



NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

**THE U.S. ARMY MEDICAL SERVICE CORPS AREA OF
CONCENTRATION MATCHING SYSTEM**

by

Diego A. Rincon

September 2020

Thesis Advisor:
Second Reader:

Robert L. Bassett
Chad W. Seagren

Approved for public release. Distribution is unlimited.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			<i>Form Approved OMB No. 0704-0188</i>	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.				
1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE September 2020		3. REPORT TYPE AND DATES COVERED Master's thesis
4. TITLE AND SUBTITLE THE U.S. ARMY MEDICAL SERVICE CORPS AREA OF CONCENTRATION MATCHING SYSTEM			5. FUNDING NUMBERS	
6. AUTHOR(S) Diego A. Rincon				
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 purpose of this study is to develop an algorithm-based talent management system for designating midgrade U.S. Army Medical Service Corps (MSC) officers to specialized areas of concentration (AOCs). The AOC matching system (AOC-MS) utilizes machine learning classifiers to predict officer aptitude for specialized AOCs, informing early career MSC recruitment and talent development efforts. AOC-MS then utilizes matching algorithms and integer linear programs to create optimal matches between officers and AOCs that meet the needs of officers and of the MSC. During a pilot program, AOC-MS outperformed the current designation system. This study will serve as the basis for MSC AOC designation reform. Its implications are applicable to any large-scale military talent management problem.				
14. SUBJECT TERMS Medical Service Corps, AOC Designation, Talent Management, job assignment, Job Aptitude Prediction, matching			15. NUMBER OF PAGES 223	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UU	

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release. Distribution is unlimited.

**THE U.S. ARMY MEDICAL SERVICE CORPS AREA OF CONCENTRATION
MATCHING SYSTEM**

Diego A. Rincon
Major, United States Army
BA, Florida International University, 2005
MA, University of Central Florida, 2009

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

**NAVAL POSTGRADUATE SCHOOL
September 2020**

Approved by: Robert L. Bassett
Advisor

Chad W. Seagren
Second Reader

W. Matthew Carlyle
Chair, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The purpose of this study is to develop an algorithm-based talent management system for designating midgrade U.S. Army Medical Service Corps (MSC) officers to specialized areas of concentration (AOCs). The AOC matching system (AOC-MS) utilizes machine learning classifiers to predict officer aptitude for specialized AOCs, informing early career MSC recruitment and talent development efforts. AOC-MS then utilizes matching algorithms and integer linear programs to create optimal matches between officers and AOCs that meet the needs of officers and of the MSC. During a pilot program, AOC-MS outperformed the current designation system. This study will serve as the basis for MSC AOC designation reform. Its implications are applicable to any large-scale military talent management problem.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

I.	THE MEDICAL SERVICE CORPS AREA OF CONCENTRATION DESIGNATION PROBLEM.....	1
A.	INTRODUCTION TO THE AREA OF CONCENTRATION DESIGNATION PROBLEM.....	1
B.	ORGANIZATION AND STRUCTURE OF THE MEDICAL SERVICE CORPS	1
C.	CAREER PROGRESSION OF MAJOR FUNCTIONAL AREA 70 COMMISSIONED OFFICERS.....	4
D.	MAJOR FUNCTIONAL AREA 70 ADMINISTRATIVE AREAS OF CONCENTRATION	9
1.	70A Health Care Administration	9
2.	70C Health Services Comptroller.....	9
3.	70D Health Services Systems Management.....	10
4.	70E Patient Administrator (PAD)	10
5.	70F Healthcare Human Resources Manager	10
6.	70H Health Services Plans, Operations, Intelligence, Security, and Training.....	10
7.	70K Medical Logistics.....	11
E.	THE CURRENT MAJOR FUNCTIONAL AREA 70 AREA OF CONCENTRATION DESIGNATION PROCESS.....	11
II.	MATHEMATICAL APPROACHES FOR SOLVING THE AREA OF CONCENTRATION DESIGNATION PROBLEM.....	17
A.	SCIENTIFIC LITERATURE REVIEW	17
1.	Assignment Problems	17
2.	Machine Learning Models for Talent Management.....	25
B.	ALGORITHM-ENABLED TALENT MANAGEMENT IN THE DEPARTMENT OF DEFENSE	27
1.	Army Functional Area (FA) Designation Process (2001).....	28
2.	U.S. Air Force Academy (USAFA) Air Force Specialty Code (AFSC) Matching Process (2001).....	29
3.	The Sailor Assignment Problem (SAP) (2006)	31
4.	U.S. Army Talent Alignment Process (ATAP) (2019)	32
5.	U.S. Army Talent Based Branching (TBB) (2019).....	33
III.	A SOLUTION TO THE AREA OF CONCENTRATION DESIGNATION PROBLEM.....	35
A.	AOC-MS TALENT DEVELOPMENT PHASE	37

B.	AOC-MS TALENT DEVELOPMENT PHASE	39
C.	AOC-MS TALENT UTILIZATION PHASE	42
IV.	EXPLORATORY ANALYSIS OF AREA OF CONCENTRATION DESIGNATION DATA.....	43
A.	MEDICAL SERVICE CORPS OFFICERS ASSIGNMENT AND EDUCATION RECORDS DATASET	43
1.	Feature Engineering	46
2.	The Area of Concentration Predictor Dataset	53
B.	FISCAL YEAR 16–18 MEDICAL SERVICE CORPS OFFICER AREA OF CONCENTRATION PREFERENCES AND PANEL VOTING RECORDS.....	73
1.	The Area of Concentration Designation Dataset	73
2.	Univariate Exploratory Analysis of the Designation Dataset.....	76
3.	Multivariate Exploratory Analysis of the Area of Concentration Designation Dataset	78
C.	MEDICAL SERVICE CORPS OBJECTIVE FORCE MODEL AOC DESIGNATION TARGETS DATASET	91
D.	MEASURES OF EFFECTIVENESS FOR THE AREA OF CONCENTRATION DESIGNATION PROCESS.....	96
V.	PREDICTING AREA OF CONCENTRATION APTITUDE USING MACHINE LEARNING CLASSIFIERS	101
A.	TRAINING AND TESTING DATA	101
B.	MULTINOMIAL LOGISTIC REGRESSION CLASSIFIER	103
1.	Model Selection	103
2.	Model Testing.....	109
C.	RANDOM FOREST CLASSIFIER	112
1.	Model Selection	113
2.	Model Testing.....	116
D.	SUPPORT VECTOR MACHINE CLASSIFIER.....	119
1.	Model Selection	120
2.	Model Testing.....	124
E.	FULLY CONNECTED NEURAL NETWORK CLASSIFIER	126
1.	Model Selection	127
2.	Model Testing.....	130
F.	MODEL RESULTS AND COMPARISON.....	133

VI.	AREA OF CONCENTRATION DESIGNATION ALGORITHMS	141
A.	THE AREA OF CONCENTRATION MATCHING SYSTEM PILOT PROGRAM	141
B.	TALENT MATCHING ALGORITHM (TMA) VARIANT SELECTION	146
C.	TALENT MATCHING ALGORITHM (TMA) STABLE MARRIAGE VARIANT	148
	1. The AOC-MS Target-Setting Algorithm (TSA).....	149
	2. Strictly Ordered Preference Lists.....	156
	3. AOC-MS TMA Stable Marriage Variant Implementation ...	158
	4. AOC-MS TMA Stable Marriage Variant Results	163
D.	TALENT MANAGEMENT ALGORITHM INTEGER LINEAR PROGRAM (ILP)	166
	1. Integer Linear Program Adaptations	166
	2. Comparing the AOC-MS ILP with the Current Process	171
	3. The AOC-MS Self-Target Setting and Designation Algorithm (STSDA)	174
E.	TALENT MATCHING ALGORITHM VARIANT COMPARISON.....	180
VII.	RECOMMENDATIONS.....	183
A.	RECOMMENDATIONS FOR THE MEDICAL SERVICE CORPS	184
B.	FUTURE WORK	185
	LIST OF REFERENCES	187
	INITIAL DISTRIBUTION LIST	197

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

Figure 1.	AMEDD Organization. Adapted from HQDA (2019).	2
Figure 2.	MFA 70 Specialties. Adapted from HQDA (2019).	3
Figure 3.	U.S. Army Ranks. Source: U.S. Army (n.d.).	4
Figure 4.	U.S. Army Echelons of Command. Source: U.S. Army (n.d.).	5
Figure 5.	AOCs 67J and 70B Career Timeline. Adapted from U.S. Army (n.d.).	8
Figure 6.	Current 70B/67J AOC Designation Process. Adapted from U.S. Army (n.d.).	12
Figure 7.	Word picture vote guidelines. Source: Department of the Army Secretariat for Selection Boards (DASSB) (2014).	14
Figure 8.	Bipartite Graph.....	18
Figure 9.	Generic Assignment Problem Network. Adapted from Google (2019).	20
Figure 10.	Sample Stable Matching Problem with Preference Matrices. Adapted from Cui (2015).	22
Figure 11.	Stable Matching Algorithm Results. Adapted from Cui (2015).	23
Figure 12.	U.S. Army Career Field Designation Model (CFD). Source: Shrimpton and Newman (2005).	29
Figure 13.	USAFA Career Field Selection Model. Source: Armacost and Lowe (2005).	30
Figure 14.	MyNavy Assignment Workflow. Source: Garrett et al. (2005).	31
Figure 15.	AOC Matching System (AOC-MS) Overview. Adapted from U.S. Army (n.d.).	36
Figure 16.	AOC-MS Talent Development Phase Overview. Adapted from U.S. Army (n.d.).	38
Figure 17.	AOC-MS Talent Matching Phase Overview. Adapted from U.S. ARMY (n.d.).	41
Figure 18.	Sample Officer Record Brief. Adapted from Peters (2020).	44
Figure 19.	Sample officer Record from Records Dataset	45

Figure 20.	Histogram of Previous Duty AOC Entry Count, Non-MFA70	47
Figure 21.	Histogram of Previous Duty AOC Entry Count, MFA70	47
Figure 22.	Histograms of Frequency Count for Unique Educational Entries in Records Dataset	49
Figure 23.	Types of Military Courses in Records Dataset	51
Figure 24.	Types of Civilian Degrees in Records Dataset	52
Figure 25.	Proportion of AOC Designations in the AOC Predictor Dataset.....	56
Figure 26.	Histogram and Kernel Density Estimate (KDE) of Non-AOC Specific Predictor Variables.	59
Figure 27.	Histogram and Kernel Density Estimate (KDE) of AOC-Specific Assignment History Predictors	60
Figure 28.	Histogram of AOC-Specific Military Education Predictors	61
Figure 29.	Histogram of AOC-Specific civilian Education Predictors	62
Figure 30.	Tri-variate Plots of AOC-specific Predictors per AOC	64
Figure 31.	Percentage of Variance Explained by Principal Component.....	66
Figure 32.	PCA Contributions per Variable for Principal Components 1–4.....	68
Figure 33.	PCA Individual Visualizations for Components 1 and 2.....	70
Figure 34.	Three-Dimensional Mapping of AOC Predictor Dataset.....	71
Figure 35.	Three-Dimensional Mapping of AOC Predictor Dataset with 70F, 70H and 70K Entries Removed	72
Figure 36.	t-SNE Three-Dimensional Visualization of AOC Predictor Dataset.....	72
Figure 37.	Designation Dataset Missing Data Rates	75
Figure 38.	Officer AOC Satisfaction Levels Excluding Failed Opt-ins.....	80
Figure 39.	Preference Level Attained by Designating Officers per AOC.....	81
Figure 40.	Level of Retention per AOC Assignment.....	82
Figure 41.	AOC Producing Course Attendance vs. AOC Designation (AOCs 70D, 70E, 70F, 70H and 70K)	83
Figure 42.	AOC Producing LTHET Selections vs. Final AOC Designation (AOCs 70A and 70C).....	84

Figure 43.	LTHET Status per AOC Designation Preference Level	85
Figure 44.	Rate of Gaining Consultant Opposition for Officers Designating into AOC	86
Figure 45.	Retention Status as a Function of Gaining Consultant Opposition	87
Figure 46.	Distribution of Word Picture Scores per Final AOC Designation.....	88
Figure 47.	Histogram and Kernel Density Estimate of Panel Scores by Level of Preference Satisfaction.....	89
Figure 48.	Kernel Density Estimate of Panel Score Distribution by Retention Status.....	90
Figure 49.	70B Inventory Difference per PYG	93
Figure 50.	AOC Designation Shortages for PYGs 2008–2011.....	94
Figure 51.	Missed and Overshot OFM Targets by PYG.....	95
Figure 52.	AOC Overages and Shortages for PYGs 2008–2011	96
Figure 53.	Sample Variable Transformation for AOC Predictor Matrix	102
Figure 54.	Predictor Effect Plots for Multinomial Logistic Classifier (00E Months through 70D Civ Ed)	106
Figure 55.	Predictor Effect Plots for Multinomial Logistic Classifier (70D Mil Ed through 70F Civ Ed).....	107
Figure 56.	Predictor Effect Plots for Multinomial Logistic Classifier (70F Mil Ed through 70K Months)	108
Figure 57.	Confusion Matrix Between Multinomial Logistic Classifier Predictions and Test Set AOC Assignments.....	109
Figure 58.	Receiver Operator Characteristic Curves for Multinomial Logistic Classifier	110
Figure 59.	Predictor Importance for Random Forest Classifier	114
Figure 60.	Two-Dimensional Representation of Random Forest Predictor Space ...	116
Figure 61.	Confusion Matrix Between Random Forest Classifier Predictions and Test Set AOC Assignments.....	117
Figure 62.	Receiver Operator Characteristic Curves for Random Forest Classifier	118

Figure 63.	Predictor Coefficient Magnitude for Support Vector Machine Classifier	122
Figure 64.	Two-Dimensional Representation of Support Vector Machine Predictor Space	123
Figure 65.	Confusion Matrix Between Support Vector Machine Classifier Predictions and Test Set AOC Assignments.....	124
Figure 66.	Receiver Operator Characteristic Curves for Support Vector Machine Classifier	125
Figure 67.	Architecture and Hyperparameters of Fully Connected Neural Network Classifier. Adapted from LeNail (2019).	128
Figure 68.	Importance for Fully Connected Neural Network Classifier Predictors	129
Figure 69.	Confusion Matrix Between Fully Connected Neural Network Classifier Predictions and Test Set AOC Assignments.	131
Figure 70.	Receiver Operator Curves for Fully Connected Neural Network Classifier	132
Figure 71.	Distribution of AOC Assignments and NN AOC Predictions.....	135
Figure 72.	Cramer's V Statistics for U and Xtest Model Predictions	138
Figure 73.	Three-Dimensional Mapping of Xtest	139
Figure 74.	Three-Dimensional Mapping of U	139
Figure 75.	AOC-MS Pilot Program Missing Data	144
Figure 76.	Discrepancy Between Preferences, Targets and Designations for AOC-MS Pilot Program.....	148
Figure 77.	Percentage Fill of OFM Targets for FY2020 AOC Designation.....	154
Figure 78.	Histogram of AOC Satisfaction Coefficients	170
Figure 79.	Histogram of Maximum AOC Satisfaction Coefficients per Officer	171

LIST OF TABLES

Table 1.	Administration-Related Civilian Degree Unique Entries in Records Dataset.....	50
Table 2.	AOC-Specific Numerical Scores Assigned to Educational Entries.....	52
Table 3.	Civilian Degree Level Numerical Weights.....	53
Table 4.	AOC Predictor Dataset Sample	54
Table 5.	Multinomial Distribution of AOC Assignments.....	56
Table 6.	AOC Predictor Dataset Descriptive Statistics.....	57
Table 7.	Principal Components of AOC Predictor Dataset	65
Table 8.	Exploratory Analysis for AOC Designation Panel Score	76
Table 9.	Exploratory Analysis for Key Categorical Variables in AOC Designation	77
Table 10.	Officer AOC Satisfaction Levels	78
Table 11.	Officer AOC Preferences as a Proportion of AOC Manpower Allocations	81
Table 12.	PYG2008-2012 Objective Force Model Snapshot	92
Table 13.	Proposed Measures of Effectiveness for AOC Designation Process.....	98
Table 14.	Sample Records from Predictor Matrix X.....	102
Table 15.	Regression Coefficients for Multinomial Logistic Classifier	105
Table 16.	Multinomial Classifier Test Set Accuracy	110
Table 17.	Sensitivity and Specificity for Multinomial Logistic Classifier	111
Table 18.	Random Forest Classifier Configuration	113
Table 19.	Random Forest Classifier Test Set Accuracy	117
Table 20.	Sensitivity and Specificity for Random Forest Classifier.....	119
Table 21.	Support Vector Machine Classification Configuration.....	120
Table 22.	Support Vector Machine Classifier Test Set Accuracy	125
Table 23.	Sensitivity and Specificity for Support Vector Machine Classifier.....	126

Table 24.	Fully Connected Neural Network Test Set Accuracy	131
Table 25.	Sensitivity and Specificity for Fully Connected Neural Network Classifier	132
Table 26.	Comparison of Classifier Performance	133
Table 27.	Pairwise Performance Comparisons Between Machine Learning Classifiers.....	134
Table 28.	Pearson GOF Test for Classifier Predictions on Unlabeled Dataset.....	136
Table 29.	Cramer’s V for U AOC Predictions	137
Table 30.	Cramer’s V for Xtest AOC Predictions	137
Table 31.	FY2020 AOC Designation Process Results.....	143
Table 32.	Sample Officer Preferences for FY2020 AOC-MS Pilot Program.....	145
Table 33.	Sample AOC Preferences for FY2020 AOC-MS Pilot Program.....	146
Table 34.	FY2020 AOC Designation Panel and AOC-MS TSA Raw Targets.....	155
Table 35.	FY2020 AOC Designation Panel Allocation Fill Percentages	155
Table 36.	AOC-MS TMA Fill Percentages	156
Table 37.	Sample Records from officer Preference Table OP.....	157
Table 38.	Sample Records from AOC Preference Table AP.	158
Table 39.	MOE Comparison Between AOC-MS TMA and Panel Designation.....	164
Table 40.	MOE Comparison for AOC-MS TMA Across PYG Fill Horizons.....	165
Table 41.	Officer Satisfaction Matrix OS Sample Records.....	167
Table 42.	AOC Satisfaction Matrix AS Sample Records.....	168
Table 43.	Satisfaction Matrix S Sample Records.....	169
Table 44.	MOE Comparison Between AOC-MS Comparison ILP and Panel Designation	174
Table 45.	AOC-MS STSDA Fixed Parameter Values	179
Table 46.	MOE Comparison for AOC-MS STSDA Across Three Parameter Configurations.....	180

LIST OF ACRONYMS AND ABBREVIATIONS

AFCS	Active Federal Commissioned Service
AMEDD	U.S. Army Medical Department
AOC	Area of Concentration
AOC-MS	AOC Matching System
AOC-MS STSDA	AOC-MS Self-Target Setting and Designation Algorithm
AOC-MS TDA	AOC-MS Talent Development Algorithm
AOC-MS TMA	AOC-MS Talent Matching Algorithm
AOC-MS TSA	AOC-MS Target-Setting Algorithm
APPD	AMEDD Personnel Proponent Directorate
FNN	Feed-forward Neural Network
HRC	U.S. Army Human Resources Command
ILP	Integer Linear Program
LTHET	Long Term Health Education Training
MFA	Major Functional Area
MLC	Multinomial Logistic Classifier
MOE	Measure of Effectiveness
MSC	Medical Service Corps
OCMSC	Office of the Chief, Medical Service Corps
OFM	Objective Force Model
PYG	Promotion Year Group
RF	Random Forest
ROC	Receiver Operator Characteristic
SMA	Stable Matching Algorithm

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

The U.S. Army Medical Service Corps (MSC) designates nearly one 100 mid-grade generalist healthcare administrator captains to specialized areas of concentration (AOCs) on an annual basis. The current MSC AOC designation process is antiquated, time-consuming, and prone to human error and bias. Furthermore, it fails to account for the force structure needs of the MSC and for the preferences of AOC leaders.

In this study, we introduce the MSC AOC matching system (AOC-MS) as a solution to the MSC AOC designation problem. AOC-MS is a three-phase, algorithm-based talent management process designed to identify, develop, assign, and retain MSC talent throughout a six-year life cycle. AOC-MS considers officer designation preferences and experiences along with the needs of the MSC to create the best possible matches between officers and AOCs. AOC-MS is a key MSC effort, underpinning two of its fiscal year 2020–2021 strategic plan initiatives.

AOC-MS phase I, the talent development phase, takes place on an annual basis between officers' 5th and 8th year of commissioned service. AOC-MS phase I uses four machine-learning classifiers to parse through officer assignment and educational records in order to recommend possible matches between officers and AOCs. Phase I of AOC-MS enables AOCs to recruit, develop, and shape their future teams starting three years ahead of AOC designation.

AOC-MS phase II, the talent matching phase, takes place during officers' 8th year of commissioned service. Phase II of AOC-MS uses a choice of a Stable Marriage Algorithm (SMA) and an Integer linear program (ILP) to find the best possible set of matches between officers and AOCs based on their mutual preferences and force structure constraints. The models provide the MSC with the flexibility to balance tradeoffs between officer preference satisfaction and MSC manpower requirements.

AOC-MS phase III, the talent utilization phase, takes place between officers' 8th and 11th year of commissioned service. Phase III of AOC-MS tracks officers post-AOC designation and analyzes continuation levels against AOC designation measures of

effectiveness (MOEs) in order to assess the relationship between AOC-MS talent development and management efforts to long-term officer retention and satisfaction and to improve the AOC-MS process and algorithms.

The MSC chief approved the conduct of a pilot program to evaluate AOC-MS's performance against current AOC designation process using six different MOEs. In the pilot program, AOC-MS outperformed the current AOC designation process in four MOEs while tying in the remaining two.

AOC-MS provides the MSC with a clear alternative to the current AOC designation system. AOC-MS maximizes value to officers, the MSC and the Army. It provides the MSC with a flexible and adaptable process which enables the combat readiness of the Army's future medical force.

ACKNOWLEDGMENTS

Theses—particularly those written during a pandemic while shelter-in-place orders are in effect—are solitary endeavors. Despite the seemingly lonely nature of the research and writing process, I’ve enjoyed the support and guidance of a community of peers and leaders at the Naval Postgraduate School and the Army MSC without whom I could have never completed this study.

I would like to thank Major Joseph Dywan, Major Joshua Perry, Captain Michael Lugo, Lieutenant Martin Nganga, Lieutenant Joseph Vranich, Lieutenant Olivia Wittman, and many other classmates for helping me get through the Operations Research curriculum, explaining concepts I failed to grasp and for pushing me every day to become a better student and analyst.

I am also indebted to the faculty of the Department of Operations Research. I would like to thank Dr. Robert Koyak, Dr. Lyn Whitaker, and Colonel Lamar Adams for teaching me concepts and techniques sustaining this study and for allowing me to use my thesis data for their class assignments.

I am especially thankful to Provost Robert Dell for his guidance, mentorship, and support for this project, particularly in its early stages and despite his promotion to Dean of the Graduate School of Operations and Information Sciences. His patience and input were key in the development and implementation of the integer linear program formulations used to solve the AOC designation problem. His support in securing funding for experience tour travel afforded me the opportunity to brief senior MSC leaders and collect data without which the study would not have been possible.

I owe a debt of gratitude to my advisor, Dr. Robert Bassett. He provided me with the guidance and freedom to pursue this study on my own terms while steering me away from pitfalls and rabbit holes. Dr. Bassett spent countless hours explaining mathematical concepts, helping me debug my code, and listening to my thoughts and ideas. I appreciated his consistent interest and enthusiasm for the study during the five quarters we worked together and for his ability to keep the study grounded in mathematical and academic rigor.

I'm also thankful for my second reader, Dr. Chad Seagren, who provided me with insightful feedback to enhance the cohesiveness of this study.

I would also like to thank the many Army leaders who made this study possible. I want to extend my gratitude to Colonels Michael Elliott and Yolonda Summons at the Office of the Chief, MSC, for believing that this study could be of benefit to the MSC and for supporting my requests for time, information, and data.

I want to thank Colonel Monica Douglas, Major Kellie Mosely and Ms. Maureen Martinez at the AMEDD Personnel Proponent Directorate for providing me with data and allowing me to examine their models and processes in order to perform this study. I'm also thankful to Lieutenant Colonel Riley Post with the Department of the Army G1 and Mr. Brian Davis at Human Resources Command for helping me understand the algorithmic foundations of the Army branching and assignment processes.

I'm thankful to Colonel Stacy Holman, Lieutenant Colonel Anette Carter, and Major Meredith Brown at the Army Human Resources Command Health Services Division for hosting me during my experience tour, for providing me with the necessary data to conduct this study, and for affording me the opportunity to conduct a pilot program in order to validate the concepts behind this study.

I'm especially thankful for Lieutenant Colonel Anthony Marinos and Mr. Pat McMurry for mentoring me in my journey to become an AMEDD Operations Analyst and most of all for suggesting I take on the AOC designation problem as a thesis project. I would not be graduating from the Naval Postgraduate School or finishing my thesis without their friendship, encouragement, and direction.

Last but not least, I would like to thank those who have sacrificed the most to enable me to complete this study. My sons, Camilo and Mateo, have sacrificed nights and weekends with limited access to their father as I worked to finish this project. I thank them for their patience and understanding and I hope that despite the long hours of work, I was present enough as a father to make a positive difference in their lives. I am most thankful of all to my wife, Kathryn, for always supporting me and always picking up the slack on our household and parenting responsibilities when I had to work. Throughout years of hard

academic work preceded by years of hard military work, she has stood by my side without complaint. I hope our next adventure as I take on the duties of the MSC personnel proponent gives me more time to spend with them and that this study—the product of our collective sacrifice—better the lives of MSC officers for years to come.

THIS PAGE INTENTIONALLY LEFT BLANK

I. THE MEDICAL SERVICE CORPS AREA OF CONCENTRATION DESIGNATION PROBLEM

This chapter provides necessary background and context for understanding the Area of Concentration (AOC) designation problem. Section A introduces this study's research problem. Section B introduces the Army medical department (AMEDD) and the role of the Army medical service corps (MSC) officer. Section C outlines the life cycle model for MSC officers, examining how AOC designation fits in an MSC officer's career path. Section D describes the seven administrative specialties available to MSC officers while Section E depicts the current AOC designation process in detail.

A. INTRODUCTION TO THE AREA OF CONCENTRATION DESIGNATION PROBLEM

As part of fiscal year 2020- 2021 MSC reform strategy, the MSC requested a quantitative study to assess its process for designating healthcare administration generalists optimally into seven administrative AOCs. The MSC asked for recommendations for updating the manually executed human-dependent designation process with an algorithm-based, automated process that fairly honors officer preferences and the needs of the MSC. We accomplished this and more, providing the MSC with a talent management system which uses machine learning classifiers, stable matching algorithms and mathematical optimization to manage MSC officers through a six-year life-cycle centered around AOC designation while empowering officers and MSC leaders to build effective medical teams supporting the Army's lethality.

B. ORGANIZATION AND STRUCTURE OF THE MEDICAL SERVICE CORPS

Organized under the Army surgeon general's (TSG) leadership and numbering 4,202, the Army MSC is the second largest corps of the Army medical department (Headquarters Department of the Army [HQDA] 2018). Figure 1 depicts the organization of the AMEDD. Each AMEDD corps is led by a corps chief, typically the most senior member of the corps.

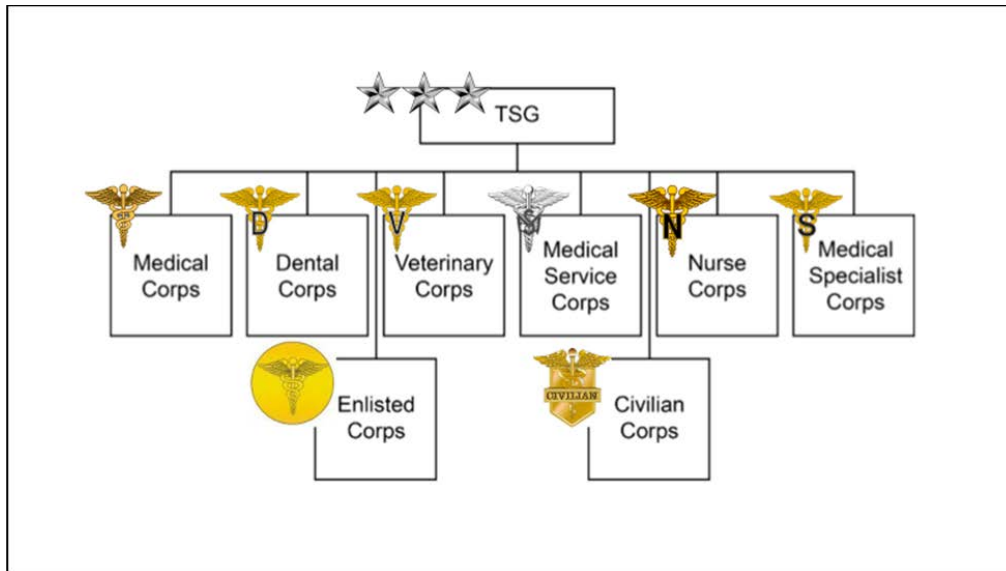


Figure 1. AMEDD Organization. Adapted from HQDA (2019).

In addition to appointing corps chiefs, TSG assigns “consultants to provide technical and professional expertise in a wide variety of healthcare disciplines” (HQDA 2019). Consultants are typically one of the most senior officers in their AOC and serve as subject-matter experts (SMEs) for their specialty, responsible for advising the Corps chief on the management, training and employment of AOC personnel. Additionally, consultants serve as community mentors for their specific AOC. In current practice, consultants have considerable influence on the assignment, educational and leadership opportunities available to officers within their AOC (HQDA 2019).

The MSC “is comprised of a wide diversity of health care administrative and scientific specialties ranging from the management and support of the Army’s health services system to direct patient care” (HQDA 2019). The MSC is organized into four medical functional areas (MFAs): administrative health services (MFA 70), medical allied sciences (MFA 71), preventive medicine sciences (MFA 72), and behavioral health sciences (MFA 73) in addition to three stand-alone Areas of Concentration (AOCs) for pharmacists, optometrists and podiatrists respectively (HQDA 2019).

MFA 70, the largest of the MSC MFAs, provides administrative health services and MEDEVAC support to the U.S. Army in garrison and deployed environments throughout the

world. The Army established the early precursors of today's MSC in the early 1800s when the need for the adept management of battlefield evacuation, health records and medical stores without burdening military physicians with non-patient care duties became apparent (Ginn 2008). The Army formally established the Medical Administrative Corps, the direct predecessor of today's MFA 70, shortly after World War I (Ginn 2008), while Congress created the MSC in 1947 (U.S. Army Medical Department Office of Medical History [AMEDD OMH] 2009). As early as 1961, the MSC had already designated officers into nine specialized administrative career fields, similar to today's AOCs: "supply, comptroller, personnel, registrar, hospital administration, medical intelligence, operations and training, aviation, warrant officer, and a noncareer field that grouped various specialties together" (Ginn 2008, p. 283). Today, the approximately 2,800 commissioned MFA 70 officers in the Army's active component (AC) are organized into the nine specialties depicted in Figure 2. The two AOCs denoted with a yellow star (AOC 70B and 67J) are entry-level specialties, while AOCs marked with a blue arrow (70A, 70C, 70D, 70E, 70F, 70H and 70K) are specialized career fields into which officers transition no later than their eighth year of commissioned service.

Medical Service Corps specialties	
MFA/AOC/MOS MFA 70	Title
	Administrative Health Services
→ 70A	Health Care Administration
★ 70B	Health Services Administration
→ 70C	Health Services Comptroller
→ 70D	Health Service System Management
→ 70E	Patient Administration
→ 70F	Health Services Human Resources
→ 70H	Health Services Plans, Operations, Intelligence, Security and Training
→ 70K	Health Services Materiel
★ 67J	Aeromedical Evacuation

Figure 2. MFA 70 Specialties. Adapted from HQDA (2019).

C. CAREER PROGRESSION OF MAJOR FUNCTIONAL AREA 70 COMMISSIONED OFFICERS

Commissioned MSC officers in MFA 70 follow the same general rank progression of other Army commissioned officers. Figure 3 shows the rank insignia, abbreviations and paygrades for Army commissioned officer ranks. The highest rank an officer in the AMEDD can attain is lieutenant general (LTG).

O-1	O-2	O-3	O-4	O-5	O-6	O-7	O-8	O-9	O-10	SPECIAL
ARMY										
										
Second Lieutenant (2LT)	First Lieutenant (1LT)	Captain (CPT)	Major (MAJ)	Lieutenant Colonel (LTC)	Colonel (COL)	Brigadier General (BG)	Major General (MG)	Lieutenant General (LTG)	General (GEN)	General of the Army (GA)

Figure 3. U.S. Army Ranks. Source: U.S. Army (n.d.).

The Army manages active component (AC) officers in annual cohorts based roughly on the fiscal year (FY) of their commissioning as second lieutenants (2LTs). These cohorts are called promotion year groups (PYGs) and determine when officers are considered for promotion, key assignments and other career milestones. The MFA 70 AOC designation process considers officers for designation based on their PYG.

Commissioned MFA 70 officers typically enter the active component upon commissioning as a 2LT from a traditional commissioning source (USMA, ROTC or OCS). MFA 70 officers are initially commissioned into either AOC 70B or AOC 67J. 70Bs serve as healthcare administration generalists while 67Js serve as MEDEVAC aviators. 70Bs make up 35% of AC MFA 70 inventory while 67Js take up approximately 11%. (AMEDD Personnel Proponent Directorate [APPD] 2020). Officers spend the first six to eight years of their career building a foundation of tactical and leadership knowledge while developing an understanding of the seven specialized administrative AOCs outlined in

Figure 2. Before promotion to major (MAJ), all MFA 70 officers designate into a specialized AOC.

Upon graduation from the AMEDD basic officer leadership Course (BOLC), newly commissioned MFA 70 2LTs perform their first assignment in an MTOE (modified table of organization and equipment) unit. MTOE units are operational units with wartime missions such as brigade combat teams (BCTs) or medical brigades while table of distributions and allowances (TDA) units are non-deployable units responsible for supporting the operational Army. Figure 4 provides a general reference of the hierarchy of U.S. Army MTOE units, along with the typical rank of the unit commander and the number of Soldiers usually encountered at each level of command. While the names of medical TDA units are different from those of MTOE units, the rank of the commander and number of Soldiers supervised are generally the same between MTOE and TDA units.

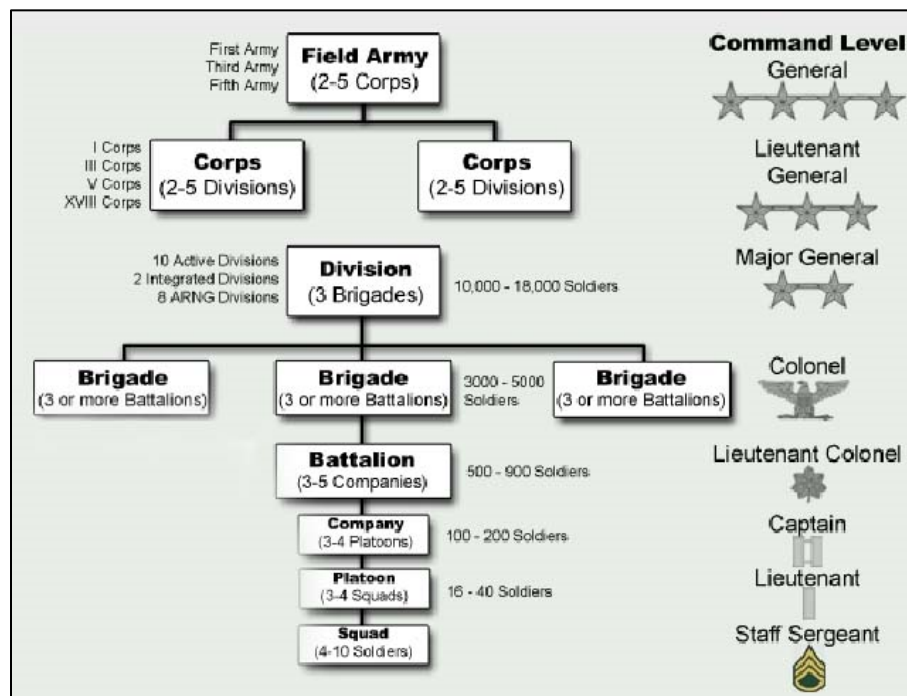


Figure 4. U.S. Army Echelons of Command. Source: U.S. Army (n.d.).

The keystone of a 70B lieutenant's first assignment is to serve as a medical platoon leader, in charge of a small tactical clinic or a medical team which is part of a larger unit.

After completing their platoon leader tour, 70Bs, now promoted to first lieutenant will often serve on a battalion staff, assisting the commander in managing a staff function, or as a medical company executive officer, second in command to the company commander. Most 70Bs, particularly those not serving in medical brigades, will have little exposure to the seven specialized AOCs depicted in Figure 2 while at their first duty station.

After three to four years on station, 70Bs will usually perform a permanent change of station (PCS) and attend the AMEDD captain's career course, which will prepare them to serve as staff officers in higher echelons of command and to serve as company commanders. While at their second duty station, 70B captains are encouraged to compete for a 12 to 24-month company command tour supervising 100 to 200 Soldiers in an MTOE or TDA unit. Officers then seek follow-on assignments in a medical brigade or fixed medical treatment facility in order to be gain experience in at least one of the seven specialized AOCs prior to mandatory designation. No later than their eighth year of commissioned service, 70B officers designate into one of the seven specialized AOCs. (HQDA 2019). The average 70B has the opportunity to work in a single specialized AOC prior to designation.

The pre-designation career path of 67J officers is similar to that of 70Bs with some key differences. First, lieutenants usually serve as aviators in general support aviation battalion air ambulance companies and do not become platoon leaders until they reach the rank of captain. 67J officers often have minimal exposure to medical treatment facilities and to the other MFA 70 AOCs until later in their career. Additionally, while 67Js may transition to one of the seven specialized AOCs, they may choose to serve as medical aviators for their entire career; however, they must be designated into a secondary specialized AOC no later than their eighth year of service (HQDA 2019).

Concurrently with AOC designation, 70B and 67J CPTs who have completed the AMEDD captain's career course have the opportunity to apply for long term health education training (LTHET). LTHET provides AMEDD officers with "full-time, DOD-subsidized (military-sponsored) health or health-related education or training in a military or civilian facility of 26 weeks or more" (HQDA 2007). While all AOCs encourage LTHET attendance and the attainment of a specialty-related graduate degree, two specialized AOCs

require it as an entry-level qualification (70A and 70C). LTHET selections have significant influence in the AOC designation process.

After AOC designation, MFA 70 officers are expected to serve the rest of their careers in their designated specialty and are seldom allowed to change AOCs. Though all MFA 70 officers are allowed to serve in senior generalist positions after designation, they are seldom allowed to work in one of the other administrative specialties. The typical career progression of 70B and 67J officers is outlined in Figure 5.

