NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

PREDICTORS OF SUCCESS AT INFANTRY TRAINING BATTALION USING COUNTERMOVEMENT JUMP METRICS

by

Cody E. Pennington

March 2021

Thesis Advisor: Marigee Bacolod
Co-Advisor: Jennifer A. Heissel

THIS PAGE INTENTIONALLY LEFT BLANK

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

1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** March 2021	**3. REPORT TYPE AND DATES COVERED** Master's thesis

4. TITLE AND SUBTITLE PREDICTORS OF SUCCESS AT INFANTRY TRAINING BATTALION USING COUNTERMOVEMENT JUMP METRICS	**5. FUNDING NUMBERS**
6. AUTHOR(S) Cody E. Pennington	
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 enlisted infantry community, all of whom acquire their training at Infantry Training Battalion (ITB), comprises approximately 15% of the Marine Corps. It is therefore concerning when, on average, 12.9% of the Marines who attend ITB fail to graduate. The majority are dropped from ITB training for four reasons: MOS Specific Physical Standards (MSPS) assessment failures, academic failure, medical injuries, and administrative issues. Of the four reasons, MSPS accounts for the majority of the failures (35.7%), followed by Academics (34.39%), Medical (23.35%), and the remainder (6.47%) for Administrative.

These statistics warrant investigation to determine what metrics can be utilized to mitigate failures. In 2019, ITB introduced a new curriculum that includes a newly developed MOS Specific Physical Standards (MSPS) assessment and force platforms to measure human kinetics and biomechanics through a Countermovement Jump (CMJ) test.

Data from multiple sources applied to econometric and machine learning models revealed that cognitive ability, demographics, physical performance, and CMJ performance are significant predictors of success at ITB. The most significant predictor turns out to be an interaction of cognitive ability and CMJ, indicating the complementarity of "brain and brawn" in determining success at ITB. Continued CMJ data collection and analysis could provide valuable insights into prediction-based schoolhouse training models.

14. SUBJECT TERMS Infantry Training Battalion, Marine Corps, Countermovement Jump, Force Platforms	**15. NUMBER OF PAGES** 93
	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.

PREDICTORS OF SUCCESS AT INFANTRY TRAINING BATTALION USING COUNTERMOVEMENT JUMP METRICS

Cody E. Pennington
Captain, United States Marine Corps
BS, The Citadel, 2014

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN MANAGEMENT

from the

NAVAL POSTGRADUATE SCHOOL
March 2021

Approved by: Marigee Bacolod
Advisor

Jennifer A. Heissel
Co-Advisor

Marigee Bacolod
Academic Associate, Graduate School of Defense Management

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The enlisted infantry community, all of whom acquire their training at Infantry Training Battalion (ITB), comprises approximately 15% of the Marine Corps. It is therefore concerning when, on average, 12.9% of the Marines who attend ITB fail to graduate. The majority are dropped from ITB training for four reasons: MOS Specific Physical Standards (MSPS) assessment failures, academic failure, medical injuries, and administrative issues. Of the four reasons, MSPS accounts for the majority of the failures (35.7%), followed by Academics (34.39%), Medical (23.35%), and the remainder (6.47%) for Administrative.

These statistics warrant investigation to determine what metrics can be utilized to mitigate failures. In 2019, ITB introduced a new curriculum that includes a newly developed MOS Specific Physical Standards (MSPS) assessment and force platforms to measure human kinetics and biomechanics through a Countermovement Jump (CMJ) test.

Data from multiple sources applied to econometric and machine learning models revealed that cognitive ability, demographics, physical performance, and CMJ performance are significant predictors of success at ITB. The most significant predictor turns out to be an interaction of cognitive ability and CMJ, indicating the complementarity of “brain and brawn” in determining success at ITB. Continued CMJ data collection and analysis could provide valuable insights into prediction-based schoolhouse training models.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

AAV	Assault Amphibious Vehicle
ASVAB	Armed Forces Vocational Aptitude Battery
AFQT	Armed Forces Qualification Test
CCRB	Course Content Review Board
CMJ	Countermovement Jump
CMP	Combat Marksmanship Program
CFT	Combat Fitness Test
DOD	Department of Defense
FY	Fiscal Year
GCT	General Classification Test
GT	General Test
HQMC	Headquarters Marine Corps
IAR	Infantry Automatic Rifle
IED	Improvised Explosive Device
ITB	Infantry Training Battalion
ITB-E	Infantry Training Battalion East
ITB-W	Infantry Training Battalion West
ITR	Infantry Training Regiment
LOE	Line of Effort
M&RA	Manpower and Reserve Affairs
MCDP	Marine Corps Doctrinal Publication
MCO	Marine Corps Order
MCRD	Marine Corps Recruit Depot
MCS	MOS Classification Standards
MCT	Marine Combat Training
MCTFS	Marine Corps Total Force System
MCTIMS	Marine Corps Training Information Management System
MOC	Marine Corps Operating Concept
MOS	Military Occupational Specialty
MOUT	Military Operations Urban Terrain

MSPS	MOS Specific Physical Standards
PFT	Physical Fitness Test
POI	Program of Instruction
ROTC	Reserve Officer Training Corps
SNCO	Staff Non-Commissioned Officer
SOI	School of Infantry
SOI-E	School of Infantry East
SOI-W	School of Infantry West
TD	Training Day
TECOM	Training and Education Command
TFDW	Total Force Data Warehouse
T&R	Training and Readiness
USMC	United States Marine Corps

ACKNOWLEDGMENTS

First, I would like to thank my wife, Danielle, and our two daughters, Brooklyn and Cali. They have provided unwavering support in all of my endeavors, often through times of uncertainty and stress. They are the foundation of my success.

Second, I would like to thank my thesis advisors, Dr. Marigee Bacolod and Dr. Jennifer Heissel, for their support, patience, and guidance during the completion of my study. Their insight, editing skills, and enthusiasm were crucial throughout my Naval Postgraduate School journey, and I will carry their wisdom with me through future endeavors.

Lastly, I would like to thank Mark White from the Office of Naval Research and Darin Peterson from the Human Performance Center at the School of Infantry for helping me develop a thesis topic that focuses on human performance and providing me with resources and points of contact to complete this study.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

> Set the mental and physical standards for Marine infantry through a mission-driven perspective that fully recognizes the demands on foot-mobile forces conducting operations in austere environments – *because superior infantry is a Marine Corps asymmetric advantage.*
>
> —*Marine Corps Operating Concept,* 2025

A. OVERVIEW

As the nation's expeditionary force in readiness, the Marine Corps focuses most of its resources on the infantry, with the remainder of the Marine Corps providing support. This idea is matched by the doctrine that "Every Marine is a rifleman," a focal point of former Commandant Alfred M. Gray, Jr., who emphasized infantry combat abilities. The infantry accounts for approximately 15 percent of the Marine Corps, all of which go through Infantry Training Battalion (ITB) or Infantry Officers Course (IOC) (Schaefer et al., 2015). It is concerning and costly when approximately 12.9 percent of Marines who attend ITB fail to graduate with their initial platoon (Infantry Training Battalion, 2019). These failures warrant investigation to determine what metrics, if any, can be utilized to mitigate failures.

In 2020, newly appointed Marine Corps Commandant General Berger released his Commandant's Planning Guidance. Within this guidance, the Commandant addresses the topic of training and education, he specifically states, "I am committed to ensuring each of you is provided the best educational opportunity available...this will require changes in how we evaluate academic performance, as well as how we annotate success, mediocrity, and potentially failure via performance evaluations" (U.S. Marine Corps, 2020). This statement holds true with the training of Marines at Infantry Training Battalion. On average, 12.5 percent of males and 48 percent of females attending the Basic Infantryman's Course at the School of Infantry-East will fail to graduate (Infantry Training Battalion, 2019). This is a steep cost at approximately $87,000.00 per trained servicemember per year (Dahlman, 2007).

In 2019, ITB introduced a new curriculum that includes a newly developed MOS Specific Physical Skills (MSPS) assessment and implemented force platforms to measure human kinetics and biomechanics through a Countermovement Jump (CMJ) test. It is important to note that the CMJ is conducted within one week of official training. The majority of Marines that attrite from ITB are due to four reasons: MSPS failures, academics failure, medical injuries, and administrative issues. Of the four reasons, MSPS accounts for the majority of the failures (35.79 percent), followed by Academics (34.39 percent), Medical (23.35 percent), and the remainder (6.47 percent).

B. PURPOSE OF THIS STUDY

The purpose of this study is to conduct a quantitative and qualitative analysis of candidates that attend ITB to determine which candidates will be successful using performance, demographic, and CMJ data. Moreover, I want to determine what factors are significant predictors of survival at ITB.

I address these questions by estimating logistic and survival regression models using data from Marine Corps' Total Force Data Warehouse (TFDW) and class data from ITB-East. Estimates from the models suggest that cognitive ability, physical performance, and CMJ metrics are the most significant success predictors. It is not only that they are independently significant, but also that cognitive and physical metrics captured by the CMJ interact together to predict success. While the CMJ is statistically significant among all models for success, it is difficult to determine if the data is meaningful and practical in an infantry training environment due to its intricacies. The impacts of physical and cognitive ability on accomplishments are not shocking; in addition, CMJ data's significance could help develop a critical screening tool for commanders in the future.

C. RESEARCH QUESTIONS

(1) Primary

- What performance, demographic, and countermovement jump predictive factors contribute to success and failure for enlisted Marines at Marine Corps basic infantry training?

(2) Secondary

- What determinants and hazards are statistically significant in the survivability of a potential ITB candidate?

- At what point(s) during the ITB course is a candidate most likely to attrite?

This research can help create a more successful candidate by minimizing failure and increasing each cycle's overall graduation rate. At a minimum, commanders can use this resource at the onset of training to determine which candidates have an increased likelihood of failure and implement preventative measures to mitigate this risk.

D. SCOPE AND LIMITATIONS

This study focuses on enlisted Marine Corps infantry candidates attending ITB from 2018 to 2019 who conducted the CMJ test. The scope is strictly enlisted Marine infantry candidates at ITB and does not include any analysis of non-infantry Marines at Marine Combat Training (MCT) or Marine officers.

E. METHODOLOGY

This study aims to develop predictive models to identify the key variables that explain a candidate's likelihood of graduating ITB. To accomplish this goal, I use the statistical software program STATA version 16.1. I utilize machine learning techniques Lasso Cross-Validation, Lasso Adaptive, Lasso Plugin, and Elastic net due to a large amount of CMJ, performance, and individual characteristic variables in my data set. These machine learning techniques prevent overfitting, helping me to identify less useful variables from my models. For each machine learning technique, I create a training and validation sample to cross-validate the results. The training sample consists of 80 percent of the total observations, while the validation sample consists of 20 percent of the total observations. I then compare each model's predictive power to the validation group to determine the preferred model. Finally, I estimate logistic regression and survival analysis models using only the preferred machine-learning model's variables to determine each factor's marginal effects on the likelihood of graduating from ITB and length of survival at ITB.

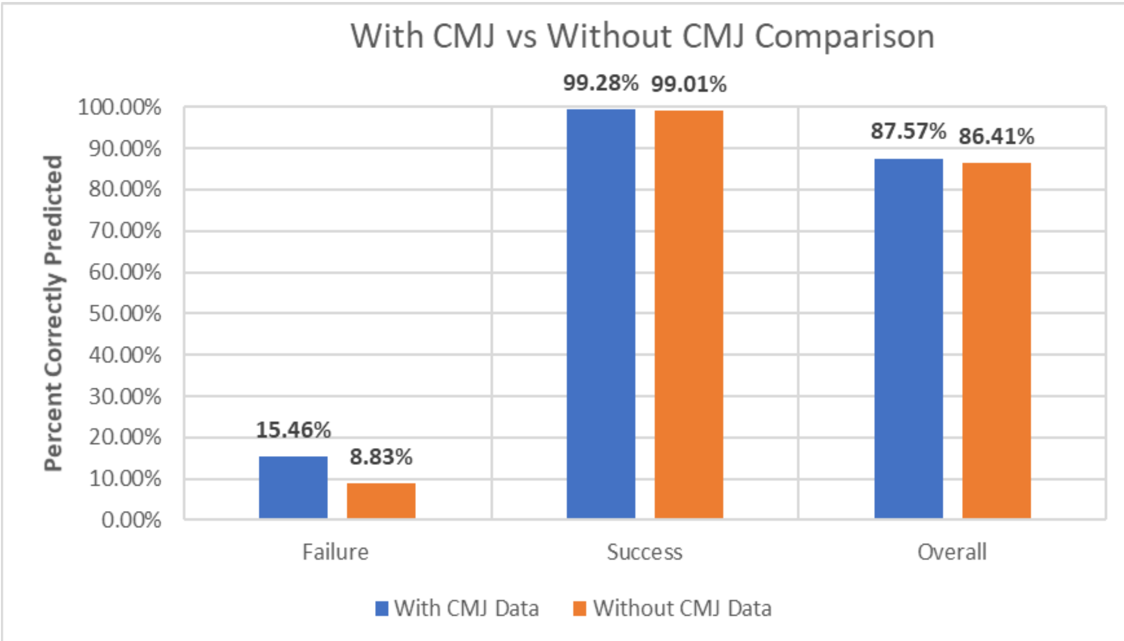
F. FINDINGS

1. Predictors of Success

Estimates from my logit model reveal that cognitive ability and CMJ performance are among the most significant predictors of ITB success. Table 1 provides a visual representation of the results. Consistent with academic literature that suggests cognitive ability is an essential factor, a candidate's AFQT score is practical and statistically significant in predicting a Marine's graduation at ITB. Likewise, the CMJ metrics provide real-time quantitative data to evaluate an athlete's execution of a skill or physical development. More important, the most interesting and statistically significant predictor in the model is the interaction of cognitive ability and CMJ. The interaction term indicates that for every one-unit increase in CRFD at 50 milliseconds, a candidate with an AFQT score of 50 or higher has 1.2 percent higher odds of graduating ITB than a candidate with an AFQT score lower than 50. These results suggest that cognitive and physical abilities complement each other and may be more prognostic of successful outcomes.

The overall results of the model suggest that CMJ metrics should be included as predictors of success at ITB. Figure 1 provides a visual representation of the predictive power with and without CMJ metrics. The model predicts approximately 1 percent better with CMJ metrics included overall. The model correctly predicted 99.28 percent of the candidates who graduate from ITB. An important caveat, however, is the model's ability to predict failure is nominal at best, only correctly predicting 15.46 percent of candidate failures from ITB. It is important to note that by including CMJ metrics, the predictive power of failure increases approximately seven percentage points. Thus, CMJ data has value. Continued CMJ data collection and analysis shows potential in prediction-based schoolhouse models.

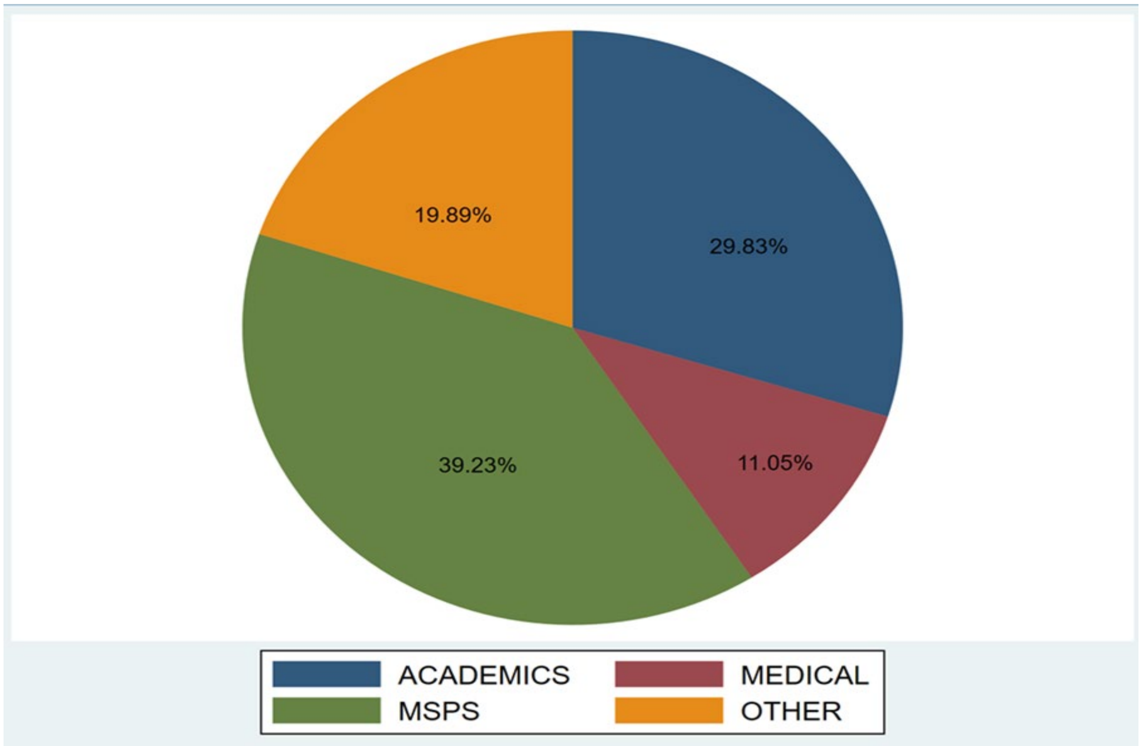
Figure 1. CMJ vs. without CMJ Comparison Model 1



2. Survival Analysis

The survival analysis provides three significant findings. First, MSPS events account for the majority of failures (35.79 percent) of all candidates that attrite, followed by academic failures (34.39 percent) and medical (23.35 percent). Put together with non-MSPS physical health reasons, physical wellbeing characterizes the key reason for failing to graduate from ITB at approximately 57 percent. Figure 2 provides a graphical representation of the attrition rates at ITB.

Figure 2. Distribution of Failure at ITB-E



Second, Figure 3 uses ITB-E data to illustrate a survival curve that suggests a significant decrease in survivability before divergence into specific MOS training. Overall, the results illustrate a significant drop in survivability around training day 10, which consists of academic and MSPS events. This makes sense given the significance of the interaction term for cognitive and physical abilities in predicting success, since the events during this time are both cognitively and physically challenging.

Last, Figure 4 presents the survival model results based on the same covariate model as the logistic regression. Much like the logistic model, cognitive ability, demographics, physical performance, and CMJ scores are the most significant contributors to survival at ITB. Again, the interaction term is the most statistically significant variable that reinforces the idea that commanders who use physical functioning measures to evaluate candidates should pay particular attention to their candidates' cognitive abilities and how these might complement each other.

Figure 3. ITB Survival Curve

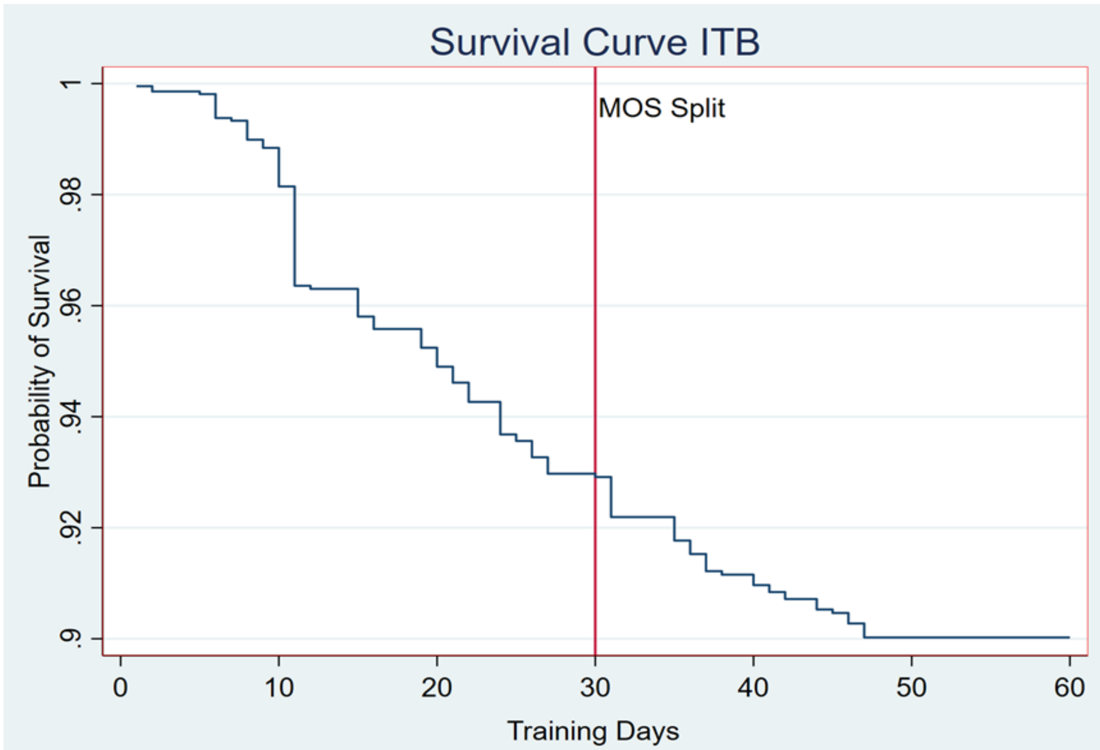


Figure 4. Survival Model Results

_t	Haz. Ratio	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
afqt_score	.9886164	.0041866	-2.70	0.007	.9804447	.9968562
height	.9054156	.03144	-2.86	0.004	.8458444	.9691822
weight	.9888592	.0053105	-2.09	0.037	.9785053	.9993227
pft_run_tm	1.000454	.000604	0.75	0.452	.9992712	1.001639
cft_manuf	1.010281	.0035274	2.93	0.003	1.003391	1.017218
cft_mtc	1.007364	.0031936	2.31	0.021	1.001124	1.013643
afqt_50_CRFD50	.9911085	.0029068	-3.05	0.002	.9854276	.9968223
StartofBrakingPhases	1.017287	.0054818	3.18	0.001	1.0066	1.028089
JumpHeightFTRelativeLandin	.9999591	.0000266	-1.54	0.125	.9999069	1.000011
rifle_qual_score	.9883747	.004307	-2.68	0.007	.979969	.9968525
ConcentricMeanForceN	.9991344	.0005064	-1.71	0.088	.9981425	1.000127
active_duty	1.576267	.4521159	1.59	0.113	.8984256	2.765523
riflehigh	.8633866	.5256408	-0.24	0.809	.261808	2.847264

G. OVERVIEW OF CHAPTERS

The research is organized into six chapters that provide background information and a detailed analysis of how I arrived at my empirical model of predictive factors of success and failure using CMJ metrics. Chapter II provides an in-depth discussion of the Marine Corps infantry's history, SOI's training mission, Marine Combat Training, and the training cycle for each infantry MOS at ITB. Chapter III analyzes prior academic research relevant to this study. Chapter IV outlines the empirical data used in this research and describes the methodology for analysis. Chapter V discusses the findings of the predictive model and survivability model of my quantitative analysis. Chapter VI concludes the research and provides recommendations for future studies.

