, Monterey, CA. http://hdl.handle.net/10945/2203.

Lin E, Spain ES, Young LV (2020) Early predictors of successful military careers among West Point cadets. *Military Psychology*, 32(6): 389–407, https://www.tandfonline.com/doi/full/10.1080/08995605.2020.1801285.

Manpower & Reserve Affairs (2021) Manpower management and officer assignments (MMOA). Presentation, FY21 road show, Quantico, VA, https://www.manpower.usmc.mil/webcenter/ShowProperty?nodeId=%2FWebCenterSpaces-ucm%2FMMRA-WC1.MANPOW126420%2F%2FidcPrimaryFile&revision=latestreleased.

Marine Corps Recruiting Command (2011) *Marine Corps Recruiting Command Enlistment Processing Manual*. MCRCP 1100.1, Quantico, VA, https://www.hqmc.marines.mil/Portals/61/Docs/FOIA/MCRCO1100.1EPM.pdf.

Marine Corps Recruiting Command (2015) Volume II: Guidebook for officer selection officers, 2016 Edition. Quantico, VA, https://www.mcrc.marines.mil/Portals/95/Officer%20Programs/References/2016%20Volume%20II.pdf?ver=2016-07-19-103214-4530.

Marine Corps Recruiting Command (2016) *Marine Corps Recruiting Command Officer Commissioning Manual*. MCRCO 1100.2A, Quantico, VA, https://www.mcrc.marines.mil/Portals/95/MCRCO%201100.2A%20(OCM).PDF.

Marine Corps Recruiting Command (2021) Officer accession working group (OAWG). Presentation, Officer accession working group, January, 2021, Quantico, VA.

Naval Service Training Command (2019) Regulation for officer development. NSTC M-1533.2D. Pensacola, FL, https://www.netc.navy.mil/Portals/46/NSTC/cmd-docs/manuals/NSTC%20M-1533.2D%20Regulations%20for%20Officer%20Development%20v18%20Final.pdf.

Naval Service Training Command (2021) About NROTC. Accessed May 1, 2021, https://www.netc.navy.mil/Commands/Naval-Service-Training-Command/NROTC/About/.

Office of the Chief of Naval Operations (2021). Memorandum of agreement between Deputy Chief of Naval Operations (Manpower, Personnel, Training, and Education) and Deputy Commandant (Manpower and Reserve Affairs) and Deputy Commandant for Combat Development and Integration. Washington, DC.

Office of the Superintendent (2021) U.S. Naval Academy midshipman summer training program. USNAINST 1530.1E, Annapolis, MD, https://www.usna.edu/AdminSupport/Inst/1000-1999/USNAINST_1530.1E_U.S._Naval_Academy_Midshipman_Summer_Training_Program.pdf.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Accessed April 6, 2021, https://www.R-project.org/.

Rogel-Salazar J (2017) *Data Science and Analytics with Python* (CRC Press, Boca Raton, FL).

Smeeton N, Sprent P (2007) *Applied Nonparametric Statistical Methods*, 4th ed. (CRC Press, Boca Raton, FL).

Szoldra P, (2014) The U.S. military has always been terrible at predicting the future. *Insider* (March 28), https://www.businessinsider.com/us-military-predictions-2014-3.

The Basic School (2018) *Standard Operating Procedures for The Basic School*. BSO 1521.1H, Quantico, VA.

Total Force Data Warehouse (2021) Total Force Data Warehouse. Accessed April 29, 2021, https://tfdw-web.mceits.usmc.mil/index.aspx.

United States Marine Corps (2021) Assistant Commandant of the Marine Corps. Accessed April 18, 2021. https://www.hqmc.marines.mil/acmc/Talent-Management/.

United States Naval Academy (2014) *The Blue and Gold Book*. (Annapolis, MD). https://www.usna.edu/BlueAndGoldBook/_files/documents/The-Blue-and-Gold-Book.pdf.

Unites States Naval Academy (2020) 2024 class portrait, https://www.usna.edu/Admissions/_files/documents/ClassPortrait.pdf.

United States Naval Academy (2021a) Steps for admission. Accessed May 1, 2021, https://www.usna.edu/Admissions/Apply/index.php#fndtn-panel1-Steps-for.

United States Naval Academy (2021b) USNA core curriculum. Accessed May 2, 2021, https://www.usna.edu/Academics/Majors-and-Courses/Course-Requirements-Core.php.

United States Naval Academy (2021c) USNA Marines summer training options. Accessed May 2, 2021. https://www.usna.edu/USMC/Summer%20Training.php.

Urech, AC (2019) Exploratory survival analysis of Department of Defense blue-collar and white-collar civilian employee attrition factors. Master's thesis, Department of Operations Research, Naval Postgraduate School, Monterey, CA. http://hdl.handle.net/10945/62312.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California