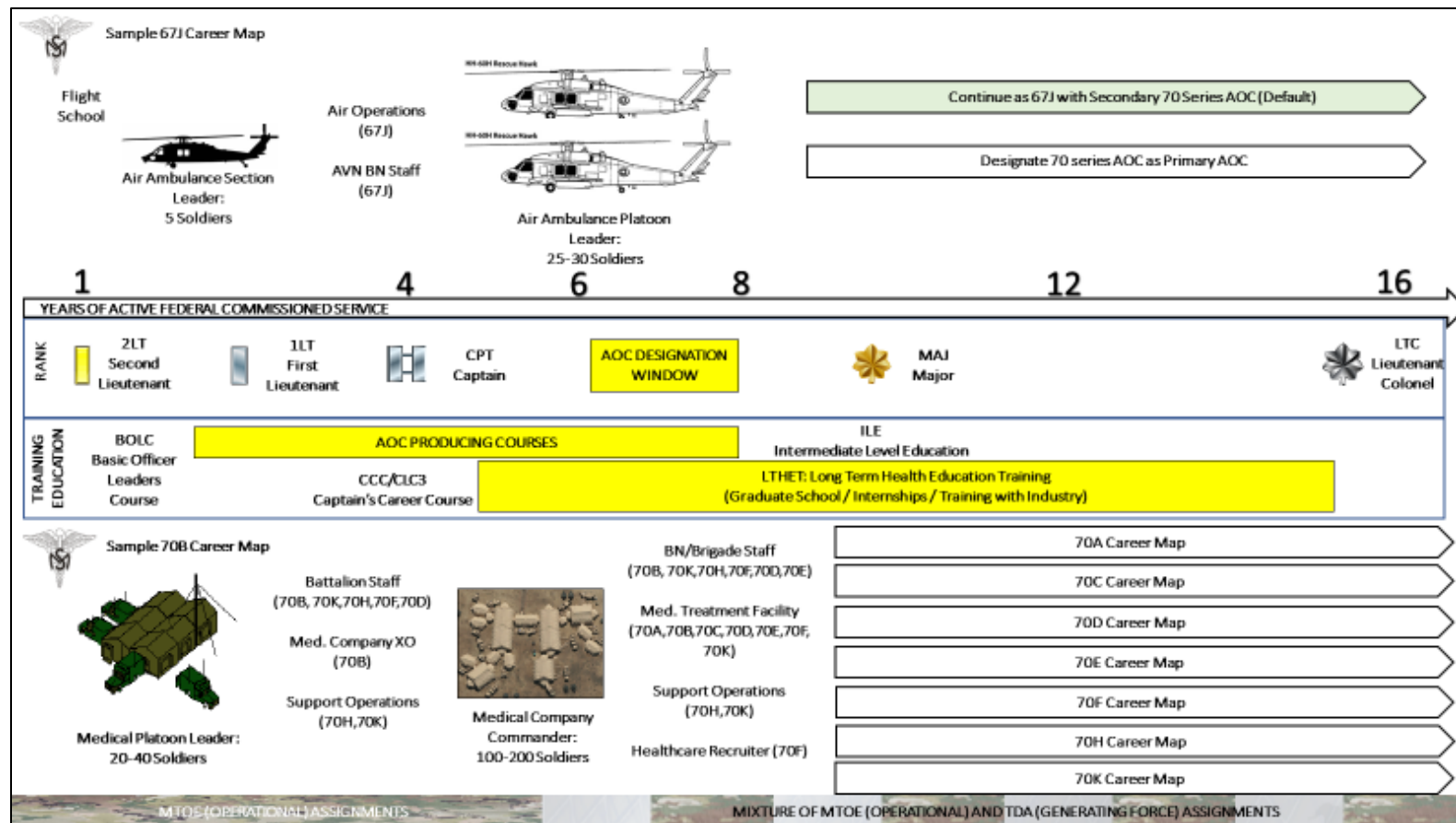


Figure 5. AOCs 67J and 70B Career Timeline. Adapted from U.S. Army (n.d.).

D. MAJOR FUNCTIONAL AREA 70 ADMINISTRATIVE AREAS OF CONCENTRATION

The seven administrative AOCs generally represent the healthcare administration competencies found in the civilian sector as described in the journal of healthcare management (Stefl 2008). Department of the Army Pamphlet 600-4 and its associated online smartbook outline the knowledge, skills and attributes (KSAs) required of MSC officers in each AOC. (HQDA 2019). All seven AOCs require officers to have completed AMEDD BOLC, CCC and AOC-specific educational requirements prior to designation. This section introduces the duties and responsibilities for each AOC along with relevant entry requirements.

1. 70A Health Care Administration

70A officers serve as hospital administrators in fixed healthcare facilities, overseeing clinical operations and accreditation standards. 70As function as integrators of healthcare delivery and administration, often serving as deputy commanders for administration (DCAs). While the AMEDD offers multiple courses in practice management and healthcare administration, officers must attain a graduate degree in business or healthcare Administration in order to be designated into AOC 70A (HQDA 2019). 70As make up approximately 5% of the AC MFA 70 inventory and are exclusively found in TDA units (APPD 2019a).

2. 70C Health Services Comptroller

70C officers are financial managers of the AMEDD. They typically serve as chief financial officers (CFOs) for MTFs and oversee payroll, accounting, commercial accounts, vouchers and claims. 70Cs manage organizational budgets and oversee the collection of funds owed to MTFs. In addition to Army finance and budget courses, 70Cs are required to have a graduate degree in healthcare comptrollership, business administration or healthcare administration (HQDA 2019). 70Cs are nearly 4% of AC MFA 70 inventory and are almost only found in TDA units (APPD 2019a).

3. 70D Health Services Systems Management

70Ds serve as AMEDD's information and technology officers. They serve as MTF chief information officer (CIO) and function as SMEs for all medical information systems. In order to be designated as a 70D, officers must complete the AMEDD medical information management course. (HQDA 2019). 70Ds are also approximately 4% of AC MFA 70 inventory, serving in both TDA and MTOE units (APPD 2019a).

4. 70E Patient Administrator (PAD)

70Es are the AMEDD's patient managers. In hospital settings PAD officers oversee patient tracking and records management; while in deployed settings, 70Es serve as medical regulators, managing the flow of patients through the military health system (MHS). Officers must complete the AMEDD patient administration course in order to be designated as a 70E. (HQDA 2019). 70Es are the smallest specialized AOC, making up nearly 2% of the AC MFA 70 inventory (APPD 2019a).

5. 70F Healthcare Human Resources Manager

70Fs serve as the AMEDD's human capital managers. They serve in fixed facilities and deployed environments as chiefs of human resources (HR), overseeing all aspects of military and civilian HR managers. 70Fs manage all aspects of personnel life cycle management including recruitment, assignments, promotions and retirements. Officers must graduate from the AMEDD healthcare human resources manager's (HHRM) course in order to be designated as a 70F. (HQDA 2019). 70Fs are about 5% of the AC MFA 70 inventory (APPD 2019a).

6. 70H Health Services Plans, Operations, Intelligence, Security, and Training

70Hs serve as the AMEDD's medical operations planners. 70Hs are mostly found in MTOE units serving as medical planners, coordinating the availability of deployed healthcare assets in support of operational Army units. In TDA units, 70Hs often serve as emergency response managers, security, training and non-clinical operations management. In order to be designated as a 70H, officers must attend the AMEDD health services plans,

operations, intelligence, security and training course (HQDA 2019). 70H is the largest of the specialized AOCs, taking up 18% of the MFA 70 inventory (APPD 2019a). 70H is most analogous to 70B work and most 70Bs will be exposed to 70H work prior to designation.

7. 70K Medical Logistics

70Ks are the AMEDD's logisticians, responsible for the management of medical supplies, medicine stocks and biomedical equipment maintenance. 70Ks manage all aspects of medical supply and logistics including the construction and management of healthcare facilities and the acquisition of medical technologies. In order to be awarded AOC 70K, officers must complete the medical logistics management course (HQDA 2019). 70K is the second largest of the specialized AOCs, making up 13% of the AC MFA 70 inventory (APPD 2019a). 70Ks serve in both MTOE and TDA settings. 70Bs are also often exposed to 70K work.

Biomedical maintenance warrant officers make up 4% of the MFA 70 inventory. Warrant officers are governed by a different set of laws, regulations and directives (APPD 2019a). As such, MFA 70 warrant officers do not participate in the AOC designation process.

E. THE CURRENT MAJOR FUNCTIONAL AREA 70 AREA OF CONCENTRATION DESIGNATION PROCESS

The MSC chief, typically a brigadier general (BG) or major general (MG) directs the MFA 70 AOC designation process. MSC Policy Memorandums 19-002 (Office of the Chief, MSC [OCMSC] 2019a) and 19-003 (OCMSC 2019b) outline the procedures, timelines and responsibilities for 70B and 67J AOC designation respectively. Figure 6 summarizes the current AOC designation process.

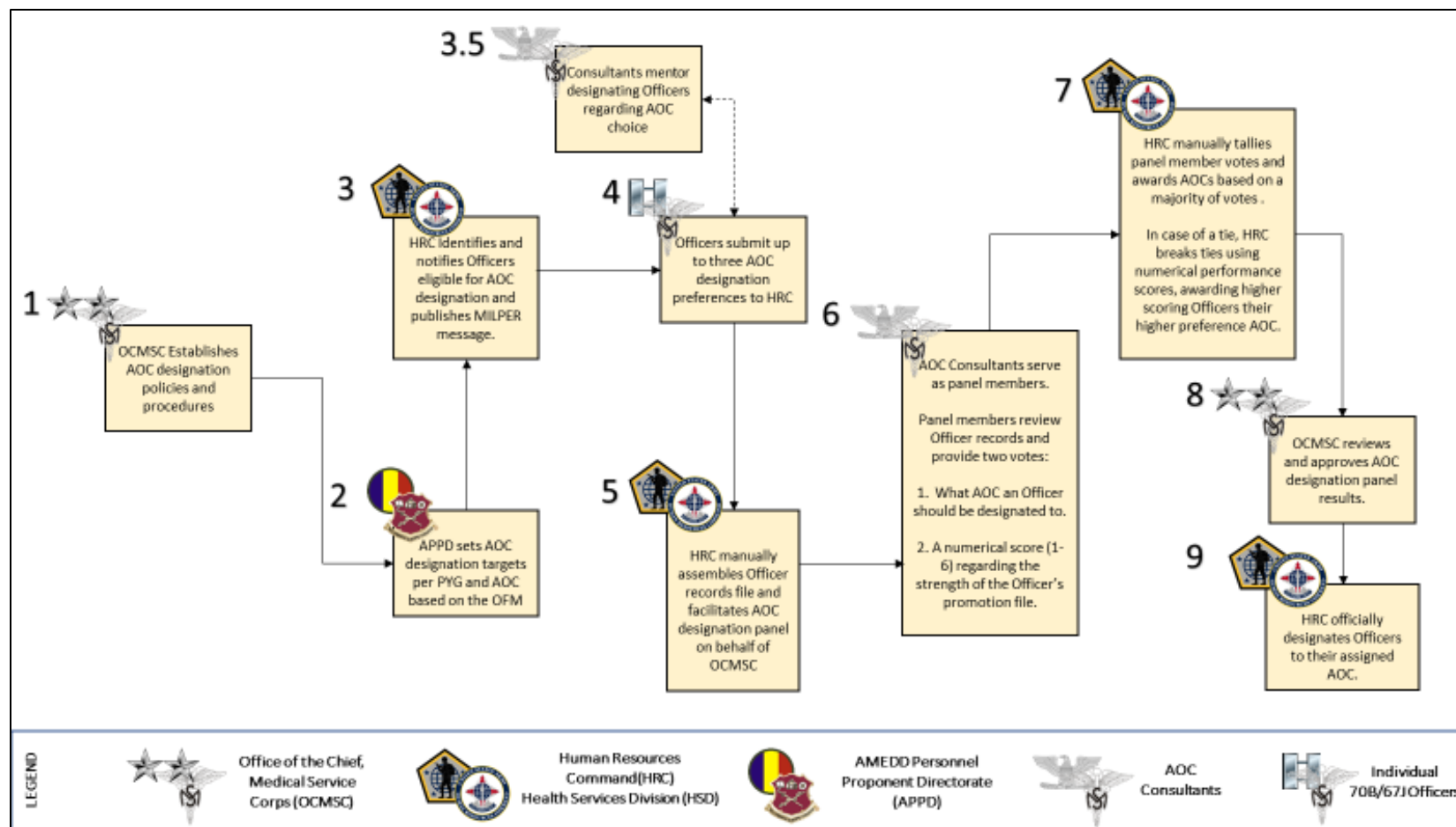


Figure 6. Current 70B/67J AOC Designation Process. Adapted from U.S. Army (n.d.).

The process begins with the AMEDD Personnel Proponent Directorate (APPD). As the branch personnel proponent for Army medicine, APPD is responsible for the “development of manpower and personnel plans, programs, and policies necessary to man, support and sustain” the force modernization of the AMEDD (HQDA 2015, p.13). APPD utilizes the objective force model (OFM) to set a target number of 70B officers per PYG to transition into every MFA 70 AOC. The OFM is an integer linear program (ILP) designed to optimize the distribution of AMEDD budgeted end strength (BES) across AOCs and PYGs in order to maintain a healthy long-term force structure with advancement opportunities for all career fields which fulfills the Army’s operational requirements. McMurry et al. (2010) provide the formulation for the ILP, which currently remains in use. APPD uses the OFM to set recruitment, promotion and AOC transition targets (APPD 2019b).

APPD uses the OFM results to provide the Army Human Resources Command (HRC) Health Services Division (HSD) with an upper and lower bound for the number of 70Bs to transition into each AOC and for 67J officers to designate a secondary AOC. HRC is the Army’s executive agent for all personnel management programs including officer assignments and record management (HQDA 2017).

HRC HSD identifies all officers eligible for designation and notifies them, providing them a deadline to submit a sheet outlining preferences for designation. Officers in their eighth year of commissioned service must list three AOC designation preferences in order of preference, whereas opt-in officers in the sixth and seventh year of commissioned service may list one to three choices. When submitting their preferences, officers provide a narrative justification for their top choice and announce their intentions for applying for AOC-producing LTHET programs (OCMSC 2019a).

Upon receiving the completed preference sheets from officers, HRC HSD compiles a file for each officer in the eighth year of commissioned service and for each officer in the sixth and seventh year who has opted in to be considered for early designation. HSD then collects each officer’s preference sheet; their officer record brief (ORB), highlighting their assignment history, awards and education; and their last five officer evaluation reports (OERs). HSD assignment officers also retrieve the un-released results of the most recent LTHET selection board and manually match the selections with the records of officers up for AOC designation.

HRC then convenes the MSC AOC designation panel on behalf of the corps chief. This virtual panel is chaired by the assistant corps chief for health services administration. The panel members are the consultants for MFA 70 AOCs. HRC HSD assignment officers provide each panel member with the records packet for each officer as well as with a manually assembled summary spreadsheet depicting the officer's preferences, demographic information and highlighting previous AOC-specific military courses as well as LTHET selections.

Panel members then complete the HRC-provided spreadsheet by providing two votes per officer. The first vote reflects which AOC the panel member believes the officer should be designated into. The second vote provides a numerical "word picture" vote. Word picture votes assess the officer's potential for future leadership as an Army officer by using his or her evaluations to 'paint a picture' of the officer's potential for future leadership. Based on the word picture, panel members assign a numerical score ranging from 1 to 6+. Figure 7 provides panel member guidelines for rendering word picture votes.

I/AZ HARD VOTE WORD PICTURE		
6 +/-	Top Few Superior Performer Superior Potential	
5 +/-	Select Ahead of Peers Outstanding Performer Outstanding Potential	
4 +/-	Select with Peers Solid Performer Good Potential	
3 +/-	Select if there is a Requirement Average Performer/Potential	FULLY QUALIFIED
2 +/-	Do not Select Weak Performer	NOT FULLY QUALIFIED
1	Show Cause	

21

Figure 7. Word picture vote guidelines. Source: Department of the Army Secretariat for Selection Boards (DASSB) (2014).

HSD Assignment officers collect the voting sheets from each panel member and manually join them. They then count the panel member votes for each AOC and award an AOC designation to an officer if a simple majority of panel members voted for that AOC. In the case of ties between panel member votes, the consultants for the destination AOCs work an agreement between themselves in regards to which AOC the officer should receive. If the consultants cannot reach an agreement, the Corps chief serves as the final tie breaker. In cases of ties between officers for the same AOC, HSD Assignment officers break ties using the average “word picture” vote score from all panel members. HSD Assignment officers designate the officers with higher average scores to their higher-preference AOCs. In the case of opt-in officers, HSD will designate only if a majority of panel members vote to assign to an officer-preferred AOC. Otherwise, the opt-in officer retains their current AOC. Once this process is complete, HSD forwards the results to the Corps chief’s office for approval. After the results are approved, HSD assignment officers update officer records to reflect their new AOC (OCMSC 2019a).

While the current AOC designation process matches officer’s first preferences near 80% of the time, there are multiple areas which can be improved using OR techniques. First, the target-setting process at APPD does not account for the inventory of officers available for designation and only considers future force structure needs. Additionally, the panel’s current business rules do not penalize or prohibit failures to meet AOC designation targets. In years when the inventory of officers is insufficient to meet these AOC targets, less-preferred AOCs often receive shortages which threaten the long-term health of the force structure. An algorithmic allocation of shortages and overages can ensure the future viability of all MFA 70 AOCs.

The current process is extremely time consuming for the HRC HSD Assignment officers who manually assemble packets, parse through vote sheets, tally panel member votes and break ties. Automating steps in this process will prevent human error, save valuable man-hours and provide faster results.

Additionally, the current paradigm for panel member voting does not provide consultants with the opportunity to shape and influence the AOC they’re responsible for overseeing. It fails to leverage the consultant’s expertise as AOC SME by weighing all

votes equally, with votes from panel members without AOC-specific knowledge potentially overriding the votes of an AOC consultant. Algorithms which account for consultant preferences instead of tallying AOC votes will empower consultants to manage entry into their own AOC without asking them to manage others.

Though the current system does a good job of matching officer preferences to AOCs, this is not by design, and likely due to panel member attempts to honor officer desires. This reliance on panel member voting philosophies addresses force structure concerns and talent-AOC matching only in an ad-hoc manner. Should consultants adopt different voting rationales, the rate of preference matching will likely be lower. An algorithm-based designation process can deliberately honor officer preferences while creating ideal matches between officers and AOCs while ensuring designation targets are met.

Furthermore, the ‘horse trading’ approach for tie breaking ties is inefficient, subject to personal bias and untransparent. Algorithm-based designation solutions can automatically break ties in a manner that is fair and transparent to consultants and officers.

Lastly, the current process is approximately 15 years old and fails to leverage the latest technological and algorithmic advances to manage MSC talent. An updated, technology-enabled designation process will support the Army Chief of staff’s vision of “taking the Army from the way we manage our people, from a data-poor industrial age system, to an information age system that is data-rich so we can make better decisions about our people” (McGee 2019).

II. MATHEMATICAL APPROACHES FOR SOLVING THE AREA OF CONCENTRATION DESIGNATION PROBLEM

This chapter reviews current literature to place this study in the larger academic and military talent management context. Section A reviews the scholarly record in order to highlight algorithmic approaches for solving problems like the AOC designation problem. Section B then surveys current Army and Department of Defense (DOD) algorithm-based solutions to large-scale assignment problems.

The remaining chapters of this study analyze the results of the AOC designation process and improve it utilizing OR techniques.

A. SCIENTIFIC LITERATURE REVIEW

AOC designation is primarily a problem of matching officers to AOCs in some optimal manner. This kind of problem is called an assignment problem. Two scholarly approaches for solving these problems are optimization and stable matching algorithms. There exists an established body of scientific literature detailing both approaches. We introduce general concepts for solving assignment problems in Section 1.

AOC designation can also be approached as a talent management problem. The use of machine learning models to identify assignment-relevant employee attributes is an emerging and developing field of study which the MSC can use to shape and inform AOC designation. We explore such applications of machine learning models in Section 2.

This project leverages the strengths of machine learning models and assignment problem algorithms in order to develop an AOC designation process which maximizes value for MSC officers and for the Army.

1. Assignment Problems

Assignment problems are mathematical problems which find the best allocation of resources to activities. In the AOC designation problem, the resources are officers and the activities are AOC vacancies.

Consider a bipartite graph divided into two disjoint sets: A group of workers I and a group of jobs J connected by edges $X_{i,j}$, where $X_{i,j}$ has one end in set I and one end in set J . This bipartite graph is depicted in Figure 8.

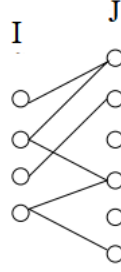


Figure 8. Bipartite Graph

A matching M is a subset of edges such that each worker $i \in I$ and job $j \in J$ appears exactly once.

There are two primary approaches for understanding and solving assignment problems: Linear programming (LP) and stable marriage algorithms (SM).

a. Linear Programs

The problem of finding optimal matches according to some objective criteria is a variant of the minimum cost network flow problem (Burkard et al. 2012). Equation 1 shows a generic integer linear program (ILP) formulation for optimally matching workers from set I to jobs in set J using the combination matrix $X = X_{i,j}$, where the binary variable $X_{i,j} = 1$ if worker i is matched to job j and $X_{i,j} = 0$ otherwise. In order to find optimal assignments, we assign a cost $c_{i,j}$ for every possible pairing between i and j according to some objective decision criteria such as preferences, needs of the hiring entity, monetary cost or job aptitude.

The objective function seeks to optimize the overall value of assignments (Burkard et al. 2012) by summing across every possible pairing.

$$\begin{aligned}
& \min \sum_{i \in I} \sum_{j \in J} c_{i,j} X_{i,j} \\
& \sum_{j \in J} X_{i,j} = 1 \quad \forall i \\
& \sum_{i \in I} X_{i,j} = 1 \quad \forall j \\
& X_{i,j} \in \{0,1\}
\end{aligned}$$

Equation 1. Generic Assignment Problem ILP Formulation

For an assignment scheme to be considered optimal, it must optimize the overall value of job-assignment matches in the entire set while satisfying the constraints that each worker must be assigned to only one job. The weights given to the cost criteria determine the problem's optimal solution.

We can also express the assignment problem as a minimum cost network flow problem, with flow supplied by worker node i to fill job demand in node j flowing through arcs (i,j) at cost $c_{i,j}$. Specialized algorithms such as the successive shortest paths algorithm and cycle-canceling algorithms, which are designed to take advantage of the network structure, can solve this type of network flow problem efficiently and quickly without the need for a linear program solver (Ahuja et al. 1993). Figure 9 depicts the network construct of a job assignment problem.

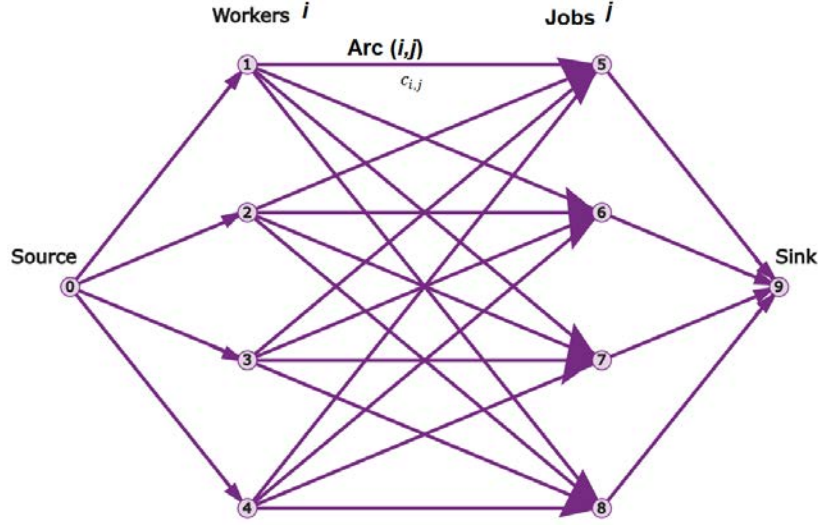


Figure 9. Generic Assignment Problem Network.
Adapted from Google (2019).

Both approaches are mathematically equivalent and yield optimal assignment schemes; however, the ILP construct accommodates the inclusion of additional policy constraints in the AOC designation problem. ILPs are also flexible enough to deal with situations where the number of officers available to designate does not match the number of required AOC designations. Chapter VI details an ILP formulation and implementation to solve the AOC designation problem.

b. Stable Marriage Problems

We can also solve the AOC designation problem as a stable marriage problem (SM) where we attempt to find the best marriage (matching) scheme between “men” (workers) and “women” (jobs) by considering their ordinal preference lists.

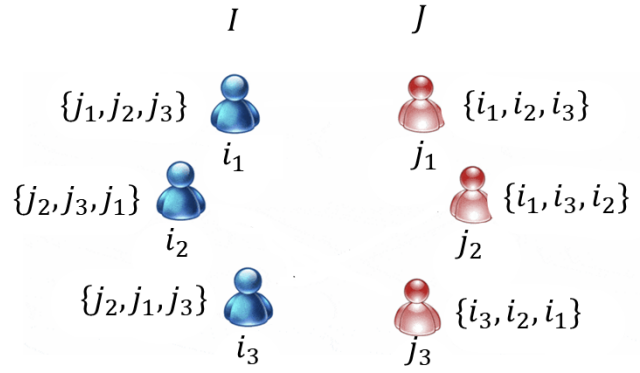
SM algorithms yield stable assignments schemes. In a marriage problem, for a matching to be considered stable, there must be no man-woman pair where the man and woman would rather be with each other than with their matched partner (Iwama and Miyazaki 2008). In 1935, Philip Hall outlined the necessary conditions for creating stable matches in equally sized sets (Hall 1935).

Consider a stable marriage problem matching workers $i \in I$ to jobs $j \in J$, where the preference matrix $P_{i,j}$ represents the ordinal preferences $p_{i,j}$ of workers $i \in I$ for jobs $j \in J$ such that $p_{i,j} = \{1, 2, 3, \dots, |J|\}$. Then $p_{i,j} = 1$ if and only if job j is worker i 's most favored assignment and $p_{i,j} = |J|$ if job j is worker i 's least preferred assignment.

Likewise, the preference matrix $P_{j,i}$ represents preferences of jobs $j \in J$ for workers $i \in I$ such that $p_{j,i} = \{1, 2, 3, \dots, |I|\}$ and $p_{j,i} = 1$ if and only if worker i is job j 's most favored hire. We consider a set of marriages A stable if and only if $\forall a_{i,j} \in A \nexists (a_{x,y}, a_{u,v})$ with blocking pair $(x \in I, v \in J)$ such that $p_{x,v} < p_{x,y}$ and $p_{v,x} < p_{v,u}$.

Building on Hall's work, D. Gale and L.S. Shapley described the SM problem in their 1962 paper, *College Admissions and the Stability of Marriage*, providing a polynomial time algorithm for solving such problems (Iwama and Miyazaki 2008). Gale and Shapley's work serves as the basis for a number of large-scale assignment applications, including the Nobel-Prize-in-Economics winning National Resident Matching Program (NRMP), which matches U.S. medical students with residency sites and fellowship positions (National Resident Matching Program 2020).

In Gale and Shapley's version of the SM problem, there are an equal number of men and women who must marry a member of the opposite sex. Each individual rank-orders the members of the opposite sex from most to least preferred. Figure 10 depicts a basic SM sample matching three workers $\{i_1, i_2, i_3\}$ with three jobs $\{j_1, j_2, j_3\}$ and provides the preference matrices $P_{i,j}$ and $P_{j,i}$ for this matching problem while Algorithm 1 provides pseudocode for solving the stable marriage problem.



$$P_{i,j} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 2 & 1 & 3 \end{bmatrix} \quad P_{j,i} = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \\ 3 & 2 & 1 \end{bmatrix}$$

Figure 10. Sample Stable Matching Problem with Preference Matrices.
Adapted from Cui (2015).

Algorithm 1. Stable Matching Algorithm Pseudocode.
Adapted from Osipenko (2019).

StableMatching

```

Initialize all  $i \in I$  and  $j \in J$  to free
while  $\exists$  free man (worker)  $i$  who still has a woman (job)  $w$  to propose to {
   $j$  = first woman on  $i$ 's list to whom  $i$  has not yet proposed
  if  $i$  is free
    ( $i, j$ ) become engaged
  else some pair ( $i'$ ,  $j$ ) already exists
    if  $j$  prefers  $i$  to  $i'$ 
       $i'$  becomes free
      ( $i, j$ ) become engaged
    else
      ( $i', j$ ) remain engaged

```

In this simple implementation of the Gale-Shapley algorithm, i_1 proposes to j_1 . Since i_1 is j_1 's first choice and $p_{i_1, j_1} = p_{j_1, i_1} = 1$, she accepts and they become permanently engaged.

Then, i_2 proposes to j_2 , who tentatively accepts as j_2 is free. Now i_3 proposes to j_2 . Since j_2 prefers i_3 and $p_{j_2, i_3} < p_{j_2, i_2}$, j_2 becomes free and gets engaged with i_3 . Since i_2 is now free again, he proposes to j_3 , to whom he was not proposed. j_3 is free and accepts the proposal, becoming engaged to i_2 . This is a stable solution to the matching problem as we can see that there is no blocking man-woman pair that mutually prefers each other to their current partner. Figure 11 shows the results of the stable matching problem outlined in Algorithm 1.

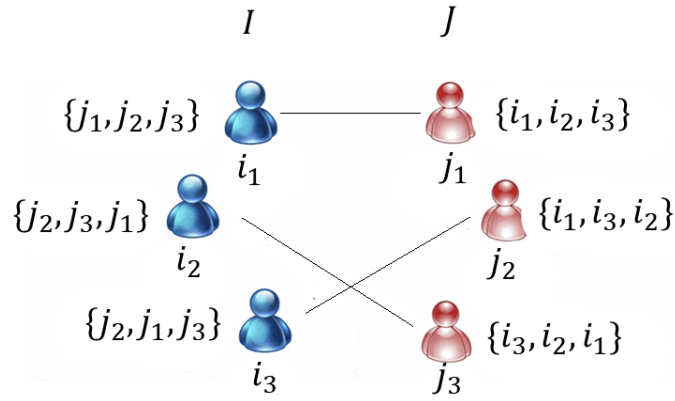


Figure 11. Stable Matching Algorithm Results. Adapted from Cui (2015).

Gale and Shapley proved that if the number of men and women are the same, there is at least one way of creating stable matches (Gale and Shapley 1962). There may be multiple stable solutions to the same assignment problem. Gale and Shapley define stable assignments as optimal “if every applicant is at least as well off under it as under any other stable assignments.” Basic implementations of the Gale-Shapley algorithm provide solutions which are man-optimal but which may not be woman-optimal (Iwama and Miyazaki 2008). The man-optimal and woman-optimal solutions are the same only when there is a single set of stable matches (Gale and Shapley 1962).

In the 1990s, anecdotal information from the first implementation of the NRMP algorithm (where hospitals proposed assignments to medical students) showed that students could misrepresent their preferences in order to receive a more student-optimal

solution. Roth and Peranson revised the NRMP algorithm by making the students propose assignments instead, guaranteeing student-optimal matches. Computer simulation experiments showed that due to the nature of the student-proposing assignment problem, where many students apply to the same hospitals, the effects of strategic misrepresentation by hospitals were negligible at best and seldom provided hospitals better matchings than they would receive by presenting their true preferences (A Roth and Peranson 1999). Romero-Medina provides an algorithm for creating “sex-equal” matches by minimizing “envy difference” between men and women (Romero-Medina 2001). Gupta et al. provide a game-theoretic analysis of how women can optimally cheat by hiding their true preferences in men-proposing versions of the SM problem. They also provide an algorithm for finding a Nash-stable match for a woman if we know her true preference list and all other women state their true preferences (Gupta et al. 2015).

Given the MSC chief’s stated intent (2020) to produce officer-optimal matches to the extent possible, we conceive of an officer-proposing SM problem where 70B and 67J officers propose assignments to AOCs, guaranteeing officer-optimal assignment schemes and assuming that consultants have no incentive for manipulating their preference lists.

Applying the Gale-Shapley algorithm to AOC designation requires significant adaptations such as accounting for ties, incomplete preference lists and unequal size sets. In Chapter VI we cover these adaptations (Iwama and Miyazaki 2008).

c. Comparison of ILP and SM Approaches to Assignment Problems

In 1993, Roth et al. showed links between the ILP and SM approaches to solving the basic assignment problems. They asserted that when the worth of an (i,j) match $c_{i,j}$ is a cardinal utility function of the ordinal preferences of agents i and j , the “matchings in the core of the marriage model can also be described by a linear programming problem whose extreme points are all integer” (Roth et al. 1993).

Despite the similarities, the LP formulations of the minimum cost network flow problem and the stable marriage problem are fundamentally different and they yield different assignment schemes. Optimality and stability are not the same: The optimal assignment schemes yielded by the network flow problem fundamentally vary according

to the choice of $c_{i,j}$ while the stable marriage algorithm guarantees stable assignments based on ordinal preference lists.

The modifications required of the network flow problem ILP formulation and the Gale-Shapley algorithm in order to accommodate the AOC designation problem ensure both approaches yield different results. The AOC designation process proposed by this study utilizes both ILPs and SM algorithms. Chapter VI analyzes and both implementations of the AOC designation problem, utilizing the strengths and drawbacks of each approach to provide a range of designation outcomes to the MSC.

2. Machine Learning Models for Talent Management

Machine learning refers to a variety of tools built for understanding data (James et al. 2013). Computational advances in the last quarter century have enabled a generation of data scientists to create mathematical models which find patterns in datasets and generalize them without being provided with specific instructions (Domingos 2012).

Machine learning models can be supervised or unsupervised. We train supervised models to estimate an output based on a series of inputs. Unsupervised models, by contrast, estimate structure and patterns in the data without being provided an output (James et al. 2013). There are dozens of different kinds of machine learning models, many of which sort observations into categories based on their attributes. These models are called classifiers.

Machine learning classifiers “predict the probability of each of the categories of a qualitative variable, as the basis of making the classification” (James et al. 2013). For every observation i , most classifiers produce a numerical prediction $\eta_{i,j} \in \mathbb{R}$ for each class j . Equation 2 shows the softmax function for a theoretical model classifying n observations into k classes, with $\eta_{i,j}$ serving as the model-generated numerical prediction matching observation i to class j and $\hat{p}_{i,j}$ reflecting the softmax-generated probability that observation i belongs to class j . The use of the softmax function guarantees that the sum of the predicted probabilities for all k classes equals 1.

$$\hat{p}_{i,j} = \frac{e^{\eta_{i,j}}}{\sum_{j=1}^K e^{\eta_{i,j}}}$$

Equation 2. Softmax function estimating probability that observation i belongs to class j

The majority of machine learning models learn by minimizing a loss function. Loss functions provide a measure of how much model-predicted responses differ from actual responses in the training data. In the case of multi-class problems such as AOC prediction, most—but not all—classifiers use the multinomial deviance or multinomial cross entropy loss function (Whitaker 2020). Equation 3 shows the multinomial deviance loss function L .

$$L = -\frac{1}{n} \sum_{i=1}^n \left[\sum_{j=1}^K y_{i,j} \left(\ln \frac{e^{\eta_{i,j}}}{\sum_{j=1}^K e^{\eta_{i,j}}} \right) \right]$$

$$y_{i,j} = \begin{cases} 1 & \text{if observation } i \text{ belongs to class } j \\ 0 & \text{otherwise} \end{cases}$$

Equation 3. Multinomial Deviance Loss Function for generic classification problem

The loss observed for any particular observation depends on how well the model-predicted probabilities match the observation's actual label. For observation i , if $y_{i,j} = 1$ and the model correctly predicts a high probability for category j , i.e., if $\hat{p}_{i,j} = 0.9$, then $-\ln \hat{p}_{i,j} = 0.105$. However, if the model incorrectly assigns a low probability, $\hat{p}_{i,j} = 0.1$, then $-\ln \hat{p}_{i,j} = 4.6$. Machine learning algorithms find the configuration which minimizes the value of the loss function across the training set.

Scientific literature regarding the use of machine learning models to predict job aptitude and assist recruitment efforts is scarce and most is relatively recent. Literature published earlier than 2012 largely comes from the field of psychology and focuses on the use of simple linear regression with cognitive ability test scores as predictors of job aptitude

(Ree et al. 1994). In 2012, Erel et al. used gradient-boosted classification trees and neural network classifiers to aid corporations in selecting candidates for boards of directors and to shed light into corporate decision-making processes. In 2016, Zimmerman et al. provided a successful framework for machine-learning enabled corporate recruiting, utilizing natural language processing (NLP) to parse through candidate resumes and machine learning algorithms to match qualifications to specific positions. providing “the right tools to objectively judge and rank candidates ... [and] improve the quality and speed...of the recruitment process.”

In the context of AOC designation, we use supervised classifiers to match officers to AOCs based on their assignment history and educational background. We explore these models as analytical tools and informative aids to support the AOC recruitment and preferencing process rather than as prescriptive models directing AOC assignments.

B. ALGORITHM-ENABLED TALENT MANAGEMENT IN THE DEPARTMENT OF DEFENSE

For the past three decades, the DOD has recognized the need to implement technology-enabled solutions to manage its manpower and personnel in order to maximize organizational efficiency. All services have followed a similar progression from centralized, industrial-age human resources (HR) systems where servicemembers of a certain rank and career field are mutually interchangeable and where their preferences are largely irrelevant (Rotich 2015), to the algorithm-enabled HR processes of today which seek to maximize servicemember satisfaction and leverage their talents while maintaining service readiness and priorities.

In the 1980s, the Army used large-scale linear goal-seeking programs to create the Army manpower long range planning system (AMLRPS), which provides twenty-year-horizon force-planning across the entire service in order to assess the long-term impacts of personnel policies and funding levels (Gass et al. 1988). While the mathematical methods existed before the 1980s, the computational complexity of these models made it impossible for DOD planners to provide congress with regular forecasts (Holz and Wroth 1980). The

AMEDD OFM, which we introduced in Chapter I and which produces the annual MSC AOC designation targets, is based on the AMLRPS (McMurry et al. 2010).

As we see in this section, assignment models used throughout the DOD prior to 2000 were primarily greedy algorithms which make locally optimal assignments but are not guaranteed to yield globally optimal assignment schemes (Greedy Algorithms 2011)

1. Army Functional Area (FA) Designation Process (2001)

In the early 2000s, the Army designed its first algorithmic talent management process when it overhauled the Functional Area (FA) designation policy for Army basic branch officers. The Army FA designation process is highly analogous to the MSC designation process.

The Army officer corps is composed of basic branches, such as infantry, armor, artillery, etc.; and special branches requiring specialized education, such as the AMEDD branches, the judge advocate general (JAG) and chaplain corps. The FA designation process transitions mid-career basic branch officers to specialized FAs such as foreign area officer, operations research, strategy and electronic warfare. Unlike the MSC AOC designation process, basic branch officers may opt to remain in their basic branch (HQDA 2019b).

The 1990s precursor to the FA designation process was very similar to the current AOC designation process: An annual selection board convened to evaluate qualified officer records and rank-order them, producing one order of merit list (OML) per FA. Officers with higher places in the OML were awarded their first choice FA until all FA slots were filled via a manual process utilizing greedy preferences.

In order to cope with the complexity of new FAs and changes in officer management policy, Army analysts with the office of the deputy chief of staff G-1 developed the career field designation (CFD) model, an network optimization program which “minimizes the sum of [OML] scores while meeting flow-balance constraints to ensure that the number of officers designated into each career field meets the requirements”

(Shrimpton and Newman 2005, p. 234). Figure 12 shows the network representation of the CFD model for designating i officers into j functional areas.

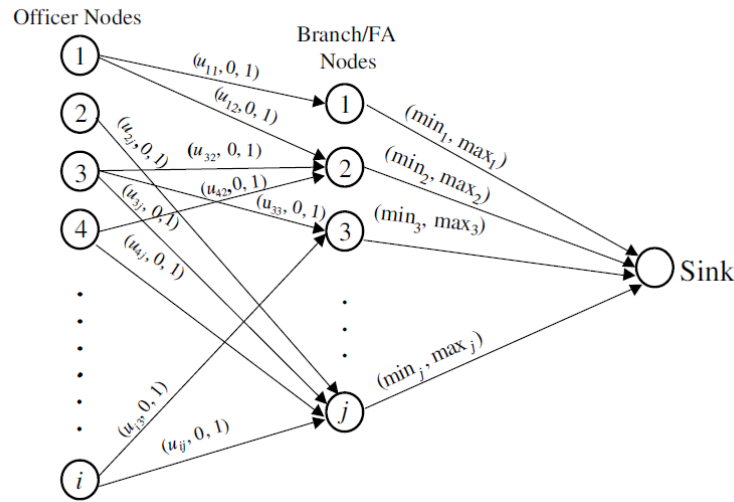


Figure 12. U.S. Army Career Field Designation Model (CFD). Source: Shrimpton and Newman (2005).

While the FA designation model tangentially accounts for officer preferences, Shrimpton and Newman claim success in satisfying first preferences for 85% of officers. As in most of the algorithm-based processes we review in this section, senior leaders review model results and adjust them if they deem necessary (Shrimpton and Newman 2005).