II. BACKGROUND

This chapter provides background on the Marine Corps infantry community from its history to the transition at the School of Infantry (SOI). I provide an overview of the different military occupational specialties (MOS) that attend the Infantry Training Battalion (ITB), the associated training schedule, and the requirements related with each. The last section provides a brief overview of the Human Performance Center (HPC) located at the SOI and the critical role it plays in understanding and predicting human performance in the Marine Corps Infantry.

A. THE MARINE CORPS INFANTRY

"Every Marine is, first and foremost, a rifleman. All other conditions are secondary" (U.S. Marine Corps, 2012). This quote, spoken by General Alfred M. Gray, 29th Commandant of the Marine Corps, encapsulates the idea of the Marine Corps infantry. The Marine Corps is a direct descendant of the British Royal Marines, and was established by the Continental Marine Act of 1775 (U.S. Marine Corps, 2012). The Marine Corps infantry came to realization when "two battalions of Continental Marines were formed on 10 November 1775 in Philadelphia, Pennsylvania, as a service branch of infantry troops capable of fighting at sea" with the same skills as soldiers on land (U. S. Marine Corps, 2012).

The Marine Corps infantry is the heart and soul of the ground combat element. In its most simple form, the Marine Corps rifle squads' mission is to "locate, close with, and destroy the enemy by fire maneuver, or repel the enemy assault by fire and close combat" (U. S. Marine Corps, 2020b). Infantry Marines are trained on the full spectrum of warfare and must evolve with requirements, technology, and the changing environment.

Historically, success on the battlefield was measured through the attrition of the enemy. Essentially, the military would "throw more troops and material at a problem" as a tactic to win a battle. After the Vietnam War, the Marine Corps realized that traditional methods to win battles were no longer viable. In order to win battles, the Marine Corps had to change tactics. It adopted a new concept of warfare known as "maneuver warfare." This concept was published as doctrine in 1989 as Fleet Marine Force Manual-1 (FMFM-1). In this doctrine, maneuver warfare was a method of thinking that taught infantrymen to be effective in guerrilla-style insurgencies and conventional conflicts. Due to this complex approach, the change increased the demand for strategic thinking among infantry Marines.

After the Cold War, infantry tactics evolved even more with the update of FMFM-1 to Marine Corps Doctrinal Publication-1 (MCDP-1) Warfighting written by Marine Corps General Charles Krulak. The goal of MCDP-1 Warfighting is to provide a mental framework for how infantry leaders view conflict. It seeks to provide a method of thinking about how to make the enemy surrender. It is about understanding the adversary as a "human being." General Krulak emphasizes the importance of modern warfighting philosophy by stating,

> Very simply, this publication describes the philosophy which distinguishes the U.S. Marine Corps. The thoughts contained here are not merely guidance for action in combat but a way of thinking. This publication provides the authoritative basis for how we fight and how we prepare to fight. This book contains no specific techniques or procedures for conduct. Rather, it provides broad guidance in the form of concepts and values. It requires judgment in application. (Krulak, 1984)

Even today, the nature of war continues to evolve, and so must the infantry. Under the guidance of the 38th Commandant General Berger, the Marine Corps is reverting to its original roots and reassuming its role as the nation's naval-expeditionary force in readiness. In his planning guidance, General Berger outlines the Marine Corps objectives he wishes to obtain by 2030. His primary focus is to design the Marine Corps of the next 25 years as prescribed in the National Defense Strategy (NDS), National Military Strategy (NMS), Defense Planning Guidance (DPG), with a focus on innovative ideas and modernization of the Marine Corps. Specifically, he states, "We must transform our traditional models for organizing, training, and equipping the force to meet new desired ends, and do so in full partnership with the Navy" (U.S. Marine Corps, 2020). For the service to meet the Commandant's guidance, it is especially important to ensure the Marine Corps has a systematic process to determine the most critical metrics in selecting and developing infantrymen for the future fighting force.

B. THE SCHOOL OF INFANTRY

According to SOI-E historic records, before 1953, "there was no formal infantry training in the Marine Corps, and all Marines received combat training at recruit training" depots in San Diego, CA, or Parris Island, SC (SOI-E, n.d.). As noted on the SOI-E website, in 1953 infantry training regiments were established at Camp Geiger, NC, and Camp Pendleton, CA. These regiments became the first phase of initial military training for enlisted Marines after boot camp regardless of MOS. Since the initial training pipeline is split between coasts, Marines from geographical areas east of the Mississippi River attend infantry training at SOI-East, while Marines west of the Mississippi River attended infantry training at SOI-West (MCB Camp Lejeune, n.d.).

According to SOI-W, the SOI is the "second stage of initial military training for enlisted Marines after boot camp" (SOI-W, n.d.). Marines who receive the infantry MOS receive training at ITB. According to the SOI-W official website, "the SOI trains riflemen, infantrymen, and assault amphibian crewmen in MOS skills across the infantry training continuum" (SOI-W, n.d.). It also produces Marine combat instructors who train students to become leaders and carry out their MOS on the battlefield. All non-infantry Marines receive Marine common skills training at Marine Combat Training (MCT) Battalion. Figures 5 and 6 depict the command structures of SOI-West and SOI-East.

Figure 5. School of Infantry-West Command Structure. Source: Dove and Richmond (2017).

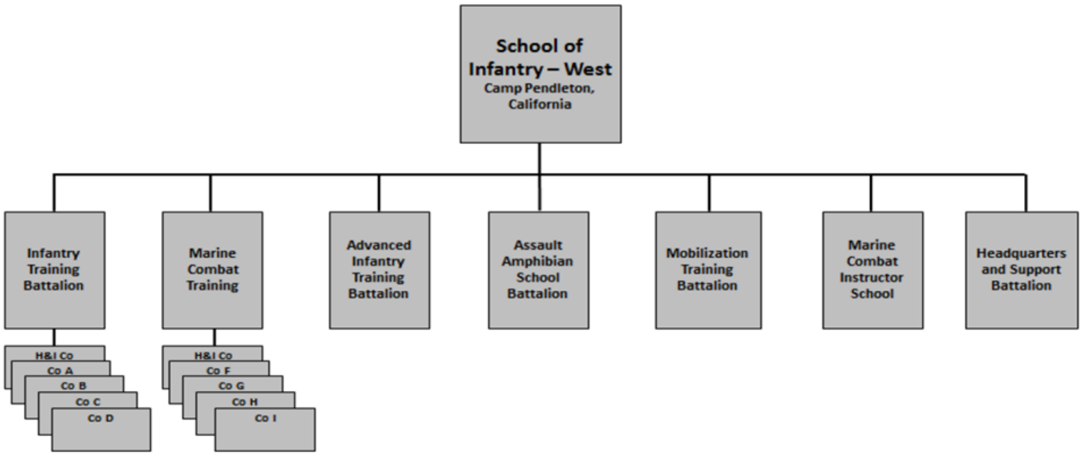
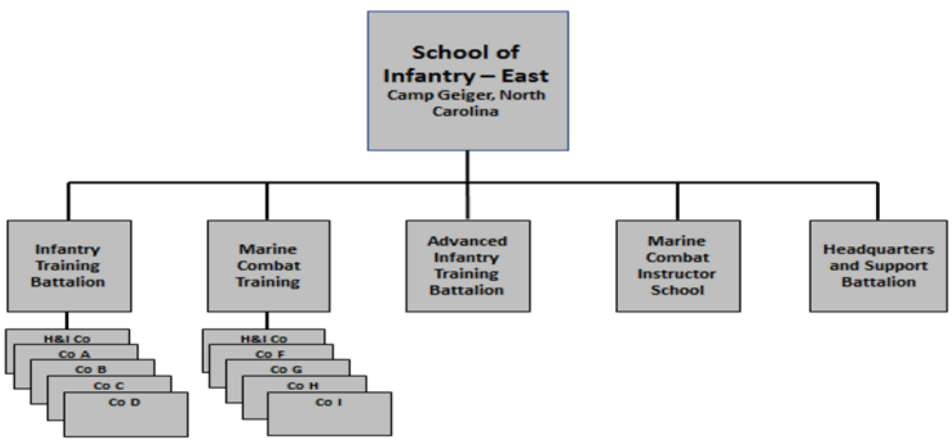


Figure 6. School of Infantry-East Command Structure. Source: Dove and Richmond (2017).



C. MARINE COMBAT TRAINING BATTALION

MCT Battalion trains all non-infantry Marines in the knowledge and skills of a basic rifleman. According to MCT's official website, MCT generates "Marine riflemen to possess a foundational understanding of their role in applying the Marine Corps' warfighting ethos, core values, basic tenets of maneuver warfare, leadership responsibilities, and mental, moral, and physical resiliency" (Marine Combat Training, n.d.). MCT Battalion consist of two lines of operations:

1. Entry-Level Training

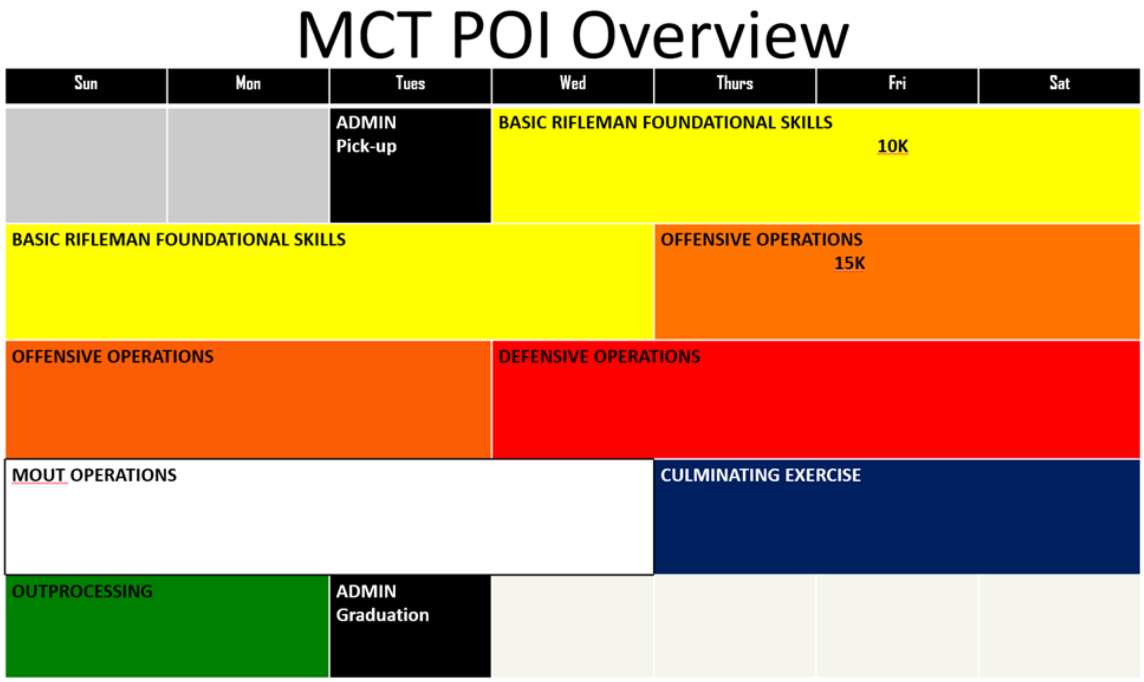
MCT battalion focuses on advancing the entry-level transformation during a 29-day training period where combat instructors teach, train, and mentor Marines in weaponeering, field skills, and basic warfighting fundamentals.

2. Combat Instructors

MCT battalion focuses on continuing the combat instructor's professional growth via the combat instructor developmental roadmap, career education, and PMOS advancement.

The MCT Plan of Instruction consist of 29 training days with 244 academic hours, 80 administrative hours, and multiple training events. There are generally 36 to 39 training cycles per year. An example training calendar is shown in Figure 7. After MCT, Marines who graduate should be confidently proficient in performing basic riflemen tasks and capable of enhancing their follow-on unit's combat readiness.

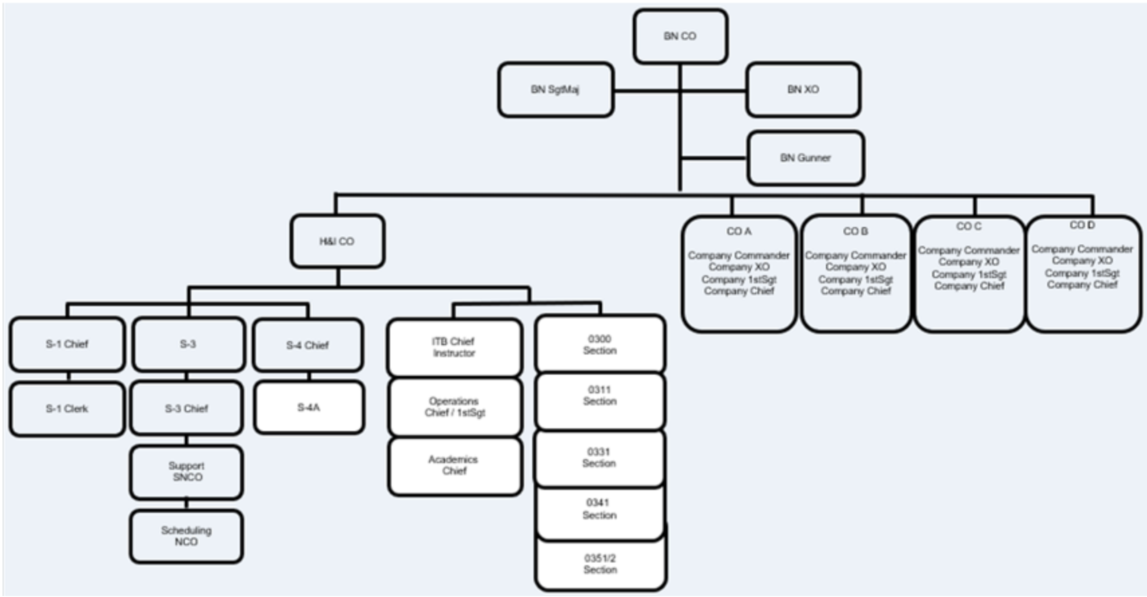
Figure 7. MCT POI Example. Source: Marine Combat Training (n.d.).



D. INFANTRY TRAINING BATTALION

According to ITB, their mission is to "train, develop, and certify Marines as riflemen, machine gunners, mortarmen, infantry assaultmen, and anti-tank missilemen to provide qualified infantrymen for service in the Fleet Marine Force" (FMF) (Infantry Training Battalion, n.d.). To ensure Marines are ready to serve, ITB trains per the Marine Corps Training and Readiness Manual. To graduate ITB, a Marine must successfully pass all required examinations and demonstrate a Marine infantrymen's requisite character. Figure 8 shows a sample task organization chart of the structure of ITB.

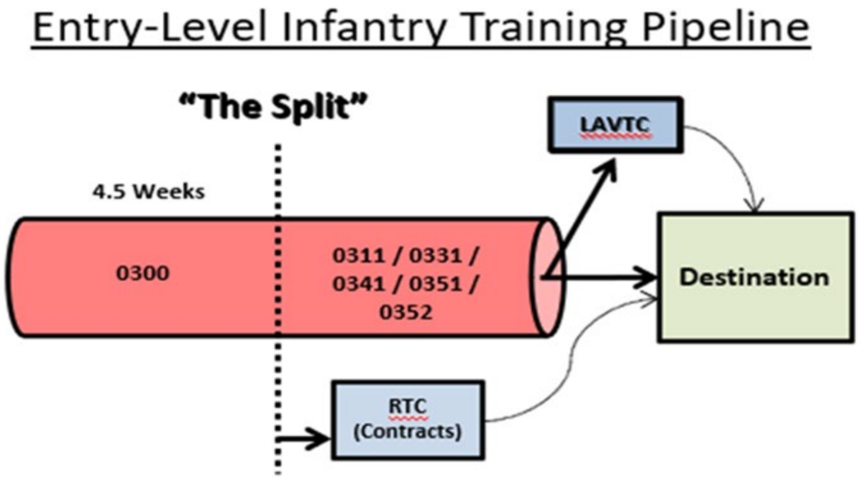
Figure 8. ITB Command Structure. Source: SOI-E (n.d.)



1. ITB Training Cycle

The ITB training cycle consists of approximately 16 cycles per fiscal year, typically spread among four training companies. Each training cycle is approximately nine weeks in duration (Infantry Training Battalion, 2019). As shown in Figure 9, all infantry Marines train together for four and a half weeks for the ITB training curriculum's 0300 common skills portion. At the completion of the initial 0300 training, each Marine branches off into their specific MOS training pipeline. This consists of 0331 machinegunner, 0341 mortarman, 0351 assaultman, and 0352 missileman. The MOS specific training consists of 22 training days for all of the beforementioned MOSs.

Figure 9. ITB Training Pipeline Split. Adapted from SOI-E (n.d.)



2. 0300 Plan of Instruction

Acceptance into the 0300 MOS depends on the infantry candidates' Physical Fitness Test (PFT) and Combat Fitness Test (CFT) scores at recruit training. The minimum standards are based on event performance and a gender-neutral evaluation to authenticate MOS grouping into the 0300 MOS (Department of the Navy, 2020). The minimums are shown below in Table 1:

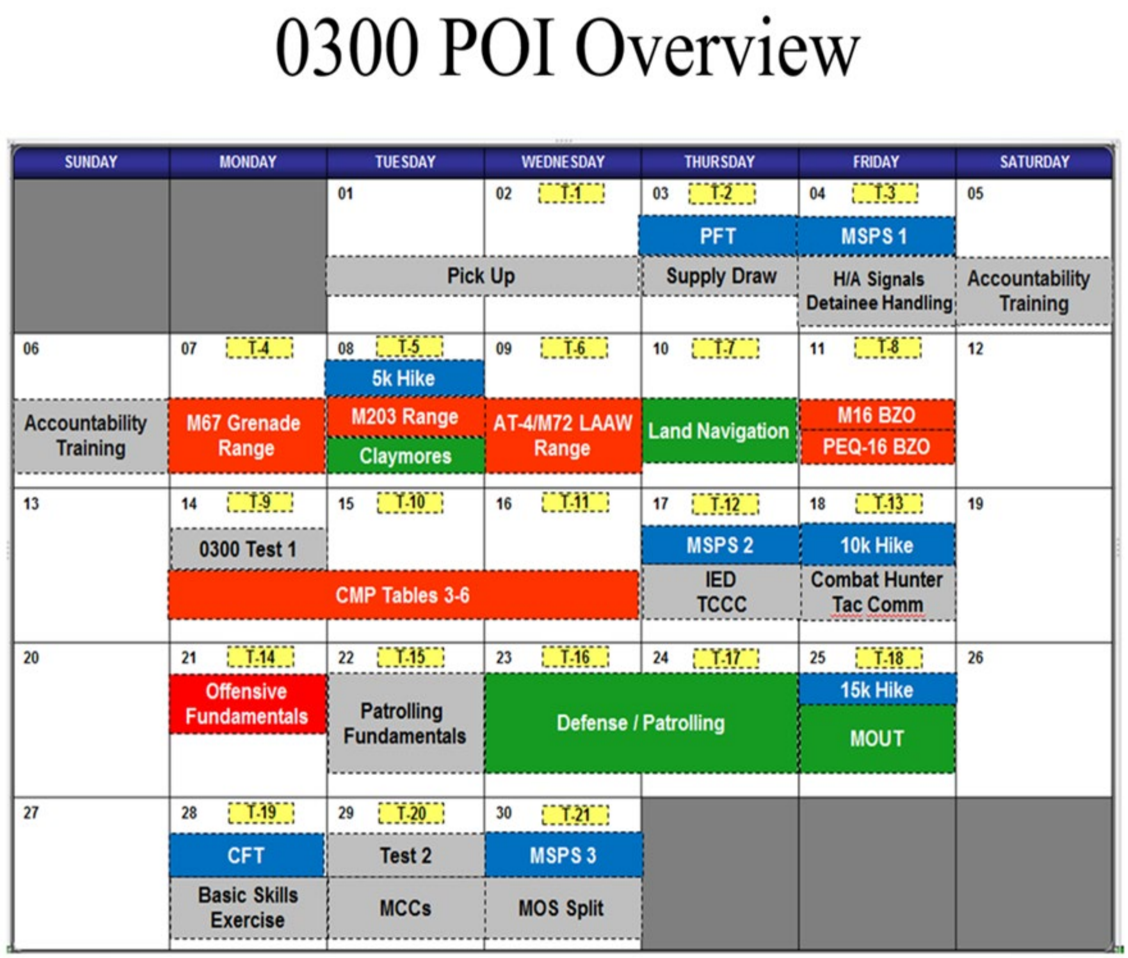
Table 1. ITB Minimum Physical Requirements. Source: Department of the Navy (2020)

Event	**Minimum**	**Unit of Measurement**
Pull-Ups (PFT)	6	Repetitions
3 Mile Run (PFT)	24:51	Min:Sec
Manuever Under Fire (CFT)	3:12	Min:Sec
Movement-to-Contact (CFT)	3:26	Min:Sec
Ammo Can Lifts	60	Repetitions

All candidates who desire to become infantryman are also subject to aptitude requirements. At a minimum, candidates must obtain a 90 or above on the General Technical (GT) score of the Armed Services Vocational Aptitude test (ASVAB).