2. U.S. Air Force Academy (USAFA) Air Force Specialty Code (AFSC) Matching Process (2001)

In 2001, the U.S. Air Force Academy (USAFA) developed an ILP model to aid in the designation of graduating cadets to air force specialty codes (AFSC). We display this model in Figure 13. As with the Army's FA designation model, the AFSC matching process replaced a manual process following greedy preferences. Likewise, two separate panels score and rank cadets amongst each other. These ranks then serve as inputs into the mathematical designation model (Armacost and Lowe 2005).

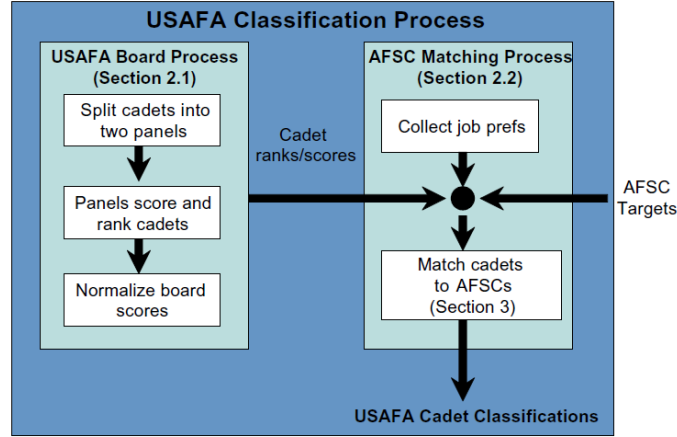


Figure 13. USAFA Career Field Selection Model. Source: Armacost and Lowe (2005).

The ILP formulation in Equation 4 solves the AFSC matching problem. It is equivalent to the generic ILP formulation we introduced in Equation 1 except for the inclusion of penalized slack variables s_j^+ and s_j^- . These variables reflect the number of cadet overages and shortages for AFSC j in situations where the number of graduating cadets is not sufficient to meet the Air Force's manpower needs (Armacost and Lowe 2005).

$$\begin{aligned}
 \max \quad & \sum_{i \in I} \sum_{j \in J} c_{ij} y_{ij} - \sum_{j \in J} (d_j^+ s_j^+ + d_j^- s_j^-) \\
 \text{subject to: } & \sum_{j \in J} y_{ij} = 1 \quad i \in I, \\
 & \sum_{i \in I} y_{ij} - s_j^+ \leq u_j \quad j \in J, \\
 & \sum_{i \in I} y_{ij} + s_j^- \geq l_j \quad j \in J, \\
 & y_{ij} \in \{0, 1\} \quad i \in I, \quad j \in J, \\
 & s_j^+, s_j^- \geq 0 \quad \text{and integer} \quad j \in J
 \end{aligned}$$

Equation 4. USAFA Career Field Matching Model Formulation. Source: Armacost and Lowe (2005).

Armacost and Lowe conceived of a flexible model, with different options for setting the value of the cost coefficient $c_{i,j}$ as a function of cadet preferences and cadet class ranks to “provide multiple solutions quickly...[allowing] the Academy’s decision makers [to] determine the trade-offs and classification policy they prefer” (Armacost and Lowe 2005, p. 847).

3. The Sailor Assignment Problem (SAP) (2006)

In the mid-2000s, the U.S. Navy instituted a web-enabled job assignment algorithm for enlisted Sailors. Currently called MyNavy Assignment (MNA), the system uses an online marketplace which “provides the Sailor information about the jobs’ suitability and displays metrics of how this job will affect his career progression...and other relevant details” (Garrett et al. 2005). Sailors apply for positions they are qualified for and gaining commands rank Sailors for their open positions. The workflow for the assignment process is depicted in Figure 14.

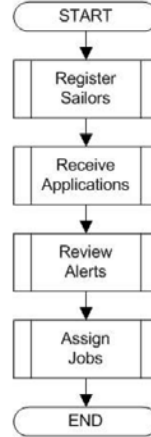


Figure 14. MyNavy Assignment Workflow. Source: Garrett et al. (2005).

Once MNA receives all applications and preferences, analysts solve an ILP formulation similar to the one introduced in Equation 1, where the cost coefficient $\mathcal{F}_{i,j}$ is a function of “the Sailor’s ability to perform the requirements of the position...and estimates the monetary cost of assigning the Sailor to that position” (Garrett et al. 2005). While the

Navy considered using a stable marriage algorithm, the requirement to consider PCS and training costs in addition to Sailor and command preferences required the use of the ILP formulation shown in Equation 5 (Vannucci et al. 2006).

$$\sum_{i=1}^N \sum_{j=1}^M \mathcal{F}_{i,j} d_{i,j},$$

subject to

$$\sum_{i=1}^N d_{i,j} \leq 1, \quad \forall j \in \{1, 2, \dots, M\}$$

and

$$\sum_{j=1}^M d_{i,j} \leq 1, \quad \forall i \in \{1, 2, \dots, N\},$$

Equation 5. Sailor Assignment Problem Formulation. Source: (Garrett et al. 2005).

Detailers at Navy Personnel Command review the results of the algorithm and make final decisions regarding Sailor assignments. Such a system is in the development stages for Navy officers; however, only enlisted Sailors currently participate in MNA (Stoner and Campbell 2019).

4. U.S. Army Talent Alignment Process (ATAP) (2019)

In 2019, after two years of small-scale testing, the Army introduced the army talent alignment process (ATAP). Implemented by HRC, ATAP manages the distribution of Army officers throughout the force. In the Assignment Interactive Module 2.0 (AIM2), a web-enabled marketplace similar to the Navy's MNA, Army officers eligible for moving upload resumes to supplement their officer record brief (ORB) and view positions for which they are eligible. Officers rank-order the positions from most to least-desired. Units with vacancies act as hiring entities, reviewing officer resumes and rank-ordering officers eligible to fill their vacancies (U.S. Army Talent Management Task Force [USATMTF] 2019).

The Army talent alignment algorithm (ATAA) is an officer-proposing Gale-Shapley-based SM algorithm which attempts to find stable matches between officers and units. While HRC has published no scientific literature regarding the ATAA, during a December 2019 visit, HRC operations analysts stated the algorithm is equipped to handle incomplete preference lists and ties. HRC assignments branches are instructed to honor algorithm matches, with branch chiefs having the ability to override marketplace matches in special occasions. During the first large-scale application of the ATAA in the Fall of 2019, 55% of officers received their first assignment choice while 80% received an assignment in their top 10% of ranked assignments (Kimmons 2020).

ATAP is a relatively closed system. In the AIM portal, officers can't see how a unit ranked them but they can see whether a unit has ranked them. Units also can see officers who have ranked their position in the top 10% of their preference list. Informally, officers and units interact in order to gauge their standing in one another's preference list and adjust those lists in order to achieve a one to one match.

HRC has invested considerable effort into collecting knowledge, skills and behaviors (KSBs) information from officers and units in order to enable future AIM2 functionality. Officers self-profess their KSBs while units outline the KSBs required for vacancies. HRC hopes to use statistical learning models to recommend officers to units and vice-versa in order to enhance matches in future personnel movement cycles (USATMTF 2019).

5. U.S. Army Talent Based Branching (TBB) (2019)

The United States Military Academy (USMA) and the Army Reserve officer Training Corps (ROTC) are two major commissioning sources for new U.S. Army officers. As of 2019, both have turned to a talent-based branching model following a similar construct to the ATAP (O'Connor 2019). The talent-based branching (TBB) model replaced older greedy algorithms using cadet OML scores to determine their basic branch of assignment.

The TBB model collects resume and KSB information from cadets and from basic branches, using a statistical model to help cadets identify branches they may be compatible

for and vice versa. As with ATAP, cadets rank order all 17 possible branch assignments while branches rank all cadets into one of three tiers from most to least preferred (U.S. Army Talent Management Task Force 2019).

According to the TBB program manager at Headquarters, Department of the Army G1 during a December 2019 telephone conversation, the TBB algorithm is a Gale-Shapley-based SM algorithm. Given the tier-based rankings (as opposed to ordinal ranking) provided by branches, the TBB model expects ties between cadets and breaks them using OML scores, matching better-ranked Cadets with higher preferences.

USMA provides cadets and branches with a limited time window in which both can see how they have been ranked by the other party. Both cadets and branches have the ability to adjust their preference lists if so desired.

USMA implemented the model with the graduating class of 2020, with branches improving their KSB-matching rate by approximately 20% while close to 80% of graduating Cadets received their top branch choice (U.S. Army Talent Management Task Force 2019).

Over the course of 2019, senior MSC leaders, including the current corps chief, deputy corps chief, APPD director and HRC MS branch chief, have become familiar and comfortable with the ATAA and TBB algorithms, both of which use the Gale-Shapley construct to solve talent management problems such as MSC AOC designation.

In this chapter, we've explored a variety of algorithmic and mathematical methods for approaching large-scale talent management problems. We've also examined a variety of algorithmic talent management applications currently utilized in the DOD. These studies and processes serve as the building blocks for the MSC AOC matching system.

III. A SOLUTION TO THE AREA OF CONCENTRATION DESIGNATION PROBLEM

This chapter presents an overview of our proposed solution to the AOC designation problem. We outline a vision for the AOC designation process and policy in order to frame the following chapters of this study, which are more technical in nature. For clarity and consistency, we use the present tense when referring to the proposed process despite the fact that the process has not been officially approved by the MSC or implemented.

The AOC Matching System (AOC-MS) is a technology-rich process leveraging the latest algorithmic developments and machine learning techniques to create optimal matches between MFA 70 officers and AOCs. AOC-MS is a three-phase process designed to identify, develop, assign and retain MSC talent throughout a 6-year life cycle.

A key difference between AOC-MS and the current AOC designation process is that there is a single designation opportunity for every PYG during their 8th year of AFCS; therefore, there are no early opt-in opportunities under the proposed process. A single designation is necessary for the AOC-MS talent matching algorithms to deliver optimal results for the entire PYG. Theoretically, AOC designation could take place at any point between an officer's selection for promotion to captain and selection for promotion to major (between the 3rd and 10th year of AFCS). This is purely a policy question and the quantitative methods we outline in this study are agnostic to this choice; however, earlier designation carries risk for officers and the MSC, as officers could be designated to AOCs with specific graduate education requirements like 70A (health care administration) or 70C (health services comptroller) prior to finishing company command or being selected for attendance to LTHET. These officers could be at risk for attrition into other AOCs with less strict qualifications, creating force structure imbalances. Later designation provides the MSC with flexibility to fill AOC-specific captain billets with 70Bs or 67Js while giving officers time to develop qualifications for their desired AOCs prior to designation. Figure 15 presents an overview of AOC-MS.

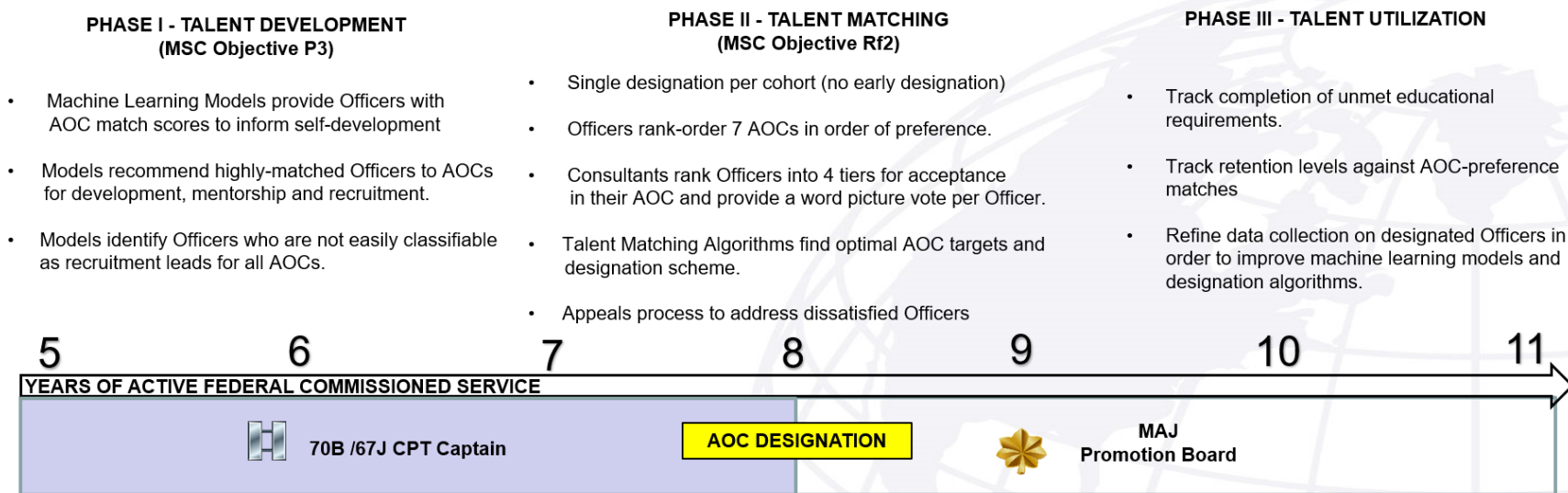


Figure 15. AOC Matching System (AOC-MS) Overview. Adapted from U.S. Army (n.d.).

A. AOC-MS TALENT DEVELOPMENT PHASE

The talent development phase of AOC-MS takes place on an annual basis between the PYG's 5th and 8th year of AFCS. Analysts at APPD use the AOC-MS talent development algorithm, using one of four machine learning classifiers which assesses HRC-provided officer records, resumes and KSBs to provide an AOC match score per officer for every AOC, indicating how their assignments, educational history and self-professed skills match those of officers who have already designated into that AOC. Chapter V explains how the AOC-MS talent development algorithms work in detail.

A web-enabled portal provides officers with their AOC match score, helping inform their future AOC preferences. Furthermore, the portal guides officers in seeking development opportunities which may more closely align them with their desired AOCs. Likewise, APPD analysts provide consultants with a list of officers with high AOC match scores, enabling senior AOC leaders to establish mentorship relationships. Consultants use AOC Match lists to grow interest in their AOC, developing prospective talent by encouraging attendance to AOC-producing courses and recommending AOC-specific assignments.

Additionally, the machine learning models identify 'blank slate' officers whose AOC match scores are relatively equal across AOCs. These lists provide consultants with opportunities to develop officers whose career paths have not steered them towards a particular AOC. OCMSC tracks and manages consultant outreach and recruitment efforts in order to track their effectiveness. Measures of effectiveness include long term retention of contacted officers and how closely the distribution of officer AOC designation preferences matches the MSC's manpower requirements.

The talent development phase of AOC-MS helps officers and AOCs to develop well-informed preference lists, increasing the possibility of individually optimal matches during AOC designation time. Furthermore, AOC-MS provides a formal framework for mentorship between senior and junior MSC officers, improving morale while positioning 70B and 67J captains for future success as AOC-designated majors and lieutenant colonels. Figure 16 provides an overview of the first phase of AOC-MS.

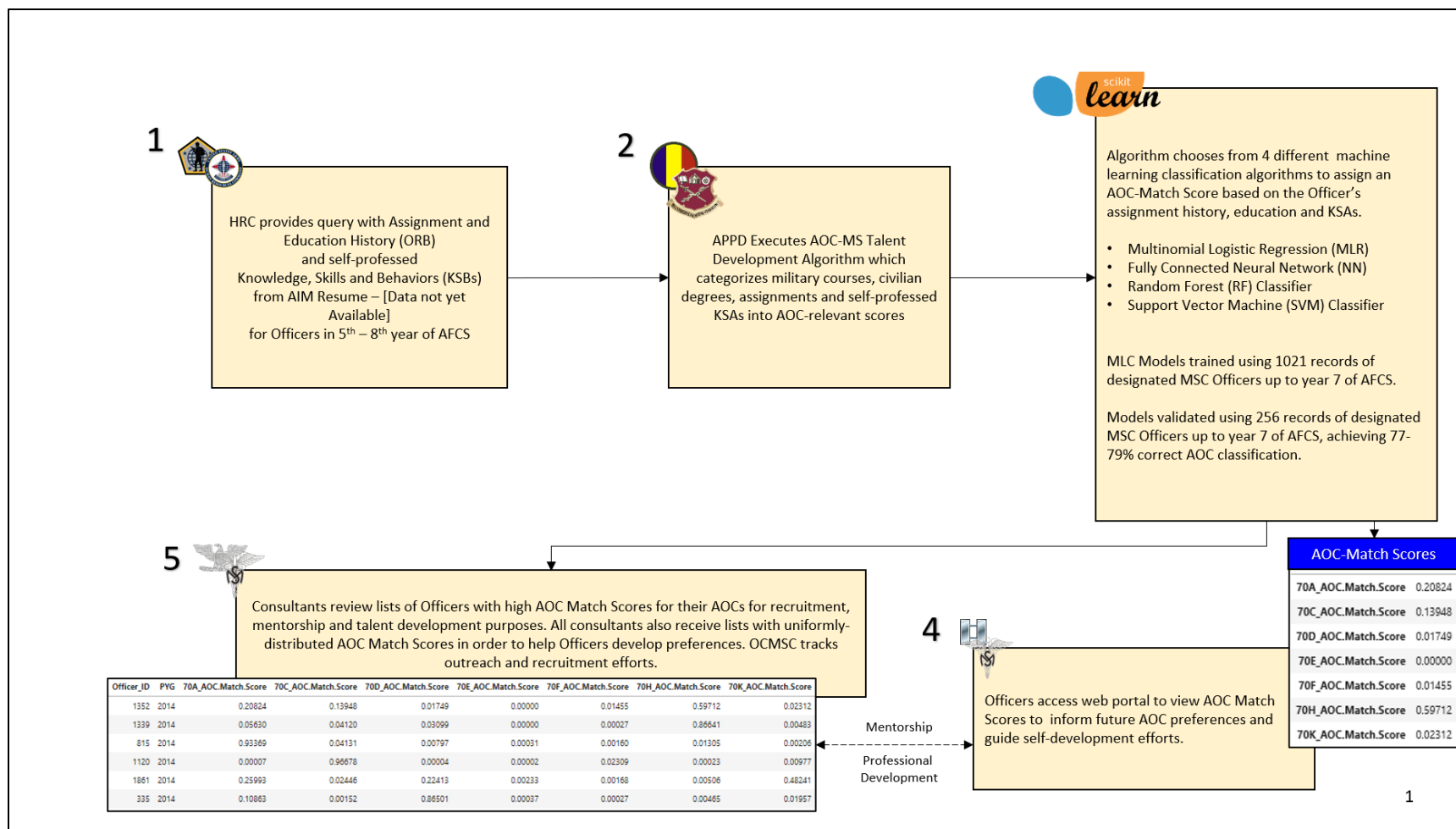


Figure 16. AOC-MS Talent Development Phase Overview. Adapted from U.S. Army (n.d.).

B. AOC-MS TALENT DEVELOPMENT PHASE

The second phase of AOC-MS takes place during officers' 8th year of AFCS. This is the actual AOC designation phase, during which we create optimal matches between officers and AOCs.

This phase begins with HRC identifying officers in their 8th year of AFCS and notifies them of actions required for them to participate in AOC-MS. Officers utilize a web-enabled portal to rank-order the seven MFA 70 specialized AOCs in order of preference. HRC analysts use a parsing algorithm to read officer preference sheets and automatically generate voting sheets for panel members, saving HRC assignment officers hours of data entry.

HRC then convenes the AOC-MS designation panel on behalf of OCMSC. The panel members representing the seven gaining AOCs review officer records and rank them into four preference tiers, identifying officers who the consultants believe will be good fits for their particular AOC. The ranking structure is one that consultants are familiar with, as it is based on the senior rater ranking structure for Army noncommissioned officer evaluation reports (HQDA 2019a). Consultants may rank no more than 24% of officers as most preferred, and they may rank up to the remaining 50% of officers as highly preferred (i.e., if 10% of officers are ranked as most preferred, up to 40% may be ranked as highly preferred). Consultants may rank officers in the two remaining tiers, (preferred and least preferred) with no restrictions.

As with the current AOC designation process, all panel members review officer records and provide a numerical word picture vote from 1 to 6.5 as described in Figure 7. HRC analysts receive panel member vote sheets and merge officer and consultant preferences.

Once officer and AOC preferences are set, analysts at HRC and APPD analyze the distribution of officer preferences against the distribution of manpower billets provided by the OFM utilizing a combination of nonparametric and parametric statistical methods in order to select a variant of the AOC-MS talent matching algorithm (TMA). In PYGs where the distribution of preferences and billets is similar, phase II of AOC-MS uses the first

variant of the AOC-MS TMA, consisting of an integer linear program to adjust OFM targets and a stable marriage algorithm to find stable matches between officers and AOCs based on their mutual preference. In PYGs where the distributions of preferences and OFM billets are significantly different, phase II of AOC-MS uses the second variant of the AOC-MS TMA, an integer linear program which simultaneously adjusts designation targets and satisfies mutual preferences between officers and AOCs, allowing the MSC the flexibility to balance the risk in not filling AOC billets with the risk in not honoring officer preferences. Both variants of the AOC-MS talent matching algorithm ensure designation targets meet 70B inventory available for designation, accounting for force structure imbalances in previous PYGs when distributing shortages or overages. Chapter VI outlines the AOC-MS talent matching algorithm in detail.

Once the algorithm has produced optimal matches, OCMSC reviews the model designations. If deemed necessary by OCMSC, HRC analysts adjust algorithm parameters in order to provide options to the MSC chief. Once the corps chief approves designations, HRC officially designates officers into their assigned AOC by updating their official record.

Finally, AOC-MS accounts for the fallibility of algorithms by providing officers with a 90-day window with which to appeal designation results. An appeal process ensures the MSC has the ability to address unsatisfied officer concerns while providing information which to assess and improve AOC-MS policy and algorithms. Figure 17 provides an overview of the talent matching phase of AOC-MS.

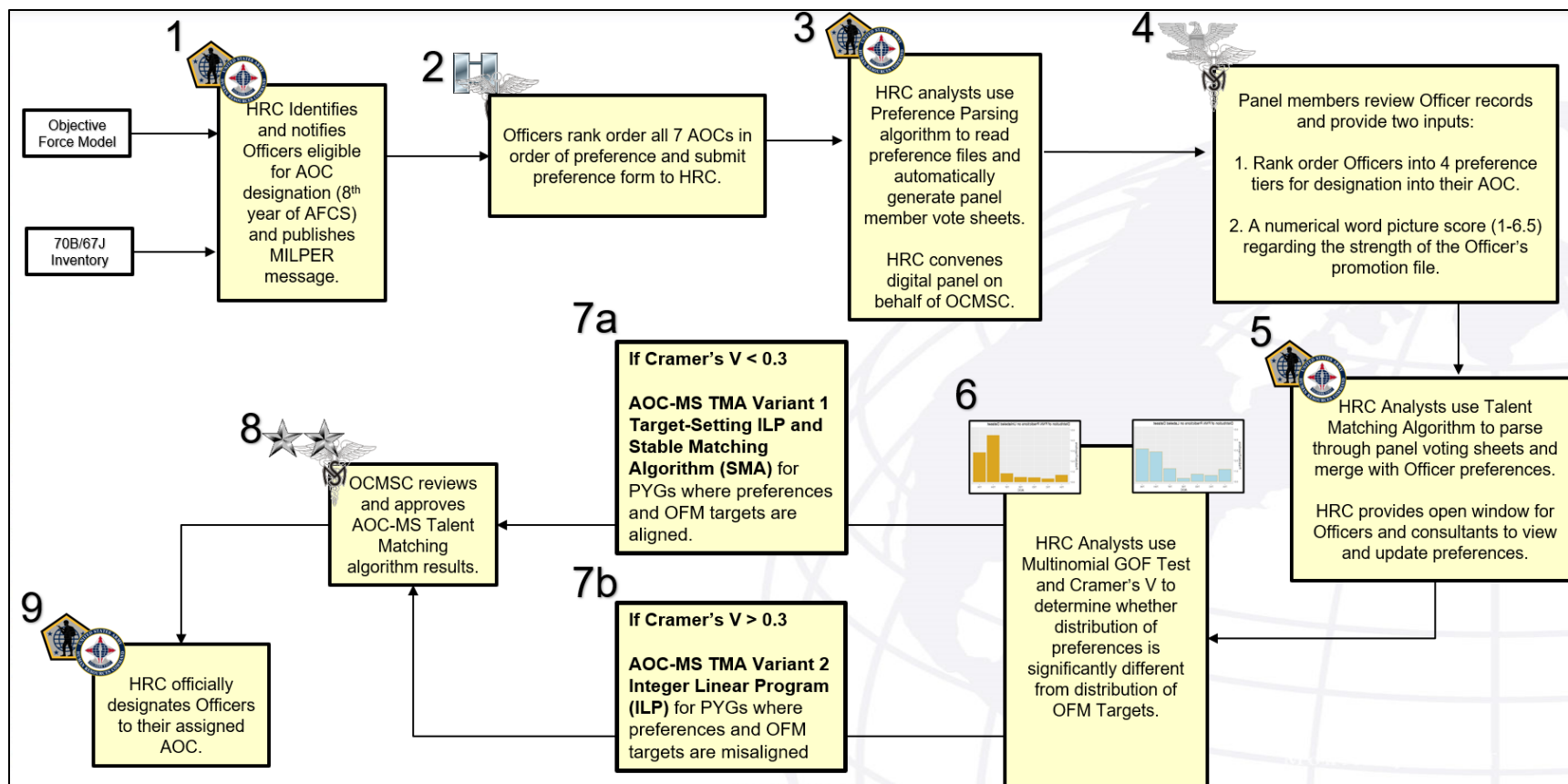


Figure 17. AOC-MS Talent Matching Phase Overview. Adapted from U.S. ARMY (n.d.).

C. AOC-MS TALENT UTILIZATION PHASE

During the final phase of AOC-MS, OCMSC in coordination with APPD and HRC tracks designated officers to ensure unmet educational requirements are completed within a timeframe specified by MSC policy. Additionally, APPD tracks retention levels against AOC-preference matches in order to assess how AOC-MS talent development and talent matching efforts relate to long term officer retention and satisfaction. Lastly, analysts continue collecting data on designated officers in order to retrain machine learning models and improve the AOC-MS designation process and algorithms.

IV. EXPLORATORY ANALYSIS OF AREA OF CONCENTRATION DESIGNATION DATA

HRC and APPD have provided three datasets for the purposes of analyzing the current AOC designation process and designing a new one. APPD operations analysts sanitized all records of personally identifiable information (PII) by, replacing officer names and DOD identification numbers with randomly generated officer ID numbers.

Section A of this chapter explores the assignment and educational of records of 1277 active component MSC officers provided by HRC. A transformed version of this dataset serves as the predictor matrix for all machine learning classifiers used in phase I of AOC-MS.

Section B and Section C evaluate the current AOC designation process and identify possible measures of performance. Section B reviews the records for the AOC designation panels executed during fiscal years 2016, 2017 and 2018 by HRC, assessing how well the current designation process satisfies officer and AOC stakeholder needs. Section C examines the AOC designation targets provided by the MSC objective force model (OFM) and evaluates the current process's ability to sustain MSC OFM manpower requirements.

Lastly, Section D provides a summary assessment of the current AOC designation process and proposes a framework for evaluating AOC-MS.

A. MEDICAL SERVICE CORPS OFFICERS ASSIGNMENT AND EDUCATION RECORDS DATASET

The first dataset contains the assignment and military/civilian educational records of all 1924 active component MFA 70 commissioned officers in the grades of captain through colonel, including 67Js and 70B, as of 20 August 2019. For brevity, we will refer to this dataset as the records dataset.

The purpose of the records dataset is to assess how pre-designation officer assignments and educational experiences relate to AOC designation. While the dataset contains records for all MFA 70 officers, it is nonetheless incomplete as there are no records for officers who left in the active component after designation.

Randomly assigned ID#		Rank		Start Date for Previous Officer Assignments (1 st to last)			Duty Title for Previous Officer Assignments			Civilian Degree Information					Military Course Info.				
										Permanent Designated AOC									
ID#	PERMANENT GRADE CODE	BASIC BRANCH	UNIT NAME_CURRENT	CURR ASGN AREA STA NAME_CURRENT	PREV ASGN AREA FROM DATE	PREV ASGN AREA NO OF MO	PREV ASGN AREA DUTY TITLE	PREVIOUS ASGMT STATION	PREV DUTY AOC	FUNCTIONAL AOC	CIVILIN EDUCATIN LEVEL CD	CIVILIAN EDUCATION DEGREE	CIVILIAN EDUCATION SCHOOL	CIVILIAN EDUCATN SPECIALTY	MILITARY SCHOOLING COURSE	DATE MIL CRS COMPLETED	PY		
107	CPT	MS	0014 MD HSP COMBAT SUPPORT	FT BENNING	20151030	15	BN MEDICAL OFFICER	FT BLISS	70B	70B	K	BACHELOR OF SCIENCE	VA, G.MASON U, FAIRFAX	NUTRITION	AIRBORNE	20130614	201		
107	CPT	MS	0014 MD HSP COMBAT SUPPORT	FT BENNING	20170110	1	BN MED OFFICER (FWD AF)	FT BLISS	70B	NA	NA	NA	NA	NA	ARMY MEDL DEPT AMEDD OFF BAS	20151030	201		
107	CPT	MS	0014 MD HSP COMBAT SUPPORT	FT BENNING	20170224	7	SQDN MEDO (FWD AF)	FT BLISS	70B	NA	NA	NA	NA	NA	UN MOVE OFF CRS	20180209	201		
107	CPT	MS	0014 MD HSP COMBAT SUPPORT	FT BENNING	20170916	8	EXECUTIVE OFFICER	FT BLISS	70B	NA	NA	NA	NA	NA	ARMY MED DEPT AMEDD CPT CRER	20190515	201		
107	CPT	MS	0014 MD HSP COMBAT SUPPORT	FT BENNING	20180529	10	AIDE-DE-CAMP	FT BELVOI	05A	NA	NA	NA	NA	NA	NA	NA	201		
107	CPT	MS	0014 MD HSP COMBAT SUPPORT	FT BENNING	20190408	1	KNOWN LOSSES	FT BELVOI	67A	NA	NA	NA	NA	NA	NA	NA	201		

KEY

Demographic Data

Assignment Data

Civilian Ed Data

Military Ed Data

Start Date for Previous Assignments

AOC Code for Previous Assignment

Promotion Year Group (PYG)

Figure 19. Sample officer Record from Records Dataset

The same line may hold an officer's current assignment, their last military course and their most recent civilian degree information. By contrast, there may be lines with a large number of NA columns as an officer may have more assignments than courses or vice versa.

Additionally, as each officer has a different number of previous assignments, each record has a different number of lines and the lines are only tied together by the unique officer ID.

The most significant shortcoming in the records dataset is its lack of insight into officers who are no longer in the active component MSC, as some could have resigned from active duty or transferred to another branch due to a poor AOC designation.

1. Feature Engineering

In order to transform the records dataset into a numerical predictor matrix that is usable by machine learning classifiers, we transform each multiline officer record into a single-line record utilizing the Pandas (McKinney et al. 2010) and NumPy (Oliphant 2006) libraries for Python (Python Core Team 2019).

a. Assignment Histories

In order to identify assignments relevant to AOC designation, we isolate the pre-designation careers of officers in the records dataset. Assignment histories are coded with an AOC code (i.e. 67J, 70A, 70B, etc.). While there are 12 common AOC codes for MFA 70 officer jobs, there are 174 unique AOC job codes values in the records dataset, with most of them being restricted to fewer than 10 entries. The histograms in Figure 20 and Figure 21 contrast the frequency count for non-MFA70 AOCs in the records dataset against the frequency count of the 12 MFA 70 AOCs. For example, close to 40 non-MFA 70 AOCs have only one entry in the dataset while a single MFA 70 AOC has over 4000 entries.

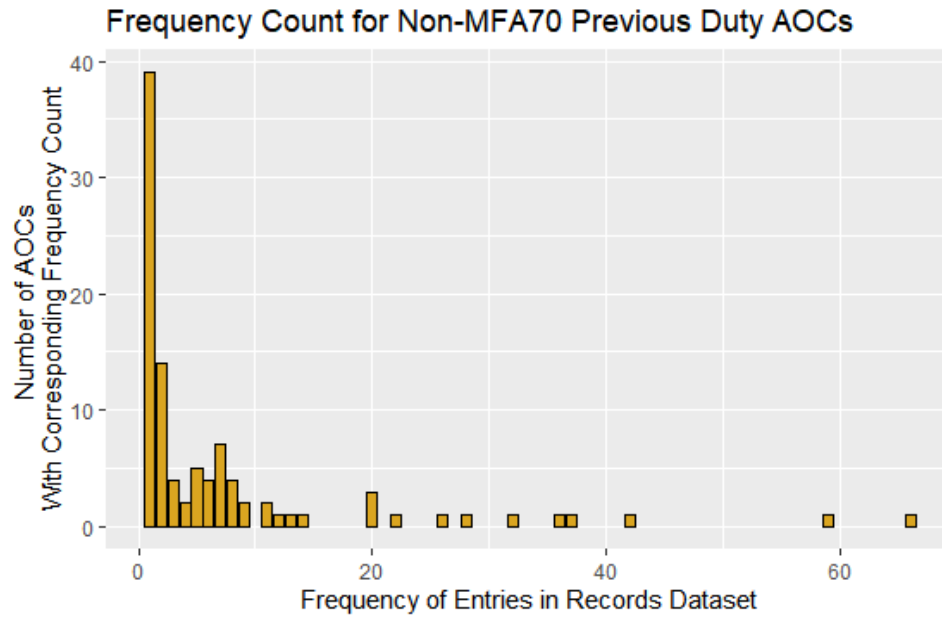
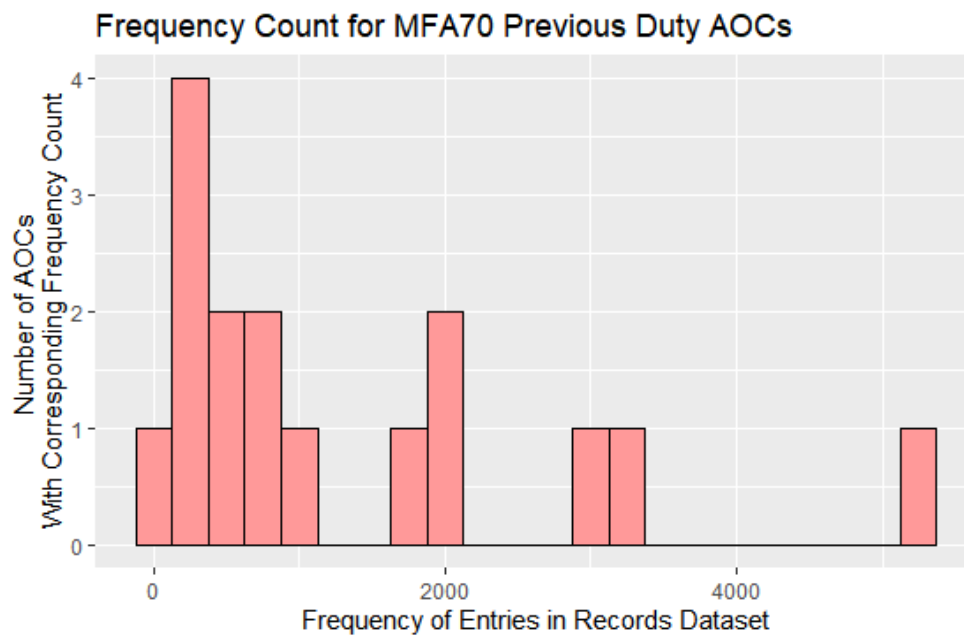


Figure 20. Histogram of Previous Duty AOC Entry Count, Non-MFA70



70F healthcare human resources officer. As such, we replace all 42A entries with 70F entries. We likewise recode non-MFA 70 AOCs without an analogous MFA 70 specialty as healthcare administration generalist assignments (70B).

Since we're interested only in how pre-designation careers relate to AOC designation, we parse through all officer assignment histories in order to identify start dates for initial officer assignments (different for all officers) and go forward to the last assignment start date before their 7th year of service. We count the number of months that an officer spent working under each AOC code prior to designation in order to use them as numerical predictors for AOC assignments.

b. Military and Civilian Education

Military and civilian education entries are free text entries and are not coded. There are 933 unique military course entries and 334 unique civilian degree entries for 1934 officers.

Most educational entries are unique values and some are simple variations of each other. For example, Table 1 displays a sample of entries for administration-related civilian degrees in the records dataset and the histograms in Figure 22 display the frequency of count for civilian degrees and military courses in the records dataset. For example, there are close to 800 military courses and near 250 civilian degrees counting with fewer than a handful of entries.

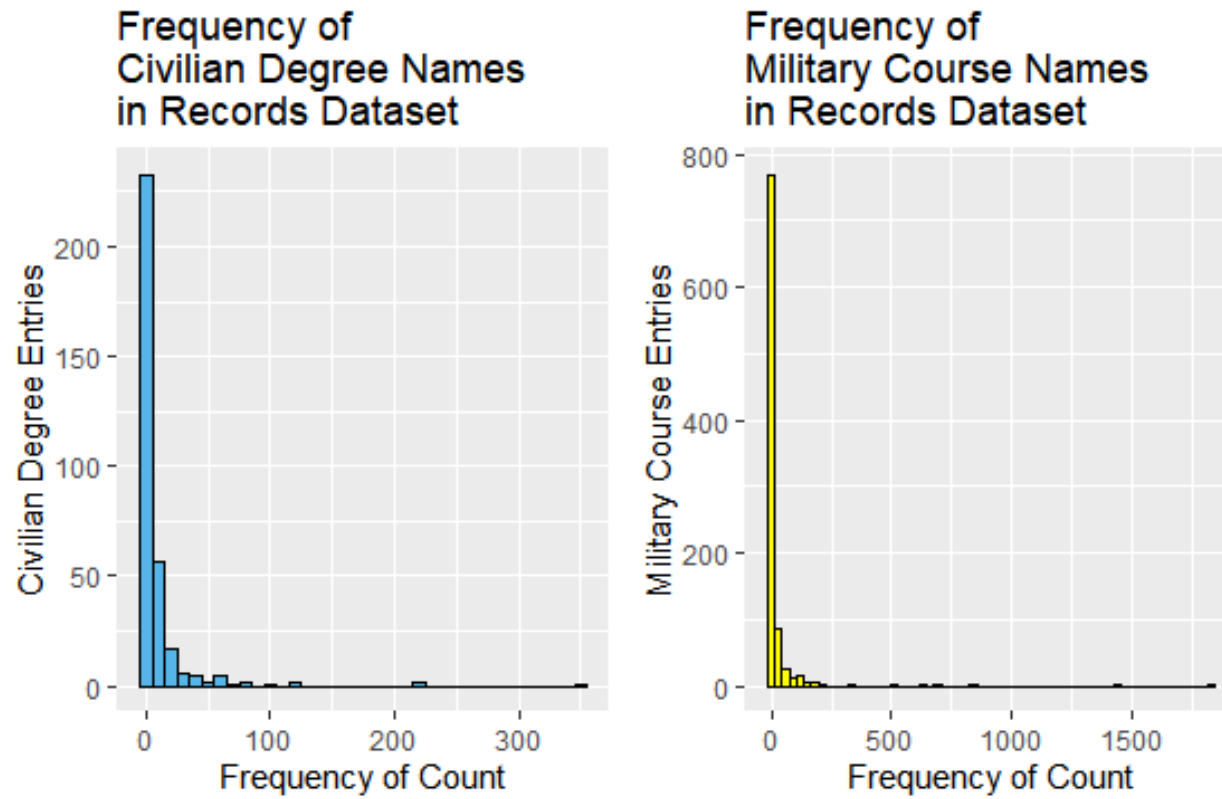


Figure 22. Histograms of Frequency Count for Unique Educational Entries in Records Dataset

Table 1. Administration-Related Civilian Degree Unique Entries in Records Dataset

Administration-Related Degree Unique Entries Records Dataset

BUSINESS ADM	ADMIN/MGT	LEADERSHIP DEVELOPT
HEALTH SVCS ADMIN	M/ADV MIL STU	ADULT, OCC & CONT ED
ADMINISTRATION	HIGH ED ADMIN	CORRECTNL ADMIN
ADMIN PUBLIC	ADMIN EDUC	FITNESS LEADERSHIP MGMT
LEADERSHIP AND MANAGEMENT	ADV MGMT PROG	PARK & REC ADMIN
ORGANIZATIONAL LEADERSHIP	PERS MGT/ADMIN	ADPS-BUSINESS
PUBLIC POLICY ADMIN	STRATEGIC LEADERSHIP	AVN BUS ADMIN
HLTH CARE ADMIN	ADULT LEARNING/LEADERSHIP	ADMIN OF JUSTICE AND SECURITY
PUBLIC ADMINISTRATION	GLOBAL LEADERSHIP	ADMIN LDRSH

In order to transform the free text information into a numerical value related to AOC assignments, we utilize regular expressions to classify civilian degrees and military courses according to AOC-specific educational requirements as outlined in Department of the Army Pamphlet 600-4 (HQDA 2019). For example, the health informatics systems management AOC 70D requires the health information management course and prefers informatics or computer science related degrees. Figure 23 displays the 44 categories into which we classified military courses. Figure 24 displays the 32 categories into which we classified civilian degrees. The majority of entries in both datasets (over 500 for military education and over 300 for civilian degrees) were not associated with a particular MFA 70 AOC.

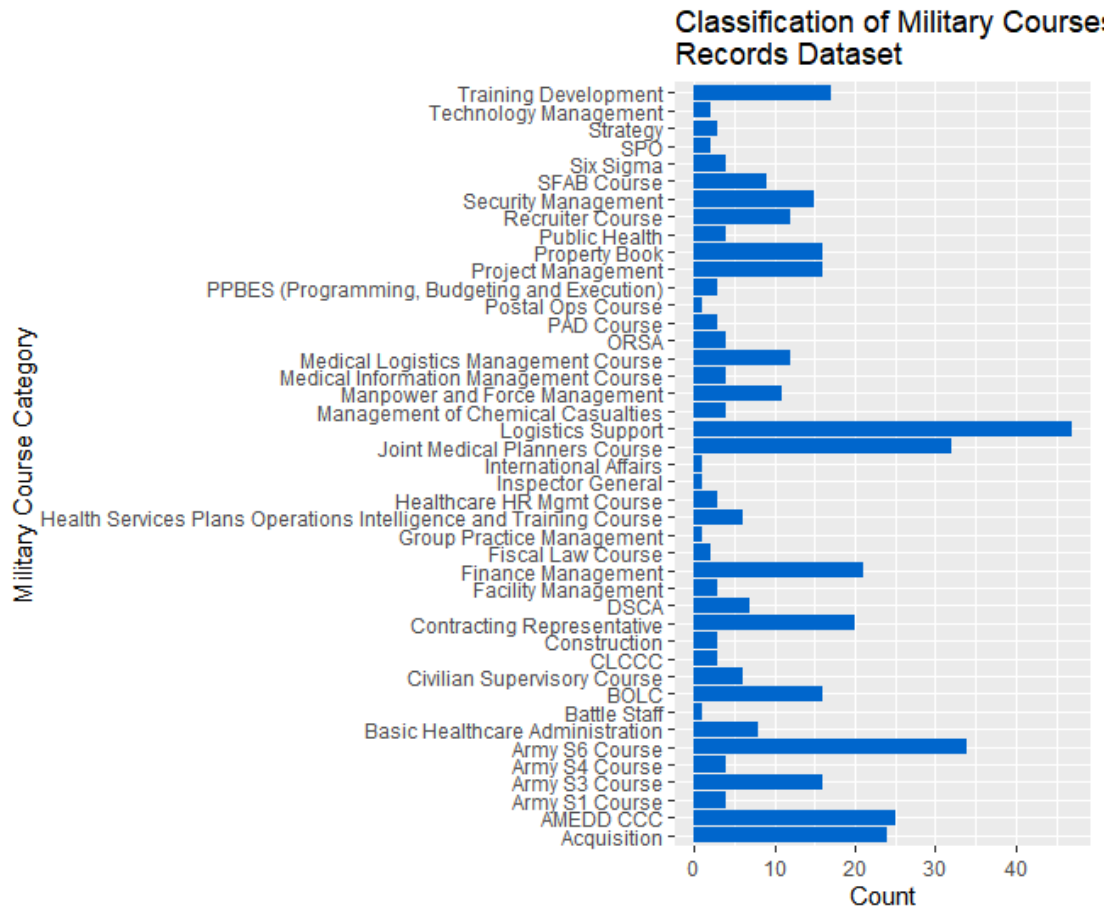


Figure 23. Types of Military Courses in Records Dataset

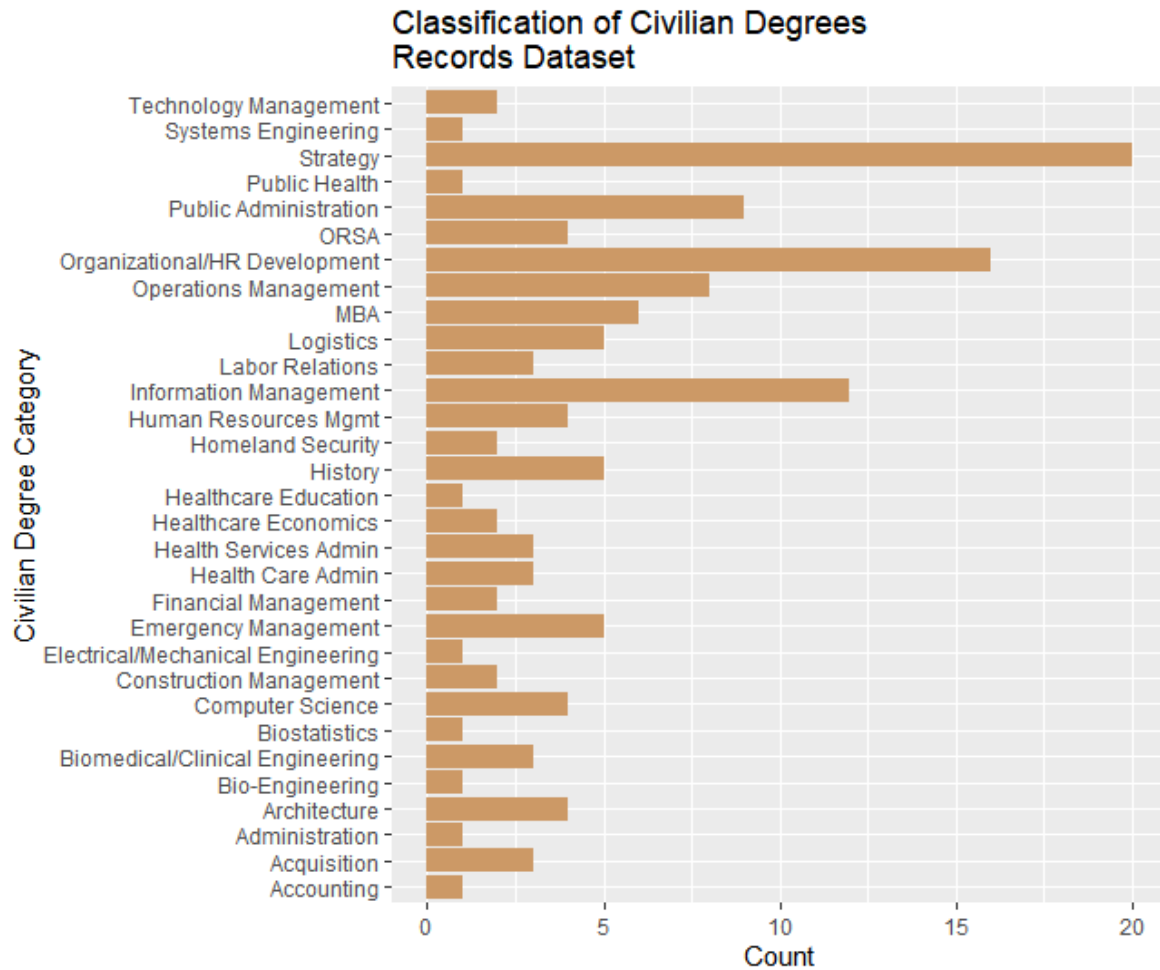


Figure 24. Types of Civilian Degrees in Records Dataset

We then assign numerical values per AOC for each educational entry, depending on its relevance to the AOC as stipulated in DA Pam 600–4 smartbook. Table 2 displays the possible numerical scores assigned to educational entries.

Table 2. AOC-Specific Numerical Scores Assigned to Educational Entries

Numerical Scores Assigned to Educational Entries According to DA Pam 600-4 Smartbook AOC Educational Requirements

	Required	Desired	Not Stipulated
Score	1	0.5	0

We further weigh civilian degree entries according to the degree level. Table 3 displays the weights assigned to each degree level.

Table 3. Civilian Degree Level Numerical Weights

Numerical Weights Assigned to Civilian Degree Entries According to Degree Level				
	Associate	Bachelor	Master	Doctoral
Weight	0.25	0.5	1	1.25

Finally, we parse through each officer's record to isolate pre-designation educational entries and assigned a numerical score per entry, adding them up to arrive on a military and civilian education coefficient per officer and per AOC.

2. The Area of Concentration Predictor Dataset

The result of the data manipulation referenced in Section A, part 1 of this chapter is a data-frame with a single row per officer, featuring numerical counts of months and educational scores for each AOC.

In the AOC predictor dataset, there are two demographic (officer ID and promotion year group) fields not used for prediction, 26 numerical predictors and one response variable reflecting the officer's permanently designated AOC. Table 4 displays the record for a single officer from the AOC predictor dataset for reference. For readability, the record is displayed over three rows.

Table 4. AOC Predictor Dataset Sample

Demographic Information – Not used for prediction		Response Var.		Non-AOC Specific Predictors				AOC Specific Predictors		
Sample Officer Record										
Officer ID	PERMANENT AOC	PYG	00E Months	05A Months	67A Months	67J Months	70A CIV ED	70A MIL ED	70A Months	
564 582	70F	2001	0	13	0	0	0	1.5	0	
70B Months	70C CIV ED	70C MIL ED	70C Months	70D CIV ED	70D MIL ED	70D Months	70E CIV ED	70E MIL ED	70E Months	
564 27	0	1.5	0	0	0.5	0	0	0.5	0	
70F CIV ED	70F MIL ED	70F Months	70H CIV ED	70H MIL ED	70H Months	70K CIV ED	70K MIL ED	70K Months		
564 0	1.5	8	0	0.5	26	0	0.5	0		

a. Multinomial Distribution of AOC Designations

We first analyze the response variable reflecting AOC assignments. Since AOC assignments can fall into one of $a \in A$ categories where $A = \{70A, 70C, 70D, 70E, 70F, 70H, 70K\}$, AOC assignments occur with a probability $p_{a \in A}$. We estimate that AOC assignments follow the multinomial distribution depicted in Equation 6.

$$AOC \sim Mult(p_{70A}, p_{70C}, p_{70D}, p_{70E}, p_{70F}, p_{70H}, p_{70K})$$

Equation 6. Multinomial Distribution of AOC Assignments

Equation 7 reflects the sample estimate for A , the multinomial distributions of AOC Assignments. We used the sample proportions of AOC assignments for the 1277 designated MFA 70 officers in the records dataset. Figure 25 graphically depicts A while Table 5 reflects 95% confidence intervals for the proportion estimates using the Sison & Glaz method (Sison and Glaz 1995).

$$AOC \sim Mult(\hat{p}_{70A} = 0.15, \hat{p}_{70C} = 0.08, \hat{p}_{70D} = 0.06, \hat{p}_{70E} = 0.06, \hat{p}_{70F} = 0.1, \hat{p}_{70H} = 0.31, \hat{p}_{70K} = 0.24)$$

Equation 7. Sample Estimate of AOC Assignment Multinomial Distribution

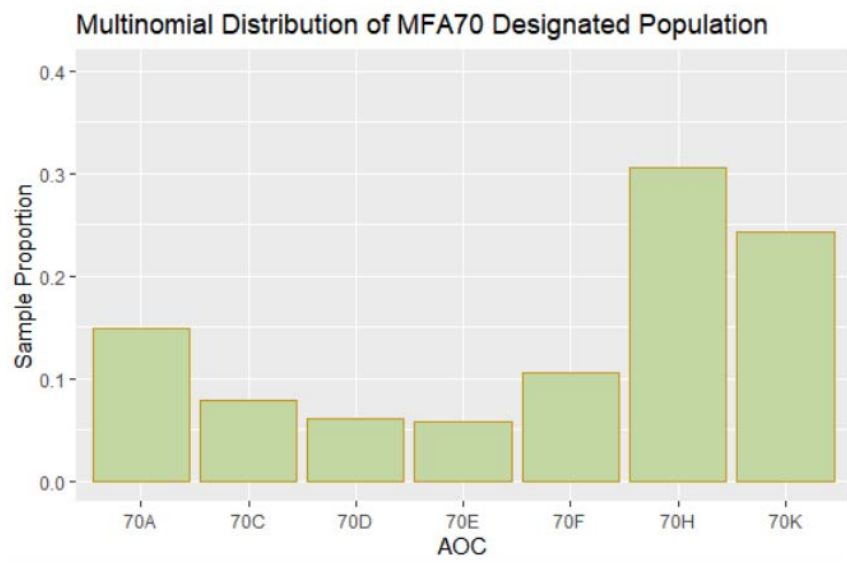


Figure 25. Proportion of AOC Designations in the AOC Predictor Dataset

Table 5. Multinomial Distribution of AOC Assignments

95% Confidence Intervals for AOC Proportions							
	70A	70C	70D	70E	70F	70H	70K
est	0.149	0.079	0.061	0.058	0.106	0.305	0.242
lwr.ci	0.121	0.052	0.034	0.031	0.078	0.278	0.215
upr.ci	0.177	0.107	0.089	0.086	0.134	0.334	0.270

We can observe that 70H is the largest AOC with approximately one third of the designated population. It is closely followed by 70K with approximately one quarter of the population and by 70A with approximately 15% of designated officers. The largest of the low density AOCs is 70F with a tenth of the population. The remaining AOCs (70C, 70D and 70E) hold between 6 and 7 percent of the designated population.

b. Univariate Analysis of Predictor Variables

In the AOC predictor dataset, there are two demographic fields (officer ID and promotion year group) not used for prediction and 26 numerical predictors. There are five non-AOC specific numerical predictors, reflecting the months spent in 00E assignments (student status), 05A assignments (AMEDD immaterial assignments—usually company command or aide-de-camp assignments), 67A assignments (MFA 70 immaterial—

typically executive officer or special staff assignments), 67J assignments and 70B assignments. The remaining 21 predictors are split evenly amongst the seven specialized AOCs, with an entry for the number of months worked in each AOC, one for military education scores and one for civilian education. Table 6 displays summary statistics for the 26 predictor variables.

Table 6. AOC Predictor Dataset Descriptive Statistics

Summary Statistics for AOC Predictor Dataset					
	Mean	SD	Median	Min	Max
00E Months	2.8199	6.7782	0.00	0	71.000
05A Months	14.7533	13.3770	17.00	0	67.000
67A Months	6.3383	11.7001	0.00	0	82.000
67J Months	5.4824	20.2392	0.00	0	123.000
70A CIV ED	0.4513	0.5622	0.00	0	2.750
70A MIL ED	2.2056	0.9540	2.00	0	7.000
70A Months	3.2999	10.0348	0.00	0	97.000
70B Months	31.3853	19.6940	31.00	0	99.000
70C CIV ED	0.5998	0.7356	0.00	0	3.500
70C MIL ED	2.1641	1.0442	2.00	0	14.000
70C Months	1.2796	7.2541	0.00	0	106.000
70D CIV ED	0.1858	0.3172	0.00	0	2.000
70D MIL ED	0.8015	0.4984	0.50	0	5.000
70D Months	1.7087	8.7403	0.00	0	86.000
70E CIV ED	0.3217	0.3720	0.25	0	1.750
70E MIL ED	0.8900	0.5376	0.50	0	4.000
70E Months	2.2060	9.1385	0.00	0	94.000
70F CIV ED	0.2879	0.3957	0.00	0	2.000
70F MIL ED	0.9424	0.6414	0.50	0	5.000
70F Months	5.0525	12.5844	0.00	0	82.000
70H CIV ED	0.3348	0.4710	0.00	0	3.125
70H MIL ED	1.1143	0.7114	1.00	0	4.000
70H Months	12.0752	15.7545	5.00	0	94.000
70K CIV ED	0.4518	0.5671	0.00	0	2.750
70K MIL ED	1.4076	1.1020	1.00	0	9.500
70K Months	11.9460	21.6825	0.00	0	110.000

All predictor variables feature count-like data and therefore follow practically discrete distributions. Figure 26 displays the distribution of non-AOC specific assignment predictors. Figure 27 shows the distribution of AOC-specific assignment predictors, while Figure 28 and Figure 29 feature the distribution of military and civilian educational entries respectively.

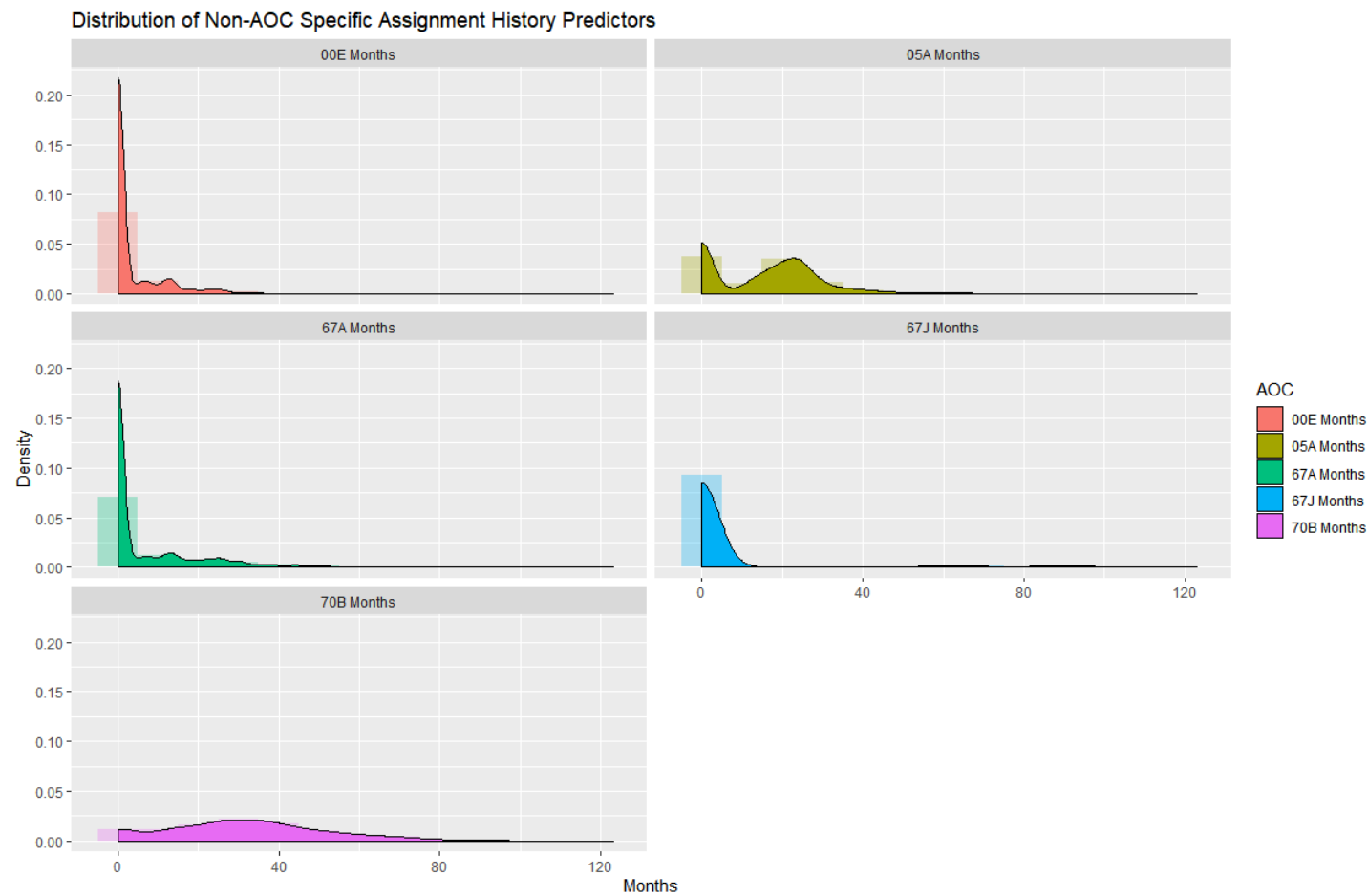


Figure 26. Histogram and Kernel Density Estimate (KDE) of Non-AOC Specific Predictor Variables.

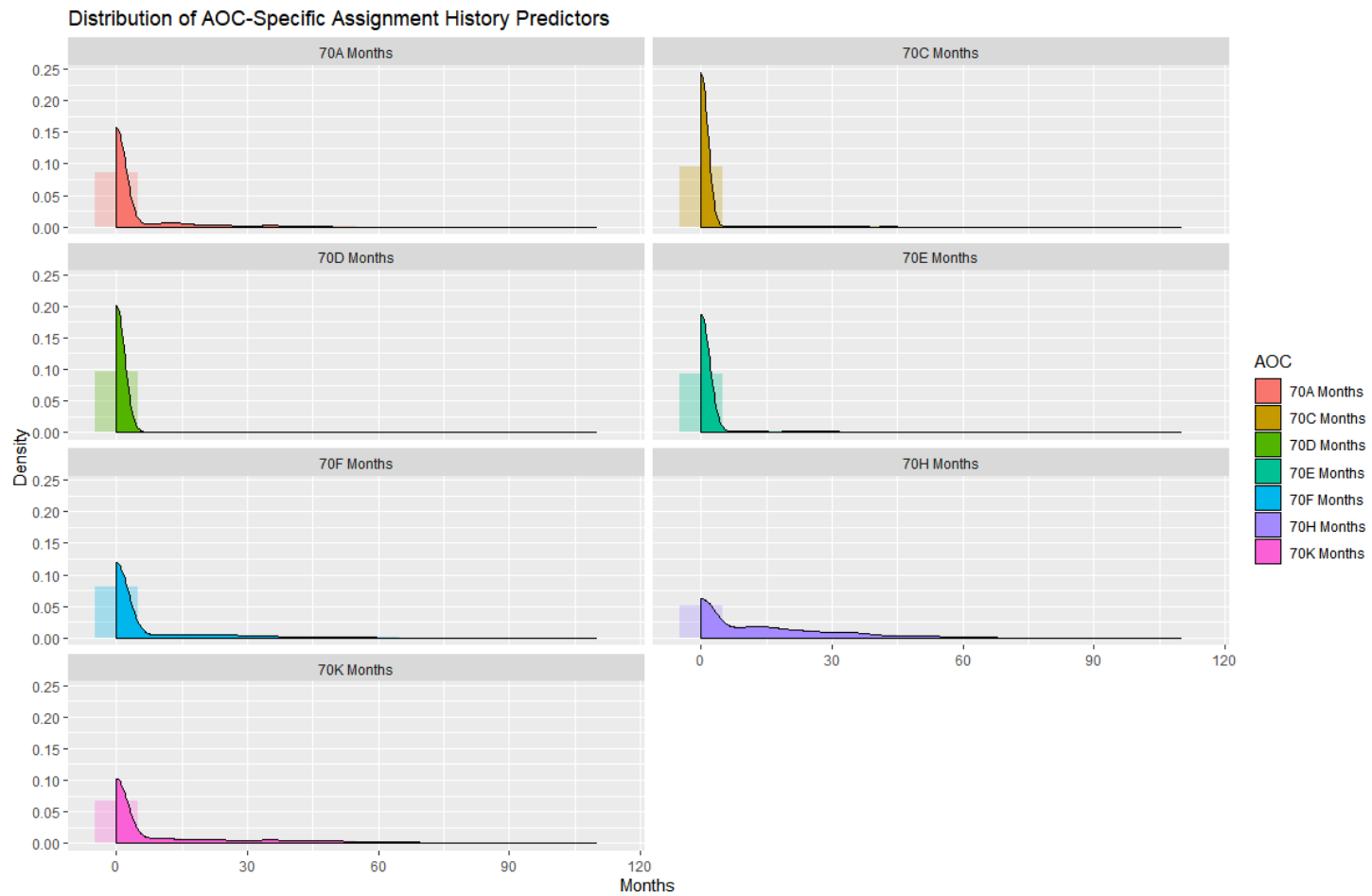


Figure 27. Histogram and Kernel Density Estimate (KDE) of AOC-Specific Assignment History Predictors

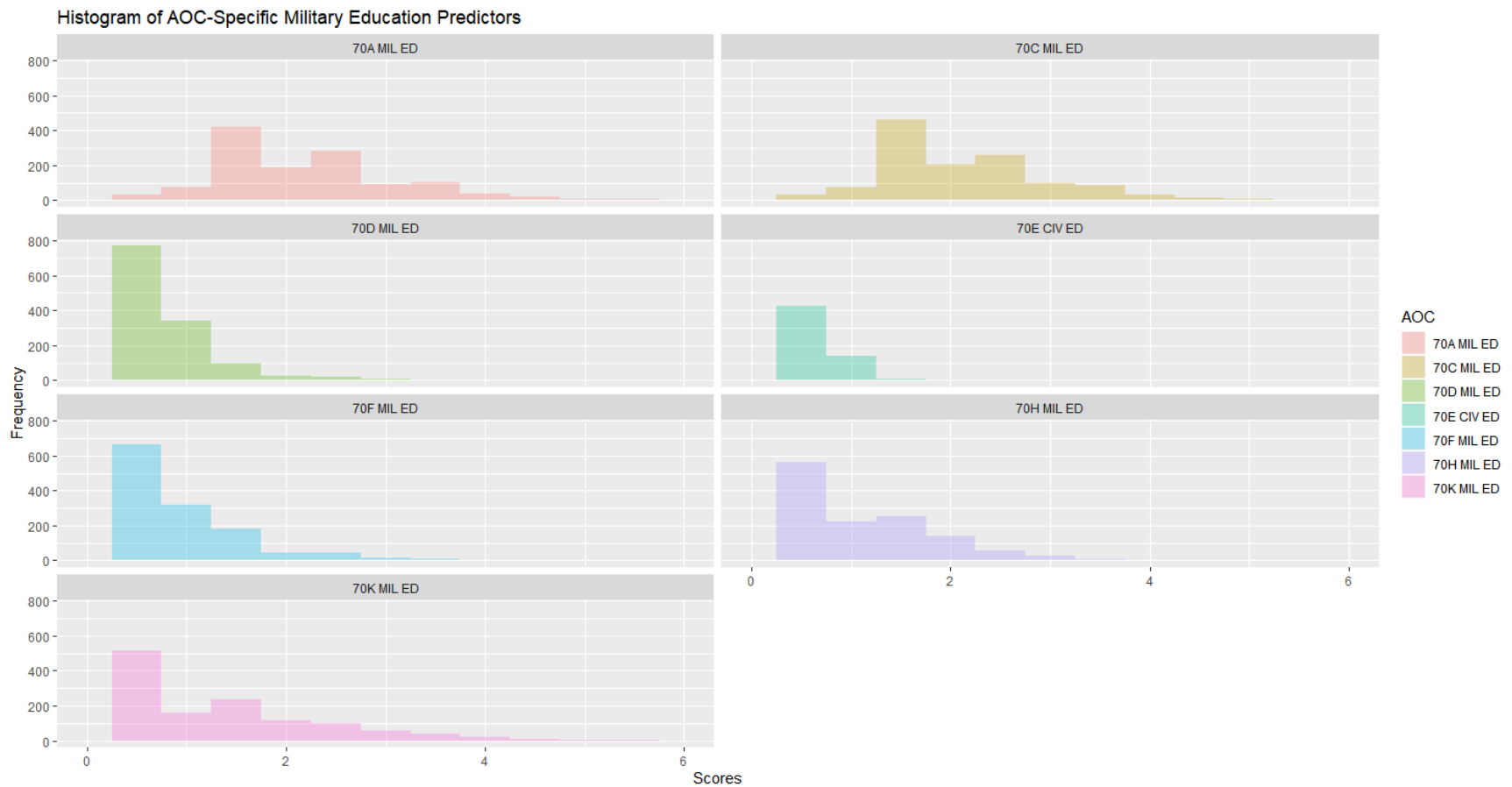


Figure 28. Histogram of AOC-Specific Military Education Predictors

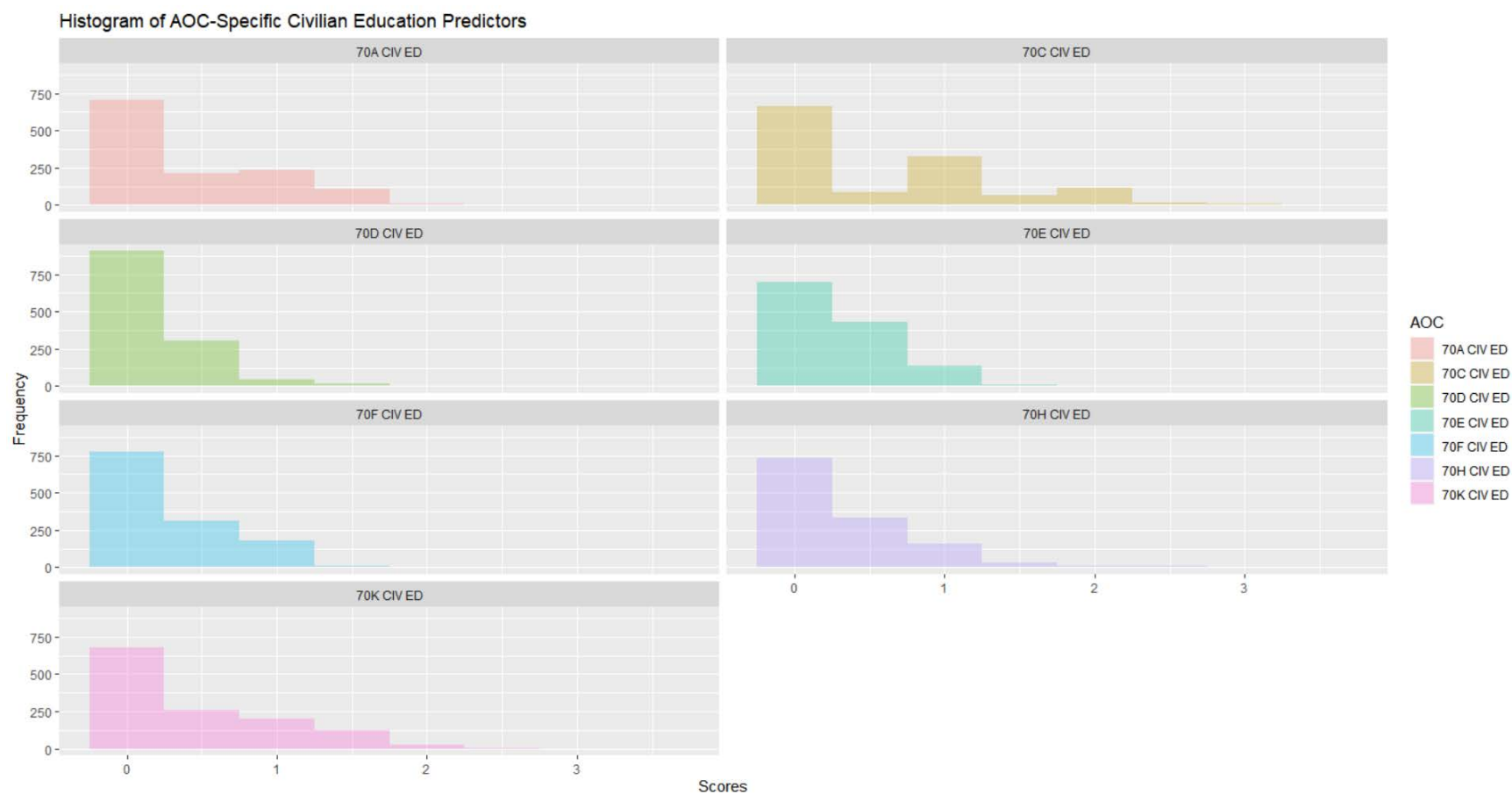


Figure 29. Histogram of AOC-Specific civilian Education Predictors

All predictor variables exhibit unimodal right-skewed distributions. Of these distributions, 20 have a median near zero. In practice this means that with the exception of generalist assignments, officers designating into different AOCs have different experiences from each another. For example, officers who designate as 70F are more likely to work in 70F assignments prior to designation while the majority of non-70F officers will likely have zero months worked in 70F assignments. Section c explores the relationship between AOC designation and AOC-specific predictors in the dataset more deeply.

The remaining predictors with non-zero medians relate to common experiences that the majority of officers of any MFA 70 AOC will experience in their pre-designation careers. For example, 70B and 05A are generalist assignments that most officers hold prior to designation. Since 70H is the largest and most common specialized AOC, the majority of officers have either designated as 70H or have worked in a 70H assignment prior to designation. Additionally, since 70A accepts a variety of administration and business-related degrees while other AOCs have narrower educational requirements, the majority of officers have some 70A civilian education credit.

c. Three-Dimensional Analysis of AOC-Specific Predictors

In the previous section we theorized that there exists a relationship between AOC designation and the three predictors associated with the AOC. The three-dimensional analysis displayed in Figure 30 features a scatterplot per AOC. In each scatterplot, the lower right rear octant reflects low values for the three AOC-related predictors, while the upper left front octant reflects high values for all three predictors. The datapoints are color-coded with green reflecting officers who are not designated to the AOC and purple referencing officers designated to the AOC.

Relationship Between Months Spent in AOC, Military and Civilian Education and AOC Designation

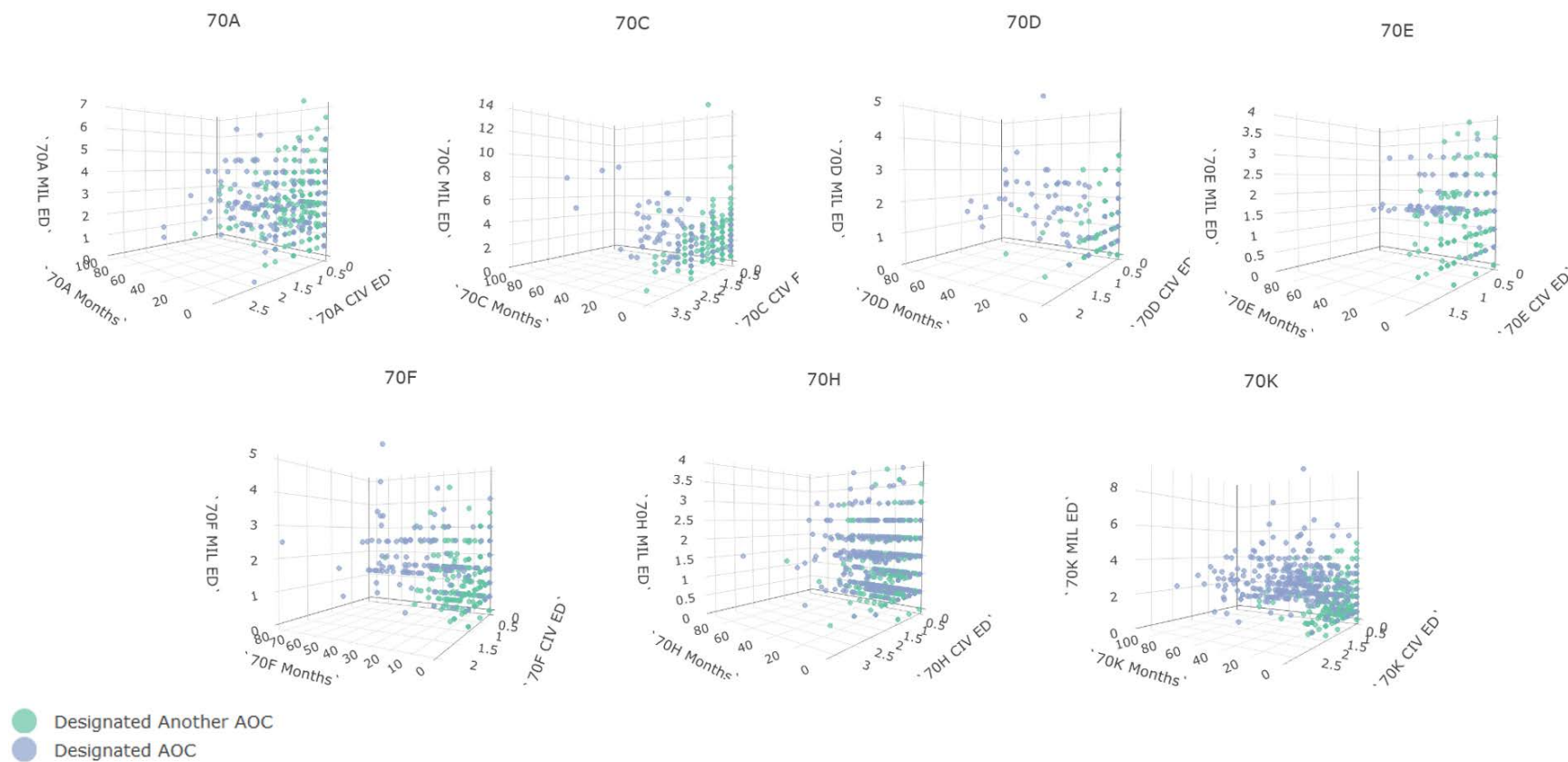


Figure 30. Tri-variate Plots of AOC-specific Predictors per AOC

In all scatterplots, we observe a clear separation between designated officers and non-designated officers. Furthermore, the two groups appear to be approximately linearly separable for all AOCs. In general, designated officers have higher values in two out of the three predictors, with very few having high values in all three predictors.

d. Multidimensional Analysis of Predictor Variables

We utilized dimensionality-reduction techniques in order to best understand the multidimensional relationships and patterns in the 26-dimensional AOC Predictor dataset. We started by centering and scaling the AOC predictor dataset and performing principal component analysis (PCA). Table 7 displays a summary of the PCA results.

Table 7. Principal Components of AOC Predictor Dataset

Principal Component Analysis Summary for AOC Predictor Matrix

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13
Standard deviation	2.127	1.793	1.354	1.296	1.219	1.164	1.137	1.104	1.074	1.042	1.006	0.953	0.921
Proportion of Variance	0.174	0.124	0.071	0.065	0.057	0.052	0.050	0.047	0.044	0.042	0.039	0.035	0.033
Cumulative Proportion	0.174	0.298	0.368	0.433	0.490	0.542	0.592	0.639	0.683	0.725	0.764	0.798	0.831
	PC14	PC15	PC16	PC17	PC18	PC19	PC20	PC21	PC22	PC23	PC24	PC25	PC26
Standard deviation	0.893	0.873	0.866	0.685	0.598	0.528	0.505	0.464	0.426	0.403	0.291	0.229	0.155
Proportion of Variance	0.031	0.029	0.029	0.018	0.014	0.011	0.010	0.008	0.007	0.006	0.003	0.002	0.001
Cumulative Proportion	0.862	0.891	0.920	0.938	0.952	0.962	0.972	0.981	0.988	0.994	0.997	0.999	1.000

The first 18 of 26 principal components explain 95% of the variance in the dataset. Figure 31 shows the percentage of dataset variance explained by each of the 26 principal components.