Additionally, all candidates are subject to an Initial Physical Assessment (IPA) and Countermovement Jump (CMJ) test within the first week of training. Figure 10 is a depiction of the 0300 plan of instruction:

Figure 10. 0300 POI Overview Sample. Source: Infantry Training Battalion (n.d.)

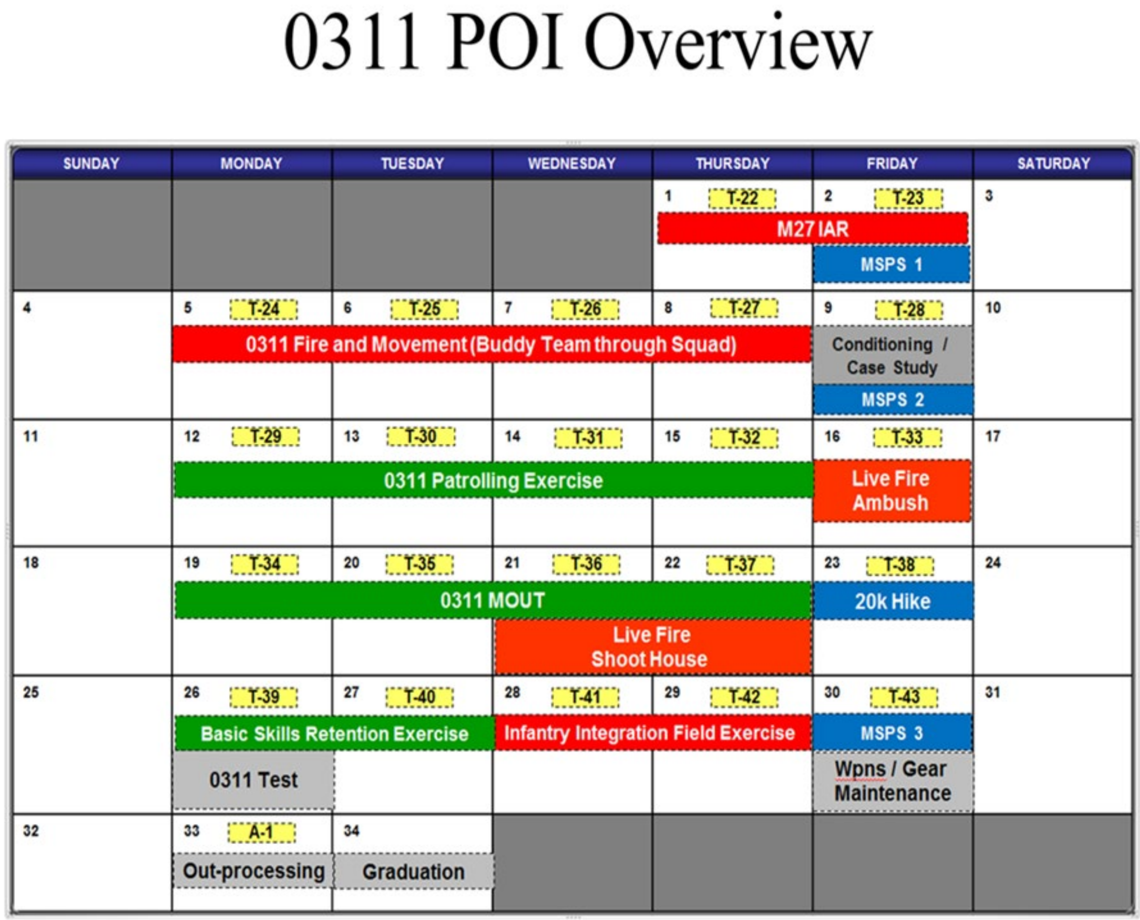


3. 0311 Plan of Instruction

The majority of Marines attending ITB will attempt to become a rifleman. According to the T&R manual, Marines with an 0311 riflemen MOS "employ the M16M4/ A4 service rifle, the M203 grenade launcher, and the M27 Infantry Automatic Rifle (IAR)"

(Department of the Navy, 2020). The 0311 MOS is the foundation of the Marine infantry organization. Figure 11 is a depiction of the 0311 plan of instruction:

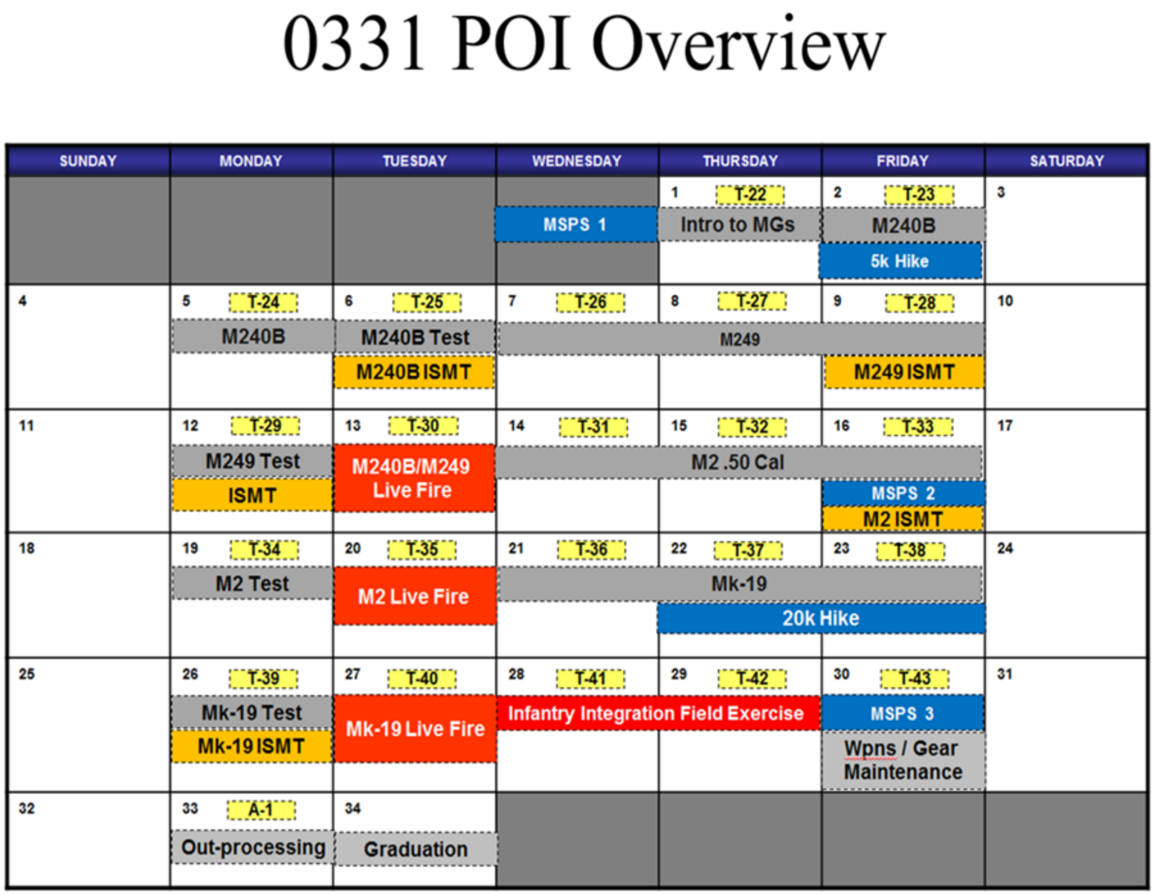
Figure 11. 0311 POI Overview Sample. Source: Infantry Training Battalion (n.d.).



4. 0331 Plan of Instruction

All Marines who attend the 0331 machinegunner course will be "responsible for the tactical employment of the 7.62 mm medium machine-gun, the 50-caliber machine gun, 40mm heavy machine-gun, and their support vehicle" (Department of the Navy, 2020). According to the training and readiness (T&R) manual, "machine gunners provide direct fire in support of infantry rifle squads/platoons/companies" (Department of the Navy, 2020). Figure 12 is a depiction of the 0331 plan of instruction:

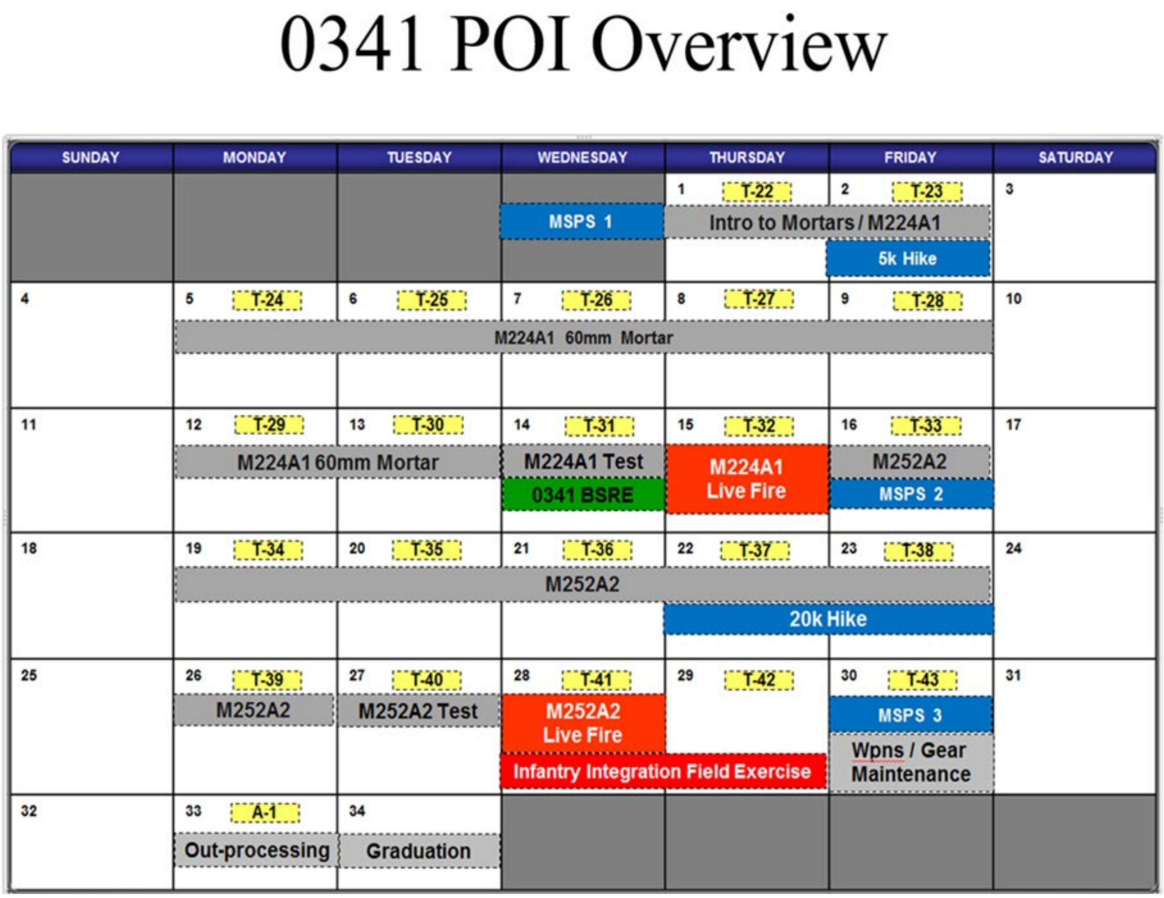
Figure 12. 0331 POI Overview Sample. Source: Infantry Training Battalion (n.d.).



5. 0341 Plan of Instruction

According to the T&R manual, candidates who desire to become mortarmen are "responsible for the tactical employment of the M224 (60mm) light mortar and M252 (81mm) medium mortar. Mortarmen provide indirect fire supporting the infantry and Light Armored Reconnaissance (LAR) units" (Department of the Navy, 2020). Mortarman are generally placed with the weapons platoon, 81mm mortar platoons, and LAR companies (Department of the Navy, 2020). Figure 13 is a depiction of the 0341 plan of instruction:

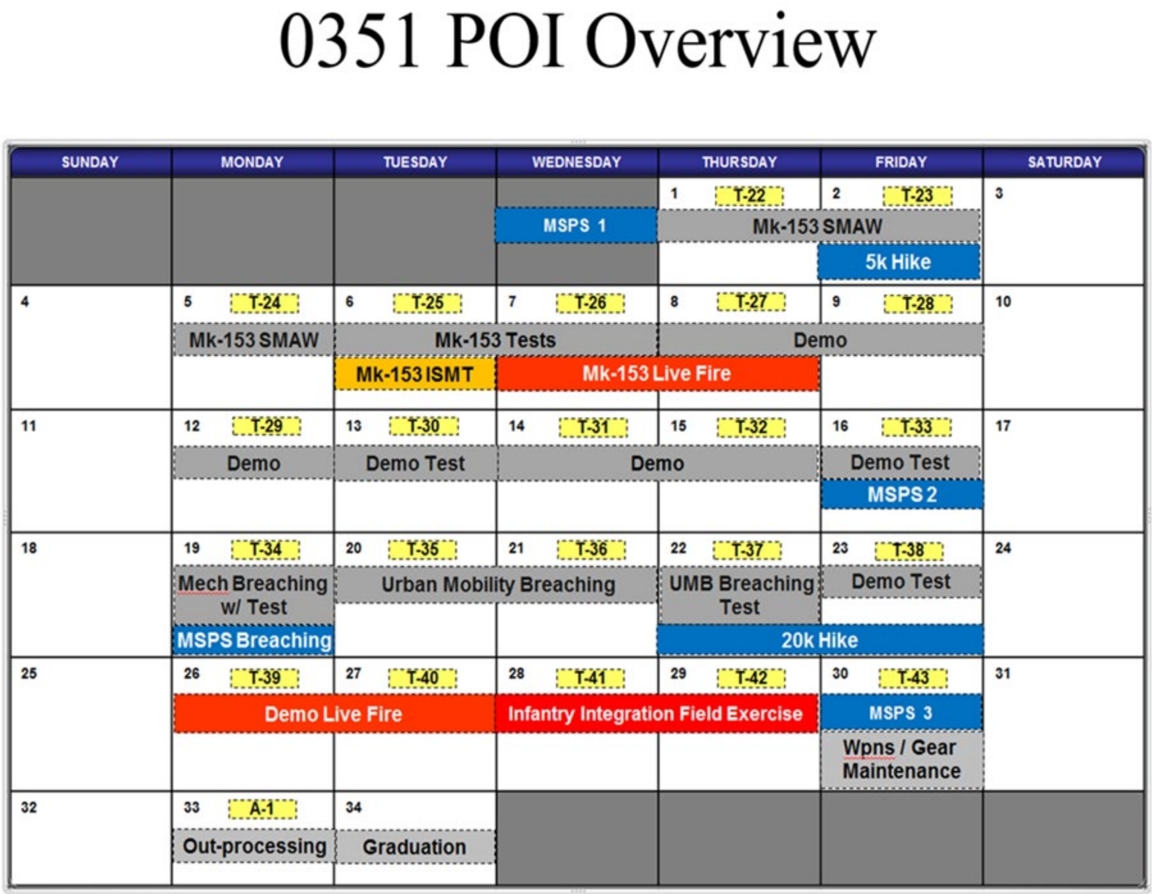
Figure 13. 0341 POI Overview Sample. Source: Infantry Training Battalion (n.d.)



6. 0351 Plan of Instruction

Candidates who desire to become an infantry assault Marine "employ rockets, the Anti-Personnel Obstacle Breaching System (APOBS), and demolitions" (Department of the Navy, 2020). According to the T&R manual, "assault Marines provide rocket fire against fortified positions supporting the rifle squads, platoons, and companies within an infantry battalion" (Department of the Navy, 2020). In addition to standard prerequisites for other infantry MOSs, infantry assault Marines are required to have a GT score of 100 or higher, normal color vision that is correctable to 20/20 (Department of the Navy, 2020). Figure 14 is a depiction of the 0351 plan of instruction:

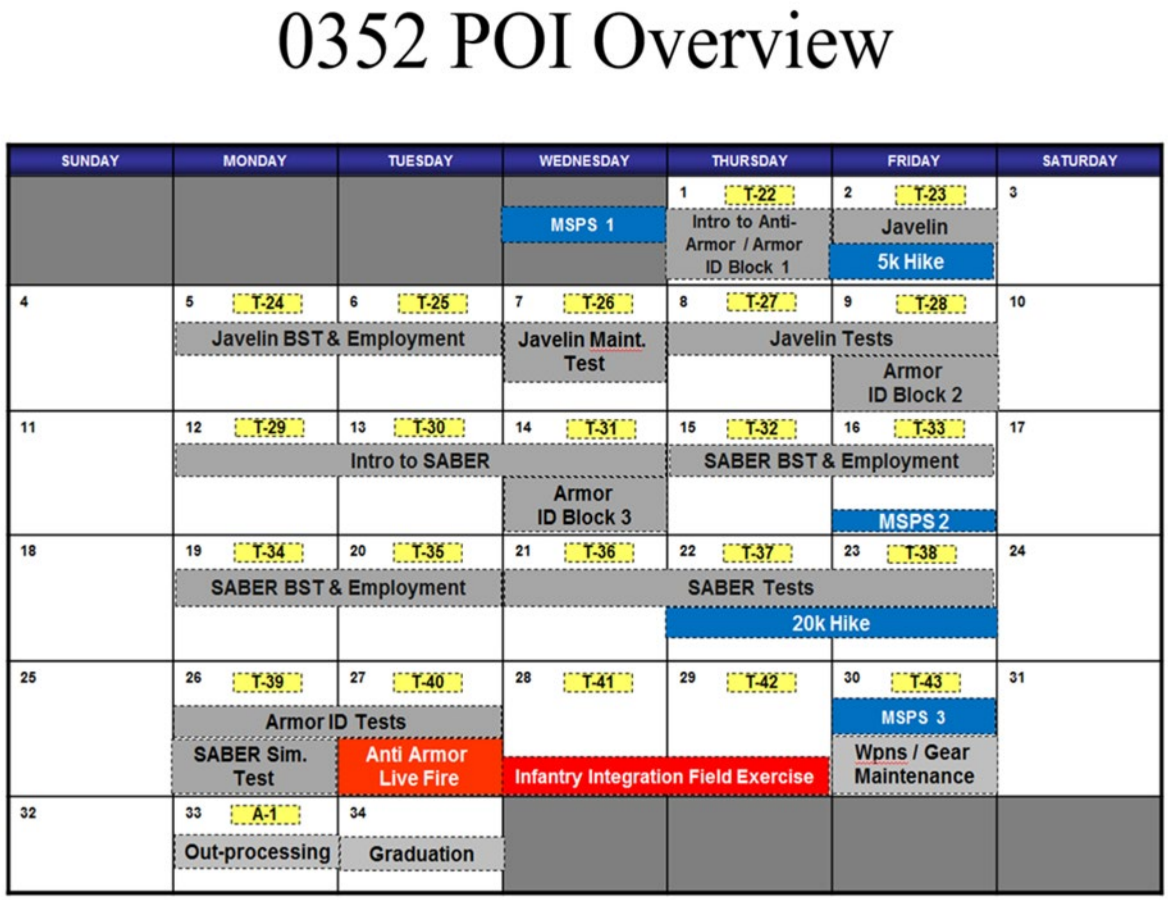
Figure 14. 0351 POI Overview Sample. Source: Infantry Training Battalion (n.d.)