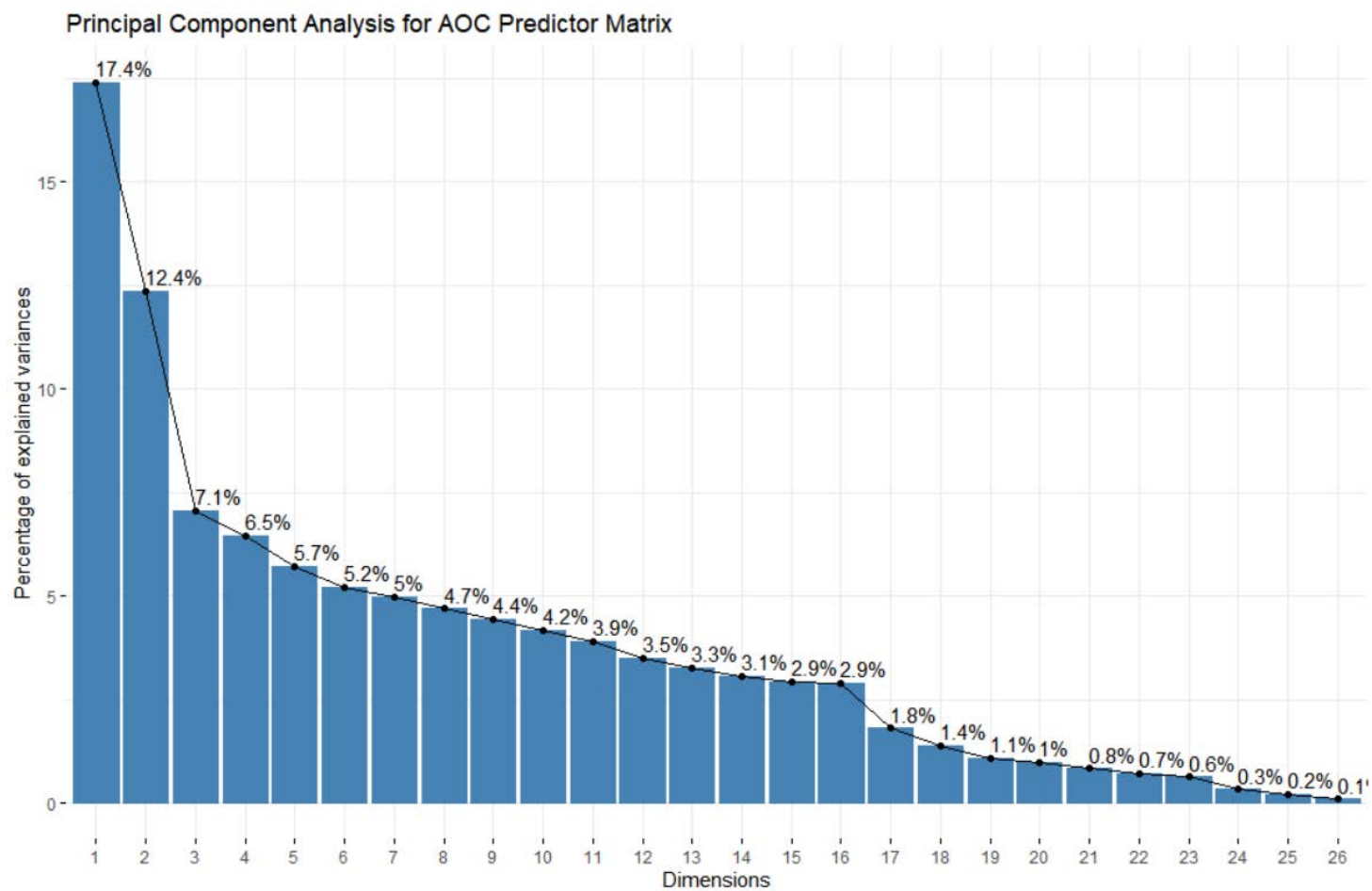


Figure 31. Percentage of Variance Explained by Principal Component

Figure 32 shows two-dimensional plots of the first two pairs of principal components, displaying the contribution of each predictor variable. civilian education entries are the primary contributors for the first principal component, which explains 17.4% of the dataset's variance. The civilian education entries for 70A and 70C have the largest loadings for the first principal component.

Military education entries contribute the most to the second principal component, explaining 12.4% of the dataset's variance. The military education entries for 70D, 70E, 70K, 70H and 70F contribute the most to this principal component. For the third and fourth principal components, the most variables with the largest loadings were assignment entries for 70F, 70K and 70H.

The relative magnitude of the loadings for the first few principal components make sense in the context of the problem. According to DA Pam 600–4, 70A and 70C are the two AOCs with minimum civilian education entry requirements, while 70D, 70E, 70K, 70H and 70F have military education entry requirements (HQDA 2019). The composition of the two most important principal components suggests that the educational requirements for each AOC explain approximately 30% of the dataset's variance.

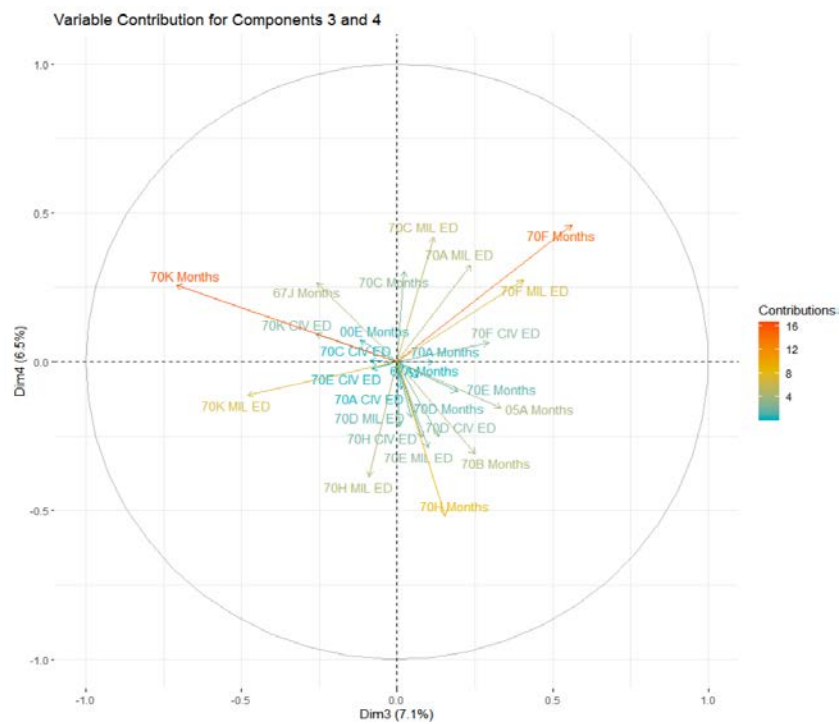
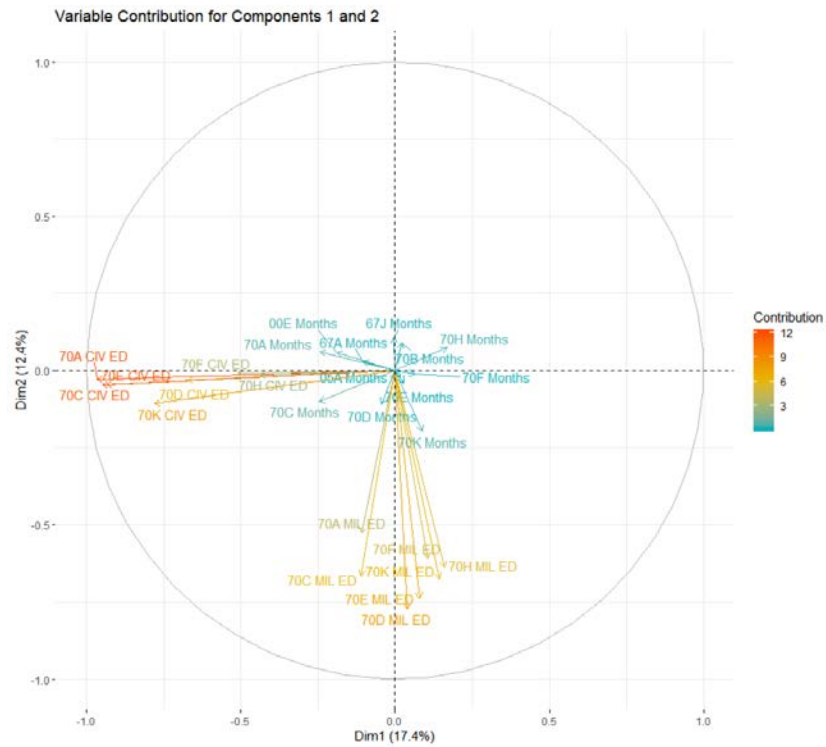


Figure 32. PCA Contributions per Variable for Principal Components 1–4.

In order to best understand how the first two principal components reflect separation between officers by AOC, we plot the PCA entries for a random sample of 100 officers in Figure 33. We can observe some concentration of officers by designated AOC within the first two principal components, suggesting the officer records could be separable in higher dimensions.

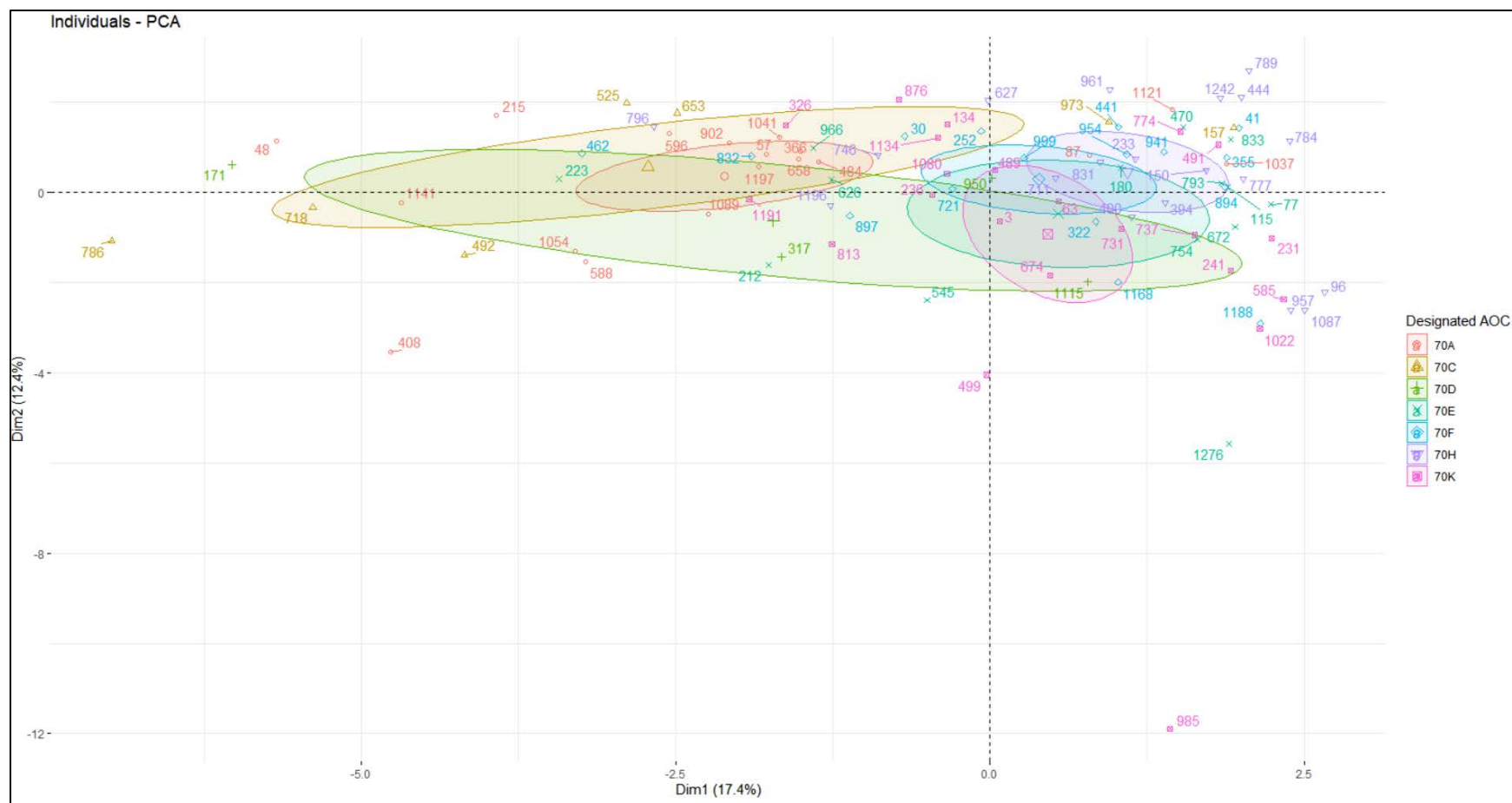


Figure 33. PCA Individual Visualizations for Components 1 and 2

We continue the analysis by mapping the centered and scaled AOC predictor dataset down to a three-dimensional Euclidean space using classical multi-dimensional scaling (MDS). Figure 34 shows the three-dimensional mapping of the AOC predictor dataset colored by AOC of assignment. We can observe separation between 70Hs, 70Ks and 70Fs. Figure 35 removes the mapping for these three AOCs and shows that 70A, 70C, 70D and 70E officers also form clusters in certain parts of the plot, although to a smaller degree. This separation also makes sense in the context of the AOC designation problem, as the majority of AOC-specific captain billets are for 70H, 70K and 70F officers.

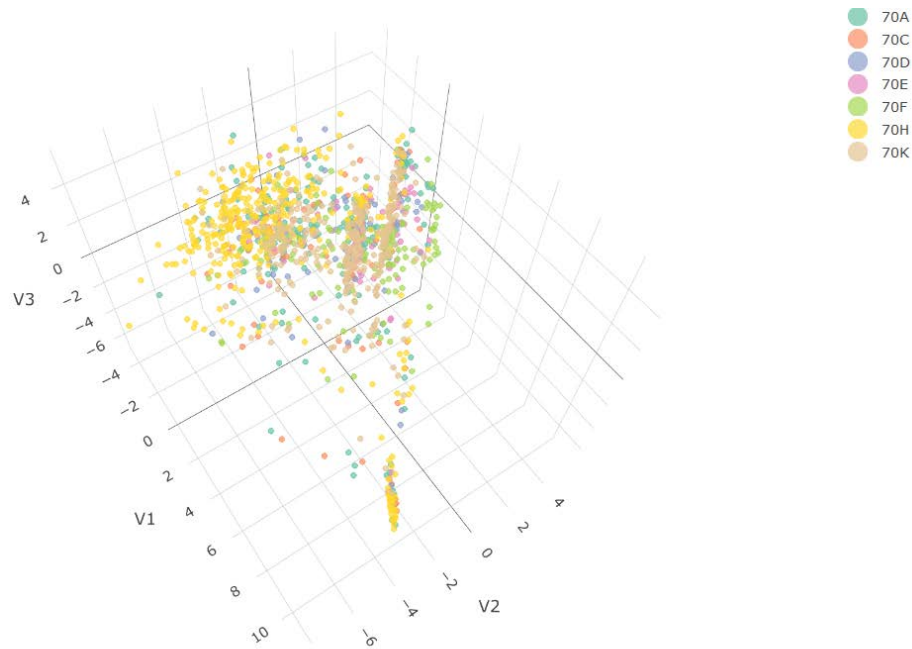


Figure 34. Three-Dimensional Mapping of AOC Predictor Dataset

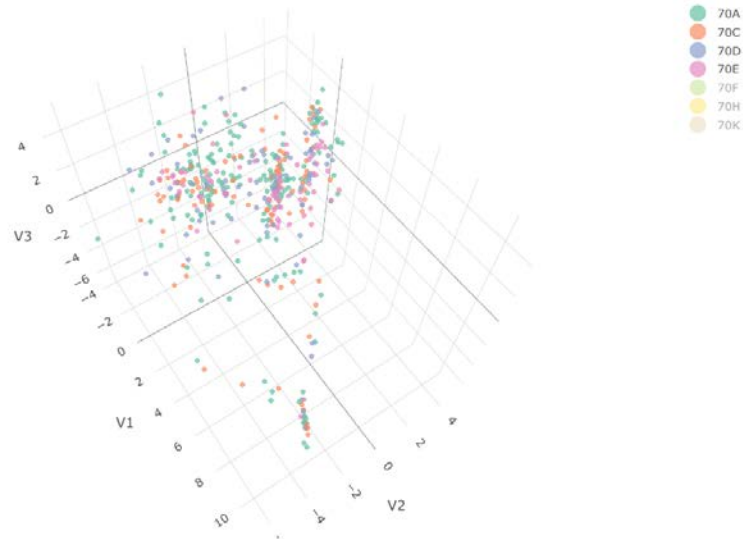


Figure 35. Three-Dimensional Mapping of AOC Predictor Dataset with 70F, 70H and 70K Entries Removed

We also visualize the dataset in three-dimensions using T-distributed stochastic neighbor embedding (t-SNE) and a perplexity parameter of 5. Figure 36 shows more pronounced separation patterns using this visualization technique than with MDS, with all AOCs separating into clear clusters.

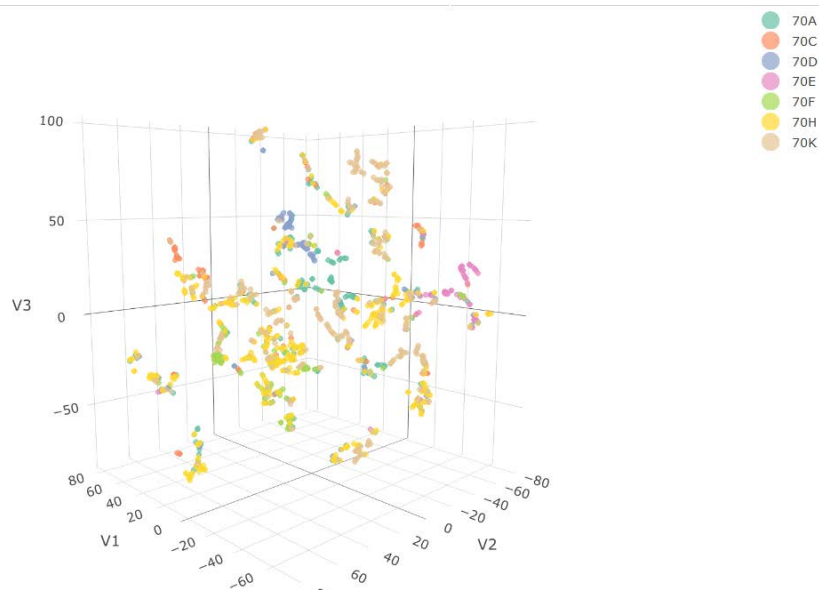


Figure 36. t-SNE Three-Dimensional Visualization of AOC Predictor Dataset

Finally, in order to train the machine learning classifiers, we randomly split the AOC Predictor dataset into a training set with 80% of the records (1021) and a test set with 20% of the records (256).

B. FISCAL YEAR 16–18 MEDICAL SERVICE CORPS OFFICER AREA OF CONCENTRATION PREFERENCES AND PANEL VOTING RECORDS

1. The Area of Concentration Designation Dataset

The second dataset contains the AOC preferences and panel voting records for the AOC designation processes executed between FY2016 through FY2018. The dataset consists of one Excel spreadsheet per fiscal year.

Each year, the HRC 70B Assignments officer assembles the vote sheets upon reviewing AOC preferences submitted by officers and the OFM targets provided by HRC. The 70B Assignments officer then retrieves the record for each officer and manually populates the officer's preferences and demographic data including whether the officer has been accepted into or participated in an LTHET program. The 70B Assignments officer also provides each consultant with OFM designation targets per AOC.

The AOC consultants for AOCs 67J, 70A, 70B, 70C, 70D, 70E, 70F, 70H and 70K utilize the spreadsheets to render AOC designation votes and panel scores for each officer. Finally, the 70B Assignments officer compiles the vote sheets and manually tallies AOC votes, recording the AOC with the majority vote as the final designation AOC. There are sixteen different tabs per spreadsheet, consisting of one vote sheet for each of the nine consultants, and three designation sheets for 70Bs and 67Js, with one for the promotion year in mandatory designation and two for the opt-in year groups.

The purpose of analyzing the designation dataset is to assess the effectiveness of the current AOC designation process, establish measures of performance for the AOC designation process and compare its results with those of AOC-MS. The spreadsheets are designed for ease of visual inspection but not for data analysis. Tab names, column headers and column orders are inconsistent across sheets and workbooks. Similar information is captured in different manners in different spreadsheets. Therefore, the designation dataset required extensive manipulation in order to be analyzed.

The result of the manipulation is a single data-frame with 438 designation records for 417 MSC captains. The duplicate designation records refer to officers who opted in unsuccessfully on a particular FY but weren't designated until another FY. The designation dataset also provides the only available insight into officers who have left active duty, with 66 such records in the dataset.

There are 59 possible columns per each officer's designation record. The columns contain demographic information, consultant votes, panel scores, designation data and post-designation analysis conducted by HRC. Despite the extensive manipulation, the consolidated designation dataset features a missing data rate of nearly 10%, as reflected in Figure 37, created with the missing no package for Python (Bilogur 2018). The majority of the missing data related to missing preferences, lack of LTHET data and missing votes by the 70F consultant.

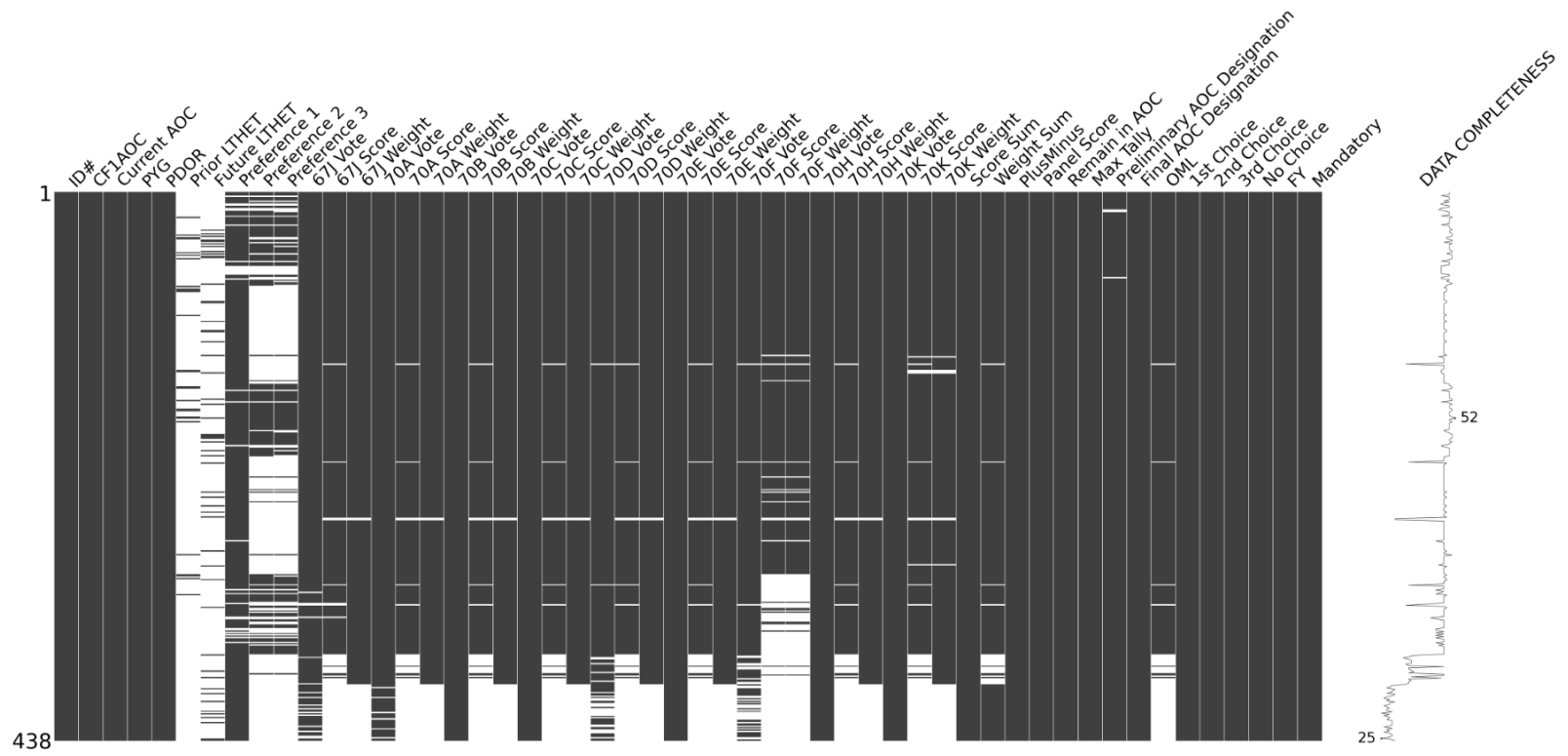


Figure 37. Designation Dataset Missing Data Rates

2. Univariate Exploratory Analysis of the Designation Dataset

Given the large number of variables and the fact that we will not use the designation dataset for prediction, we focus our analysis on selected variables which are instrumental for assessing the effectiveness of the current AOC designation process. The majority of datapoints are categorical in nature.

Table 8 provides summary statistics for the AOC Designation Panel Score utilizing the summary tools (Comtois 2020) exploratory analysis package for R while Table 9 displays proportions for 14 categorical variables using the skimr (Waring et al. 2020) package.

Table 8. Exploratory Analysis for AOC Designation Panel Score

Variable	Number Missing	Complete Rate	Mean	SD	Min	25th Percentile	Median	75th Percentile	Max	Histogram
64 Panel.Score	71	0.838	4.57	0.655	2.11	4.22	4.67	5	6	

The most relevant numerical variable is the panel score, which is the average of the word picture scores rendered by each of the nine panel members when assessing officer performance. In theory, the scores could range from 1 to 6.5; however, the actual scores for the cohort exhibit a right-skewed distribution ranging from just above 2 to 6, with a median of 5.

Table 9. Exploratory Analysis for Key Categorical Variables in AOC Designation

No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
1	Current.AOC [factor]	1. 67J 2. 70B	59 (13.5%) 379 (86.5%)		438 (100%)	0 (0%)
2	PVG [factor]	1. 2008 2. 2009 3. 2010 4. 2011 5. 2012	75 (17.1%) 106 (24.2%) 145 (33.1%) 69 (15.8%) 43 (9.8%)		438 (100%)	0 (0%)
3	LTHET [factor]	1. (Empty string) 2. 70A 3. 70C 4. 70D 5. 70F 6. 70H 7. 70K 8. 70K70K	0 (0.0%) 24 (30.8%) 15 (19.2%) 2 (2.6%) 5 (6.4%) 8 (10.3%) 24 (30.8%) 0 (0.0%)		78 (17.81%)	360 (82.19%)
4	Preference.1 [factor]	1. 67J 2. 70A 3. 70C 4. 70D 5. 70E 6. 70F 7. 70H 8. 70K	16 (3.9%) 79 (19.4%) 31 (7.6%) 28 (6.9%) 36 (8.8%) 38 (9.3%) 87 (21.4%) 92 (22.6%)		407 (92.92%)	31 (7.08%)
5	Preference.2 [factor]	1. 67J 2. 70A 3. 70C 4. 70D 5. 70E 6. 70F 7. 70H 8. 70K	9 (5.7%) 42 (26.8%) 12 (7.6%) 10 (6.4%) 15 (9.6%) 15 (9.6%) 29 (18.5%) 25 (15.9%)		157 (35.84%)	281 (64.16%)
6	Retention [factor]	1. Not Retained 2. Retained as of Dec 2019	66 (15.1%) 372 (84.9%)		438 (100%)	0 (0%)
7	Participation [factor]	1. Did Not List Preferences 2. Listed Preferences	31 (7.1%) 407 (92.9%)		438 (100%)	0 (0%)

No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
1	Preference.3 [factor]	1. 67J 2. 70A 3. 70C 4. 70D 5. 70E 6. 70F 7. 70H 8. 70K	5 (3.4%) 29 (19.5%) 20 (13.4%) 14 (9.4%) 19 (12.8%) 20 (13.4%) 19 (12.8%) 23 (15.4%)		149 (34.02%)	289 (65.98%)
2	Final.AOC.Designation [factor]	1. 67J 2. 70A 3. 70B 4. 70C 5. 70D 6. 70E 7. 70F 8. 70H 9. 70K	18 (4.1%) 63 (14.4%) 33 (7.5%) 36 (8.2%) 33 (7.5%) 30 (6.8%) 41 (9.4%) 111 (25.3%) 73 (16.7%)		438 (100%)	0 (0%)
3	Preference Level [factor]	1. First Choice 2. Second Choice 3. Third Choice 4. Needs of the MSC	348 (79.5%) 18 (4.1%) 7 (1.6%) 65 (14.8%)		438 (100%)	0 (0%)
4	Max.Tally [factor]	1. 3 2. 4 3. 5 4. 6 5. 7 6. 8 7. 9	2 (0.5%) 7 (1.6%) 34 (7.8%) 44 (10.1%) 65 (14.8%) 122 (27.9%) 164 (37.4%)		438 (100%)	0 (0%)
5	Mandatory [factor]	1. 0 2. 1	241 (55.0%) 197 (45.0%)		438 (100%)	0 (0%)
6	Unanimity Level [factor]	1. 3 Votes or Fewer 2. 4-6 Votes 3. 7 Votes or More	43 (9.8%) 231 (52.7%) 164 (37.4%)		438 (100%)	0 (0%)
7	LTHET.Mismatch [factor]	1. LTHET Mismatch 2. LTHET AOC Awarded	4 (5.1%) 74 (94.9%)		78 (17.81%)	360 (82.19%)

Table 9 displays summary statistics for key categorical variables. Overall, 86% of designating officers come from AOC 70Bs while approximately 14% are 67Js. The designated officers come from PYGs 2008 through 2012, with the majority of officers coming from PYGs 2009 and 2010. 82% of officers had not been selected for LTHET at the time of AOC designation. Of the officers selected for LTHET, 5% were designated to an AOC which differed from that of the LTHET selection. 55% of officers designated during their mandatory window while the remaining 45% opted into early designation.

Panel members reached 53% of AOC designation decisions with 4–6 votes, while 37% of decisions were nearly unanimous. Fewer than 10% of decisions were decided by 3 votes or fewer, suggesting there is general consensus amongst panel members regarding designation decisions.

Just under 85% of designating officers remained on active duty as of December of 2019. 93% of designating officers listed AOC preferences, while 7% chose not to actively participate in the designation process.

Table 10 reflects how well the current AOC designation process honors officer preferences. Officers received their first choice almost 80% of the time; however, nearly 15% of officers were not designated into one of their three preferences. The majority of these were failed early opt-ins where were designated to remain 70Bs.

Table 10. Officer AOC Satisfaction Levels

First Choice	Second Choice	Third Choice	Needs of the MSC
0.795	0.041	0.016	0.148

3. Multivariate Exploratory Analysis of the Area of Concentration Designation Dataset

Univariate analysis of the AOC designation dataset suggests two qualitative measures of performance for the AOC designation process: officer preference satisfaction and retention rates. However, these measures of performance are officer-centric and fail to account for how the process meets MSC force structure needs or

whether the AOC match is good for the Army. Multivariate analysis of the AOC designation dataset helps answer these questions by examining the relationships between variables and assessing the current process's effectiveness in satisfying officer and MSC stakeholder needs.

Figure 38 provides additional context for the information presented in Table 10. The figure shows that when excluding officers who opted in for early designation but remained 70Bs or 67Js, the current designation process awarded officers their top AOC preference 85% of the time, the second preference 4% of the time and their third preference 1% of the time. Approximately 8% of officers were designated according to the needs of the MSC. However, 97% of those designated according to the needs of the MSC failed to list preferences for AOC designation. For the remainder of this chapter, unless explicitly stated, all analysis will exclude failed opt-ins

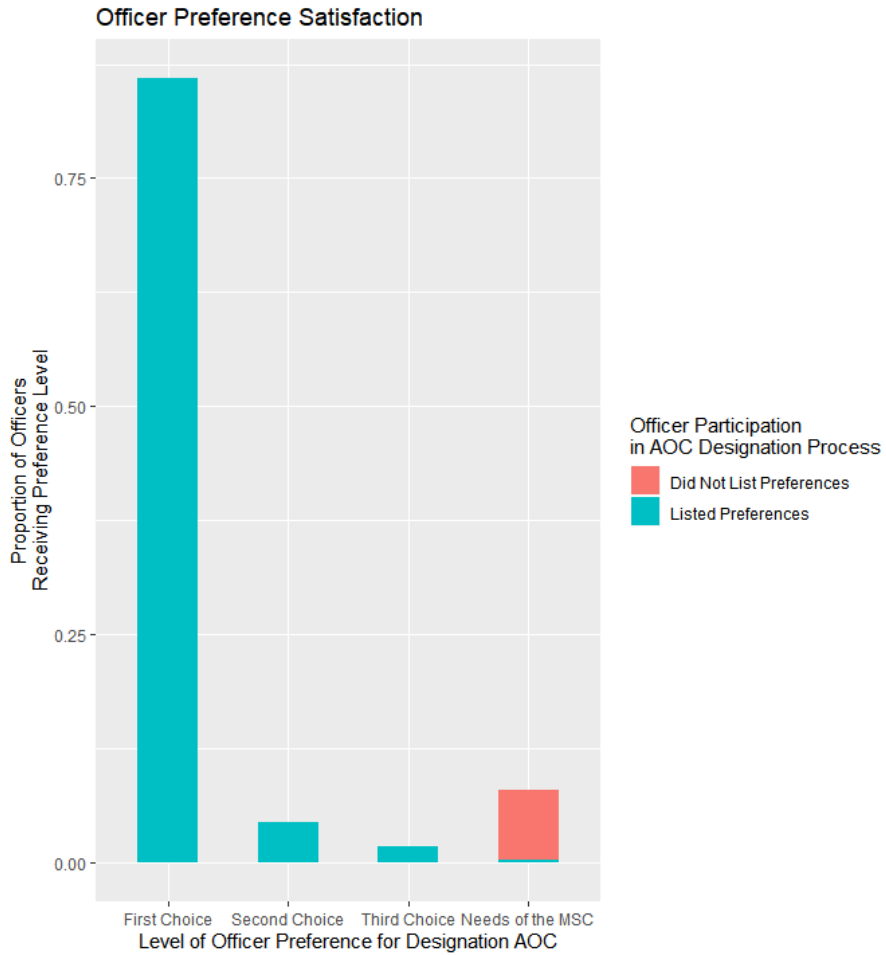


Figure 38. Officer AOC Satisfaction Levels Excluding Failed Opt-ins

a. AOC Popularity

We next examine how officer preferences shape the AOC designation process. Table 11 displays the rate at which officers rank a particular AOC as their first choice compared to the proportion of force structure requirements for that particular AOC. 70A, 70E, 70C and 70K are clearly popular amongst 70Bs and 67Js. Other AOCs, such as 70F, 70D and 70H are not preferred at sufficient rates to support their manpower requirements.

Table 11. Officer AOC Preferences as a Proportion of AOC Manpower Allocations

Popularity Rates for MFA70 AOCs

70A	70E	70C	70K	70F	70D	70H
1.52	1.25	1.02	0.934	0.895	0.828	0.776

Figure 39 further explores the relative relationship between officer preferences and AOC designation, suggesting some differences between AOCs. While the majority of designees for each AOC received their first choice, a greater proportion of officers designated according to the needs of the MSC into 70H and 70D than into other AOCs, while almost all 70A and 70K designees received their first choice.

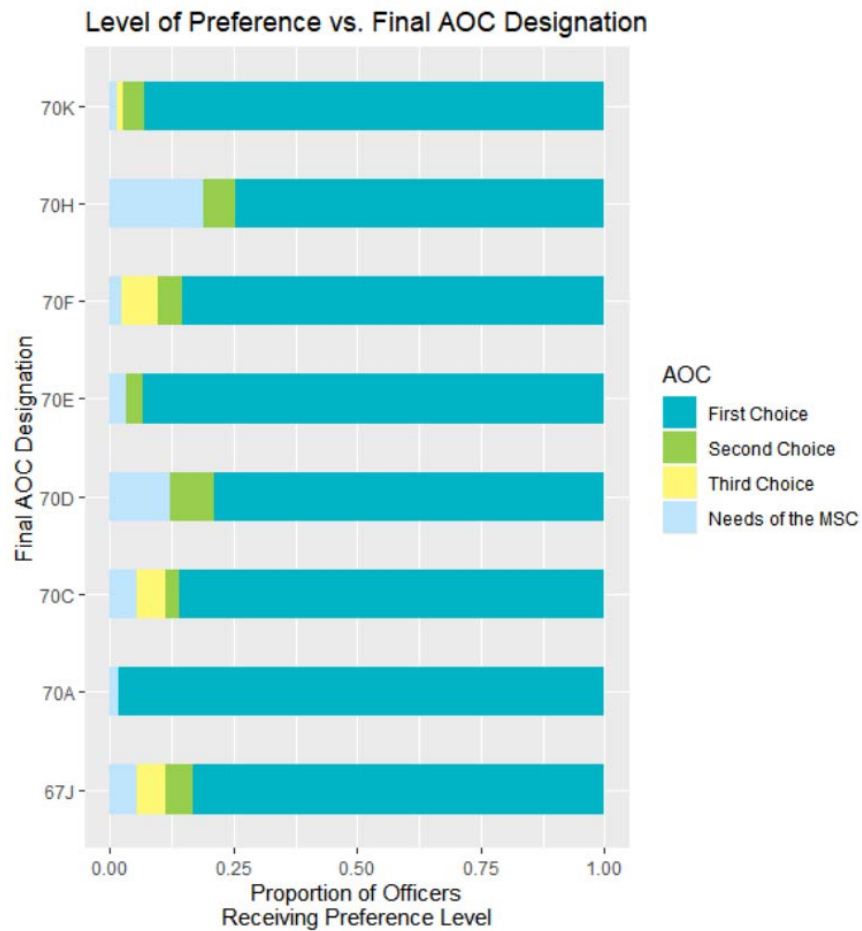


Figure 39. Preference Level Attained by Designating Officers per AOC

AOCs with high proportions of unwilling officers are more likely to face future retention challenges and force structure imbalances. Figure 40 shows that the popularity pattern does appear to be positively correlated with retention. AOCs 70D and 70H, with the highest proportion of unwilling designees, also faced the most significant retention challenges, with a near 75% retention rate. By contrast, popular AOCs like 70A and 70K feature retention rates above 90%.

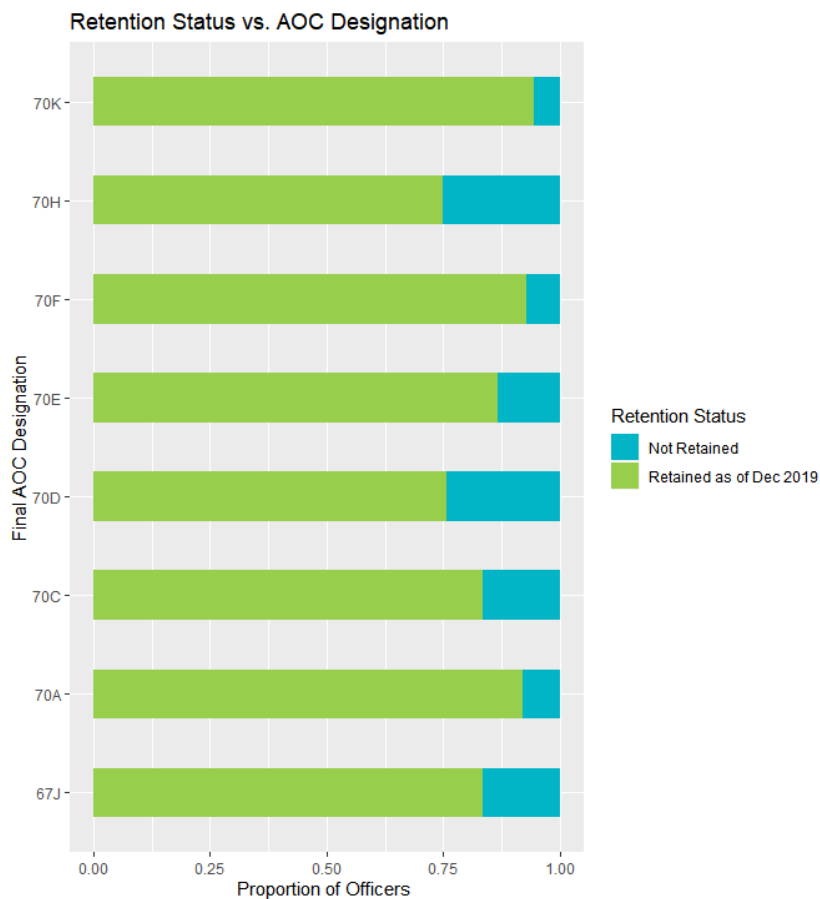


Figure 40. Level of Retention per AOC Assignment

By allowing officers to list preferences for all AOCs instead of for only the top three, AOC-MS gives the MSC more opportunities to honor officer wishes. By providing AOCs with a framework for recruiting officers and shaping their future preferences, AOC-MS seeks to minimize the popularity gap between AOCs.

We can assess the effectiveness of recruitment efforts in phase I of AOC-MS in bridging the popularity gap between AOCs by gauging pre-designation AOC interest. One measure of pre-designation interest is attendance to AOC-producing courses for AOCs 70D, 70E, 70F, 70H, and 70K, which list AOC-producing courses as requirements for AOC designation and LTHET applications for AOCs 70A and 70C, which require graduate educations (HQDA 2019).

Figure 41 examines whether officers attended AOC-producing courses prior to applying for designation. It shows that officers designating into AOCs 70K and 70E generally had completed the respective AOC-producing course prior to designation, while AOCs 70D and 70H featured large proportions of officers designating without prior attendance to the AOC-producing courses. This difference appears to relate to AOC popularity amongst officers and to the relatively large proportions of officers designated into less popular AOCs according to the needs of the MSC.

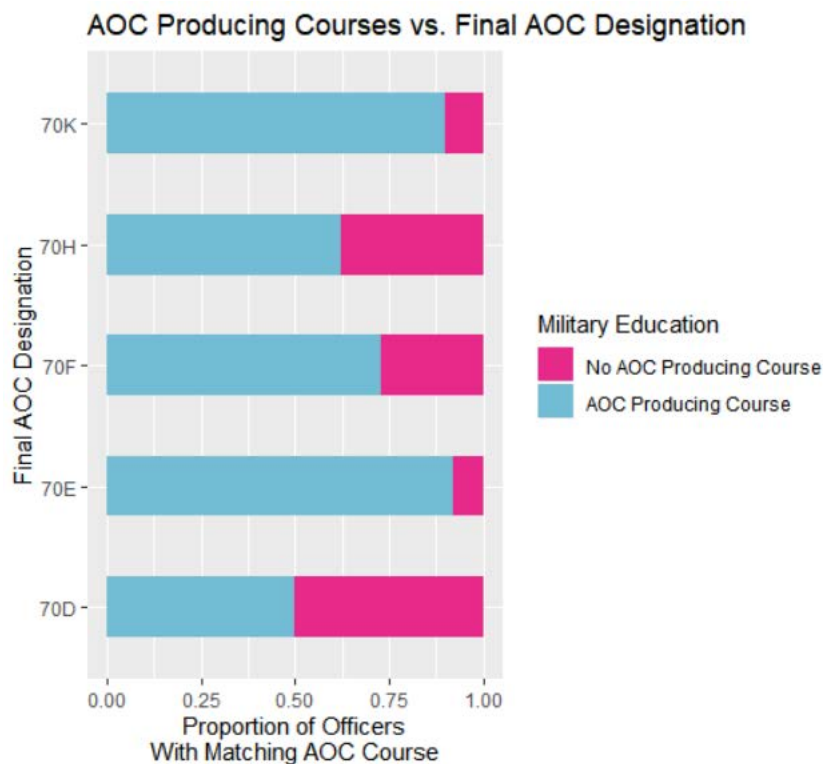


Figure 41. AOC Producing Course Attendance vs. AOC Designation (AOCs 70D, 70E, 70F, 70H and 70K)

Figure 42 shows that officers designating into AOCs 70A and 70C generally have not applied to LTHET for their designating AOC. Designating officers who have not applied for or achieved AOC educational requirements carries some risk that the officers will not be able to attain the graduate degrees required for service in AOCs 70A and 70C. The MSC can help bridge the popularity gap between AOCs by requiring LTHET applications for high demand AOCs like 70A, which suggest a high level of commitment by applying officers.

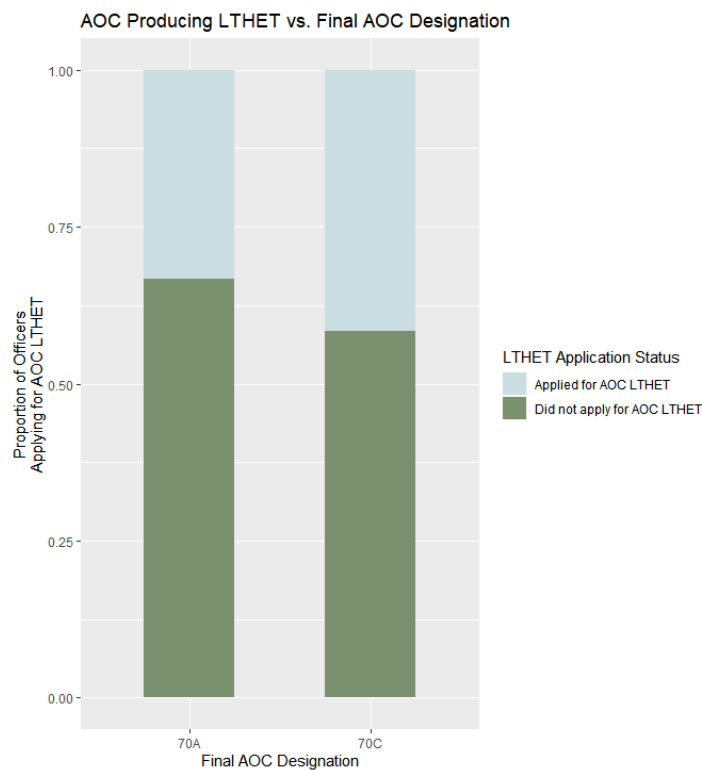


Figure 42. AOC Producing LTHET Selections vs. Final AOC Designation (AOCs 70A and 70C)

The recruitment efforts recommended as part of phase I of AOC-MS are designed to allow consultants the opportunity to shape AOC perceptions and preferences. Taking pre-designation AOC-producing course attendance and LTHET applications as a measure of interest in AOCs, the MSC may track these metrics over time as indicators of the success of AOC-MS phase I.

Last, Figure 43 examines the connection between LTHET selections, AOC assignments and officer preference levels. The overwhelming majority of officers selected for LTHET were designated into the AOC associated with the LTHET program they were selected for. This is by design, as panel members are provided with a list of past and future LTHET selections so that the MSC may recoup its investments of time and money on officers’ professional educations. The bar plot shows that in all cases where officers were designated to AOCs different from those of their LTHET selections, they were awarded their top choice, suggesting a panel preference towards granting officer their first preferences even at the expense of return-on-investment for LTHET selections.

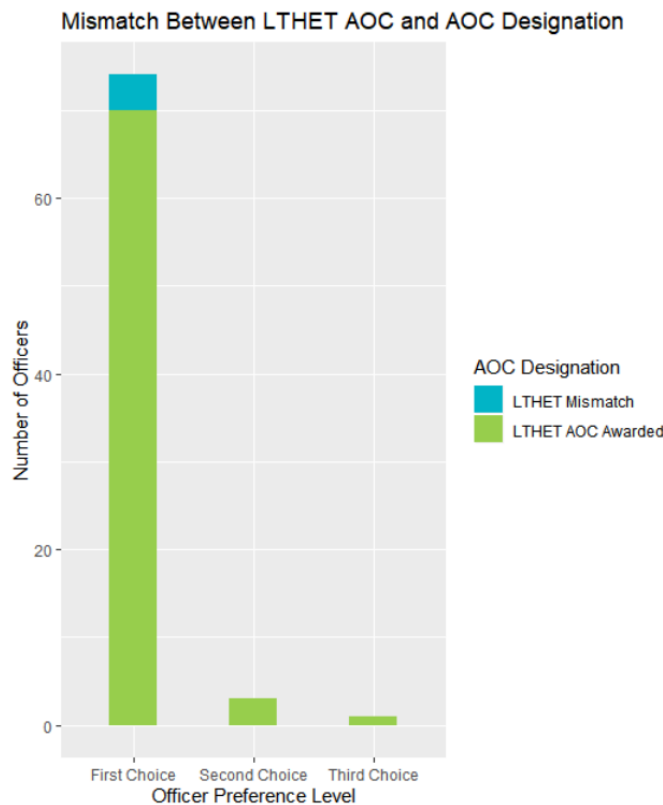


Figure 43. LTHET Status per AOC Designation Preference Level

b. Consultant Preferences

While the current AOC designation process lacks a mechanism for directly capturing consultant preferences, we can infer some information about consultant desires

based on their panel votes. Figure 44 examines the relationship between an officer's final AOC designation and the vote of the gaining AOC consultant (i.e., did the 70E consultant vote against designating a particular officer as a 70E while other consultants who are not subject-matter experts in the 70E AOC voted for a 70E assignment). In the aggregate, this happened approximately 24% of the time, with AOCs such as 70H and 70C experiencing higher rates. Based on this information, we can assume the consultant preference satisfaction rate to be approximately 76%. This metric matters. When consultants from other career fields frequently override gaining consultants, officers are assigned to AOCs against the advice of the Army's appointed subject matter expert for that specialty. These officers could presumably lack the skills and attributes necessary for success in that AOC. By providing consultants with a direct mechanism for outlining their preferences, AOC-MS can improve consultant satisfaction rates.

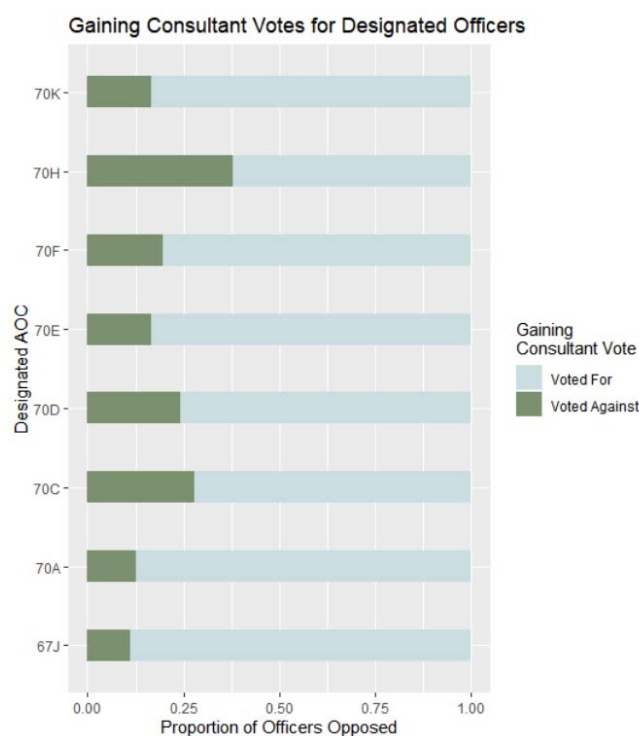


Figure 44. Rate of Gaining Consultant Opposition for Officers Designating into AOC

Figure 45 suggests that preference matching may matter to officers as well. Approximately 42% of officers who listed AOC preferences but were designated against the gaining consultant's wishes remained on active duty as of December of 2019. By contrast, over 80% of officers designated with the gaining consultant's consent remained in the MSC. On the other hand, when officers failed to list preferences and were designated according to the needs of the MSC, consultant preferences didn't seem matter. Over 62% of officers meeting these criteria departed the MSC after AOC designation. AOC-MS attempts to mitigate this disparity by seeking mutual matches between officers and consultant preferences, maximizing value to the officer and the MSC while increasing retention potential.

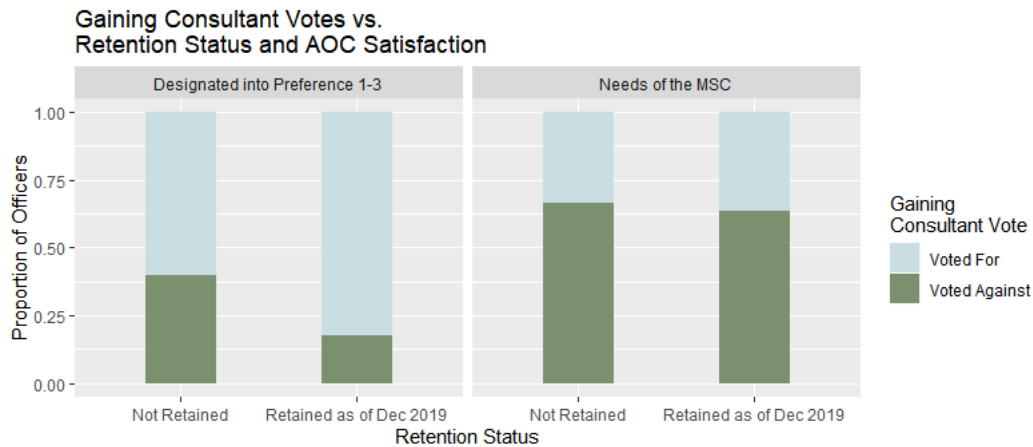


Figure 45. Retention Status as a Function of Gaining Consultant Opposition

c. *Panel Word Picture Scores*

The analysis of AOC Panel Word Picture Scores in Figure 46 suggests that scores generally follow a slightly right-skewed unimodal distribution with a median between 4 and 5. This score is the mean of the scores submitted by each panel member for each officer. The distribution of panel scores for 70Bs who opted in early but were not designated is significantly lower than the distribution for those who designated into a specialized AOC. While similar, the distribution of panel scores seems to vary with AOC assignment, with wider spreads among larger AOCs such as 70F, 70H and 70K.

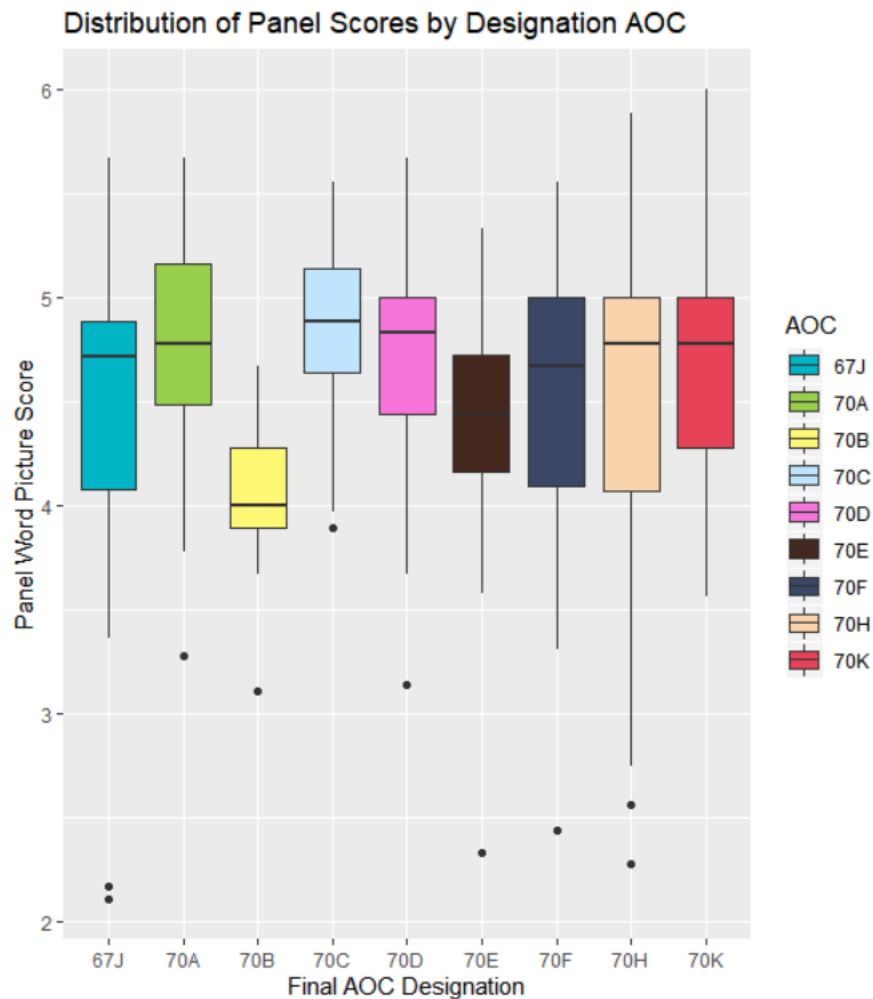


Figure 46. Distribution of Word Picture Scores per Final AOC Designation

Figure 47 explores the relationship between panel scores and officer satisfaction levels. The distribution of panel scores for officers who received their first or second preference is centered near 5 while the distribution of scores for officers receiving their third choice or being designated according to the needs of the MSC is centered near 4, suggesting a correlation between officer performance and AOC satisfaction. We can assess the fairness of the AOC designation process towards AOCs by ensuring that the distribution of panel performance scores is similar between AOCs.

While in theory, the current AOC designation process only factors performance when breaking ties between consultant votes, Figure 47 suggests that in practice, panel members may be willing to reward better performers by granting them their AOC of choice.

AOC-MS avoids conflating aptitude with performance by using panel scores only as tie-breakers for officers in the same preference tiers.

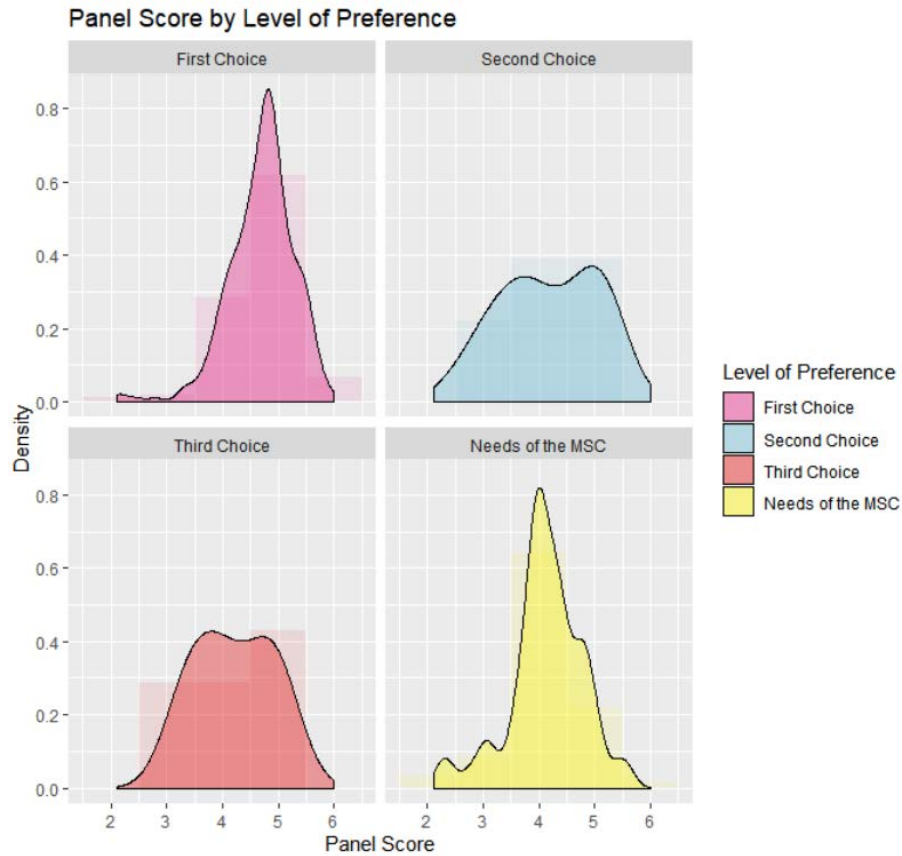


Figure 47. Histogram and Kernel Density Estimate of Panel Scores by Level of Preference Satisfaction

Figure 48 compares the distribution of panel scores between officers who have been retained on Active Duty and those who have left the MSC. The distribution of scores appears slightly higher for those remaining on Active Duty, implying correlation between performance and retention.

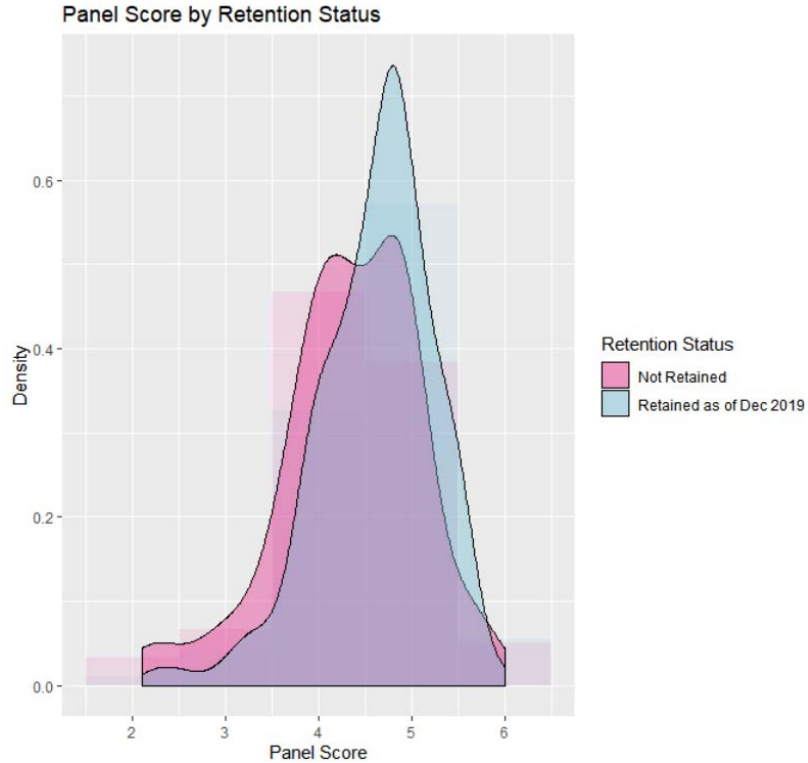


Figure 48. Kernel Density Estimate of Panel Score Distribution by Retention Status

After analyzing the AOC Designation dataset, we’ve developed several measures of effectiveness for contrasting the current designation process with AOC-MS. We can assess the effectiveness of phase I of AOC-MS by tracking AOC popularity rates as well as the rate of pre-designation AOC-producing LTHET applications and course completions. We can evaluate phase II of AOC-MS by tracking the rate at which officers and consultants are awarded their top preferences. We can also gauge the fairness of AOC-MS phase II towards AOCs by ensuring performance scores are equitably distributed amongst all AOCs. The Kruskal-Wallis non-parametric test can help determine whether there are significant differences in performance amongst AOCs for any given designation scheme (Oja 2008). However, we still have no method for determining how well the current AOC designation process ensures the long-term health of the MSC force.

C. MEDICAL SERVICE CORPS OBJECTIVE FORCE MODEL AOC DESIGNATION TARGETS DATASET

We turn to the MSC objective force model (OFM) introduced in Chapter I and Chapter II to appraise how well the current AOC designation process meets the force structure needs of the MSC.

The OFM is an ensemble of integer programming models advising AMEDD leaders regarding how many officers per rank and AOC each PYG should have in order to project officer grade structure over a 30-year horizon (McMurry et al. 2010). As explained in Chapter I, APPD utilizes OFM calculations to determine the AOC transition targets for the annual AOC designation process.

The MSC OFM is a subset of the OFM model ensemble. It is an ILP with over four-thousand variables solved in Microsoft Excel using the Frontline Premium Solver (McMurry 2019). The MSC OFM optimizes accession and promotion targets for the entire MSC, managing 30 PYGs across 22 AOCs and six ranks in four separate MFAs. Describing and analyzing the output of the MSC OFM is well outside the scope of this study. Rather, in this section we analyze OFM AOC designation targets for MFA 70 in conjunction with the AOC Designation results explored in Section B of this chapter, concentrating in PYGs 2008–2012, for which we have designation records.

In order to analyze the OFM’s role in the AOC designation process and to assess how well the AOC designation process meets the force structure needs of the MSC, we summarize relevant target data for PYGs 2008–2012. This summary is contained in the OFM Summary Dataset, pictured in its entirety in Table 12.

Table 12. PYG2008-2012 Objective Force Model Snapshot

Objective Force Model Summary Dataset

PYG	Statistic	70A	70C	70D	70E	70F	70H	70K	Total	Available 70B	Available Delta	AOCs Short	AOCs Over	Overall Short
2008-2012	Inventory	65	38.00	33.000	36.00	53.00	140.00	123.000	488.00	NA	NA	NA	NA	
2008-2012	OFM Auth	65	34.00	38.000	28.00	47.00	113.00	128.000	453.00	NA	NA	NA	NA	
2008-2012	Starting Delta	0	4.00	-5.000	8.00	6.00	27.00	-5.000	35.00	NA	NA	2	4	Above OFM Designation Target
2008-2012	Percentage Fill	1	1.12	0.868	1.29	1.13	1.24	0.961	1.08	NA	NA	NA	NA	
2012	Min Target	1	2.00	8.000	4.00	8.00	13.00	4.000	40.00	30	-10	1	NA	Below OFM Designation Target
2008-2012	Ending Delta	1	6.00	3.000	12.00	14.00	40.00	-1.000	75.00	NA	NA	1	NA	
2010	Min Delta	2	-1.00	-1.000	2.00	1.00	7.00	2.000	10.00	NA	NA	2	5	Above OFM Designation Target
2010	Min Target	4	5.00	3.000	1.00	4.00	16.00	6.000	41.00	51	10	2	5	Above OFM Designation Target
2009	Min Target	7	6.00	8.000	1.00	2.00	13.00	11.000	48.00	45	-3	4	3	Below OFM Designation Target
2009	Min Delta	0	1.00	2.000	-2.00	-2.00	1.00	-1.000	-1.00	NA	NA	4	3	Below OFM Designation Target
2008	Min Target	5	6.00	5.000	4.00	7.00	10.00	7.000	44.00	56	12	1	5	Above OFM Designation Target
2008	Min Delta	2	0.00	1.000	1.00	-1.00	5.00	4.000	12.00	NA	NA	1	5	Above OFM Designation Target
2011	Min Target	3	7.00	2.000	2.00	9.00	20.00	5.000	48.00	46	-2	3	2	Below OFM Designation Target
2011	Min Delta	0	-2.00	0.000	1.00	-4.00	-1.00	4.000	-2.00	NA	NA	3	2	Below OFM Designation Target
2008-2012	Years Short	0	2.00	1.000	1.00	3.00	1.00	1.000	NA	NA	NA	NA	NA	
2008-2012	Years Over	4	2.00	3.000	3.00	1.00	3.00	3.000	NA	NA	NA	NA	NA	

Using the MSC OFM to set AOC designation targets is a sound personnel practice; however, it has a significant shortcoming in that the OFM is primarily a manpower management model. Its objective function optimizes accessions and promotions over a thirty-year horizon while adhering to Army personnel policies and constraints.

By its nature, the OFM is an idealist model. It only considers on-hand officer inventory indirectly. The AOC designation targets provided by the model are designed to ensure a certain number of officers reach the rank of colonel in thirty years. As a result, it may prescribe AOC designation targets which are infeasible given the available 70B inventory. Figure 49 shows that this scenario has happened in three out of the five PYGs we examine in this section. However, despite shortages in individual PYGs, the MSC had 7 surplus officers over the span of the PYGs.

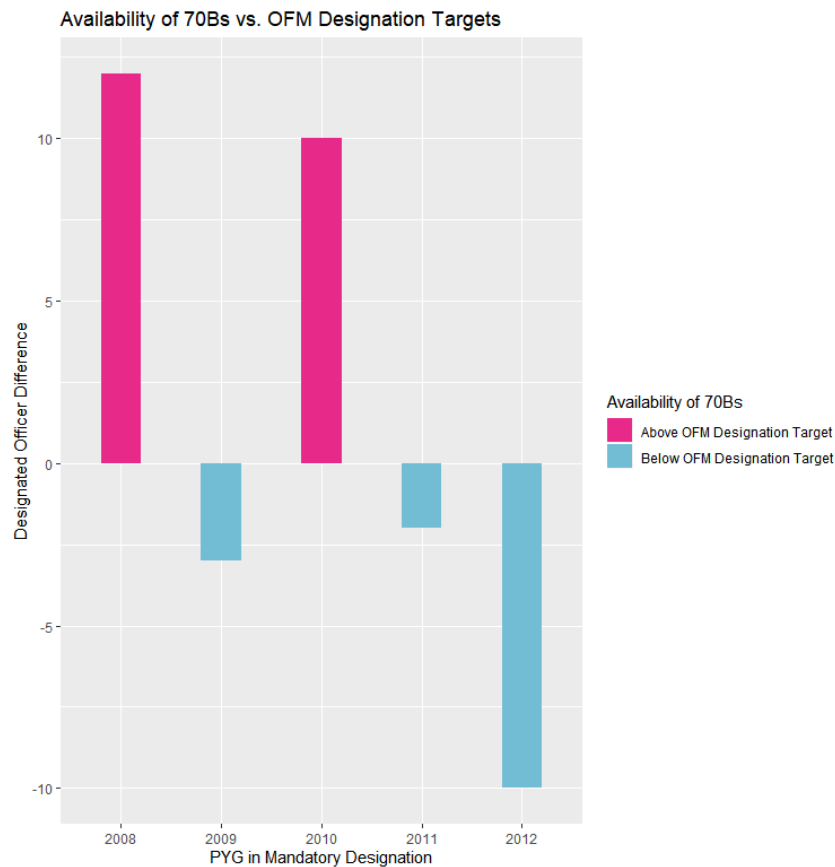


Figure 49. 70B Inventory Difference per PYG

This situation places the burden of allocating shortages and overages on the panel members, who are voting blindly, without knowing how other panel members are voting and who do not have mathematical tools at their disposal to find an optimal allocation of the available inventory to AOCs. The result is a haphazard allocation of officers to AOCs across PYGs which fails to rectify misallocations.

Despite an overall surplus of officers, the AOC designation process has not adjudicated the shortages and overages evenly across all AOCs. For example, 70A did not miss its OFM targets in any of the designation cycles while 70F missed its OFM targets three times. Figure 50 displays the number of PYGs with missed OFM targets for each AOC. Only PYGs 2008–2011 are included as the AOC designation process for 2012 has not been adjudicated as of the time of writing.

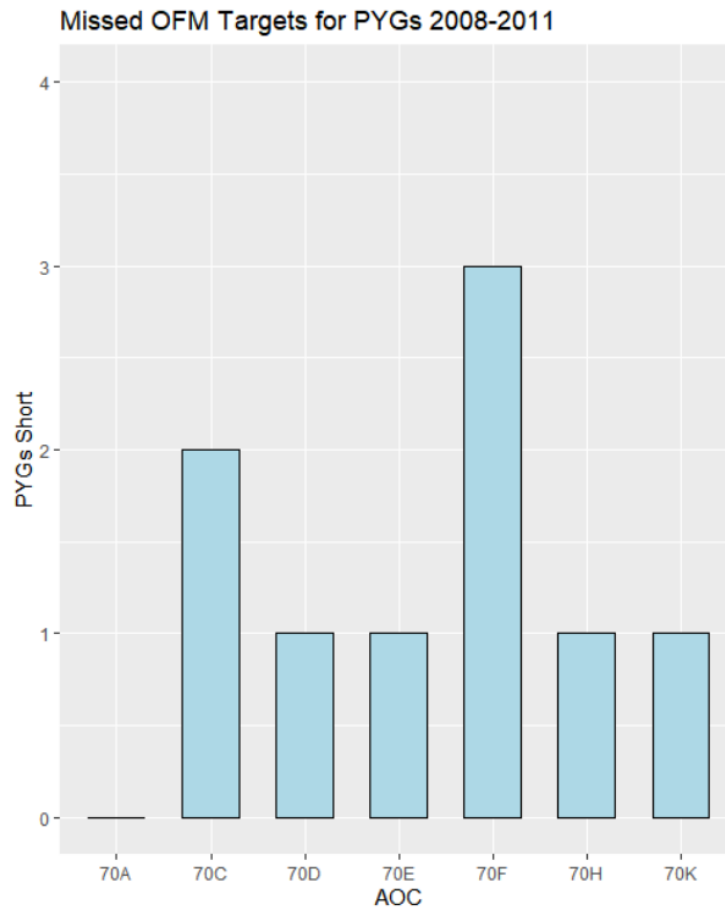


Figure 50. AOC Designation Shortages for PYGs 2008–2011

The current designation process also failed to equitably distribute shortages within PYGs. In all PYGs, regardless of whether there was an overall shortage or overage, at least one AOC missed its minimum OFM targets while at least two AOCs received an overage above maximum OFM targets. Figure 51 displays how many AOCs missed or overshot their targets over the course of four designation cycles.

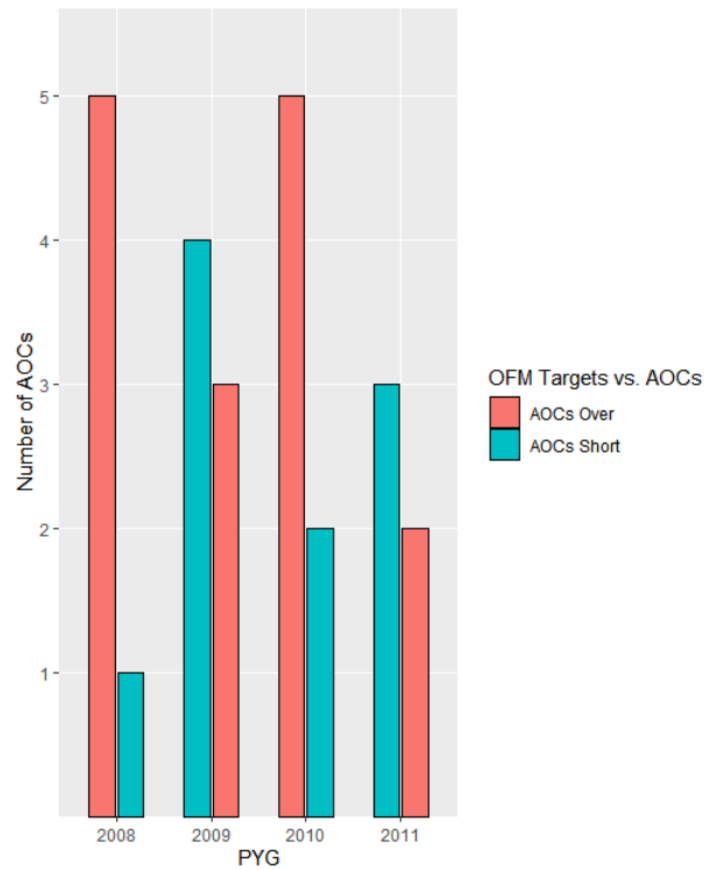


Figure 51. Missed and Overshot OFM Targets by PYG

Figure 52 aggregates the information presented in the previous two graphics to display how each AOC performed over the four PYGs. 70K and 70D failed to fill 100% of OFM slots, with 70D filling below 90% of slots while on the other hand, 70H and 70E are filled near 130%.

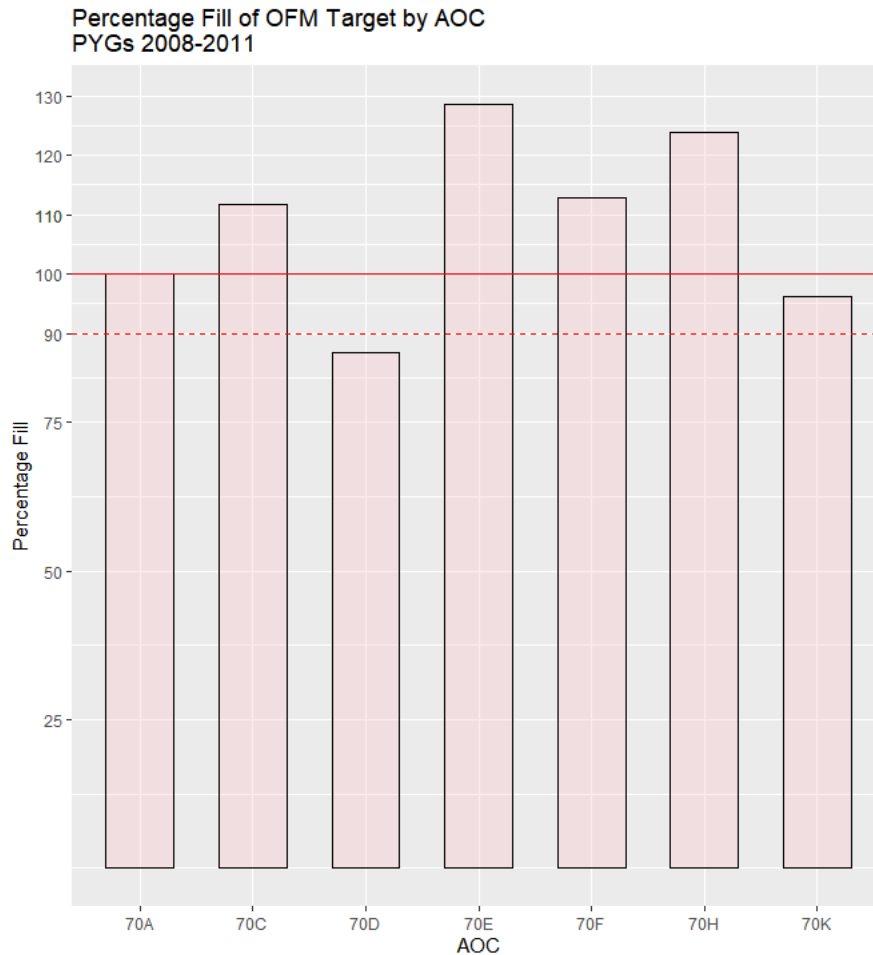


Figure 52. AOC Overages and Shortages for PYGs 2008–2011

D. MEASURES OF EFFECTIVENESS FOR THE AREA OF CONCENTRATION DESIGNATION PROCESS

In an ideal scenario, the number of available officers and AOC billets would match perfectly with officer preferences. In such a scenario, there would be no need for an AOC designation process: All officers would be designated to their desired AOC and all AOC billets would be filled.

However, in this chapter we have established that the number of officers available for designation may not always match the number of AOC billets required and that officer prefer certain AOCs over others. These differences illustrate the need for an algorithm-based designation process like AOC-MS which manages tradeoffs between officer

satisfaction, consultant satisfaction and force structure concerns in a fair and impartial manner. In order to ensure the designation process is effective and equitable to all stakeholders, we need a set of performance measures that evaluate all relevant aspects of the process.

Table 13 displays a proposed set of measures of effectiveness. While some of the measures are long-term indicators, we use applicable measures to compare the current AOC designation process with AOC-MS in Chapter VI of this study.

Table 13. Proposed Measures of Effectiveness for AOC Designation Process

	Measure of Effectiveness	AOC-MS Phase	AOC-MS Measures of Effectiveness		How Measured
			Goal	Current Process	
1	AOC Popularity	Phase I - Talent Development	Minimize	0.265	Standard Deviation of AOC popularity scores.
2	DA Pam 600-4 Requirements	Phase I - Talent Development	Maximize	0.621	Proportion of Officers with AOC-producing course completion or LTHET acceptance.
3	Officer Preference Satisfaction	Phase II - Talent Alignment	Maximize	0.836	Proportion of Officers receiving top two AOC choices.
4	Consultant Preference Satisfaction	Phase II - Talent Alignment	Maximize	0.765*	Proportion of designating Officers in consultant's top two preference tiers.
5	Panel Score Distribution	Phase II - Talent Alignment	Above 0.05	0.109	KW Test p-value testing whether score distribution is same for all AOCs.
6	OFM Target N-Year Percentage Fill	Phase II - Talent Alignment	Maximize	0.935	Mean of OFM target fill proportion minus standard deviation of OFM target fills.
7	Matching Stability	Phase II - Talent Alignment	1	Not Available	1 if matching is stable. 0 if there exists at least one blocking pair.
8	Four-Year Retention	Phase III - Talent Utilization	Maximize	0.849	Proportion of Officers remaining in MSC four years after designation

The first two measures assess phase I of AOC-MS. We explore disparities in popularity between AOCs. In order to do so, we seek a designation process which minimizes the standard deviation of AOC popularity scores introduced in Section B of this chapter. Likewise, the AOC designation process should seek to maximize the rate at which designating officers meet pre-designation requirements. We obtain this rate by taking the mean the AOC-producing course rates for AOCs 70D, 70E, 70F, 70H and 70K as well as the pre-designation LTHET acceptance rates for 70A and 70C. These measures should provide AOC consultants with relevant information regarding the efficacy of their pre-designation recruitment, outreach and mentorship efforts.

The next six measures assess phase II of AOC-MS and ensure the fair distribution of officers to AOCs. The third measure is officer preference satisfaction. The AOC designation process should seek to maximize the number of officers who received their first designation preference. The fourth measure assesses consultant satisfaction with AOC designations. Since the current process does not capture this information directly, we infer this information by examining the proportion of designating officers that gaining consultants voted against, as introduced in Section B of this chapter. For AOC-MS we measure this information directly in the same manner we measure officer preference satisfaction.