7. 0352 Plan of Instruction

The antitank missile gunner is accountable for the "tactical employment of the M220E4 Tube-launched, Optically tracked, Wire-guided (TOW) weapon system, M98A1 javelin weapons system, anti-armor operations, and tactical vehicle operations" (Department of the Navy, 2020). According to the T&R manual, they are "located in the anti-armor platoon within the weapons company of infantry battalions and LAR battalions" Department of the Navy, 2020). In addition to standard prerequisites for the other infantry MOSs, antitank missile gunners must have a GT score of 100 or higher, have normal color vision, and no vehicular infractions (Department of the Navy, 2020). Figure 15 is a depiction of the 0352 plan of instruction:

Figure 15. 0352 POI Overview Sample. Source: Infantry Training Battalion (n.d.).



E. HUMAN PERFORMANCE CENTER

The Human Performance Center (HPC) located at Camp Lejeune, NC, was established in January 2017 in association with the Office of Naval Research, West Virginia University, and the Air Force Research Laboratory. The HPC's mission is to "reduce attrition and lost work-days associated with musculoskeletal injuries (MSKI) to increase individual Marines' operational readiness" (Infantry Training Battalion, 2019). The trainers at the HPC conduct a multi-disciplinary approach to training and treating Marines through innovative techniques, ground-breaking technology, and scientifically proven research. The HPC is integrated with Marine Corps Force Fitness Instructors (FFI), combat instructors, and civilian specialists. By incorporating technology, gathering subjective feedback, and information gained through evaluations, these specialists have developed protocols and programs to meet the needs of all Marines and Sailors attending the SOI.

The experts that work at the HPC perform two functions: (1) athletic training to evaluate, treat, rehabilitate, and document all Marines and Sailors upon arrival and (2) strength and conditioning to optimize human performance through education, training, and programming to Marines and Sailors located at Camp Lejeune, NC. In order to execute their duties, the Table 2 describes some of the innovative technologies currently being used at the HPC.

Table 2. HPC technology. Source: SOI-E (n.d.)

Technology	**Application**
Force Platforms	Force Platforms measure the ground reaction forces generated by a figure standing or movement impulse. It helps figure out an athlete's neuromuscular capacity and uses the results see where they are in terms of their training, and gauge how competition, injury and training affect biomechanics parameters (Lake, n.d.).
Sleep Monitoring Devices	Sleep monitoring devices help "explore causal factors of insufficient sleep and inventory known effects of sleep restriction on human performance" (Shattuck et al., 2020). These devices help identify concerns with exhaustion and sleep deficiency in military operational environments (Shattuck et al., 2020).
Heart Rate Monitors	Heart Rate Monitors help the specialist determine a Marine's physiological response to training changes and promote better functioning and a decrease in injuries. It is also a useful indicator of physiological variation and amount of exertion.
Biomodule Feedback Systems	The Biomodule Feedback Systems capture and turn raw data into usable intelligence in graphs and reports that allows specialist to interpret quickly. This allows specialist to make educated decisions in response to changing health factors and conditions.
Movement Analysis Software	Motion Analysis Software features allow specialists to capture and track Marines, products, or an entire team's subtle movements. This aids in reporting, coaching, and practice sessions.

THIS PAGE INTENTIONALLY LEFT BLANK

III. SCIENTIFIC LITERATURE REVIEW

This chapter provides a review of relevant scientific research conducted in the past. The first section reviews research on the attributes used to create a predictive model for completing the Basic Reconnaissance Course (BRC) and ITB. While research on Marine Corps high-risk training is starting to gain traction in academia, students at the Naval Postgraduate School have conducted studies on which attributes are most predictive of success in physically and mentally rigorous training. The second section looks at CMJ research and its ability to help optimize and predict human performance among elite athletes and servicemembers. The chapter concludes with a summary of previous research and links how my study will expand the Marine Corps knowledge in this field.

A. INSTITUTIONAL RESEARCH

1. Marine Corps Basic Reconnaissance Course Research

In his Naval Postgraduate School thesis, Albert Nowicki analyzed how the current requirements to attend the BRC predicted the probability of success and a BRC candidate's survivability. Nowicki (2017) believed the need for reconnaissance forces had been documented throughout history. The high attrition rate associated with the BRC warranted investigation of what metrics contributed to success. His research uses logistic regression models and survival analysis to determine the extent to which the current requirements to attend the BRC are indicators of success. Data in Nowicki (2017) comes from Total Force Data Warehouse (TFDW) and the BRC. His data includes multiple cohorts, which accounted for approximately 1,577 candidates who attended the BRC from Fiscal Year 2013 to Fiscal Year 2016.

Nowicki (2017) begins his analysis with three different logistic regression models. In the first logistic regression, Nowicki analyzes which independent variables are statistically significant in predicting success at BRC. In his regression, he included PFT score, rifle qualification score, GT score, proficiency and conduct markings, time in grade, age, marital status, combat deployments, and the number of previous attempts at completing BRC. In his analysis, his dependent variable was the binary variable GRAD. I believe logistic regression was appropriate because of the dependent variable's binary nature reflecting graduation from the BRC (GRAD=1) or attrition (GRAD=0).

Nowicki's (2017) model's results suggest that the PFT score, GT Score, and completing at least one semester of college were statistically significant predictors of success at BRC. Specifically, Nowicki (2017) found that both a 1 percent increase in PFT score or GT score accounted for a 1 percent increase in the probability of graduating from BRC. He also found that having completed at least one semester of college was associated with a 20.5 percent increase in the probability of graduating BRC (Nowicki, 2017).

In the second logistic model, Nowicki (2017) estimates the probability of graduating BRC applying the course's current prerequisites for attendance. He utilizes the same variables shown in the previous model, but creates categorical variables for both PFT and GT score. The relevant prerequisites to attend the course are as follows: must have a minimum GT score of 105 and a minimum PFT score of 225. His results suggest that having a PFT score higher than 225 is statistically significant in predicting success at the BRC. GT score is also statistically significant, but the coefficient is negative. The negative coefficient is attributed to over 95 percent of candidates in his dataset possess a GT score higher than 105, with a graduation rate of approximately 57 percent.

For the last logistic regression, Nowicki estimated the probability of graduating BRC with increased prerequisites by examining the marginal effects. His results suggest that candidates who attain a PFT Score of 275 or higher have a 28 percent higher likelihood of graduating the BRC than a lower score candidate. The GT score was also statistically significant. Nowicki found that candidates with a GT score of 115 increase their probability of graduating from the BRC by 17.8 percent, relative to candidates with a lower score.

In addition to his logistic regression models, Nowicki (2017) constructed a predictive model of survival at BRC intending to identify which events were statistically significant in the survival at the BRC. He used "training days" as his duration variable and "failure" as his failure variable. I believe this was appropriate for establishing a fitted model to display the probability of survival based on the independent variables used in the model. As of 2017, there were 14 different drop categories at the BRC. Of these 14 categories, attrition, land navigation, swim qualification, patrolling, and medical injuries are the most significant contributing factors to attrition.

Nowicki's analysis determined that training day 10 (land navigation), training day 15 (swim qualification), and training day 53 through 55 (patrolling) are the days that candidates are most likely to attrite. Training day 10 displayed the steepest drop in survivability at BRC. This made sense because candidates attend the rigorous land navigation course during this timeframe, which accounts for 14 percent of failures at BRC. He also determined that a PFT score greater than 225 and a GT score greater than 105 significantly increased the probability of survival among BRC candidates. Lastly, Nowicki (2017) found that as previous attempts increased, the probability of survival increased by approximately 22 percent. This makes sense because, with each attempt, students developed a better understanding of the course's rigors and how to overcome them.

2. Infantry Training Battalion Research

Dove and Richmond (2017) analyzed a predictive model for success under female integration. Their study focused on integrating females into combat arms military occupational specialties due to the Department of Defense lifting the restrictions during 2016. Their goal was to use quantitative and qualitative analysis to determine Marines' success and failure at ITB West and East. They utilized both logit and multinomial logit regression models using data from TFDW and ITB West. Their data includes multiple cohorts, which accounted for approximately 41,153 Marines who attended ITB West from Fiscal Year 2010 to Fiscal Year 2017.

Dove and Richmond (2017) begin their study with running a Logistic Model to determine what variables contribute to the success of graduation from ITB with the outcome variable of GRAD where graduation from ITB is (GRAD = 1) while attrition is (GRAD = 0). This is similar to the thesis completed by Nowicki (2017); however, their independent variables were different. These included the Armed Forces Qualification Test (AFQT) score, PFT, CFT, Height, Weight, and rifle qualification score.

Their analysis found that AFQT, PFT, CFT, and rifle qualification scores were all statistically significant and contributed to successfully graduating ITB. This makes sense because ITB is both physically and mentally challenging, so physical fitness and aptitude are significant contributors to success.

For their second model, Dove and Richmond (2017) chose a multinomial logistic model to examine the reasons for failure from ITB. A multinomial logistic model is appropriate to use when there is more than one type of dependent variable deemed necessary to answer a specific question. It is important to note that each of these dependent variables are nominal and is not likely to violate the assumption of independence.

Dove and Richmond (2017) concluded that MOS-Specific Physical Standards (MSPS) assessments represented the majority of the failures at 58 percent, followed by physical health at 20 percent. This interprets to roughly 80 percent of failures from ITB are physical performance or health-related. These events were statistically associated with movement-to-contact times and maneuver under fire time during the CFT and the number of pull-ups during the PFT. Their results contradict Larkin (2017), who studied the Marine Corps Advance Mortarman Course. Larkin (2017) found that "GT scores, proficiency and conduct marks, and experience as a Marine are significant determinants of success, while physical fitness is not."

B. COUNTERMOVEMENT JUMP RESEARCH

Over the last decade, force measurement technology has increased in use among elite institutions. CMJ tests are useful for examining the kinetic characteristics of an athlete's movement. The CMJ is a practical and reliable test for detecting both athletic potential and identifying areas of weakness among athletes. It provides analysts real-time quantitative data to evaluate an athlete's execution of a skill or physical development. The CMJ contains three key phases: unweighting, braking, and propulsive. A picture of the phases can be seen in Figure 16. By recording the CMJ through a force platform, an HPC analyst at SOI can determine how fast, how much force is generated, and in what direction a Marine or Sailor moves. Having this metric available may aid in leadership at the SOI in determining a Marine's neuromuscular capacity and identify where they are in terms of their training, injury recovery, and the probability of completing a training curriculum.

Figure 16. CMJ Overview. Source: Hawkins Dynamics (2019).

1. Effects of Military Training on Explosive Performance

Welsh et al. (2008) assessed the effect of intensified military field training on jump performance. The purpose of their study was to determine the ability of an unloaded jumping test in detecting decrements in physical performance using CMJ force plate metrics. The authors chose the CMJ due to its reliability as a field-expedient test that is sensitive and practical for evaluating training interventions. The study consisted of twenty-nine infantry U.S. Marines who performed 1, 5, and 30 repetitions of unloaded CMJs before and after eight days of sustained field operations to determine if there were significant detriments in human performance. The authors chose a multiple jump test to measure maximal power and fatigue within each series of tests. Marines were exposed to stressors such as continuous activity, sleep deprivation, cognitive stress, and malnutrition during the field operations.

To determine the CMJ test's utility on monitoring physical performance, the authors focused on the CMJ variables jump height, jump power, fatigue index score, and body mass index. To collect data, the authors used both a linear product transducer (LPT) and kinetic measurement system (KMS) switch mat simultaneously to cross-reference results and ensure validity. Once the data was received from the CMJ, the authors conducted a repeated-measures analysis of variance (ANOVA) to compare differences in jump power and jump height between each respective unloaded CMJ test series. Once the authors identified a significant comparison, they conducted an ANOVA Turkey's Honestly Significance Difference test to determine where the significance occurred.

The authors found that mean body mass (-4.1± 1.6 percent), fat mass (-12.7± 7.6 percent), and fat-free mass (-2.4± 1.2 percent) all declined significantly following field operations. Jump performance was also significantly affected. The mean jump power measured on both the LPT and KMS models declined on all three post-field operation repetition jumps and were statistically significant at the 95 percent level. The decrease in mean jump height was statistically significant after post-field operations on the five repetition jumps for both the LPT and KMS models; however, the one repetition jump was also statistically significant on the KMS model.

The results suggest that the CMJ test is sufficiently sensitive to detect human performance changes in a military training environment. The findings of this study support previous literature reporting that with more physical activity, human performance declines. However, this is the first study to my knowledge that utilizes CMJ metrics to measure performance in a field training military environment, which can prove beneficial due to its reliability and minimal cost.

2. CMJ Test Predicting Anaerobic Performance

The CMJ test has also been used as a human performance predictive metric. Markström and Olsson (2013) conducted research to predict sprint running performances among elite athletes. For the analysis, their goal was to determine if variables from the "1-leg drop jump (DJ), squat jump (SJ), and CMJ tests could predict sprint performances among elite athletes" (Markström and Olsson, 2013). The authors chose to use force platforms to measure these variables due to their high reliability and no need to familiarize equipment. The subjects consisted of elite sprinters, jumpers, and throwers at the collegiate track and field skill-level. Additionally, they wanted to see if sprinters and jumpers could be differentiated based on variables from 1-leg DJ, SJ, and CMJ tests (Markström and Olsson, 2013).

The research consisted of two cross-sectional sub-examinations. For the first sub-examination, subjects included five elite sprinters (one female). The second sub-examination consisted of five sprinters (one female) versus five jumpers and six sprinters versus six throwers (four females). According to Markström and Olsson (2013),

- "sprinters performed the 1-leg DJ, SJ, and CMJ tests" in a laboratory in the first sub-examination.
- On the same day, the "sprinters performed sprint performances on an indoor track."
- They sprinted for a time as follows: 0 to 25m, 10m, and 60m. The authors examined "differences in physical characteristics between the track and field athletes" for the second sub-examination group.
- The sprinters and jumpers performed the "1-leg DJ, SJ, and CMJ tests" to analyze group differences.
- Three weeks later, sprinters and throwers executed the SLJ, SJ, and CMJ tests to analyze group differences.
- The second sub-examination group executed the same sprint routine as the first group.

The authors used a single linear regression and multiple linear regression analysis approaches with models, including the beforementioned performance variables for predicting sprint performances. According to Markström and Olsson (2013), the "relationships between jump heights and sprint performances were examined" in both models—all measurements received from the force platform were relative to the subject's body weight. The authors used a Multivariate Analysis of Variance (MANOVA) to examine each group of sub-examinations' different characteristics. This method is commonly used in the health and science field due to its ease of use by allowing tests for the difference in means between two or more groups.

The authors determined that the CMJ peak force (PF) in relation to body mass was statistically significant in predicting sprint execution maximal running speed through the 10-meter and 60-meter sprint time (Markström and Olsson, 2013). Specifically, an "improvement of CMJ PF by 1-N/kg decreases 10m/60m time by .026 seconds and .158 seconds," respectively (Markström and Olsson, 2013). The maximum running velocity at the 10-meter sprint was also highly predicted by the CMJ test. According to the authors, 1cm improvement of jump height decreases 10m sprint time by .015 seconds. However, jump heights from SJ and DJ were not effective at predicting sprint execution. According to Markström and Olsson (2013), the results suggest that the "use of CMJ PF seems to be an effective predictor of sprint speed execution."

C. SUMMARY

Predicting athletic performance is a complex task to measure due to a multitude of variables that can contribute to the results. The BRC literature review suggests that the GT score, PFT score, and the number of attempts are significant contributors to success in a military training environment. However, students at ITB are typically not allowed to make multiple attempts to complete the training. If a student fails multiple times, they are normally recycled to a non-combat MOS or separated from the Marine Corps. I also believe it may have been beneficial to use a Least Absolute Shrinkage and Selection Operator (LASSO) to determine if any unidentified independent variables may have been useful in their analysis.

The ITB gender integration study relates closely to my research and has many of the same variables. The authors found that AFQT score, PFT score, CFT score, and rifle qualification score were all statistically significant contributors to success at ITB. However, their data is exclusively from ITB-West, while my data is exclusively from ITB-East, and the change of the ITB curriculum and grading requirements during 2019 may play a factor in what variables are appropriate in measuring success. Lastly, it would have been interesting if the authors conducted a survival analysis, so I could compare their results from the previous curriculum at ITB-West to the current curriculum at ITB-East.

Both BRC and ITB integration studies are similar in their mission and training in a combat-like environment. Likewise, this intrinsic relationship led to some of the same contributors to success in both courses. Based on their results, it appears that cognitive and physical performance are significant contributors to success in combat-arms type training. However, their results directly contradict Larkin (2017) in his analysis of the Advance Mortarman's Course, which found that experience and cognitive ability are the most significant contributors to training success, while physical ability was irrelevant. None of these studies used CMJ force platform data to determine if human kinetics or biomechanics also play a factor. My study will help resolve the contradiction between the literature and bridge the gap between cognitive ability and physical performance.

The CMJ research suggests that force platforms may be a useful quantitative evaluation tool to measure explosive speed and performance among athletes. Welsh (2008) found that CMJ metrics are adequately sensitive to distinguish human performance changes during military training. Their research also supports the theory that with more stressors, human performance declines. Markström and Olsson (2013) determined that the CMJ test is an effective tool for predicting sprint speed among elite athletes. Additionally, they found that the CMJ test was more reliable than the DJ and SJ, which provides insight to why CMJ data is appropriate for my research.

I use collected data from ITB and TFDW to expound upon the probability of success at ITB and determine what variables help predict Marines' survivability during the course. With the CMJ test data, determine what variables are statistically significant, meaningful, and practical. This data suggests that the CMJ test may be a useful quantitative evaluation tool to measure explosive speed and performance among athletes.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. DATA AND METHODOLOGY

This thesis brings together several data sources to characterize each ITB student and identify predictive factors for his or her success at ITB. This chapter discusses each data source by identifying the construction of the population of ITB candidates, and the cleaning, merging, and coding of data sources to allow statistical analysis. Lastly, I will discuss the quantitative approach and statistical models I use for this study.

A. DATA SOURCES

1. Total Force Data Warehouse (TFDW)

The data for my research comes from two data repositories. The first, TFDW, collects, consolidates, and stores personnel data on all Marines from accession to separation for the Marine Corps. Data from TFDW comes from multiple sources; however, for my research, I use the Marine Corps Total Force System (MCTFS), the Marine Corps Training Information Management System (MCTIMS), and the Marine Corps Recruiting Information Support System (MCRISS). Data at TFDW provides a monthly snapshot of every Marine on the last day of each month. The monthly snapshot in this study is from the month prior to a Marine beginning ITB.

The individuals included in my study are candidates who attend ITB from January 2018 to December 2019. My analysis sample from TFDW excludes 258 observations out of 1,552 for two reasons. I drop 21 observations because of inconsistent graduation data in TFDW. Second, I exclude 237 observations with no reported EDIPIs at ITB, as I could then not match these Marines with event scores. The variables that TFDW provides are listed in Table 3.

Table 3. Definitions of Variables, TFDW Data

Variable Name	Variable Description
EDIPI	Electronic Data Interchange Personal Identifier
GRADE	Current grade
AFQT_SCORE	Armed Forces Qualification Test
GCT_GT_SCORE	GT score for Enlisted, GCT Score for a Commissioned officer
AGE	As reported by record
RACE	As reported by the candidate
MOS	Military Occupational Specialty
SCHOOL COMPLETE DATE	Date of completion from MOS school
SCHOOL START DATE	Date of start for MOS school
SCHOOL_STATUS	Identifies whether or not the candidate completed school
SCHOOL_NAME	Name of school attended
HEIGHT	As reported by record
WEIGHT	As reported by record
PFT_SCORE	Physical Fitness Test total score
PFT_CLASS	PFT Class (1st, 2nd, or 3rd)
PFT_PULL_UP	PFT Pull-ups score (repetitions)
PFT_CRUNCH	PFT Crunches score (repetitions)
PFT_RUN_TM	PFT 3-mile run time (Minutes)
CFT_SCORE	Combat Fitness Test Total Score
CFT_CLASS	CFT Class (1st, 2nd, or 3rd)
CFT_MTC	CFT Movement to Contact time (Minutes)
CFT_AMMO_LIFT	CFT ammo can lift score (repetitions)
CFT_MANUF	CFT maneuver under fire time (minutes)
EDUCATION	Highest level complete
NUMBER_OF_DEPENDENTS	Number of dependents as reported by record
COMPONENT_CODE	Active duty or Reservist code
RIFLE_QUAL_SCORE	Total annual rifle qualification score (points)
RIFLE_QUAL_CLASS	Rifle qualification class (Expert, Sharpshooter, Marksman)
GENDER	Maler or Female
WAIVER_CATEGORY_DESC	Waiver Category if received
WAIVER_TYPE_DESC	Waiver type if received
HOME_OF_RECORD	Home of Record at time of enlistment

2. Infantry Training Battalion-East

The staff and Human Performance Center at ITB provide the second source of data for this study. Data from the ITB staff includes training rosters of all Marines on Training Day 1 and end of course summaries on each cohort's graduation statistics. These documents allow me to analyze which Marines dropped before attending the course, the reason for the drop, and Marines who successfully graduate the course. The Human Performance Center also provided me with Countermovement Jump metrics. My final predictive model described later in this chapter uses four of the 73 variables. The four CMJ metrics and other relevant variables that ITB provides are listed in Table 4.

Table 4. Definitions of Variables, ITB-E Data

Variable Name	Variable Description
EDIPI	Electronic Data Interchange Personal Identifier matched with TFDW data
DROP	Binary indicator of whether or not candidate graduated BRC
DROP_CATEGORY	Reason for attrite (MSPS, Legal, Medical, Other)
GRAD	Binary indicator of whether or not candidate graduated BRC
TD	The training day identified if a candidate dropped from the training
ConcentricRFD50msNs	Measures a candidate's concentric rate of force development during the concentric phase of the CMJ [ms]
StartofBrakingPhases	The beginning of the deceleration phase of a candidates CMJ [s]
JumpHeightFTRelativeLandin	A candidate's Jump Height Flight Time relative to the landing force generated. [N/s/cm]
ConcentricMeanForceN	A candidate's average force output during the concentric phase of the CMJ [N]

B. DATA CLEANING AND VARIABLE CONSTRUCTION

The merging of the two data sets from TFDW and ITB requires extensive matching and specific attention to ensure each data source includes all of the same sample population.

The following section details the construction of variables I use for the subsequent analyses.

1. TFDW Independent Variables

I use TFDW data for many of the pre-accession variables of candidates at ITB. These personal and demographic information include race, dependents, home of record information, GT Score, AFQT Score, and waiver information from the TFDW data. I create indicator variables to capture each candidate's ethnicity, such as Black or White. I also create variables that binned home of record information into U.S. geographical regions. I create binary variables to identify the various waivers prevalent in the sample population, such as waivers for legal-related offenses or medical conditions. Lastly, I create indicator variables to determine if a candidate is active duty or a reservist while attending ITB.

2. ITB Independent Variables

I use data from ITB to account for CMJ performance and explore the reasons for failing ITB. The data provided consist of multiple cohorts from 2018 to 2019 that performed the CMJ test. The data encompasses the end of course class summaries and those who graduate or fail. Variables include CMJ test metrics, drop categories, and the number of drops. Additionally, to explore whether physical abilities as measured by CMJ interacts with cognitive abilities, I create interaction terms of the AFQT Score with CMJ variables.

C. MERGING OF DATA

Following separate TFDW and ITB data cleaning and construction of variables, I merge the data sets for analysis. First, I assign each individual EDIPI a randomly-generated study-specific ID using Excel's random number generator. Next, I match each study-specific ID to an observation's EDIPI separately within the TFDW and ITB data. Following the addition of this study ID, EDIPIs and all other personal identifiable information are stripped from both data sets, leaving only the randomly-generated specific ID to identify individual observations. I then use this study ID to match the TFDW and ITB data sets. The match rate across the data is 100 percent of the observations.

D. SAMPLE RESTRICTIONS AND MISSING DATA

As mentioned earlier, I had to exclude 258 observations with no reported EDIPIs from the ITB data set, as I could then not match these Marines with event scores or TFDW data. Because of the missing EDIPIs, my final analysis sample includes 1,294 observations. Less than one percent of this sample have incomplete data points, for which I impute missing information. I first create indicator variables for any variables that have missing entries. I then impute those missing values by taking the mean of that variable among non-missing observations, and replace the missing data with the average. This technique allows me to utilize the entire matched sample and account for any missing entries. Table 5 describes the dummy variables.

Table 5. Definition of Missing Variables

Variable Name	Variable Description
x_height	=1 if missing height; 0 otherwise
x_weight	=1 if missing weight; 0 otherwise
x_pft_run_tm	=1 if missing PFT runtime; 0 otherwise
x_ConcentricMeanForceN	=1 if missing concentric mean force; 0 otherwise

E. DATA STATISTICS

1. Summary Statistics

This study includes 1,294 ITB candidates that begin the course during calendar years 2018 to 2019. Each observation is an individual attempt to complete ITB. Individual ITB candidates who fail out of one cohort may have the opportunity to try again in a subsequent cohort; however, I did not capture subsequent attempts in this data. Each observation in the data includes information for that candidate's CMJ scores while at ITB and data points (including demographics, evaluation marks, and test scores) captured prior to beginning ITB.

Table 6 provides summary statistics of this data. This analysis is 99 percent male due to the recent inclusion of women serving in infantry units and the limited amount of female applicants. As a result, there is minimal data on females attending ITB because of the very few attempts to date. The overall graduation rate for both sexes from 2018 to 2019 is 86 percent. ITB candidates are on average 19.89 years old and have an AFQT score of 60.76.

Table 6. Data Descriptive Statistics

Variable	**Mean**	**Std. Dev.**	**Min**	**Max**
GRAD	0.860	0.347	0	1
afqt_score	60.765	17.771	0	99
height	69.298	2.623	60	79
weight	164.626	22.458	104	273
pft_run_tm	2154.015	171.853	0	2833
cft_mtc	269.708	30.993	214	355
cft_manuf	230.635	23.599	147	332
TD	54.566	14.287	1	60
female	0.014	0.117	0	1
active_duty	0.873	0.333	0	1
rifle_qual~e	306.759	15.247	179	340
riflehigh	0.021	0.143	0	1
afqt_50	0.678	0.467	0	1
afqt_50_C~50	-0.752	24.290	-100	100
StartofBra~s	13.369	9.484	3.525	117.768
JumpHeight~n	3890.242	3657.856	438.019	52625.23
Conc~nForceN	1419.308	216.196	677	2521
MISS_height	0.001	0.028	0	1
MISS_weight	0.001	0.028	0	1
MISS_pft_run_tm	0.001	0.028	0	1
MISS_Concentr~N	0.002	0.039	0	1

2. Descriptive Statistics

Figures 17–26 illustrates the frequency and distribution of the continuous variables in this study.