In order to assess whether the designation process distributes talent unevenly across AOCs, we perform the Kruskal-Wallis rank sum test (Oja 2008) on the empirical distribution of panel word picture scores per AOC and report the p-value, testing the null hypothesis that the distributions of panel scores are similar across AOCs. Since our goal is to avoid distributing talent in a way that favors some AOCs at the expense of others, we look for p-values above 0.05 where we fail to reject the null hypothesis.

We also track the percentage fill of OFM minimum targets per AOC over a multi-PYG window. To create a single measure of OFM fill for the designation, we average the OFM target fill percentages per AOC and subtract the standard deviation. This measure encourages filling AOCs at close to 100% while discouraging variance amongst AOC fill rates.

We also examine whether the designation scheme yields stable matches as described in Chapter II. This is an important MOE, as under stable matching schemes, every officer is matched as well as he or she can be without disadvantaging another officer. As a result, stable matching schemes minimize regret across the designating cohort. There can exist matching schemes where a higher proportion of officers receive their top AOC choice at the expense of worse preference satisfaction for the remaining officers.

We assess AOC-MS phase III by tracking the proportion of officers that remain in the MSC four-years post-designation. However, it is important to note that this is not a highly reliable indicator, given that many factors other than AOC designation affect officer retention.

V. PREDICTING AREA OF CONCENTRATION APTITUDE USING MACHINE LEARNING CLASSIFIERS

In this chapter we explore four different machine learning classifiers in order to identify the best-performing model for supporting the recruitment and outreach efforts of AOC-MS phase I.

Section A of this chapter reintroduces the AOC predictor dataset, which we use to train and test all machine learning classifiers. Sections B, C, D and E explore the multinomial logistic regression, random forest, support vector machine and feed-forward neural network classifiers. Finally, Section F compares model results in order to select one as the classifier of choice for AOC-MS phase I.

A. TRAINING AND TESTING DATA

As explained in Chapter IV, the AOC Predictor matrix is a numerical matrix with 1277 rows and 26 columns. Since all 26 numerical predictors feature right-skewed count data, we apply a square root transformation on every predictor (O’Hara and Kotze 2010). Figure 53 depicts a sample predictor from the AOC Predictor matrix before and after transformation, showing that the right tail of the predictor’s distribution is shortened.

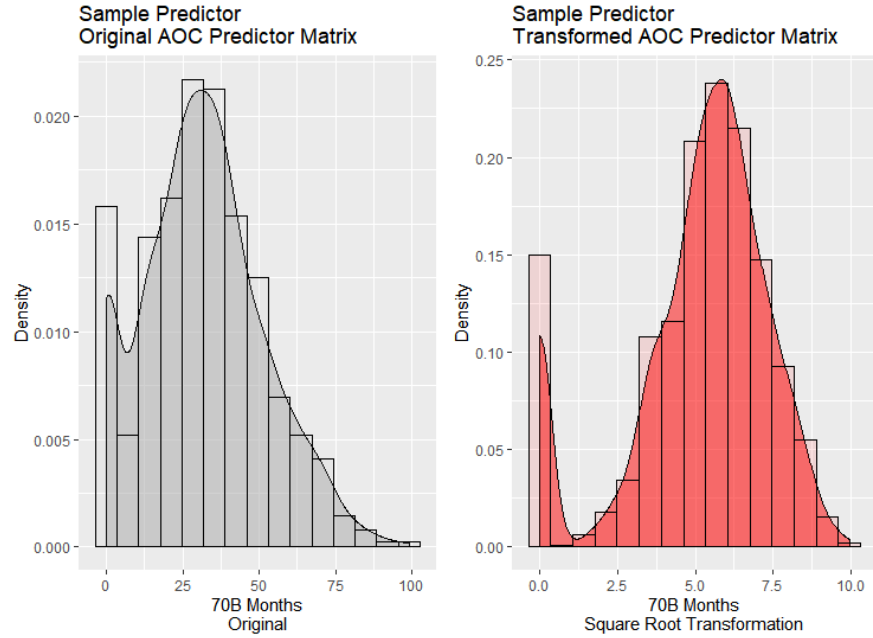


Figure 53. Sample Variable Transformation for AOC Predictor Matrix

In order to train machine learning classifiers, we split the AOC Predictor dataset into a training set with 80% of officer records and a testing set with the remaining 20% of records. We denote by $\mathbf{X}$ a matrix of 26 predictors, where rows index officers and columns index predictors describing those officers. Table 14 displays a sample of five rows and ten columns of the transformed predictor matrix.

Table 14. Sample Records from Predictor Matrix $\mathbf{X}$.

	X00EMonths	X05AMonths	X67AMonths	X67JMonths	X70ACIVED	X70AMILED	X70AMonths	X70BMonths	X70CCIVED	X70CMILED
Officer_ID										
1092	0.000	0.000	0.0	8.185	0.000	1.414	0.000	3.317	0.000	1.414
437	0.000	0.000	0.0	0.000	0.000	1.732	6.928	6.403	0.000	1.871
819	0.000	0.000	0.0	0.000	0.707	1.000	0.000	0.000	1.000	1.000
189	0.000	4.359	0.0	0.000	0.500	1.732	0.000	7.348	0.707	2.121
764	2.646	4.472	0.0	0.000	0.000	1.414	0.000	5.000	0.000	1.414

When referring to the predictor matrices for the subsets corresponding to the training and testing subsets, we will denote them by $\mathbf{X}_{train}$ and $\mathbf{X}_{test}$, respectively. By x_i , we refer to the i^{th} row of $\mathbf{X}$, i.e., the predictors associated with a certain officer.

The response vector $\mathbf{Y}$ consists of the AOC assignment information y_i for officer with the record x_i , where $y_i = a \in \{70A, 70C, 70D, 70E, 70F, 70H, 70K\}$. We define $\mathbf{y}_{train}$ and $\mathbf{y}_{test}$ corresponding to the response variables in the train/test split.

B. MULTINOMIAL LOGISTIC REGRESSION CLASSIFIER

In a multinomial logistic classifier, the log odds of an officer being assigned to a particular AOC are modeled as a linear combination of the predictor variables in the $\mathbf{X}_{train}$ matrix. Multinomial logistic regression classifiers are generally simple models which are easily interpreted. For detailed information regarding the assumptions, mathematical underpinnings, inference and diagnostics of multinomial logistic regression models, refer to the Multinomial Regression chapter of *The Handbook of Regression Analysis* (Chatterjee and Simonoff 2013).

1. Model Selection

We utilized R's `nnet` library (Venables and Ripley 2002) and Python's Scikit learn package (Pedregosa et al. 2011) to follow an iterative process for selecting the best-fitting set of predictor variables to fit our model. We first fit a saturated model using all possible predictors and two-way interactions between AOC-specific predictors. We then utilized stepwise regression in order to minimize the multinomial deviance loss function while limiting model complexity utilizing the Akaike Information Criterion (AIC) (Akaike 2011). Lastly, we regularized the model by penalizing the multinomial loss function with an L2 norm penalty in order to prevent overfitting (Faraway 2016).

The variable selection process yielded an additive model with single order effects featuring 19 predictors j and 120 coefficients $\beta_{a,j}$ arrayed over six linear combinations, with one per AOC a , where 70A is taken to be the reference class for the model. The formula for the classifier is outlined in Equation 8, where $\eta_{i,a}$ represents the log odds that officer i belongs to AOC a , and where $\eta_{i,a}$ is a linear combination of the constant $\beta_{a,0}$ and the coefficients $\beta_{a,j}$. Using the softmax function, we obtain the probability $p_{i,a}$ that officer i belongs to AOC a .

$$\eta_{i,a} = \beta_{a,0} + \sum_j \beta_{a,j} x_{i,j} \quad \forall a \in A', j \in J$$

$$p_{i,a} = Pr(y_i = a) = \frac{e^{\eta_{i,a}}}{\sum_{A'} e^{\eta_{i,a}}}$$

$$A' = \{70C, 70D, 70E, 70F, 70H, 70K\}$$

$$J = \{00E \text{ Months}, 70A \text{ Civ Ed}, 70A \text{ Mil Ed}, 70A \text{ Months}, 70C \text{ Months}, 70D \text{ Mil Ed}, 70E \text{ Civ Ed}, 70E \text{ Mil Ed}, 70E \text{ Months}, 70F \text{ Civ Ed}, 70F \text{ Mil Ed}, 70F \text{ Months}, 70H \text{ Mil Ed}, 70H \text{ Months}, 70K \text{ Mil Ed}, 70K \text{ Months}\}$$

Equation 8. Multinomial Logistic Regression Classifier Equation

The coefficients for the classifier are located in Table 15 and plots of the marginal probabilities associated with each predictor for every AOC are arrayed in Figure 54 through Figure 56. These plots are generated through the Effects package for R (Fox and Hong 2009). In the plots, we hold all other predictors constant while varying a single one. We observe how the probability of being assigned to an AOCs changes with predictor values.

We recognize some of the relationships we identified in Chapter IV. For the most part, the odds that an officer is assigned to an AOC rise with the number of months worked in it as well as with the AOC-relevant military and civilian educational scores. On the other hand, we observe a negative relationship between AOC assignment and predictors for other AOCs: The more experience an officer has in a particular AOC, the less likely it is for that officer to belong to a different one.

We can observe the results of MSC policy in the model's variable selection choices. For example, the civilian education predictors for 70H and 70K were not selected by the model. Both of these AOCs do not require civilian degrees. The variable selection process also omitted all non-AOC specific predictors such as 70B and 67J months with the exception of 00E months, reflecting the time an officer spent in a student status. Officers with high 00E counts are LTHET graduates and more likely to serve in AOCs with LTHET requirements such as 70A and 70C than in AOCs such as 70H and 70K.

Table 15. Regression Coefficients for Multinomial Logistic Classifier

Multinomial Logistic Regression Coefficients

	(Intercept)	`00E Months`	`70A CIV ED`	`70A MIL ED`	`70A Months`	`70C Months`	`70D CIV ED`	`70D MIL ED`	`70D Months`	`70E CIV ED`	`70E MIL ED`
70C	-1.43470	0.25083	-2.2100	-0.06856	-0.42632	1.17036	-1.41395	0.56307	-2.59309	3.94296	-0.73815
70D	-2.92753	-0.13616	-4.0648	-0.68566	-0.06325	-1.16995	2.38794	4.92815	0.42476	2.34666	-1.13748
70E	-0.15434	0.08207	-2.8567	-1.63709	-0.25787	-2.88252	1.61440	-4.61394	-3.03395	0.13252	5.27366
70F	-2.04214	0.04586	1.6095	0.40445	-0.22844	-2.54353	-5.08055	-4.03678	-3.02963	-2.46748	0.81993
70H	1.54167	-0.13412	-2.0887	-1.45966	-0.32401	-0.17268	-2.59218	-0.79268	-0.60737	0.85466	0.38996
70K	-2.26350	-0.04025	-2.6686	-0.37548	-0.75783	0.39389	-0.95547	-0.42678	-0.06656	2.14366	-3.05806

	(Intercept)	`70E Months`	`70F CIV ED`	`70F MIL ED`	`70F Months`	`70H CIV ED`	`70H MIL ED`	`70H Months`	`70K MIL ED`	`70K Months`
70C	-1.43470	0.09106	0.92707	-1.39349	0.12949	-1.45490	0.30946	0.05891	1.23219	0.07135
70D	-2.92753	0.25674	-0.83170	-0.15220	0.25046	-0.23879	-2.21103	0.18079	1.38647	-0.18589
70E	-0.15434	0.48867	0.67042	0.31752	-0.01623	-0.43294	-0.57065	0.14068	0.03322	-0.07579
70F	-2.04214	-0.22844	1.42913	3.30551	0.70772	-0.73107	-0.54659	0.13417	0.19925	0.13990
70H	1.54167	-0.14555	0.57745	-0.31969	0.03482	1.42095	1.20826	0.28055	1.25557	-0.22176
70K	-2.26350	0.05251	0.10821	1.52397	0.18921	-0.12295	-1.04179	-0.01364	4.76885	0.49893

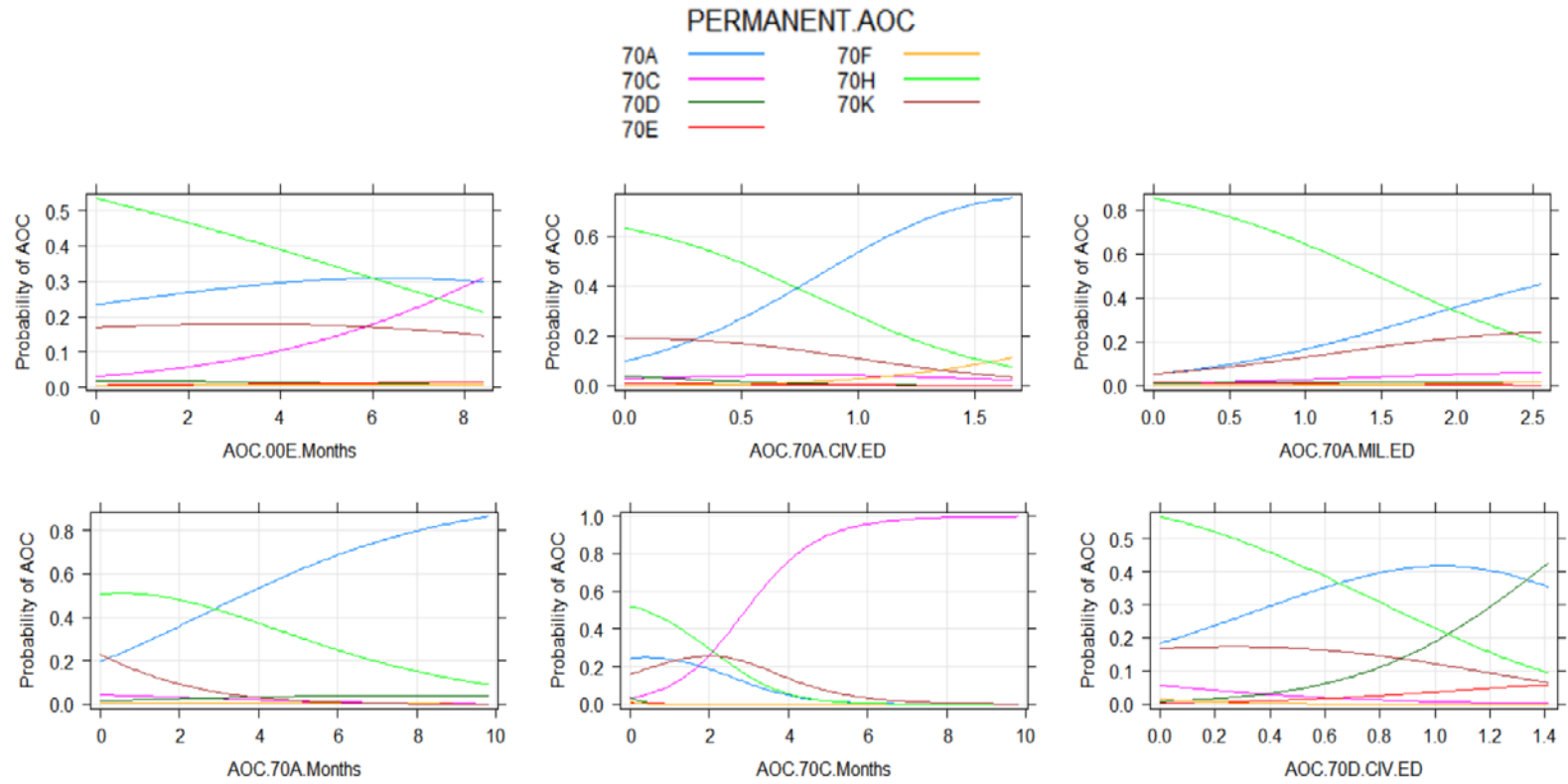


Figure 54. Predictor Effect Plots for Multinomial Logistic Classifier (00E Months through 70D Civ Ed)

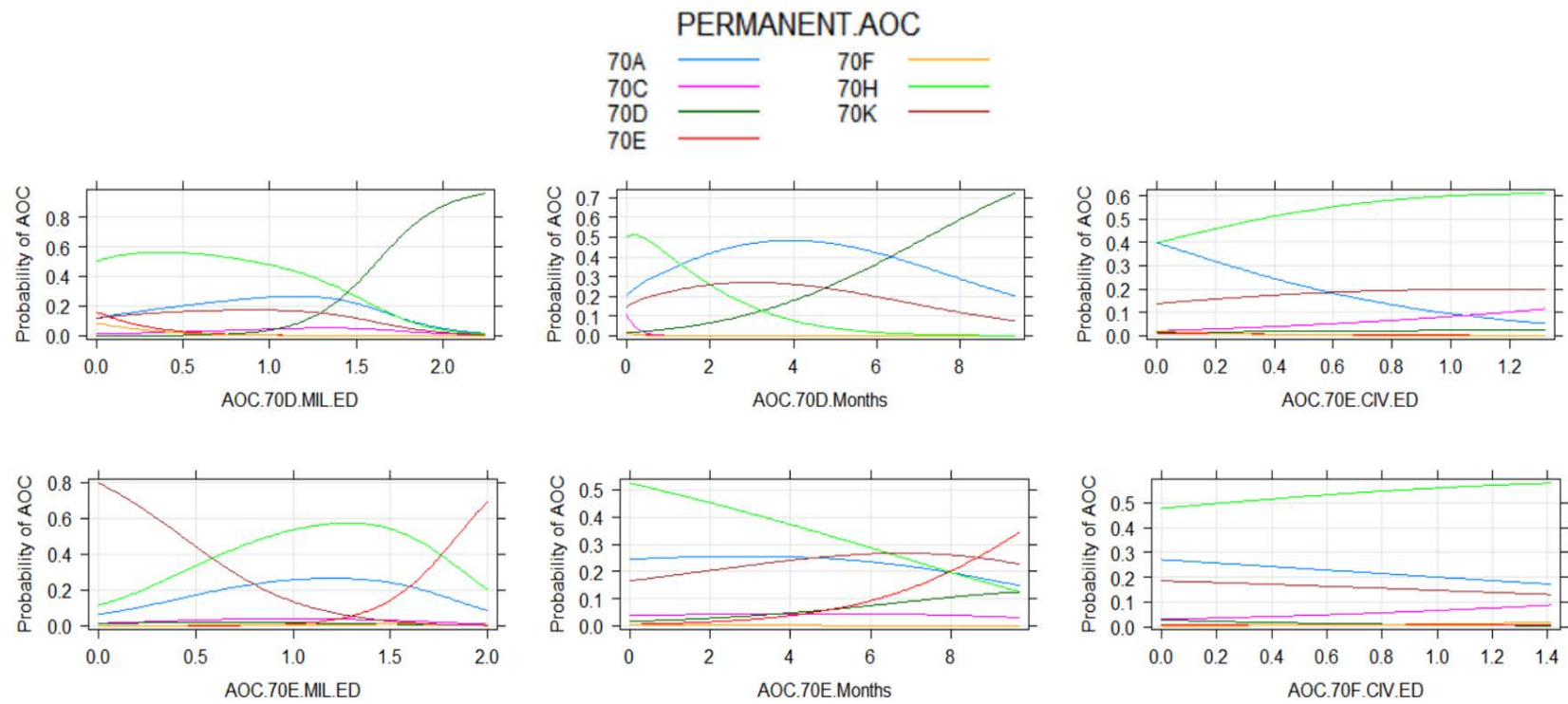


Figure 55. Predictor Effect Plots for Multinomial Logistic Classifier (70D Mil Ed through 70F Civ Ed)

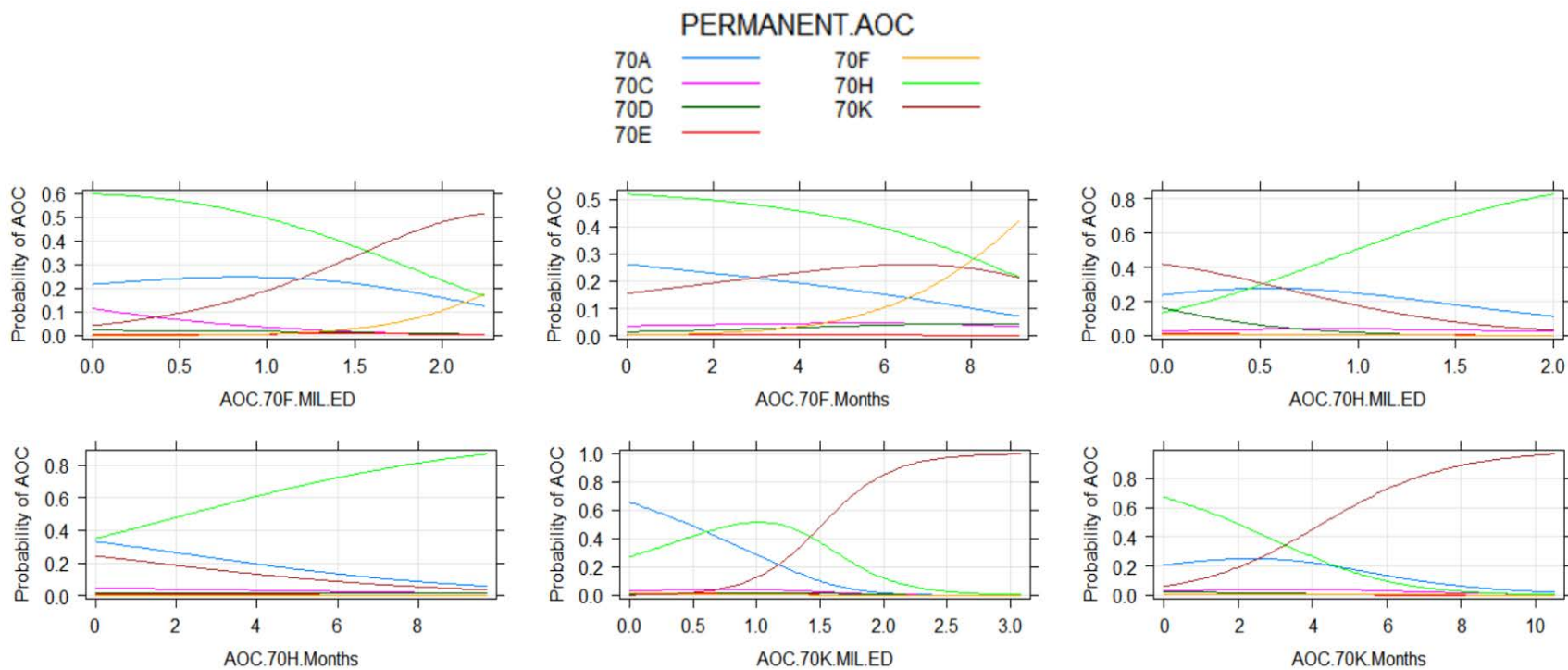


Figure 56. Predictor Effect Plots for Multinomial Logistic Classifier (70F Mil Ed through 70K Months)

2. Model Testing

In order to assess the multinomial logistic classifier's predictive abilities, we predict AOC assignments using $\mathbf{X}_{test}$. We then compare the model's predictions to the actual AOC assignments in $\mathbf{y}_{test}$. We utilize the confusion matrix function from the Caret package for R (Wing et al. 2019) and the Seaborn visualization package for Python (Waskom et al. 2017) to generate the confusion matrix heatmap in Figure 57.

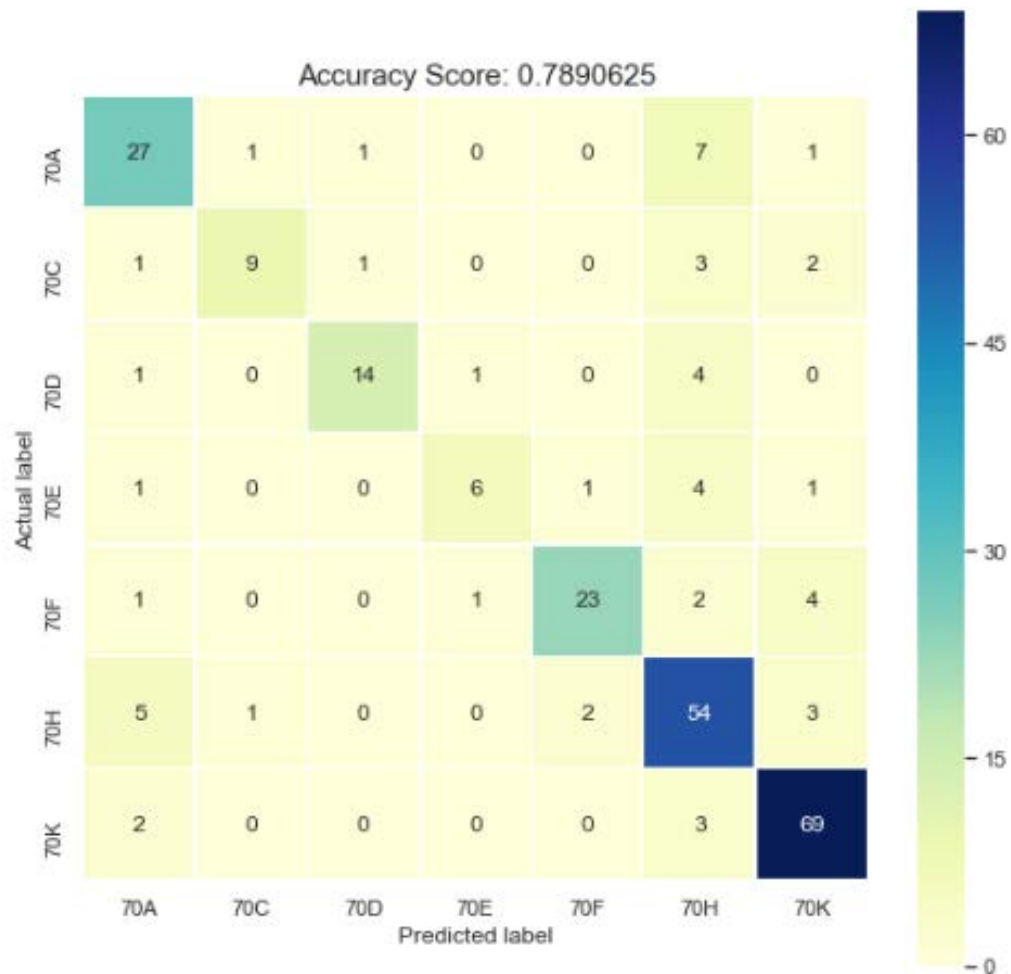


Figure 57. Confusion Matrix Between Multinomial Logistic Classifier Predictions and Test Set AOC Assignments.

The model correctly identifies officer AOC assignments close to 80% of the time. We create a 95% binomial confidence interval for this proportion, located in Table 16.

Table 16. Multinomial Classifier Test Set Accuracy

Multinomial Logistic Classifier
Testing Summary Statistics

	Value
Accuracy	0.789
95% CI Lower Bound	0.734
95% CI Upper Bound	0.837

As part of our assessment, we also check the model’s accuracy by AOC. Figure 58, built with the yellowbrick visualization package for Python (Bengfort and Bilbro 2017), shows receiver operator characteristic (ROC) curves for the multinomial logistic classifier per AOC.

ROC curves plot classifier sensitivity against specificity, with the ROC curves for better classifiers passing closer to the plot’s upper left-hand corner. Sensitivity refers to the classifier’s ability to correctly predict a particular AOC of assignment while specificity alludes to the classifier’s ability to correctly identify officers who are not assigned to a particular AOC. Table 17 displays per-AOC performance statistics for the classifier.

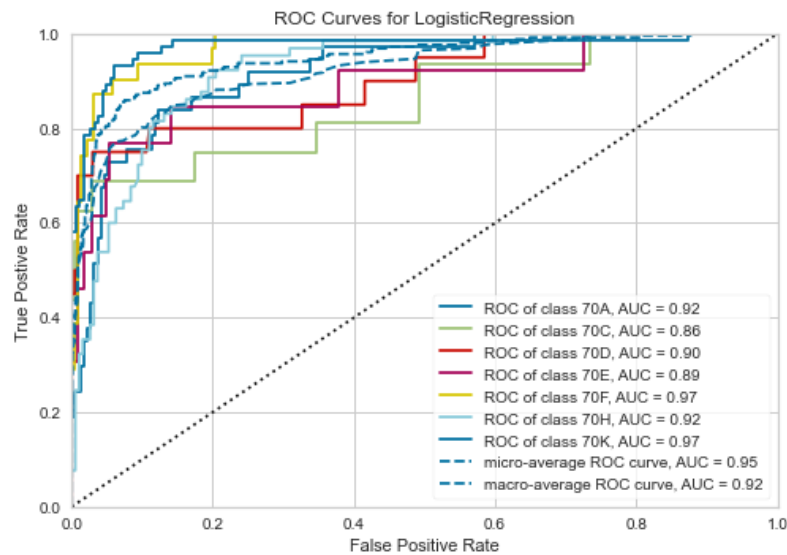


Figure 58. Receiver Operator Characteristic Curves for Multinomial Logistic Classifier

Table 17. Sensitivity and Specificity for Multinomial Logistic Classifier

Multinomial Logistic Classifier Testing per AOC							
	70A	70C	70D	70E	70F	70H	70K
Sensitivity	0.711	0.818	0.875	0.750	0.885	0.701	0.862
Specificity	0.954	0.971	0.975	0.972	0.965	0.939	0.972
No Information Rate	0.145	0.062	0.078	0.051	0.121	0.254	0.289

Given the range of possible AOC assignments, we first focus on the model's sensitivity rather than its specificity. The model struggles to classify officers into AOCs 70A, 70E and 70H, performing better with 70D, 70C and 70K. However, analyzing the model's specificity also yields important insights. AOC 70H features the lowest specificity rate on any AOC.

Misclassified officers were often classified as 70Hs or 70Ks by the model. This is a pattern we see repeated with other models in this chapter. In practice, the majority of 70Bs and 67Js have experiences compatible with a 70H or 70K career path. 70H and 70K are generally the default continuations of a 70B or 67J's career, whereas officers have to take action to posture themselves towards other AOCs.

The higher rate of misclassification for 70H is likely due to the sheer size of the AOC relative to the others. Additionally, as we explored in Chapter IV, 70H draws from a mix of willing and unwilling officers. This makes it more difficult for the model to differentiate 70Hs from officers belong to other AOCs.

As 70K has a higher relative proportion of willing 70Bs and 67J to draw from than 70H, it is more likely that officers designated to 70K have already acquired the experience and military education required to serve as a medical logistician.

The model often confuses 70As and 70Hs. 70Bs and 67Js with extensive operational experience often face a choice between continuing their operational track by designating as a 70H or 70K, or leaving it by applying to LTHET and designating as a 70A.

As a result, the pre-designation records of these officers are similar to one another, which makes it difficult for any model to differentiate between them.

Finally, we can understand the higher misclassification rate for 70Es in the context of their relative size to other AOCs. Four out of 13 70Es were classified as 70Hs, while six were correctly classified as 70Es.

The multinomial logistic classifier presents a relatively accurate yet simple model. It is more accurate than more complicated models; however, its simplicity allows for ease of interpretation.

C. RANDOM FOREST CLASSIFIER

Random forests classifiers are ensemble models which aggregate predictions from a large number of decision trees in order to make consensus predictions. In the context of AOC designation, decision trees make predictions by partitioning $\mathbf{X}_{train}$ into k non-overlapping regions R_k and placing new officer records x_i into one of the k partitions. For every decision tree t , the random forest seeks to find splits that create regions R_k which minimize entropy, a measure of class impurity in the resulting partitions. A partition R_k is impure if the officers $x_{R_k} \in R_k$ belong to different AOCs and strictly pure if for all officers $x_{R_k} \in R_k, y_{R_k} = a$, where a is a single AOC in A , the set of possible AOC assignments. Decision trees create an AOC prediction $\hat{y}_i$ based on the mode AOC $Mo(y_{R_k}) \in \mathbf{y}_{R_k}$, belonging to the majority of officers $x_{R_k} \in R_k$. Decision trees are amongst the most easily explained classification models; however, trees tend to overfit their training data. Given this propensity for overfitting, decision trees will fit noise in training data and often fail to generalize patterns in testing data.

Random forests improve the prediction accuracy and reduce variance in the predictions of individual decision trees by aggregating a large number of shallow trees. Random forests fit trees using subsets of bootstrapped observations from the training data and a randomly selected set of predictors. Random forests make predictions by averaging the predictions of the decision trees inside the forest. Random forests are relatively simple nonparametric models which require few assumptions. As a result, they are flexible, robust

and easily interpretable models. For further discussion regarding the mathematical concepts underlying random forests and forest tuning considerations, refer to Chapter 8 of *An Introduction to Statistical Learning with Applications in R* (James et al. 2013).

1. Model Selection

In order to find the ideal configuration for the random forest classifier, we follow a deliberate process of experimentation using R's `randomForest` (Liaw and Wiener 2002), `partykit` (Hothorn and Zeileis 2015) and `ranger` (Wright and Ziegler 2017) libraries as well as Python's `scikit-learn` package. We tried a variety of parameters and model configurations in order to arrive at the model parameters outlined in Table 18.

Table 18. Random Forest Classifier Configuration

Random Forest Classifier Parameters	
	Value
Bagging	TRUE
Class Weights	Balanced
Split Criterion	Entropy
Maximum Tree Depth	13
Maximum Features per Tree	Square Root
Minimum Samples per Split	5
Number of Trees in Forest	300

The predictor importance for the random forest is displayed in Figure 59. We calculate the importance of a predictor to the random forest by measuring the average decrease in impurity yielded every time a partition tree split based on that predictor value. The average impurity decrease is denoted by the red bar, while the black line shows the range of impurity decreases for the predictor in all the partition trees that make up the forest.

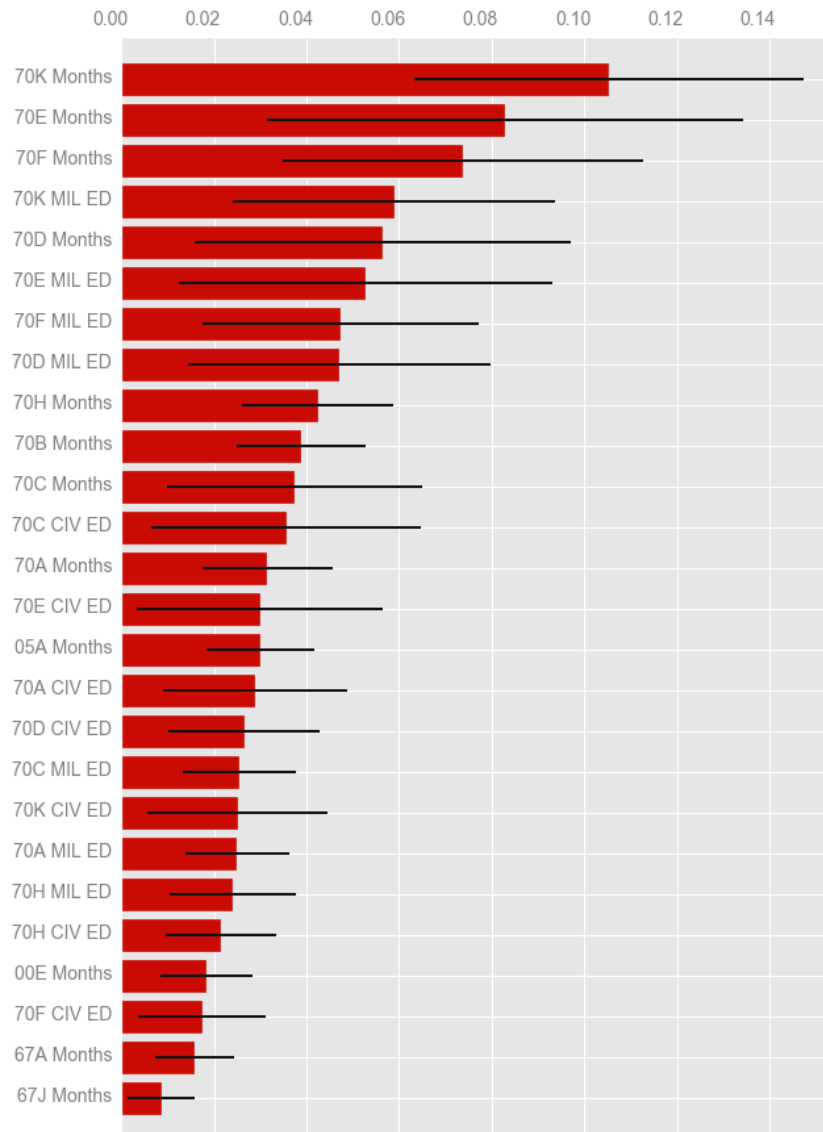


Figure 59. Predictor Importance for Random Forest Classifier

We can observe that the random forest classifier chooses different predictors than the multinomial logistic classifier, giving importance to predictors which were excluded in the multinomial classifier fitting process. In the random forest, four out of the five most influential predictors are the months served in AOCs not requiring a graduate education. The notable exception to this pattern is 70H, likely due to the fact that many non-70H officers work in 70H assignments as lieutenants and captains. The random forest also deems important non-AOC specific features such as 70B months and 05A months. This,

coupled with our predictor analysis for the multinomial logistic classifier in B.1, and our PCA analysis in Figure 31 suggests that there is correlation between predictors. For instance, more time served in a specific AOC will mean less time served in others.

Figure 60 features a two-dimensional visualization of the random forest's predictor space (Vanderplas 2016) using T-SNE (Krijthe 2015) and the random forest's training set AOC predictions. We can observe that the random forest has found clear AOC assignment patterns in the training data. As with the multinomial classifier, there are a number of 70Hs who are misclassified as 70As and a number of 70As misclassified as 70Hs for the same reasons we discussed in Section B.

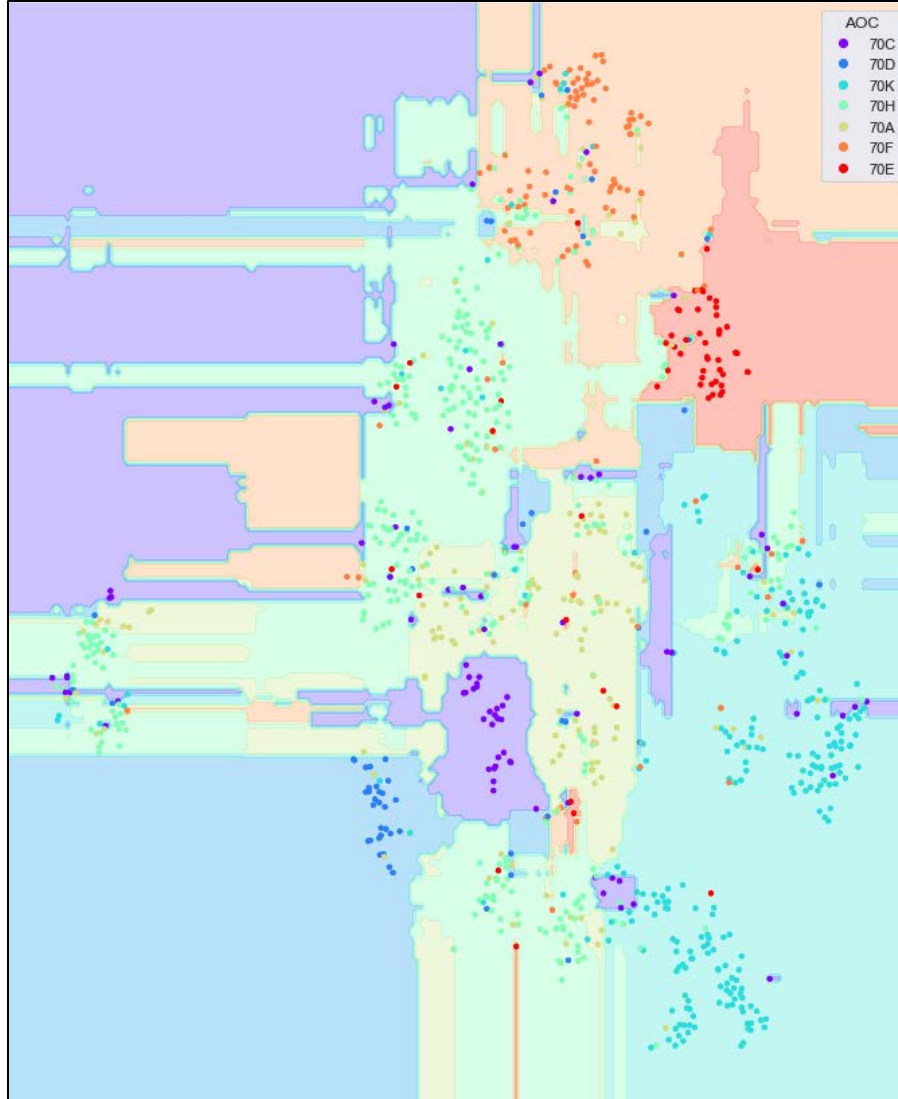


Figure 60. Two-Dimensional Representation of Random Forest Predictor Space

2. Model Testing

We assess the random forest predictions on $\mathbf{X}_{test}$ records against $\mathbf{y}_{test}$ AOC assignments. The confusion matrix for the random forest is located in Figure 61.