Figure 17. AFQT Score Distribution

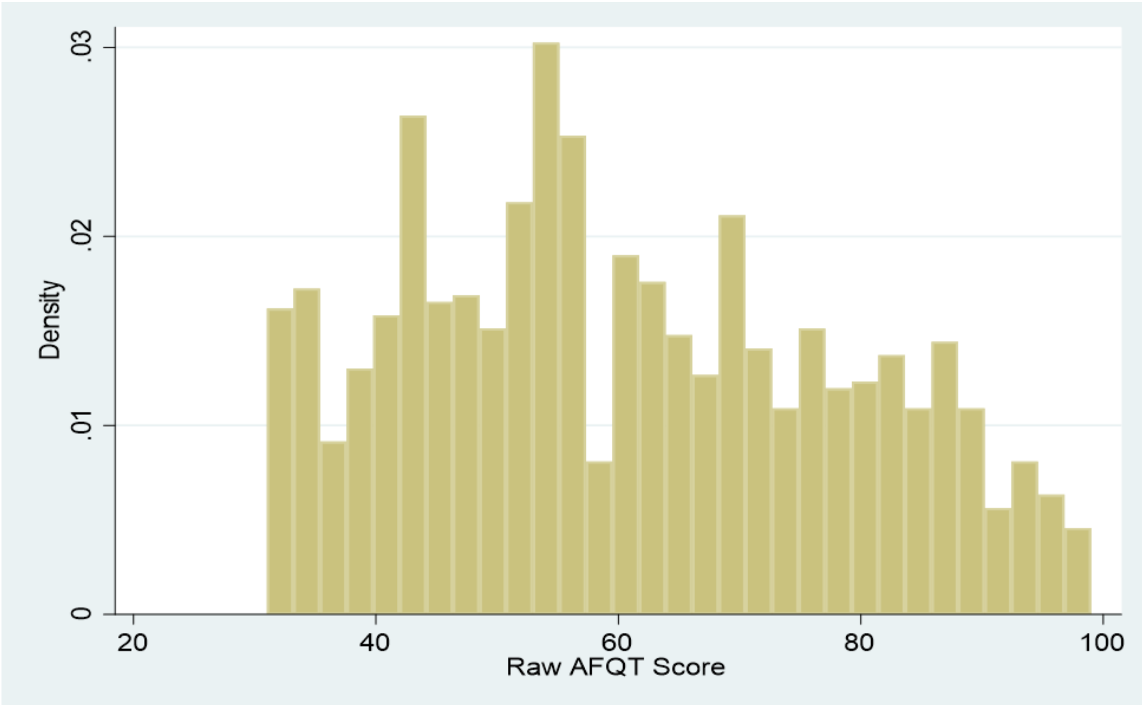


Figure 18. Height Distribution

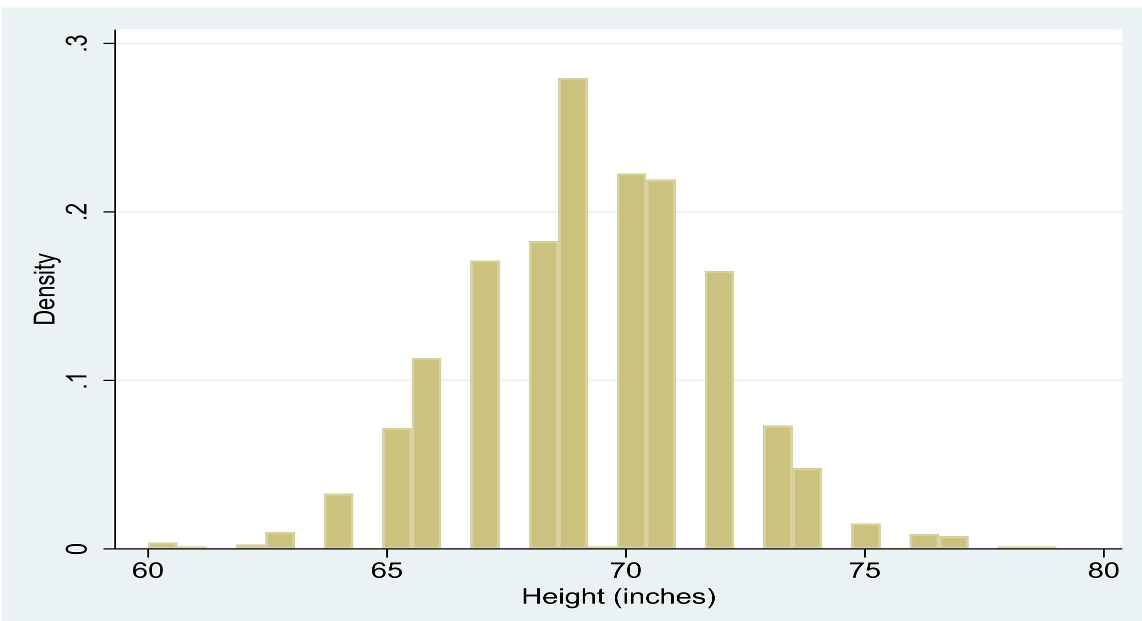


Figure 19. Weight Distribution

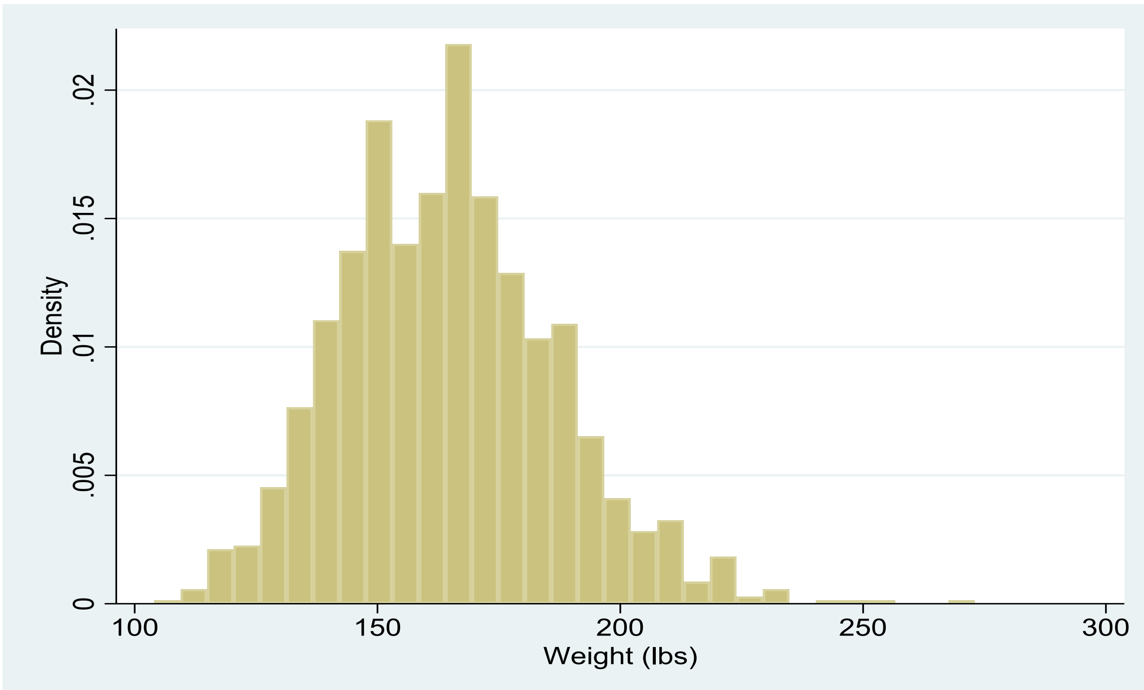


Figure 20. PFT Run-Time Distribution

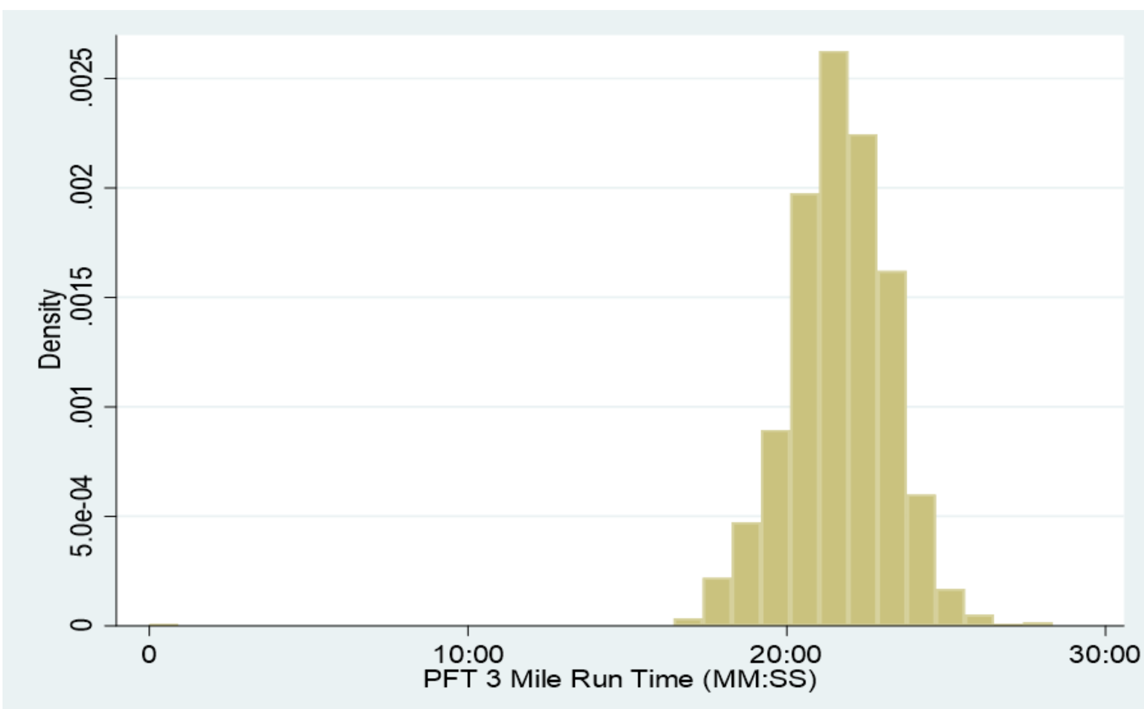


Figure 21. CFT Maneuver Under Fire Time Distribution

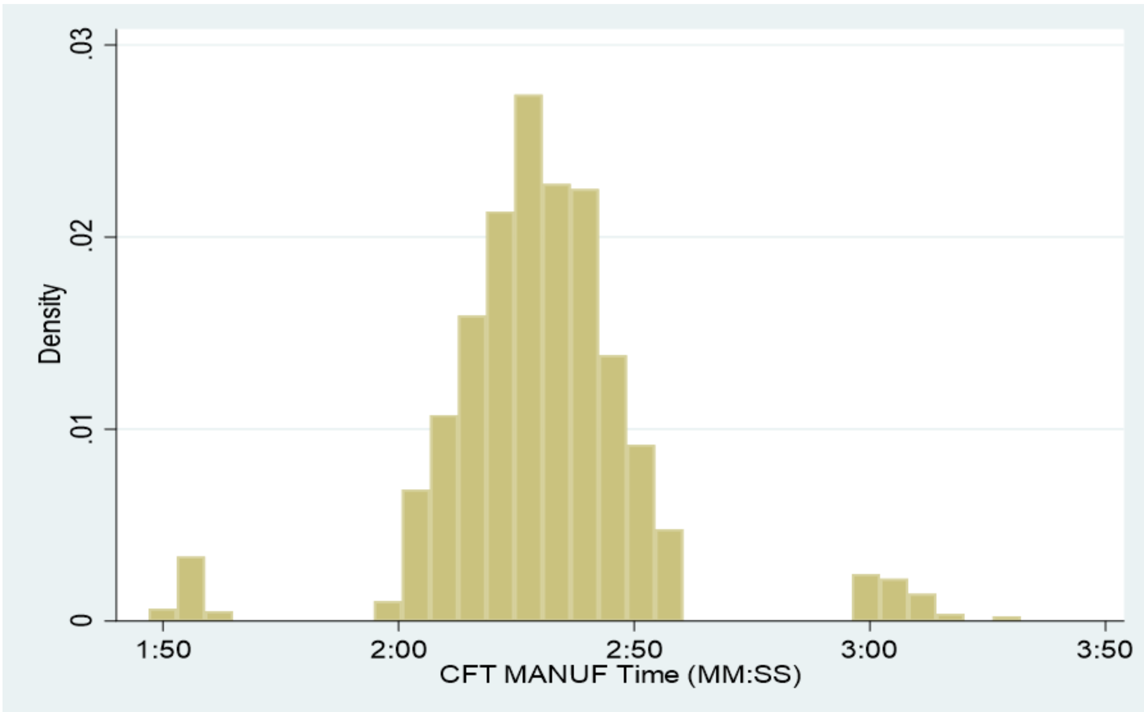


Figure 22. CFT Movement to Contact Time Distribution

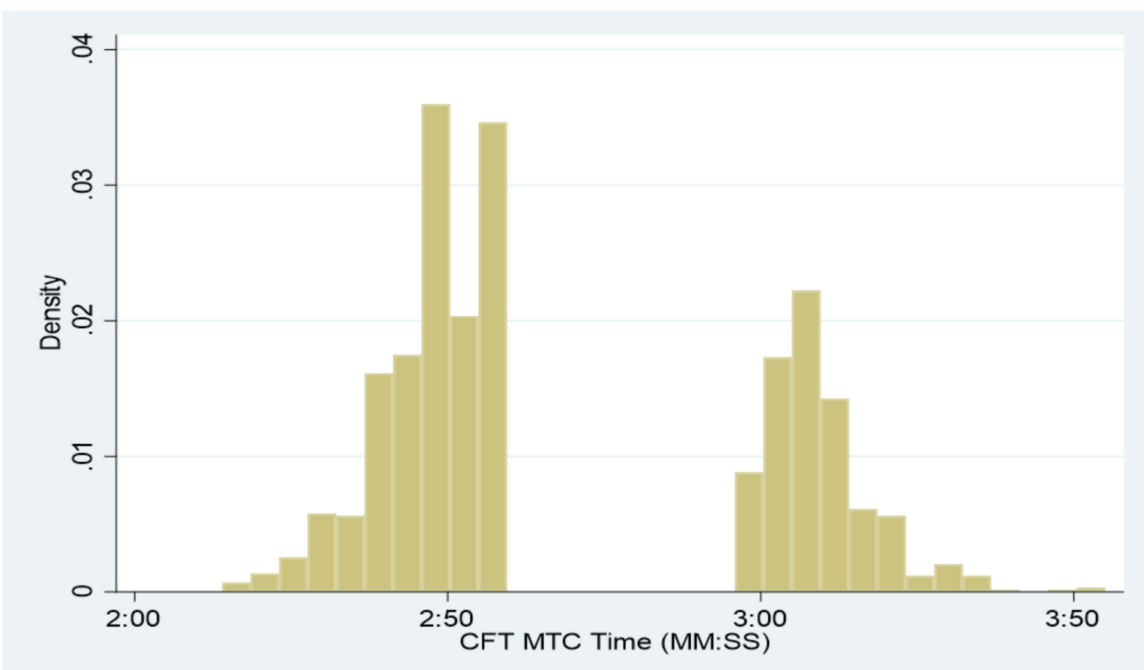


Figure 23. Rifle Qualification Score Distribution

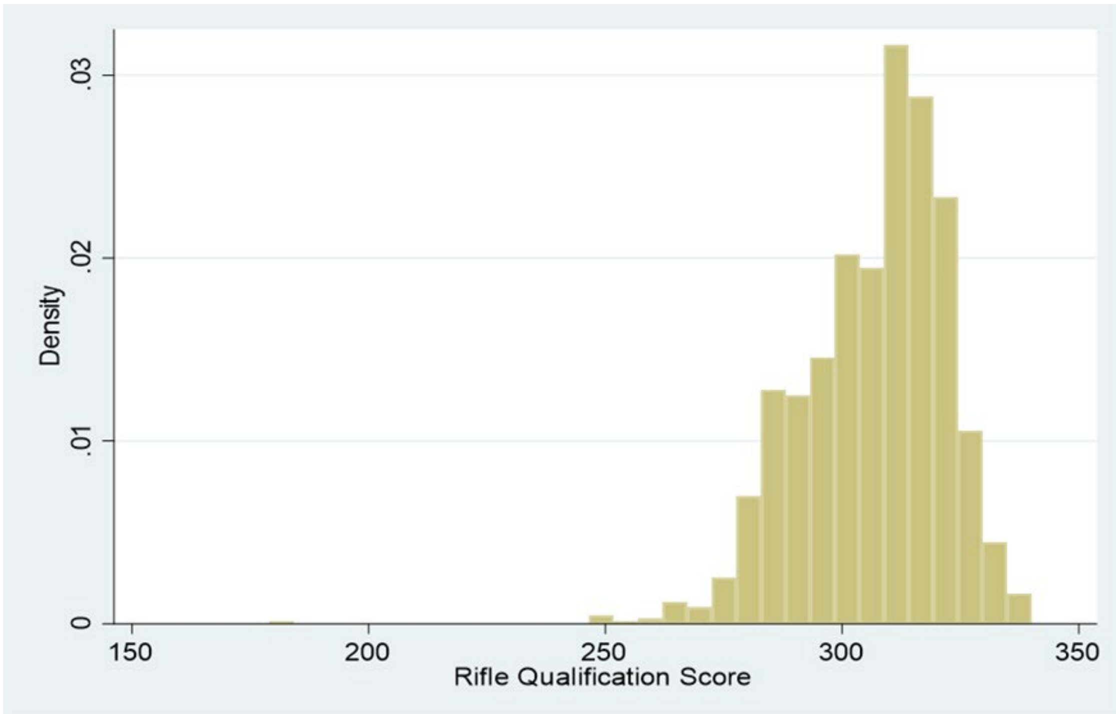


Figure 24. Start of Braking Phase Distribution

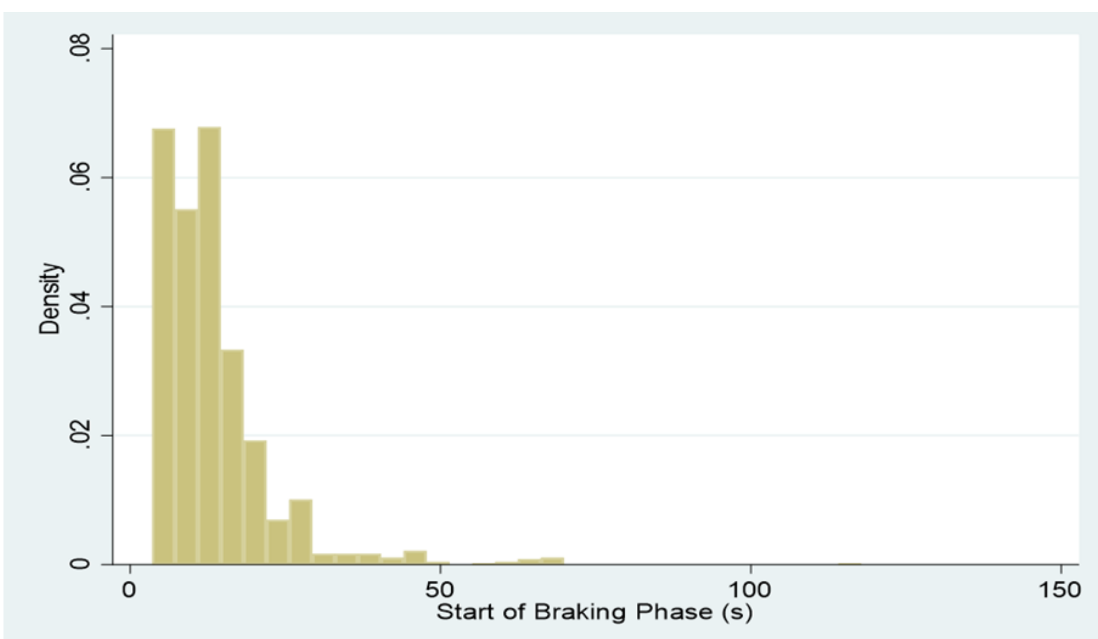


Figure 25. Jump Height Relative to Landing Force Distribution

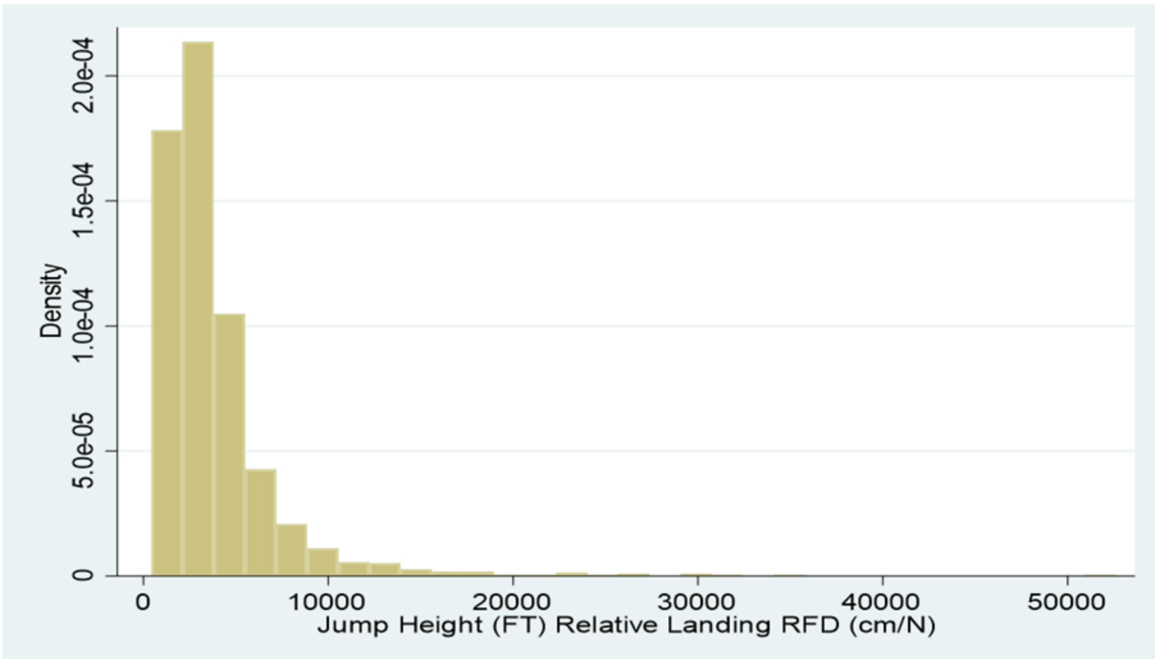
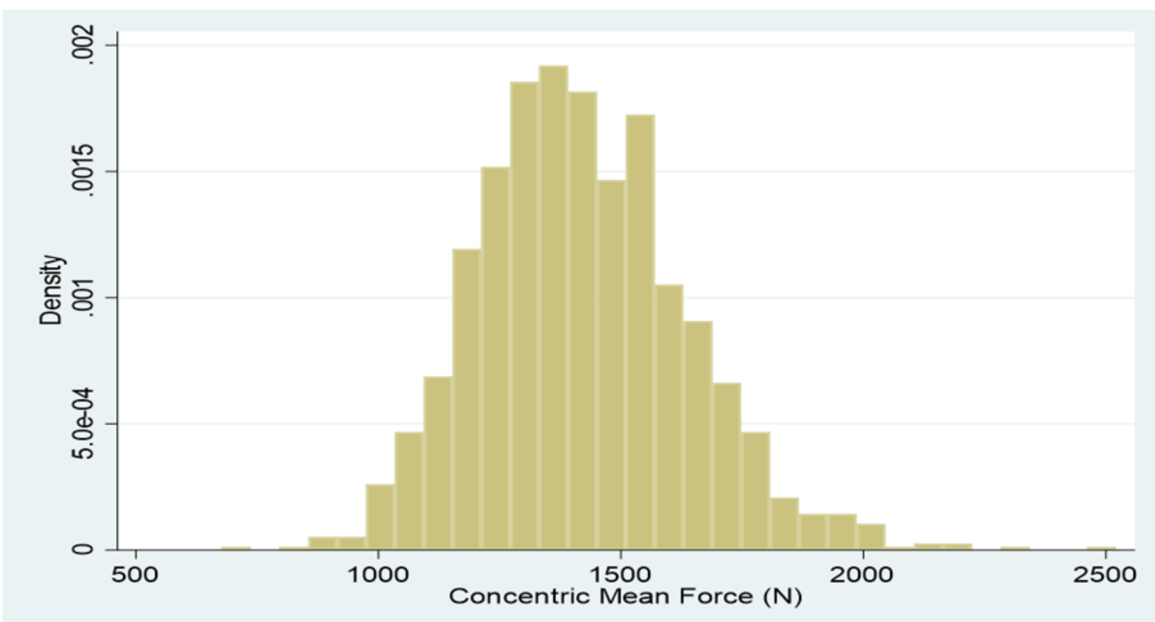


Figure 26. Concentric Mean Force Distribution



F. METHODOLOGY

1. Overview

I use STATA version 16.1 throughout this research. Logistic regression is the model I estimate in this research to characterize the factors that are significant in predicting success at ITB. Logistic regression is appropriate to use in this study because of the binary nature of the dependent variable (GRAD) reflecting graduation from ITB (GRAD=1) or attrition (GRAD=0). The logistic regression with a binary response is given by the probability of the response success:

$$P(GRAD = 1|x) = \frac{e^z}{1 + e^z}$$

Furthermore, I use the Least Absolute Shrinkage and Selection Operator (lasso) machine learning techniques to determine variable selection for my preferred model. To determine how well my model predicts success when utilizing outside data sets, I split my sample into "Training" and "Validation" subsamples. The "Training" subsample comprises 80 percent of my sample, and the "Validation" sample consists of 20 percent of my sample. By utilizing the "Training" sample, the machine learning techniques help determine which variables have the most predictive power to include in the model. I then apply these models to the "Validation" subsample to determine how well the model performs to predict success at ITB to an outside sample.

To determine the predictive probabilities, I use my preferred model to determine an observation's predicted outcome, known as $\hat{p}$. I use the commonly accepted threshold .5 to determine if an observation is expected to graduate or attrite from ITB based on my model (Wooldridge, 2015). For instance, suppose an observation's $\hat{p}$ is greater than .5. In that case, that observation is predicted to graduate from ITB, while an observation with a $\hat{p}$ less than or equal to .5 is predicted to attrite from ITB. Lastly, including too many independent variables in the model can result in "overfitting," which reduces the model's prediction accuracy due to increased estimation variability. On the other hand, including too few independent variables results in "underfitting," which introduces additional bias. Factor analysis and machine learning techniques like lasso are often utilized to balance these two types of errors.

2. Factor Analysis

Factor analysis is used mostly for reducing similar data into manageable factors. According to Bryant & Yarnold (1995), It works by "using a small set of variables (preferably uncorrelated) from an extensive collection of variables (most of which are correlated to each other)." According to Bryant & Yarnold (1995), this technique "extracts maximum common variance from all variables and puts them into a standard score." In my data set, there are 74 CMJ variables, which I attempt to use the factor analysis technique to decrease a large amount of CMJ variables into few factors. I begin by factoring in all my CMJ variables with a principal component analysis. The process creates factor loadings of the variables by removing the variance shared by the initial factor and then extracting variance for the remaining factors in chronological order (Pett et al., 2003). This process proceeds until the last factor of my analysis. I then rotate the factors to determine the variables weight for each factor and the correlation between them. Lastly, I take the relevant factors and include them in my machine learning regressions to determine if they contribute to my models' predictive power.

3. Machine Learning

According to Tibishirani (1996), lasso is a "machine learning technique that minimizes the usual sum of squared errors, bound on the sum of the coefficients' absolute values" that can assist in variable selection. Lasso is useful when utilizing a data set with many potential variables that affect an outcome, such as the data I am using in this study. Logistic lasso regression differs from ordinary logistic regression in that it adds a penalty term to the log-likelihood function to determine variable selection (Pereira et al., 2016). I achieve the penalized version I use to select variables for inclusion in the model by maximizing the log-likelihood function of (Hastie et al., 2009):

$$l_{\lambda}^{L}(\beta) = \sum_{i=1}^{n}\left[y_i x_i \beta - \log\left(1+e^{x_i\beta}\right)\right] - \lambda\sum_{j=1}^{p}\left|\beta_j\right|$$

Additionally, lasso allows me to apply different penalty parameters to determine which variables to include in my model. The penalty parameter in logistic lasso is λ or the tuning parameter. I use the most common methods to select penalty parameters through cross-validation lasso, adaptive lasso, and plug-in (Drukker, 2019). During my analysis, I use each of the tuning parameters to determine which covariates and factors I will include and exclude.

The first lasso parameter I use is cross-validation. According to Liu (2019), "lasso cross-validation selects the λ value that minimizes the out-of-sample mean squared error (MSE) of the predictions" and determines λ with minimum MSE. Cross-validation is exceptional at retaining meaningful variables but tends to include additional variables that are not required for the model. The second model I run is the plug-in based lasso. The plug-in-based lasso attempts to find the value of λ great enough to control the estimation clutter.

Additionally, unlike the cross-validation lasso, the plug-in based lasso is an excellent tool at including significant covariates and not selecting covariates that do not belong in the model. For my last lasso parameter, I use the adaptive method. Lasso adaptive is an iterative procedure of cross-validated lasso that adds larger penalty loadings on small coefficients than regular lasso ((Drukker & Liu, 2019). Covariates with large coefficients are more likely to be selected, and covariates with small coefficients are more likely to be dropped (Liu, 2019).

Elastic net is an alternative machine learning technique similar to lasso that often produces better results (Hui et al., 2005). Elastic net serves as a hybrid between the lasso and ridge regularization. According to Hastie et al. (2005), elastic net uses a "two-stage approach for each fixed λ through ridge regression coefficients and then lasso-type shrinkage along the lasso coefficient solution path." The version I chose to use in my study for inclusion in my model is (Zou & Hastie, 2005):

$$P_{\alpha}(\beta) = \frac{(1-\alpha)}{2}\|\beta\|_2^2 + \alpha\|\beta\|_1 = \sum_{j=1}^{p}\left(\frac{(1-\alpha)}{2}\beta_j^2 + \alpha|\beta_j|\right)$$

Once I estimate all lasso and elastic net logistic regressions on the training sample, I compare each model's predictive power using the validation sample. As previously discused, I use each model's estimates to determine whether an observation will or will not pass ITB. Next, I compare these predictions to the actual outcomes, and using a threshold of 0.5, determine the percentage of observations correctly predicted by each model. Table 7 displays the percent correctly predicted by each model. Most of the models were close in predictive probability; however, the adaptive lasso selection method achieved the highest predictive probability both in- and out-of-sample.

Table 7. "Validation" Sample Model Percent Correctly Predicted

Model	Training	Validation	Overall
Lasso Cross-Validation	87.03%	83.39%	85.21%
Lasso Adaptive	88.04%	85.32%	87.57%
Lasso Plugin	86.59%	83.78%	85.19%
Elastic net	87.07%	83.39%	85.23%
	= Selected Model		

4. Logistic Regression

Reviewing the machine learning technique's predictive probabilities, I determine that the adaptive lasso outperforms elastic net's predictive power. Following this determination, I construct a regression model using the variables selected by the adaptive lasso model. I estimate my preferred predictive model using 17 independent variables in the logistic regression, represented by the following equation:

$$P(GRAD = 1|x) = \frac{e^z}{1 + e^z}$$

where, the dependent variable reflects graduation from ITB (GRAD=1) or failure (GRAD=0) and

z is defined as:

$z =$ $b_0 + b_1 AFQTSCORE + b_2 HEIGHT + b_3 WEIGHT + b_4 ACTIVEDUTY + b_5 PFTRUN\ TIME + b_6 CFTMANUF + b_7 CFTMTC + b_8 RIFLEQUAL\ SCORE + b_9 RIFLEHIGH + b_{10} AFQT\_HIGH{*}CRFD50 + b_{11} BRAKINGPHASE + b_{12} JUMPHEIGHT\ RELATIVE\ TO\ LANDING + b13 CONCENTRICMEAN\ FORCE + b14 MISSHEIGHT + b_{15} MISSWEIGHT + b_{16} MISSPFT\ RUN\ TM + b_{17} MISSConcentricMeanForce$

b_0= the intercept or constant term

b_1= change in the likelihood of graduating ITB associated with a change in cognitive ability as measured by the AFQT score (holding all other variables constant)

b_2= change in the likelihood of graduating ITB associated with a change in height (inches) (holding all other variables constant)

b_3= change in the likelihood of graduating ITB associated with a change in weight (lbs) (holding all other variables constant)

b_4= change in the likelihood of graduating ITB associated with a change in service component code (Active or Reservist) (holding all other variables constant)

b_5= changes in the likelihood of graduating ITB associated with a change in PFT 3-mile run time (holding all other variables constant)

b_6= changes in the likelihood of graduating ITB associated with a change in CFT Maneuver Under Fire time (holding all other variables constant)

b_7= change in the likelihood of graduating ITB associated with a change in CFT Movement to Contact time (holding all other variables constant)

b_8= change in the likelihood of graduating ITB associated with a change in annual rifle qualification score (holding all other variables constant)

b_9= change in the likelihood of graduating ITB associated with a change in having an annual rifle qual score higher than 310 (holding all other variables constant)

b_{10}= change in the likelihood of graduating ITB associated with a change in having an AFQT score higher than 50 interacted with Concentric Rate of Force Development (holding all other variables constant)

b_{11}= change in the likelihood of graduating ITB associated with a change in having an AFQT score higher than 50 interacting with CMJ Concentric Rate of Force Development (holding all other variables constant)

b_{12}= change in the likelihood of graduating ITB associated with a change in CMJ Jump Height relative to Landing Force (holding all other variables constant)

b_{13}= change in the likelihood of graduating ITB associated with a change in CMJ Concentric Mean Force (holding all other variables constant)

b_{14}= effect of missing height measurements

b_{15}= effect of missing weight measurements

b_{16}= effect of missing PFT 3 Mile run time scores

b_{17}= effect of missing CMJ Concentric Mean Force metrics

5. Survival Model

To determine the duration of a candidate's survivability at ITB, I also estimate a cox proportional hazards model with the same predictor variables from the preferred logistic model. Instead of the dependent variable being a binary indicator for graduation, this model uses the variable "training days" as the outcome. This duration variable measures the number of days a candidate survives at ITB before dropping, ranging from days 1 to 60 (graduation), and is right-censored at 60 days. For candidates who graduate, training days equal 60, while those that attrite have training days less than 60. Observations that are not censored are indicated by the variable "drop," which is (drop=1) if a candidate fails and (drop=0) if they graduate. The equation represents the model I estimate, where h() indicates the hazard for dropping out at training day t:

$$h(t|x) = h_0(\mathrm{t})\mathrm{e}^{(x\beta_x)}$$

As indicated by this equation, survival analysis differs from logistic regression because it analyzes length of time until a particular event of interest. The hazard rate *h(t|x)* indicates the probability of failure in the next instant given they have survived up to time t given x covariates, instead of an absolute proportion. The cox proportional hazards model assumes that there are relevant covariates x that affect an entity's survivability, and that these x's affect the outcome proportionally.

I estimate the Cox Proportional Hazards model using STATA 16.1 and discuss estimates in the next chapter. I also plot the survival curves to illustrate the predictive probabilities of survival at ITB by training day. Additionally, I use the hazard and cumulative hazard, H(t), to estimate the hazardous contributions (probabilities of failure) against the established baseline. This method is appropriate for establishing a fitted model to display the survival probability based on the independent variables inputted into the model (Cleves et al., 2004).

G. CHAPTER SUMMARY

This chapter describes the data and empirical models I utilize in this study. I define the methods to clean the data sets, variable creation, and merging of the various data sets into one usable format for analysis. I provide summary and descriptive statistics that describe the data set and the sample population of the study. Lastly, I outline the methodology, machine learning techniques, the criteria for model selection, and the final logistic regression and survival models. Next, I turn to estimate the models and discuss results.

V. RESULTS AND ANALYSIS

Chapter V reports and interprets the findings from my logistic regression model and survival model. I begin this chapter by providing a comprehensive analysis of my logistic regression model's predictive probabilities and the predictive power of the model with and without CMJ metrics. Following the presentation of predictive probabilities, the next sections discuss the predictors chosen in the logistic and survival models and their roles with cognitive abilities and physical performance at ITB.

A. PREDICTIVE PROBABILITIES

Figures 27 and 28 depict the predictive probabilities of my final model. Figure 27 represents the model that includes all data sets, including the CMJ metrics, while Figure 28 compares my model with and without the CMJ metrics. The sample population for both figures consists of all 1,296 observations included in this study.

Figure 27. Model Predictive Probabilities (n=1,296)

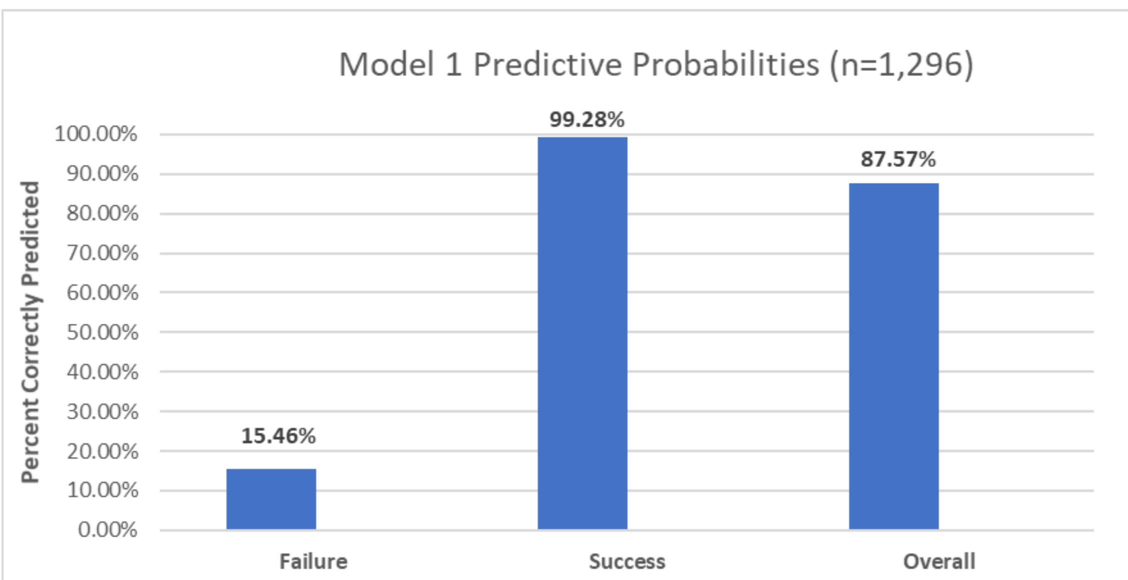
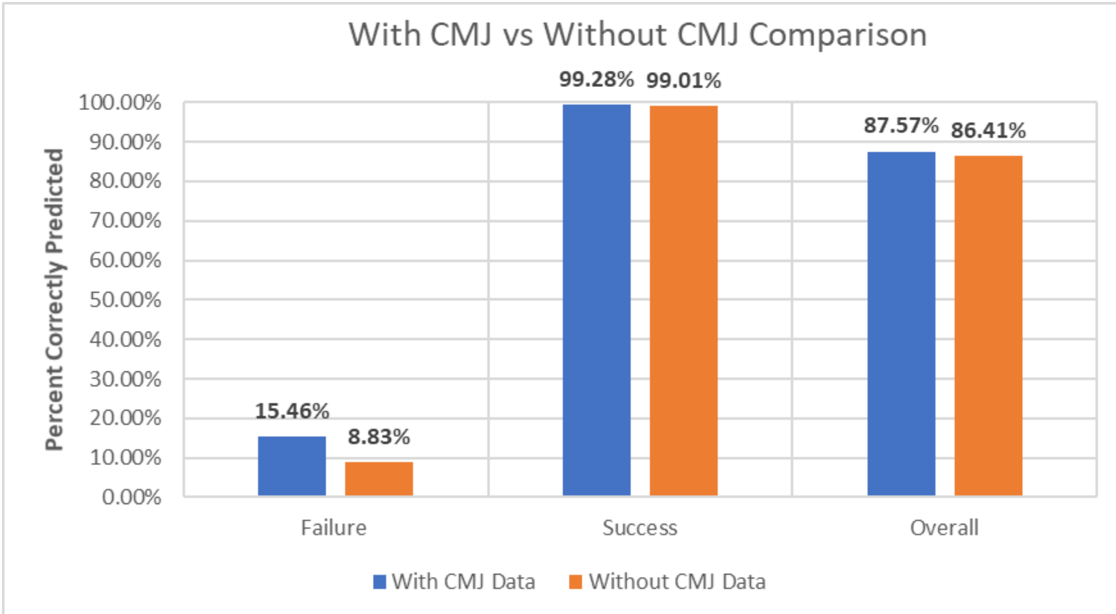


Figure 28. CMJ vs. No CMJ Predictive Probabilities (n=1,296)



The failure bin percentage represents the number of observations my model correctly predicted would fail, divided by the total number of actual failures in the sample population. The success bin percentage represents the number of observations my model correctly predicted would graduate, divided by the total number of actual graduates in the sample population. The overall bin percentage represents the total number of failures and graduates my model correctly predicts, divided by the entire sample population relevant to the model. Last, Figure 28 provides a comparison of the beforementioned bin percentages with and without CMJ metrics.

The overall results of the model suggest that it predicts well at 87.57 percent correct. The model correctly predicted 99.28 percent of the candidates who graduate from ITB. However, the model's ability to predict failure is nominal at best, correctly predicting 15.46 percent of candidate failures from ITB; however, it is important to note that by including CMJ metrics, the predictive power of failure increases approximately seven percentage points.

B. LOGISTIC REGRESSION MODEL

My logistic regression model's overall results suggest that testing, demographic, performance, and CMJ variables play a significant part in determining the success of graduating at ITB. Figure 29 reports the odds ratios of the logistic regression models I present in Chapter IV by starting with testing variables and progressively adding demographic, performance, and CMJ variables.

The coefficients in Figure 29 identify the change expected in the logistical odds when there is a one-unit change in the independent variable while holding all other variables constant.

Figure 29. Logistic Regression Odds Ratio Results

Logistic Regression Estimates

	(2) Testing	(4) Demographic	(5) Performance	(6) CMJ
AFQT Score	1.023^{**}	1.019^{***}	1.015^{**}	1.017^{**}
	(4.65)	(3.69)	(2.85)	(3.08)
Height		1.157^{***}	1.121^{**}	1.124^{**}
		(3.86)	(2.95)	(2.98)
Weight		1.020^{***}	1.021^{***}	1.015^{*}
		(4.24)	(4.19)	(2.49)
Active Duty		0.610	0.571	0.589
		(-1.62)	(-1.77)	(-1.65)
PFT Run Time			0.999	0.999
			(-0.83)	(-1.29)
CFT MANUF			0.986^{***}	0.987^{**}
			(-3.33)	(-3.12)
CFT MTC			0.992^{*}	0.993
			(-2.42)	(-1.90)
Rifle Qual			1.011^{*}	1.013^{*}
			(2.10)	(2.34)
Rifle High			1.173	1.298
			(0.24)	(0.39)
AFQT*CRFD50				1.012^{***}
				(3.43)
Start of Braking Phase				0.977^{**}
				(-2.87)
Jump Height				1.000
				(1.83)
Concentric Mean Force				1.001
				(1.49)
Observations	1,296	1,295	1,294	1,292

Exponentiated coefficients; t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

1. Cognitive Variable

The first group of testing variables includes the AFQT score. The AFQT coefficient is statistically significant at the $\alpha = 0.01$ level when regressed alone and with the inclusion of performance and CMJ variables. The odds ratio of 1.017 suggests that a candidate's odds to graduate ITB increase by nearly 2 percent for every unit increase in the AFQT score, holding all other variables constant.

2. Demographic Variables

Model 2 adds height, weight, and an indicator variable for being active duty. In this model, AFQT remains statistically significant at approximately the same magnitude as in model 1. Height is statistically significant at $\alpha = 0.01$ in predicting the success of a candidate at ITB. For every inch increase in height, a candidate's odds to graduate ITB increase by a factor of 1.124 (or 15 percent more likely per inch), holding all other variables constant. Because I assume lean muscle mass is actually what matters, I hypothesize that weight may be particularly important for predicting graduation. Indeed, for every one-pound increase in weight, a candidate's odds to graduate ITB is 1.015 higher, holding all other variables constant. Lastly, active-duty status was not statistically significant at any level in my model.

3. Performance Variables

I next add five performance variables to the model. Of the performance variables, CFT MANUF has an inverse relationship with a candidate's likelihood of graduating ITB. For every second increase on the CFT MANUF time, a candidate's odds to graduate ITB decrease by a factor of .987 (or approximately 1 percent per second) higher, holding all other variables constant. Rifle score is statistically significant at the $\alpha = 0.05$ level. For every increase in rifle score, a candidate's odds to graduate ITB is 1.013 higher, holding all other variables constant. Even though candidates do not conduct a rifle range at ITB, their rifle score from recruit training indicates a positive success factor at ITB. Lastly, CFT Movement to Contact and Rifle High are not statistically significant when including all other covariates within the model. The coefficients on the other variables remain similar to model 2.

4. CMJ Variables

The final model adds three CMJ variables (Start of Breaking Phase, Jump Height, and Concentric Mean Force) as well as an interaction of a categorical test variable with a continuous CMJ variable. The testing variable I use is AFQT50, which equals one if a candidate has an AFQT score of 50 or higher and 0 otherwise. I use the CMJ variable CRFD50, which measures a candidate's concentric rate of force development at 50 milliseconds.

Braking Phase duration is statistically significant at $\alpha = 0.01$. Specifically, for every second increased in the braking phase of the CMJ, a candidate is 0.977 times less likely to graduate from ITB, holding all else constant. This makes sense because, during this phase of the CMJ, the athlete decelerates or "brakes" their center of mass. Literature shows that elite athletes adopt explosive movement techniques to optimize force production and complete tasks with shorter duration during the braking phase (Kennedy & Drake, 2018). Jump Height and Concentric Mean Force are both statistically insignificant in my model; however, they contribute to my model's validity in a meaningful way, even at a statistically insignificant level.

The interaction term is statistically significant at $\alpha = 0.001$ level with an odds ratio of 1.012. Specifically, for every one-unit change in CRFD at 50 milliseconds, a candidate with an AFQT score of 50 or higher is approximately one time more likely to graduate ITB than a candidate with an AFQT score lower than 50.

C. SURVIVAL MODEL

Next, I use survival and duration analysis to assess empirically if a given candidate does attrite from ITB, at what time event does failure occur, and what data correlates with that candidate's duration of survival at ITB. First, I determine when and for what reason candidates attrite from ITB. Figure 30 summarizes the reasons candidates attrite. The substantial extracts are MSPS events, which account for 39.23 percent of all candidates that attrite; Academics at 29.83 percent, followed by medical and other. Table 8 depicts what days during the training cycle candidates are likely to attrite. Of significance are training days 10 through 15, which account for approximately 32 percent of the total failures from ITB.

Figure 30. Distribution of the Reasons for Failure from ITB

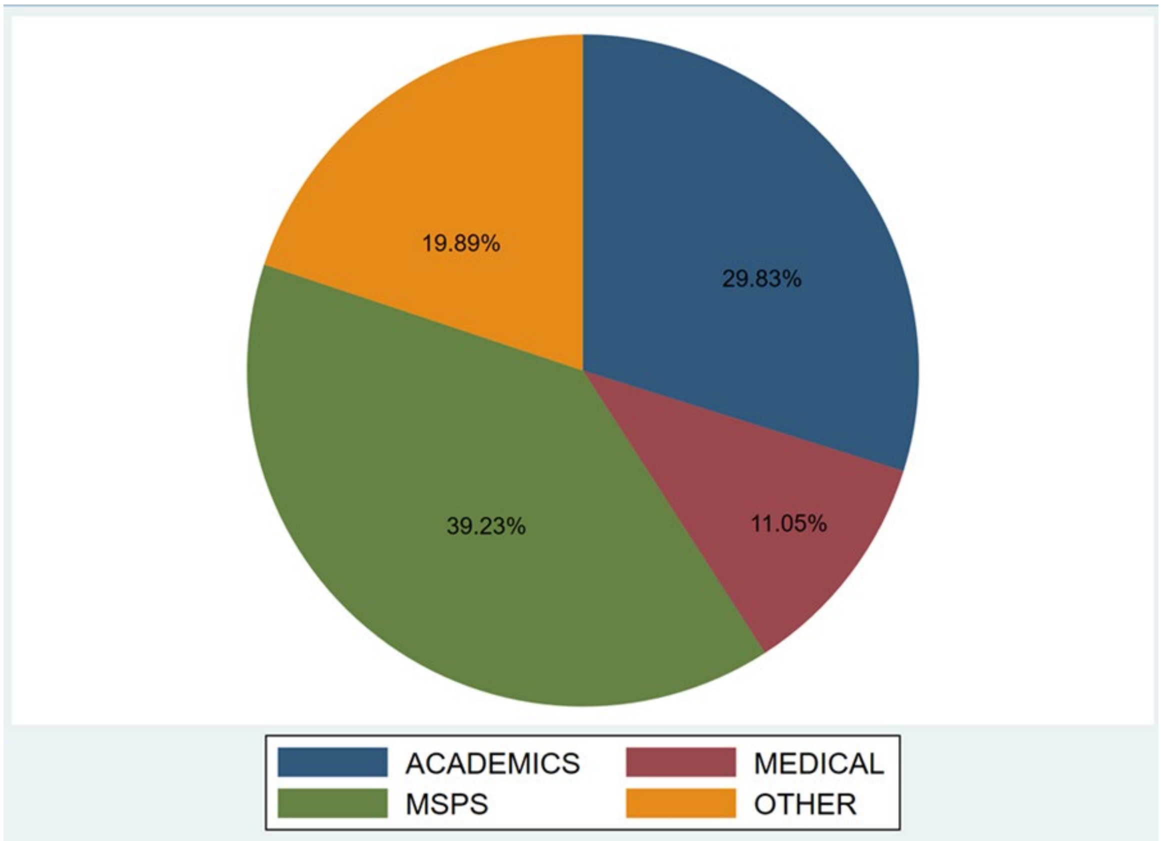


Table 8. Percentages of Attrition by Training Day

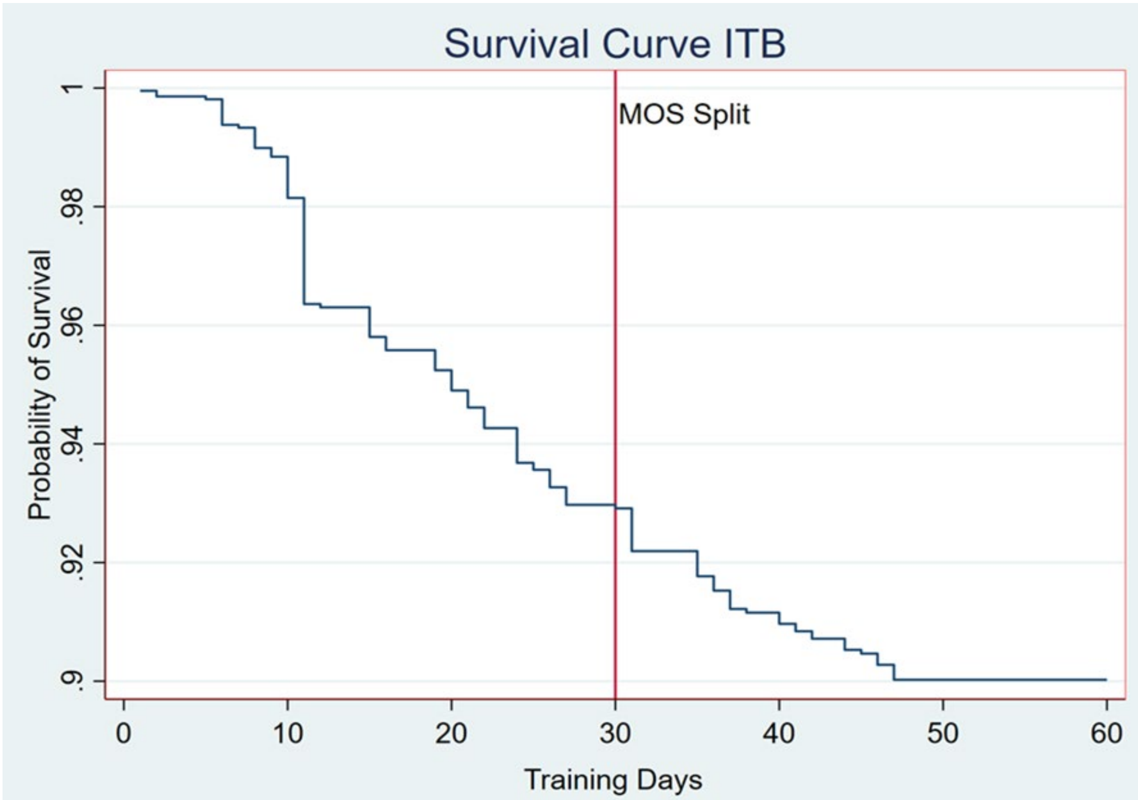
Training Days	**Freq.**	**Percent**
1	1	0.56%
2	2	1.12%
5	1	0.56%
6	9	5.03%
7	1	0.56%
8	7	3.91%
9	3	1.68%
10	14	7.82%
11	34	18.99%
12	1	0.56%
15	9	5.03%
16	4	2.23%
19	6	3.35%
20	6	3.35%
21	5	2.79%
22	6	3.35%
24	10	5.59%
25	2	1.12%
26	5	2.79%
27	5	2.79%
30	1	0.56%
31	12	6.70%
35	7	3.91%
36	4	2.23%
37	5	2.79%
38	1	0.56%
40	3	1.68%
41	2	1.12%
42	2	1.12%
44	3	1.68%
45	1	0.56%
46	3	1.68%
47	4	2.23%
Total	179	100%

Figure 31 presents the survival model results based on the same covariate model as the logistic regression I use. Much like the logistic model, cognitive ability, demographics, physical performance, and CMJ scores are the most significant contributors to survivability at ITB. Additionally, Figure 32 provides a visual representation of the survival curve for candidates attending ITB. This figure suggests a steep drop in survivability before divergence into specific MOS training. Overall, the results illustrate a significant decrease in survivability around training day 10, consisting of both academic and MSPS events. This makes sense since the events during this time are both cognitively and physically challenging.

Figure 31. Survival Analysis Results

_t	Haz. Ratio	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
afqt_score	.9886164	.0041866	-2.70	0.007	.9804447	.9968562
height	.9054156	.03144	-2.86	0.004	.8458444	.9691822
weight	.9888592	.0053105	-2.09	0.037	.9785053	.9993227
pft_run_tm	1.000454	.000604	0.75	0.452	.9992712	1.001639
cft_manuf	1.010281	.0035274	2.93	0.003	1.003391	1.017218
cft_mtc	1.007364	.0031936	2.31	0.021	1.001124	1.013643
afqt_50_CRFD50	.9911085	.0029068	-3.05	0.002	.9854276	.9968223
StartofBrakingPhases	1.017287	.0054818	3.18	0.001	1.0066	1.028089
JumpHeightFTRelativeLandin	.9999591	.0000266	-1.54	0.125	.9999069	1.000011
rifle_qual_score	.9883747	.004307	-2.68	0.007	.979969	.9968525
ConcentricMeanForceN	.9991344	.0005064	-1.71	0.088	.9981425	1.000127
active_duty	1.576267	.4521159	1.59	0.113	.8984256	2.765523
riflehigh	.8633866	.5256408	-0.24	0.809	.261808	2.847264

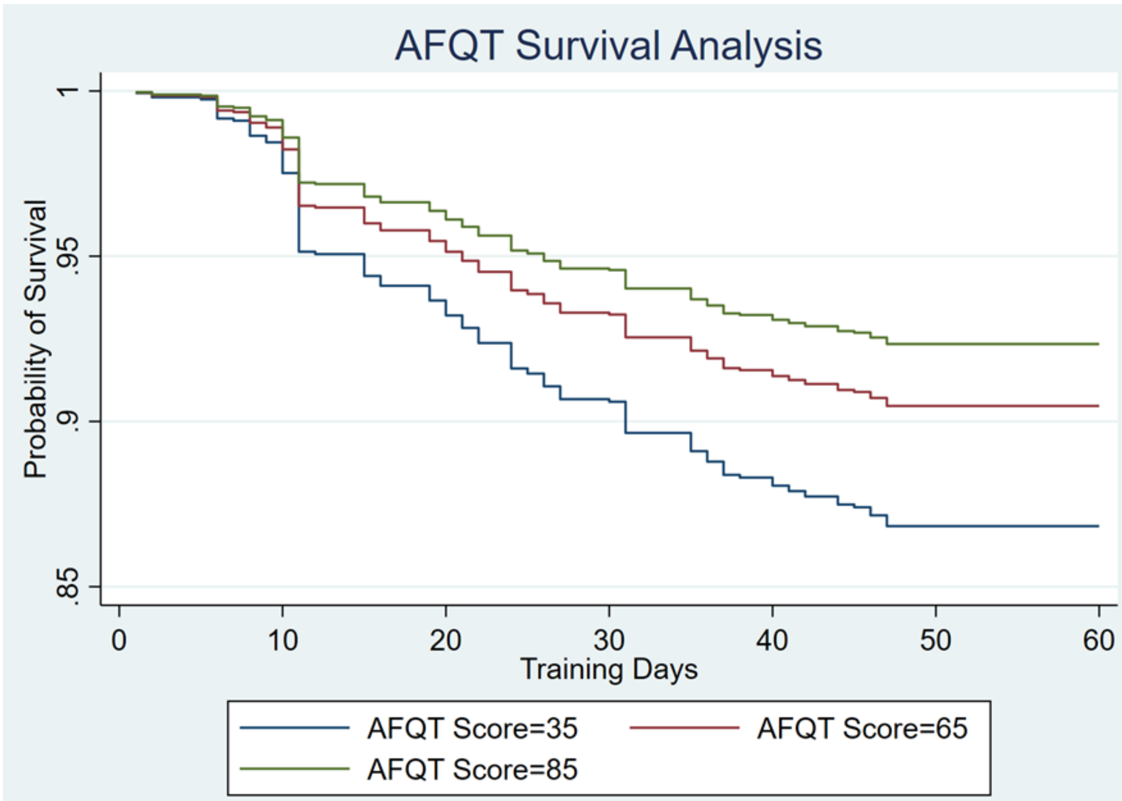
Figure 32. Survival Curve



1. Cognitive Variable

The results suggest that cognitive ability is statistically significant in predicting a candidate's survivability at ITB. A candidate who increases their AFQT score by one point is .988 as likely to attrite (or 1 percent less likely) as a candidate with an AFQT score one point lower. Figure 33 illustrates the impact of AFQT score on survivability using the mean of thirds within the data, holding all else constant. The lower bound is the mean of the bottom third AFQT Score (35), and the upper bound is the mean of the upper third AFQT Score (85). This difference expands as training duration increases, meaning that candidates with a lower AFQT score are more likely to attrite from ITB than candidates with a higher score.

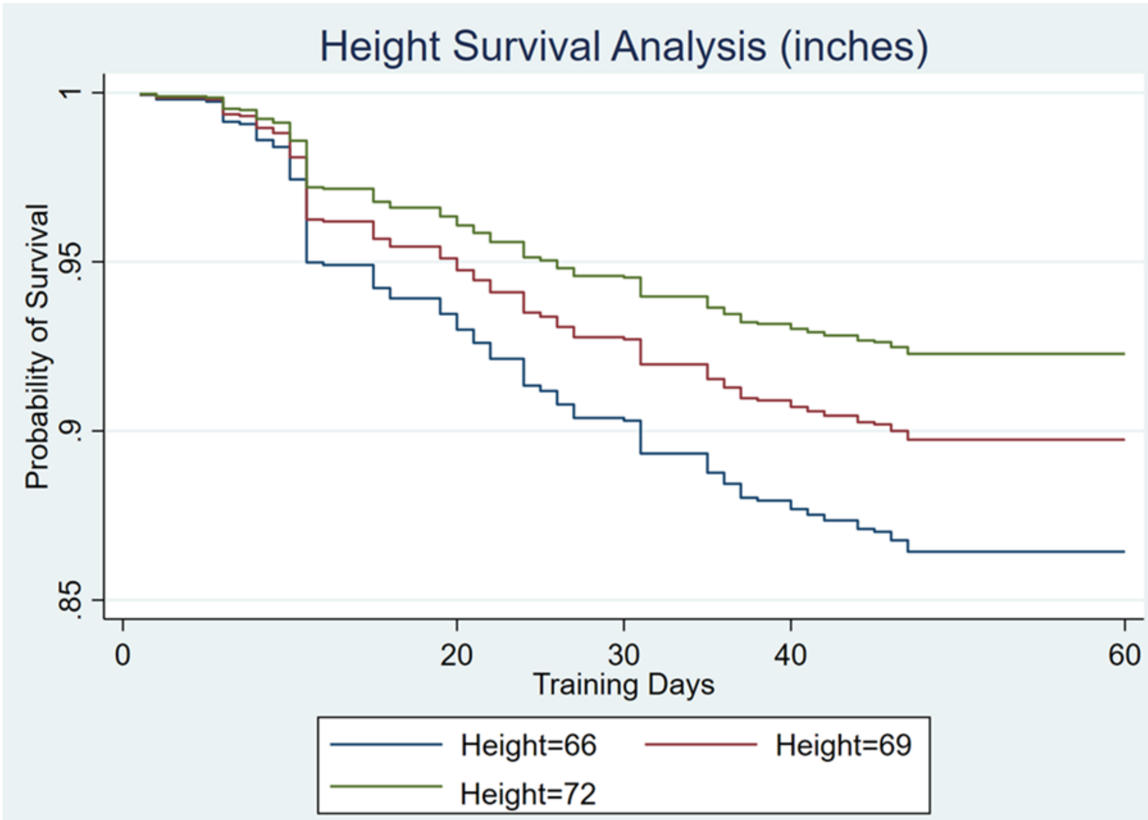
Figure 33. AFQT Survival Analysis Results



2. Demographic Variables

The results suggest that height and weight are statistically significant at the α = 0.01 and 0.00 levels, respectively. Specifically, for every inch increase in height, a candidate is .905 as likely to attrite (or 9 percent less likely) than a shorter candidate. Figure 34 provides a graphical representation of the impact of height throughout the course using the same weighting thirds tiered weighting system as previous graphs. Weight is less significant than height when predicting survivability but is meaningful none the less. As a candidate's weight increases by one pound, they are .988 less likely to attrite than a candidate who weighs a pound less. It is important to note that the Marine Corps has stringent body composition standards that determine the minimum and maximum height and weight standards allowed within the Marine Corps (Department of the Navy, 2019). If these standards were not in place, diminishing returns would be prevalent.

Figure 34. Height Survival Analysis Results



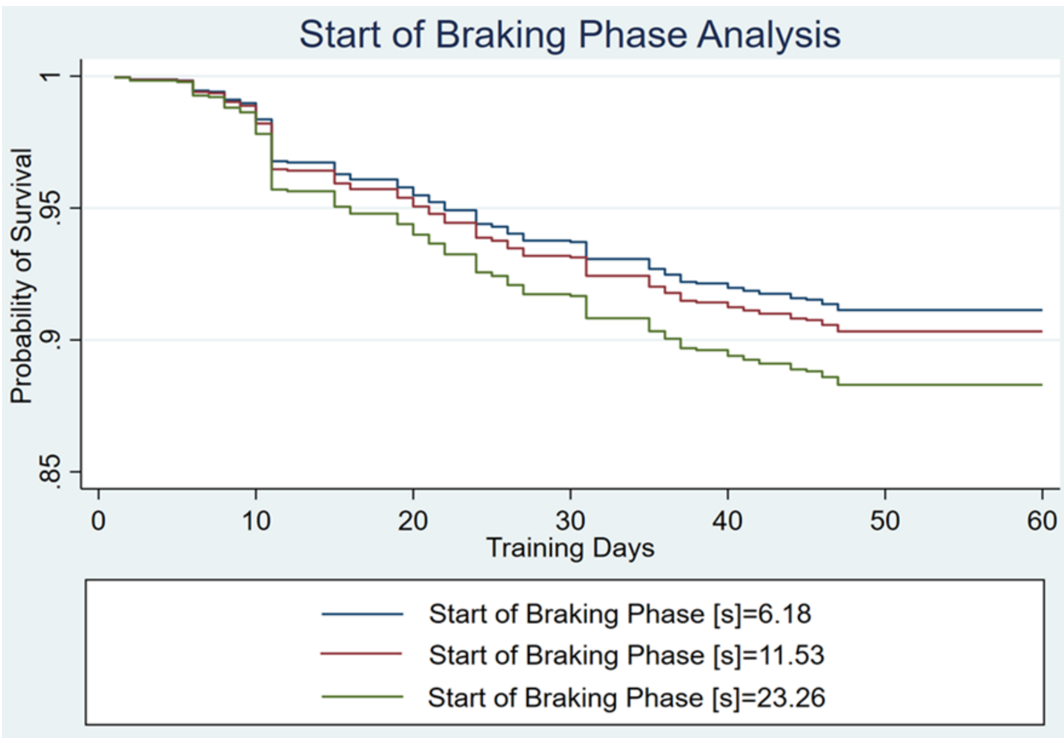
3. Performance Variables

According to the survival model, the anaerobic capacity variables (CFT MANUF and CFT MTC) are statistically significant predictors of the ITB survivability. For both variables, every second a candidate increases during the events, they are approximately one time more likely to attrite from ITB than a candidate who executes them at a faster pace. Lastly, rifle qualification score is statistically significant at the α = 0.01 level, but the magnitude is relatively small. Specifically, for every one-point increase on a rifle qualification score, a candidate is .988 times as likely to attrite from ITB as someone with a lower score. No other performance variables within the model are statistically significant.

4. CMJ Variables

Braking phase duration is also statistically significant in both models. For every second increase during the breaking phase of the CMJ, a candidate is one time more likely to attrite than a candidate that performs the movement faster. Figure 35 provides a graphical representation of the probability of survival in relation to braking phase duration in seconds. A braking phase of 11.53 seconds or less has a survivability rate of 90 percent or greater; however, once a candidate surpasses this threshold, their probability of survival drops significantly. The remaining CMJ variables were not statistically significant in the survival model.

Figure 35. Start of Braking Phase Analysis Results



5. Interaction Term

Much like the logit model, AFQT_50_CRFD50 is statistically significant at the $\alpha = 0.01$ level. Specifically, for every one-unit change in CRFD at 50 milliseconds, a student with an AFQT score of 50 or higher is approximately .99 as likely to graduate ITB than a student with an AFQT score lower than 50. This supports the idea that both cognitive ability and CMJ performance are complementary and play an essential part in survivability at the ITB.

VI. CONCLUSIONS AND RECOMMENDATIONS

A. SUMMARY

This study aims to use performance, demographic, and CMJ data to analyze what factors correlate to success and failure at ITB. My study's objective is to find quantitative evidence that identifies the characteristics that contribute to success and failure at ITB. Due to the significant cost associated with training a Marine, it is critical to determine which Marines are the most likely to be successful at the onset of training. Having meaningful and practical predictors of success will give commanders a solid framework for screening potential candidates and lower attrition rates, saving the Marine Corps' fiscal and time resources.

To achieve the study objective, I use machine learning techniques by applying different penalty parameters to determine which variables to include in my empirical models predicting success at ITB. Using these variables from the preferred model, I estimate logit regression and Cox Proportional Hazards models to estimate the effects of the selected characteristics on ITB graduation and length of survival at the school.

Estimates from my logistic and survival regression models highlight cognitive ability and physical performance are necessary attributes to complete the demanding and challenging training at ITB. These findings are supported by Dove and Richmond (2017), Larkin (2017), and Nowicki (2017). The results reviewed in Chapter V suggest that raising the minimum required AFQT Score to attend ITB will positively affect the probability of a candidate graduating. CFT anaerobic capacity events also play a role in success at ITB. Like cognitive ability, physical performance is well documented in the academic literature as a determinant to success in a military training environment. The majority of the failures at ITB are due to physically-intensive MSPS related events. The significance of these outcomes suggests that senior military leaders should focus on candidates' overall physical well-being before indoctrinating them to train.

In addition to the typical PFT and CFT measures for physical ability, this study used CMJ metrics which have not previously evaluated at ITB as predictors of success.

The CMJ uses kinetic characteristics of a candidate's movement and is a practical and reliable test for detecting both athletic potential and identifying areas of weakness among candidates. It provides immediate quantitative data to evaluate an athlete's execution of a skill or physical development. This study finds that the braking phase of the CMJ test is statistically significant in determining the probability of graduating ITB and predicting candidates' survivability.

While physical and cognitive abilities are independently significant, it is the complement of the two that is the most important predictor of success at ITB. To measure cognitive ability and CMJ metrics' complementarity, I create an interaction between high AFQT scores and CMJ test metrics. This interaction term provides the most statistically significant contributor to success in both the logit and survival models. This result suggests that it is not only physical skills that matter on its own, but that cognitive function can complement physical and motor skills to improve a candidate's function and skill performance at ITB.

Finally, while the preferred logit model predicts success very well, it struggles with accurately predicting the failures from ITB. Including CMJ metrics does significantly add to the predictive power of both success and failure predictions' accuracy (increasing the prediction of failure by nearly seven percentage points). As a result, I recommend policymakers re-examine the requirements to attend ITB and potentially alter them to increase the probability of candidates' success. At a minimum, commanders can use the CMJ data at the onset of training to determine which candidates have an increased likelihood of failure and develop preventative measures to mitigate this risk.

B. RECOMMENDATIONS FOR FUTURE RESEARCH

1. Further ITB Research

The majority of failures at ITB occur between training days 10 and 15 prior to the MOS split. This is partly due to the newly developed MSPS grading criteria established in 2019 and the physically challenging nature of ITB. With 32 percent of failures happening within this timeframe, there is concern the right candidates are not being recruited. Moreover, the Marine Corps is testing the newly developed Infantry Marine Course (IMC) that is scheduled to replace ITB in the near future, focusing on building cognitive capability among future candidates. Further research should be conducted to determine if the failure rate significantly changes with IMC or if there are more indicative prerequisites that can target candidates with the highest probability of completing IMC.

2. Expand Population Set

My population set was limited due to the recent inclusion of CMJ metrics as part of the ITB screening process and the COVID-19 pandemic, preventing CMJ test execution. Because of this, there is limited data on CMJ metrics. It would be beneficial to the Marine Corps to reexamine critical predictors of success and failure when more Marines have attempted the new POI at ITB/IMC.

3. Continuing and Expanding the Use of CMJ Test

The findings regarding CMJ data lead me to believe that CMJ data collection warrants continued gathering at ITB. An expanded CMJ data set has the potential to aid future research in accounting for factors not easily captured by existing Marine Corps collection procedures for physical skills. For example, expanding CMJ metrics to all entry-level school houses to include MCRD basic training will create a research opportunity to assess performance demographics across the Marine Corps that can affect service-wide policy. Additionally, CMJ metrics are a useful predictor of injuries among servicemembers (Kodeish et al., 2015). I recommend conducting further research into the CMJ test and its ability to predict injuries among servicemembers attending initial training.

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Bryant, F. B., & Yarnold, P. R. (1995). Principal-components analysis and exploratory and confirmatory factor analysis. In *Reading and understanding multivariate statistics.* American Psychological Association. 99–136.

Cleves M., Gould W., & Gutierrez, R .G., (2008). An introduction to survival analysis using Stata. *Journal of Statistical Software, 13*(1). https://jstatsoft.org/article/view/V012b01

Dahlman, C. J. (2007). The cost of a military person-year: A method for computing savings from force reductions. *National Defense Research Institute*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.140.911&rep=rep1&type=pdf

Department of the Navy (2019). *Marine Corps body composition and military appearance program*. (MCO 6110.3A). https://www.fitness.marines.mil/Portals/211/documents/BCP%2020190502%20.pdf

Department of the Navy. (2020). *Infantry training and readiness manual* (MCO P3500.72A). https://www.marines.mil/Portals/1/Publications/NAVMC%203500.44D%20Infantry%20T-R%20Manual.pdf?ver=2020-05-21-104009-273

Dove, J. M., Richmond, B. A. (2017). *Infantry training battalion: A predictive model for success under female integration* [Master's thesis, Naval Postgraduate School]. NPS Archive: Calhoun. https://calhoun.nps.edu/handle/10945/56908

Drukker, D. (2019). *An introduction to the lasso in Stata*. The STATA Blog. https://blog.stata.com/2019/09/09/an-introduction-to-the-lasso-in-stata//

Drukker, D., & Liu, D. (2019*). A Plug-in for poisson lasso and a comparison of Partialing-out poisson estimators that use different methods for selecting the lasso tuning parameters*. Queens University. https://www.econ.queensu.ca/sites/econ.queensu.ca/files/csplugin.pdf

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning data mining, inference, and prediction* (2nd ed.). Springer Publishing. https://doi.org/10.1007/978-0-387-84858-7

Lake, J. (n.d.). *So, you've heard about force plates... A Gentle Introduction to Using Force Plate*. Hawkindynamics.com. https://hawkindynamics.com/blog/so-youve-heard-about-force-plates...-a-gentle-introduction-to-using-force-plates-by-dr-jason-lake

Infantry Training Battalion. (n.d.). *Infantry training battalion mission*. http://www.trngcmd.marines.mil/Units/West/School-of-Infantry- West/Infantry-Training-Battalion/

Infantry Training Battalion., (2019). Basic infantryman program of instruction. *Training and Education Command*. 1–99.

Kennedy, R. A., & Drake, D. (2018). Is a bimodal force-time curve related to Countermovement Jump Performance? *Sports*, *6*(2). https://doi.org/10.3390/sports6020036

Krulak, V. H. (1984). *First to fight: An inside view of the U.S. Marine Corps*. Naval Institute Press. 1–272

Lake, J. (n.d.). *So, you've heard about force plates... A Gentle Introduction to Using Force Plate*. Hawkindynamics.com. https://hawkindynamics.com/blog/so-youve-heard-about-force-plates...-a-gentle-introduction-to-using-force-plates-by-dr-jason-lake

Larkin, J. P. (2017). *Advanced infantry training: an empirical analysis of (0341) Mortarman success while attending advanced mortarman course*. [Master's thesis, Naval Postgraduate School]. NPS Archive: Calhoun. https://calhoun.nps.edu/handle/10945/56752

Liu, D. (2019). Using Stata 16's lasso features for prediction and inference. Stata Corp.https://www.stata.com/meeting/china19_Shanghai/slides/china19_Liu.pdf

Marine Combat Training. (n.d.). Marine combat training battalion. http://www.trngcmd.marines.mil/Units/West/School-of-Infantry-West/Marine-Combat-Training-Battalion/

Marine Corps Operating Concept 2025. (2016). How an expeditionary force operates in the 21st century. https://www.mcwl.marines.mil/Portals/34/Images/MarineCorpsOperatingConceptSept2016.pdf

MCB Camp Lejeune. (n.d.). *History*. http://www.lejeune.marines.mil/visitorsV/History.aspx

Markström, J. L., & Olsson, C. J. (2013). Countermovement jump peak force relative to body weight and jump height as predictors for sprint running performances: (In) homogeneity of track and field athletes? *The Journal of Strength & Conditioning Research, 27*(4), 944–953. https:// 10.1519/JSC.0b013e318260edad

Nowicki, A. C. (2017). United States Marine Corps basic reconnaissance course: predictors of success [Master's thesis, Naval Postgraduate School]. NPS Archive: Calhoun: http://hdl.handle.net/10945/53025

Pereira, J. M., Basto, M., & Silva, A. F. (2016). The logistic lasso and ridge regression in predicting corporate failure. *Procedia Economics and Finance, 39*, 634–641. https://doi.org/10.1016/S2212-5671(16)30310-0

Pett, M. A., Lackey, N. R., & Sullivan, J. J. (2003). Making sense of factor analysis. Sage Publishing. https://dx.doi.org/10.4135/9781412984898

Shattuck, N. L., Matsangas, P., Mysliwiec, V., & Creamer, J. L. (2020). *The Role of Sleep in Human Performance and Well-Being*. Oxford University Press. NPS Archive: Calhoun: http://hdl.handle.net/10945/65640

SOI-E. (n.d.). https://www.trngcmd.marines.mil/Units/South-Atlantic/SOI-E/Units/Infantry-Training-Battalion/History/

SOI-W. (n.d.). https://www.trngcmd.marines.mil/Units/West/SOI-W/igphoto/200216093

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of Royal Statistics Society*, *58*(1), 267–288.

U.S. Marine Corps. (1997). Warfighting (MCDP1). https://www.marines.mil/Portals/1/Publications/MCDP%201%20Warfighting.pdf

U.S. Marine Corps. (2012). "Every Marine a Rifleman" begins at recruit training. *Marine Corps Training and Education Command*. http://www.tecom.marines.mil/News/NewsArticleDisplay/tabid/5055/Article/528587/every-marine-a-rifleman-begins-at-recruit-training.aspx

U.S. Marine Corps. (2020a). *Force design 2030*. https://www.hqmc.marines.mil/Portals/142/Docs/CMC38%20Force%20Design%202030%20Report%20Phase%20I%20and%20II.pdf?ver=2020-03-26-121328-460

U. S. Marine Corps. (2020b). *Marine rifle squad* (MCIP 3–10). https://www.marines.mil/Portals/1/Publications/MCIP%20310A.4i%20wChg1.pdf?ver=2020-05-22-085455-150

Welsh, T., Alemany, J., Montain, S., Frykman, P., Tuckow, A., Young, A., & Nindl, B. (2008). Effects of intensified military field training on jumping performance. *International Journal of Sports Medicine*, *29*(1), 45–52. https://doi.org/10.1055/s-2007-964970

Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach* (6th ed.). Cengage Learning.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of The Royal Statistical Society Series B-Statistical Methodology*, *67*, 301–320.

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