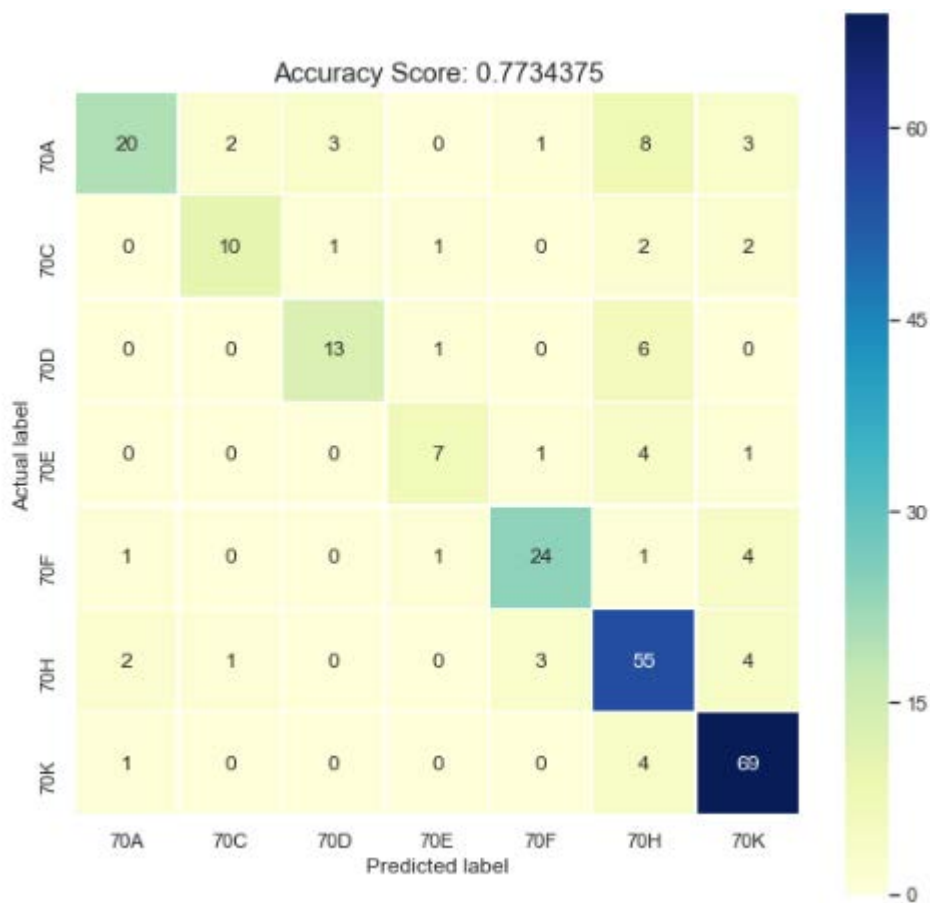


Figure 61. Confusion Matrix Between Random Forest Classifier Predictions and Test Set AOC Assignments.

The random forest achieves a 77% correct test set classification rate. Table 19 displays overall statistics for the random forest's testing performance.

Table 19. Random Forest Classifier Test Set Accuracy

Random Forest Classifier
Testing Summary Statistics

	Value
Accuracy	0.773
95% CI Lower Bound	0.717
95% CI Upper Bound	0.823

The random forest misclassifies officers in similar pattern to the multinomial classifier. However, it fares worse for 70Cs, 70Ds and 70Hs. Like the multinomial logistic classifier, the random forest tends to mix 70A and 70H officers, with both AOCs featuring the lowest specificity in the dataset. The model predicted well for 70K, perhaps owing to the importance of the 70K Months predictor in building the forest. Figure 62 displays ROC curves for the random forest classifier per AOC, while Table 20 shows statistics for per AOC performance.

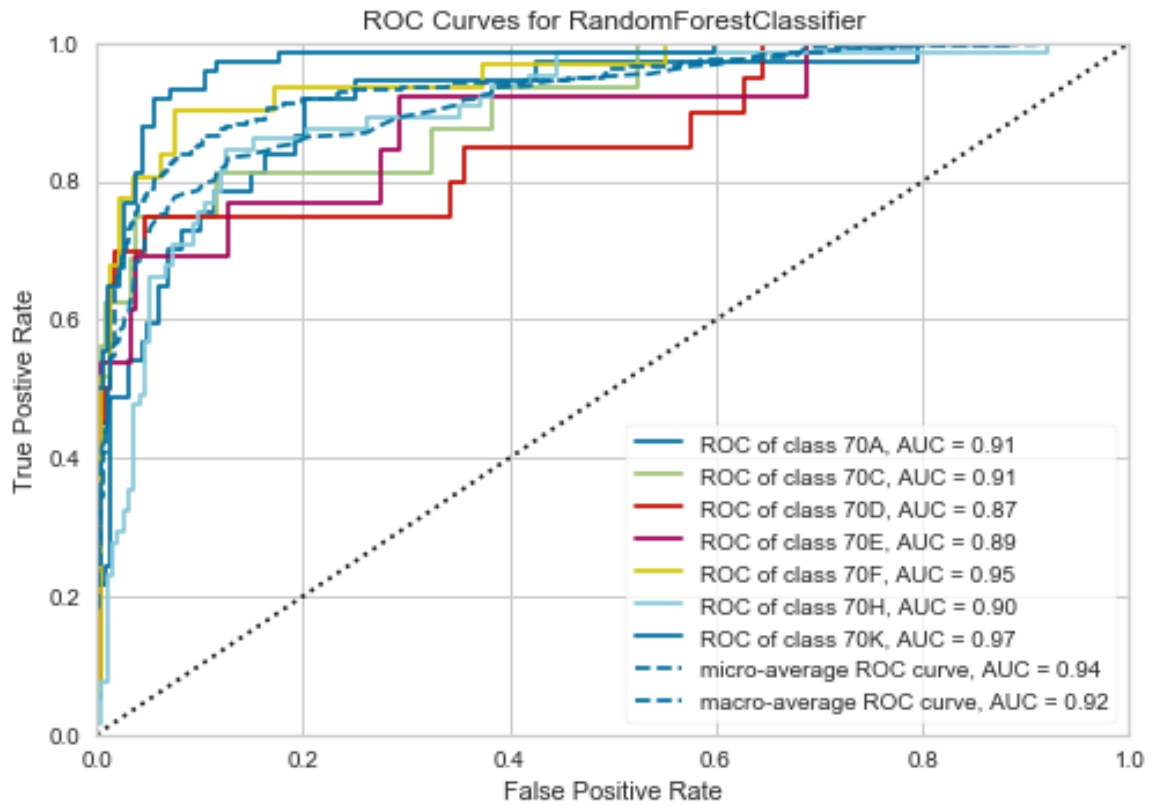


Figure 62. Receiver Operator Characteristic Curves for Random Forest Classifier

Table 20. Sensitivity and Specificity for Random Forest Classifier.

Random Forest Classifier Testing per AOC

	70A	70C	70D	70E	70F	70H	70K
Sensitivity	0.833	0.769	0.765	0.700	0.828	0.688	0.831
Specificity	0.927	0.975	0.971	0.976	0.969	0.943	0.971
No Information Rate	0.145	0.062	0.078	0.051	0.121	0.254	0.289

With nearly overlapping 95% confidence bounds, the multinomial classifier and the random forest perform similarly when classifying $\mathbf{X}_{test}$ observations. Although both models have different approaches and predictor choices for classifying the data, the models classify in similar patterns. The random forest presents an easily interpretable alternative to the multinomial classifier. In Section F of this chapter, we utilize statistical tests to better understand how the models compare to one another.

D. SUPPORT VECTOR MACHINE CLASSIFIER

Support vector machine classifiers are powerful machine learning algorithms that are effective in separating data in high dimensional spaces. Support vector machines are extensions of the maximal margin classifier model which separates a p -dimensional space using a $(p - 1)$ -dimensional hyperplane in order to perform binary classification. The classifier finds the optimal hyperplane $\mathbf{X}_{train}\beta$ to separate the data so that the observations on either side of the hyperplane are in the same class. To do so, support vector machines minimize the hinge loss function $L(\mathbf{X}_{train}, \mathbf{y}_{train}, \beta)$.

Support vector machines enlarge the feature space to high dimensions using kernels, which are functions that quantify the similarity between observations x_i and x'_i , expressed as $K_i(x_i, x'_i)$. The support vector machine classifier is expressed as a linear combination of kernel functions across the support points $S \in \mathbf{X}_{train}$ for which $K_i(x_i, x'_i) \neq 0$.

Support vector machines can use linear, polynomial, radial and sigmoid kernels. Nonlinear kernels extend the predictor space to sufficient dimensions where they can be linearly separated. In order to extend support vector machines to handle a multi-class problem like the AOC designation problem, we can rely on two approaches, which are the one-versus-one approach and one-versus-all approach. In the one-versus-one approach, the support vector machine classifies officers $x_i \in \mathbf{X}_{train}$ by comparing the AOCs $a \in A$ to each other pairwise, performing $\binom{|A|=7}{2} = 21$ comparisons. The support vector machine then classifies x_i according to the AOC assignment most often found during the pairwise comparisons. In the one versus all approach, the support vector machine classifies officers x_i by comparing each of the AOCs a against the remaining AOCs, performing $|A| - 1 = 6$ comparisons. The support vector machine then classifies x_i according to the comparison with the highest coefficients on its hyperplane. For further information regarding support vector machines parameters and the hinge loss function, refer to Chapter 9 of *An Introduction to Statistical Learning with Applications in R* (James et al. 2013).

1. Model Selection

Using the R package e1071 (Meyer et al. 2019) as well as Python's Scikit Learn, we experimented with a variety of kernel functions and parameters as well as multiclass approaches and regularization parameters. The parameters for the support vector machine resulting from our investigation process are located on Table 21.

Table 21. Support Vector Machine Classification Configuration

Support Vector Machine Classifier Parameters	
	Value
Kernel	Linear
Lambda Parameter	0.03
L2 Regularization Parameter	0.7
Class Weight	Balanced
Multiclass Approach	One Versus One

The formulas for the hinge loss function with an L2 norm regularization penalty (1), the linear kernel (2) and for the support vector machine classifier (3) are located in Equation 9. For officer records x_i and x'_i , the linear kernel is the sum across the predictors $p \in P$ of the product between the observation $x_{i,p}$ and $x'_{i,p}$. The support vector machine classifier $f(x \in \mathbf{X}_{train})$ is expressed as a linear combination of kernel functions across the support vectors in S .

$$\min_{\boldsymbol{\beta}} \sum_{i \in \mathbf{X}_{train}} \max(0, 1 - y_i(x_i \boldsymbol{\beta})) + \lambda \sum_{p \in P} \beta_p^2 \quad (1)$$

$$K_i(x_i, x'_i) = \sum_{p \in P} x_{i,p} x'_{i,p} \quad (2)$$

$$f(x \in \mathbf{X}_{train}) = \beta_o + \sum_{i \in S} K_i(x_i, x'_i) \quad (3)$$

Equation 9. Equations for Support Vector Machine Classifier

Given that we're using a linear kernel, we can interpret the absolute value of the coefficients $|\beta_p| \in \boldsymbol{\beta}$ as indicative of predictor importance in a similar way that we assessed predictor importance for the random forest (Rakotomamonjy 2003). In order to assess predictor importance across the model, we average the coefficients across the affine functions for each AOC. Predictor coefficient magnitude for the support vector machine classifier is displayed in Figure 63.

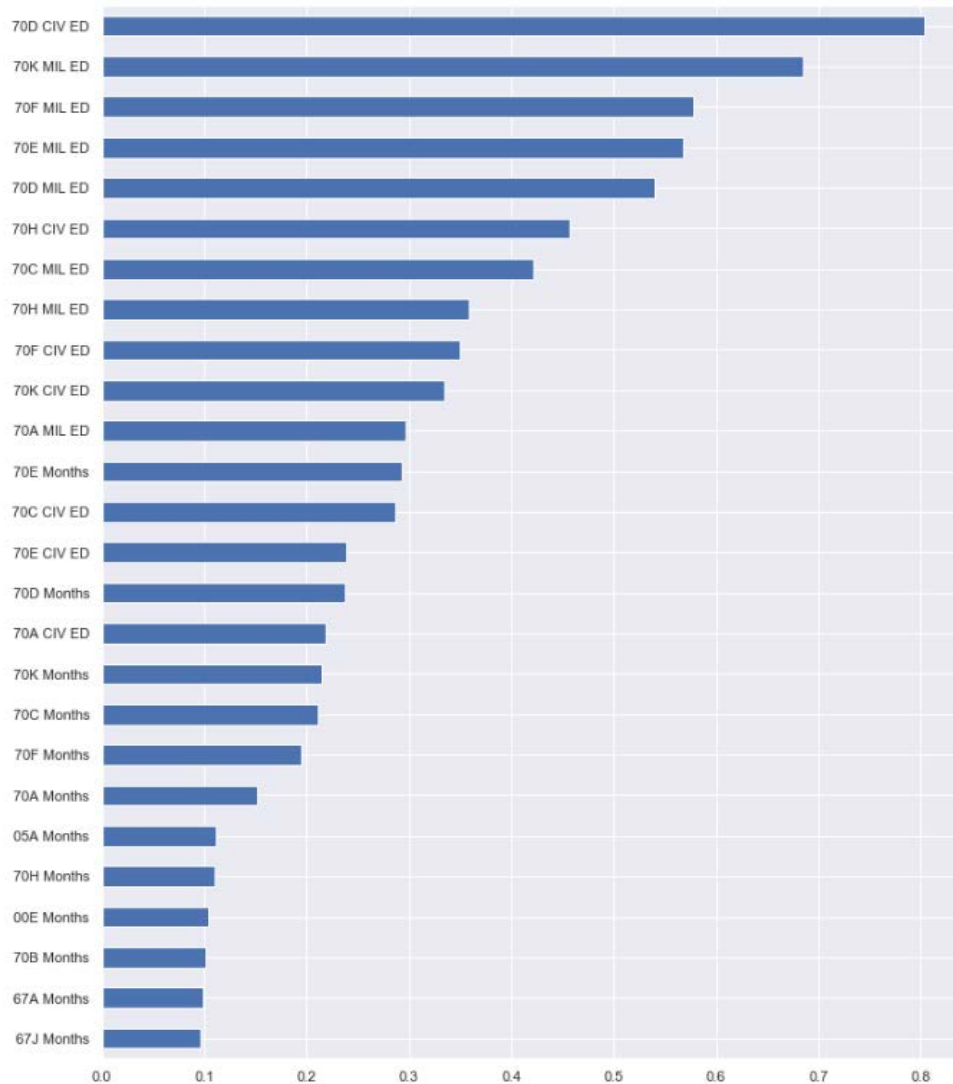


Figure 63. Predictor Coefficient Magnitude for Support Vector Machine Classifier

As the random forest did, the most influential predictors were the military education coefficients for the AOCs not requiring a graduate education, with the exception of 70D civilian education. While this predictor seems counterintuitive given that 70D does not require civilian education, the information technology educational requirements tied to the 70D AOC are so specific that the predictor could be informative about who is a 70D. The least influential variables in this model were the non-AOC specific predictors, as with the other models.

In Figure 64 we visualize the support vector machine's response surface in two dimensions using t-SNE. Given that we're representing the 26-dimension predictor space in two dimensions, the response surface is represented using a radial basis kernel, which approximates the multidimensional nature of the AOC Predictor dataset, where a linear kernel works well to fit the data. We observe classification patterns similar to those observed with the random forest. We can see that the support vector machine assumes 70H predictions take the majority of the response.

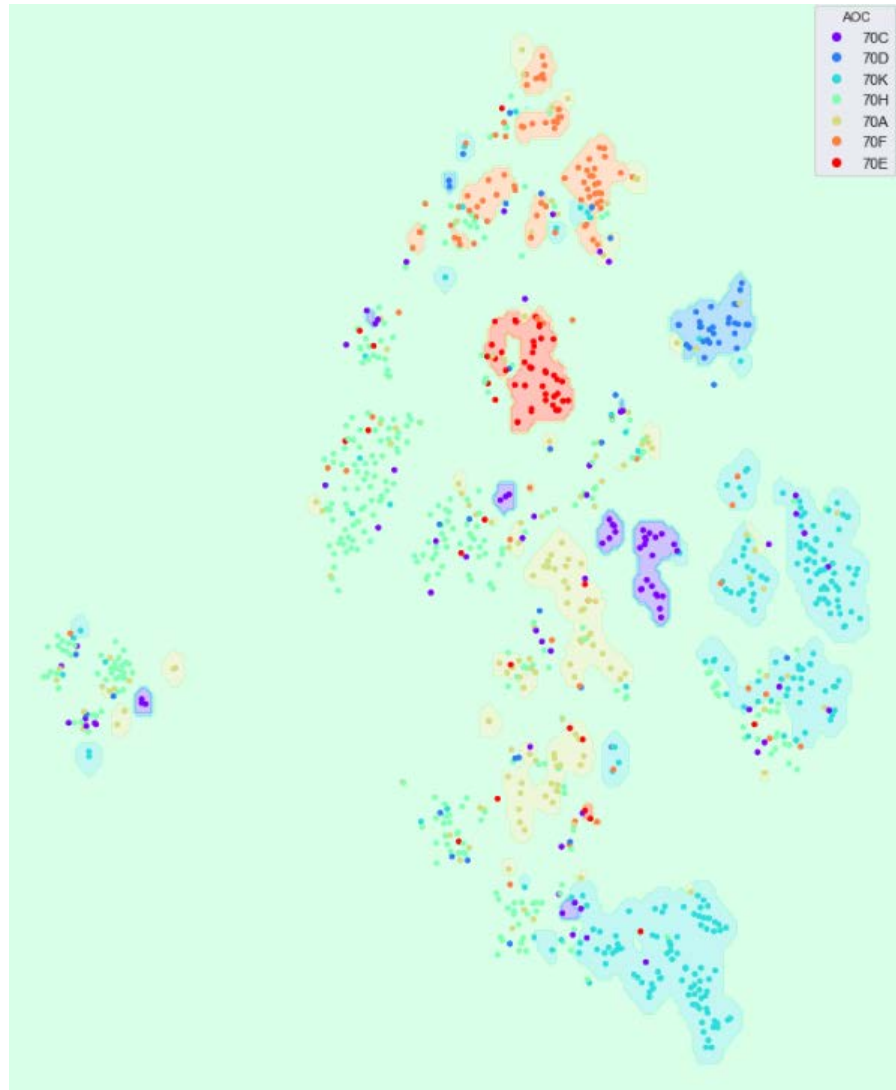


Figure 64. Two-Dimensional Representation of Support Vector Machine Predictor Space

2. Model Testing

When predicting AOC assignments for the records of $\mathbf{X}_{test}$, the support vector machine classifier performs slightly better than the random forest but slightly worse than the multinomial logistic classifier, with 78% of records correctly classified. Figure 65 displays the confusion matrix between the support vector machine's predictions and the AOC assignments in y_{test} .

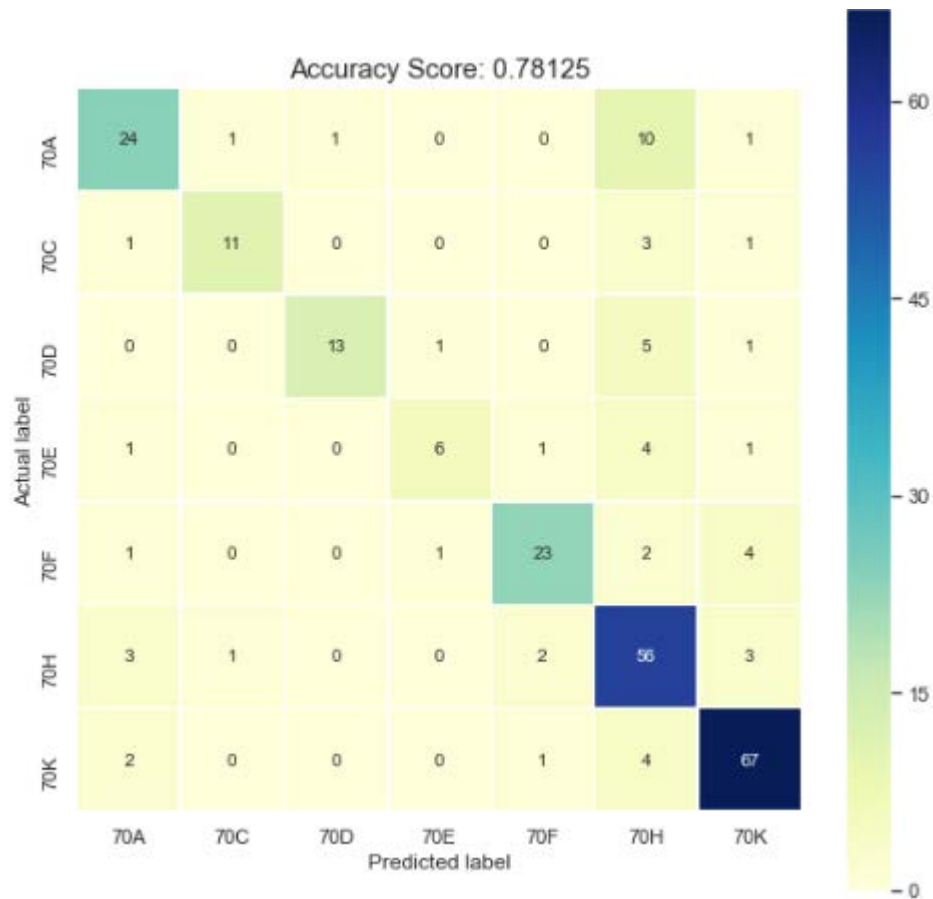


Figure 65. Confusion Matrix Between Support Vector Machine Classifier Predictions and Test Set AOC Assignments.

In Table 22, which shows overall statistics for assessing the support vector machine classifier, we see that the confidence interval for this model almost completely overlaps with those of our previous two models.

Table 22. Support Vector Machine Classifier Test Set Accuracy

Support Vector Machine Classifier Testing Summary Statistics	
	Value
Accuracy	0.781
95% CI Lower Bound	0.726
95% CI Upper Bound	0.830

By AOC, the support vector machine also struggles to predict well for 70Es and 70As, while 70Hs again features the lowest specificity of the AOCs. This model also confuses 70Hs and 70As, and assigns a significant number of officers from other AOCs to 70H. Overall, we see similar patterns to those found in the previous models, which suggests that the model shortcomings reflect variability in the data. Figure 66 displays ROC curves for the support vector machine classifier per AOC, while Table 23 shows statistics for per AOC performance.

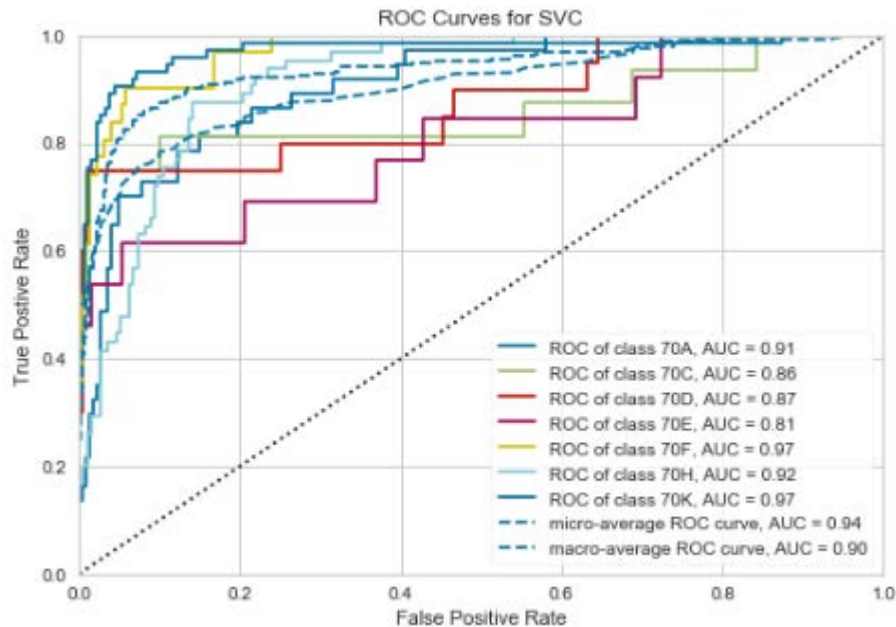


Figure 66. Receiver Operator Characteristic Curves for Support Vector Machine Classifier

Table 23. Sensitivity and Specificity for Support Vector Machine Classifier

Support Vector Machine Classifier Testing per AOC							
	70A	70C	70D	70E	70F	70H	70K
Sensitivity	0.750	0.846	0.929	0.750	0.852	0.667	0.859
Specificity	0.942	0.979	0.971	0.972	0.965	0.948	0.961
No Information Rate	0.145	0.062	0.078	0.051	0.121	0.254	0.289

Overall, we see similar patterns to those found in the previous models, which suggests that the model shortcomings reflect variability in the data. Although slightly more complex than the random forest and the multinomial logistic classifier, the support vector machine classifier with a linear kernel presents an interpretable and well-performing alternative to the first two models we explored.

E. FULLY CONNECTED NEURAL NETWORK CLASSIFIER

Neural networks are inspired by the structure of the human brain, where layers of information processing nodes (neurons) communicate with each other to model complex relationships. Information processes from input nodes to model outputs, often through a variety of hidden layers, where every node in one layer is connected to every other node in its adjacent layers. Nodes capture information as linear combinations of their inputs and then pass the information through nonlinear activating functions, which generate outputs that then serves as an input to the next layer of nodes. During training, neural networks adjust the weights of the inputs to nodes in order to minimize the loss function.

There are multiple network architectures available to modelers, primarily feedforward or fully connected neural network (FNN); convolutional neural networks (CNN), which are tailored towards image recognition; and recurrent neural networks (RNN), which process sequential data. Modelers can tune a variety of hyperparameters when training neural networks, including the number of hidden layers and nodes within layers, the activation function for each layer, the learning rate for the network and a variety of regularization parameters to avoid overfitting the training data.

Compared to the other models we have explored, neural networks are more complex and less easily interpreted, though they often classify more precisely than simpler models. For an in-depth explanation of how neural networks work, refer to *Deep Learning* (Kelleher 2019).

1. Model Selection

We experimented with a variety of network configurations and hyperparameters in order to find the best-performing neural network for predicting AOC assignments. We utilize the scikit learn, keras (Chollet 2015) and TensorFlow (Abadi et al. 2015) packages for Python to train the models.

During our model training and selection process, smaller and simpler FNNs generally outperformed more complex ones, which tend to overfit the training data even when employing multiple regularization schemes. Even with regularization and slowed learning to avoid overfitting the data, model validation accuracy tended to stall or worsen for more complex model architectures.

The result of our model selection process is a FNN with a single hidden layer of 10 nodes. We utilize the leaky rectified linear unit (ReLU) activation function (Jain 2020) for hidden layer outputs and the softmax activation function for the final output nodes, yielding probabilities of AOC assignments. We also utilize both L1 and L2 regularization weight penalties for the hidden layer as well as dropout regularization at a rate of 20%. The structure and hyperparameters of the resulting model are displayed in Figure 67.

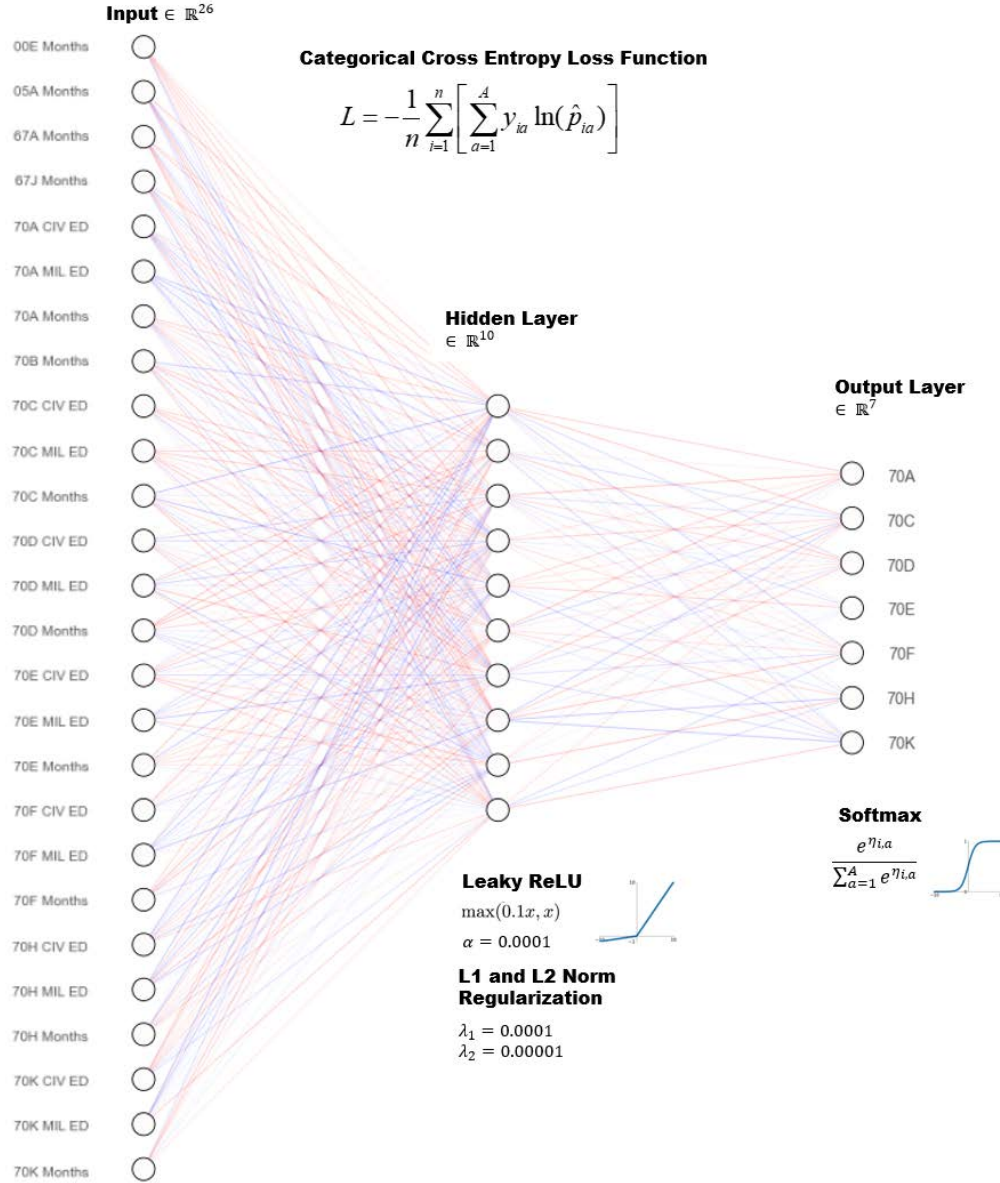


Figure 67. Architecture and Hyperparameters of Fully Connected Neural Network Classifier. Adapted from LeNail (2019).

In order to identify the importance of predictors in the FNN classifier, we use the ELI5 package (Korobov and Lopuhin 2019) for Python. Permutation importance helps estimate how much model validation accuracy decreases when we randomly remove the predictor values from the model. Figure 68 displays the mean permutation importance for the FNN in the blue bars, while the black lines display the range of accuracy decreases obtained every time a predictor variable is removed.

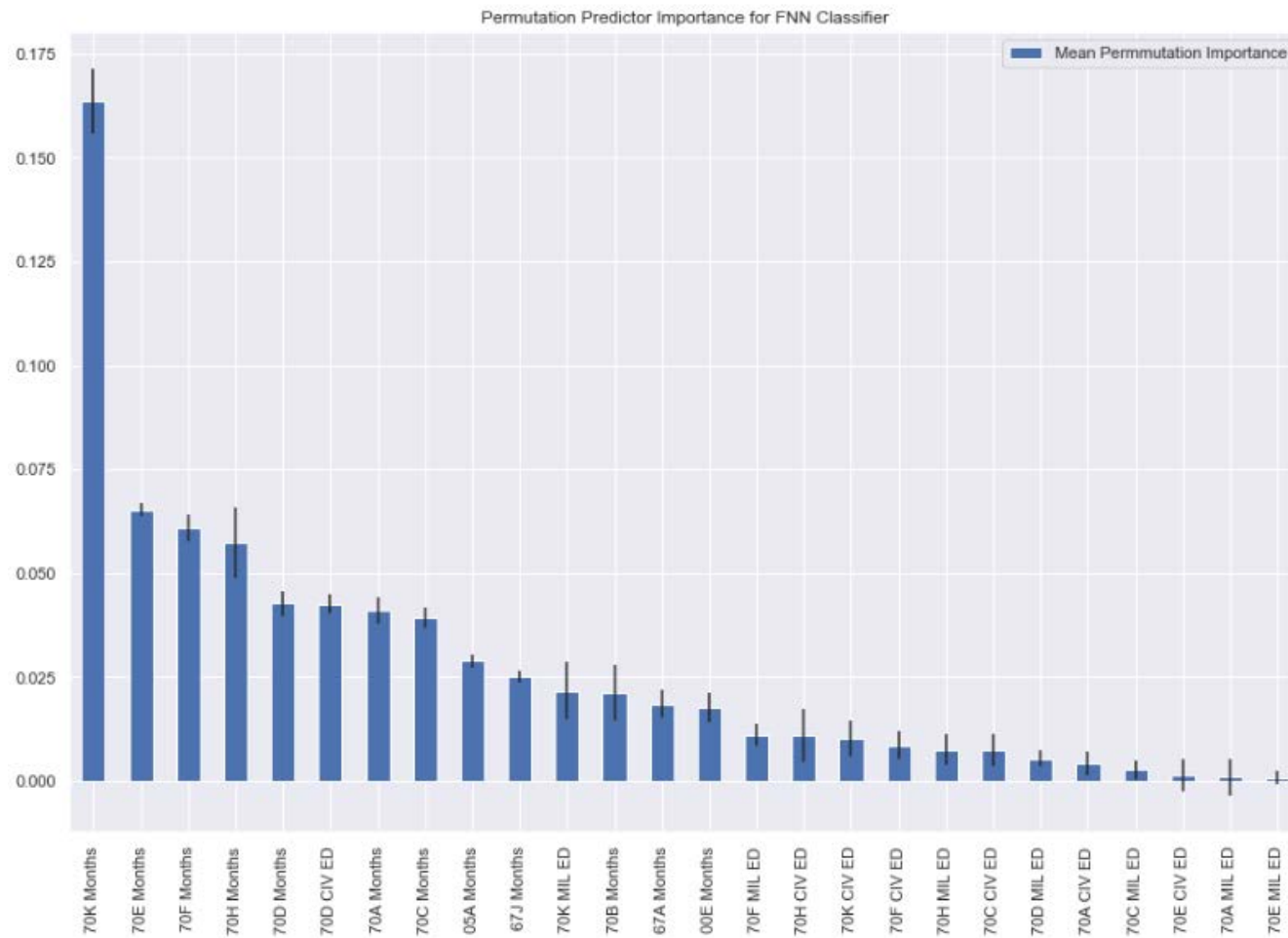


Figure 68. Importance for Fully Connected Neural Network Classifier Predictors

Like the random forest, the FNN classifier determines 70K Months to be the most informative predictor, and holds all AOC-specific month counts to be the most important. Like the support vector machine, the FNN classifier deems 70D civilian education entries significant above any other educational entries; however, unlike the three simpler models we've explored, the model deems non-AOC specific month counts to be more significant than the majority of educational coefficients. This variable selection seems to suggest that the model is able to make AOC classifications based off the month count predictor and may not get much additional classification information from educational entries, unlike the support vector machine classifier, which assigns larger coefficients to the educational entries while neglecting month counts.

2. Model Testing

Despite its complexity, the most accurate FNN classifier we fit fails to outperform the simpler models we covered earlier in the chapter. With a test set accuracy of 78%, the FNN performs as well as the support vector machine, better than the random forest and worse than the multinomial logistic classifier. Nonetheless, the 95% confidence interval for all four models overlap almost perfectly. Figure 69 displays the confusion matrix comparing neural network predictions against $\mathbf{y}_{test}$ AOC assignments while Table 24 shows a 95% confidence interval for the FNN's correct classification rate.

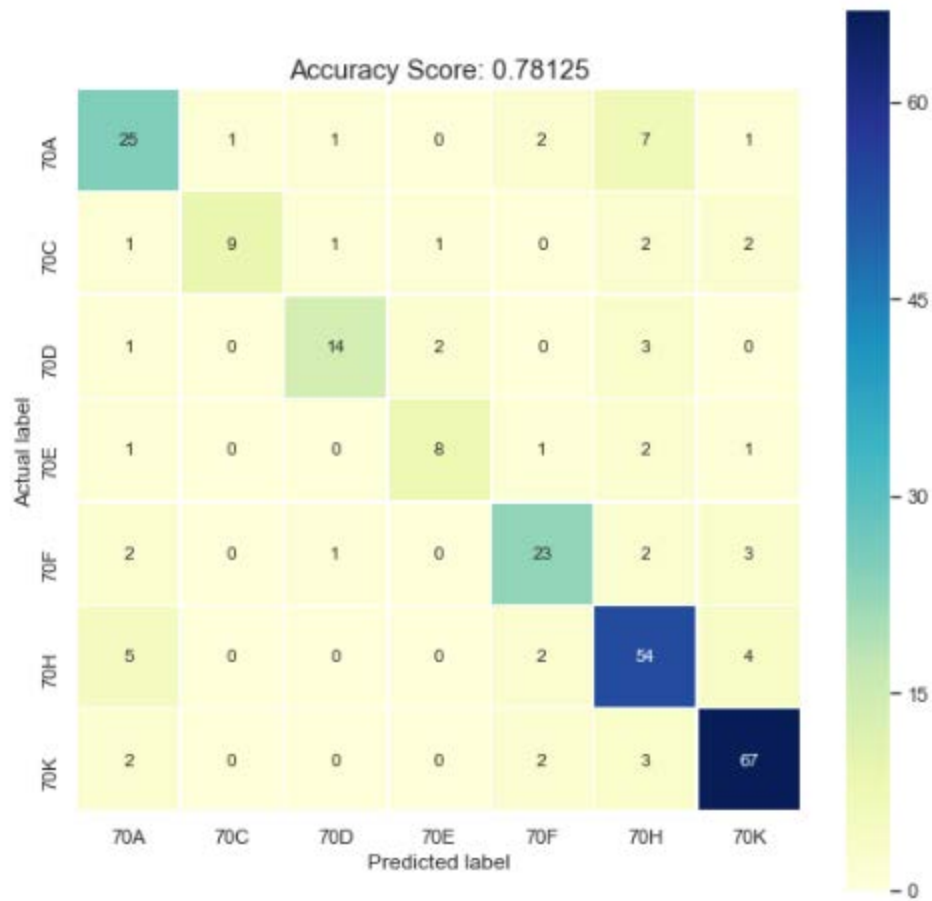


Figure 69. Confusion Matrix Between Fully Connected Neural Network Classifier Predictions and Test Set AOC Assignments.

Table 24. Fully Connected Neural Network Test Set Accuracy

Fully Connected Neural
Network Classifier Testing
Summary Statistics

	Value
Accuracy	0.781
95% CI Lower Bound	0.726
95% CI Upper Bound	0.830

The misclassification patterns and per-AOC performance are similar to those of other classifiers; however, the FNN tends to fare worse in sensitivity and better in specificity for all AOCs than the three models we presented earlier in the chapter. Figure 70 displays ROC curves for the support vector machine classifier per AOC, while Table 25 shows statistics for per AOC performance.

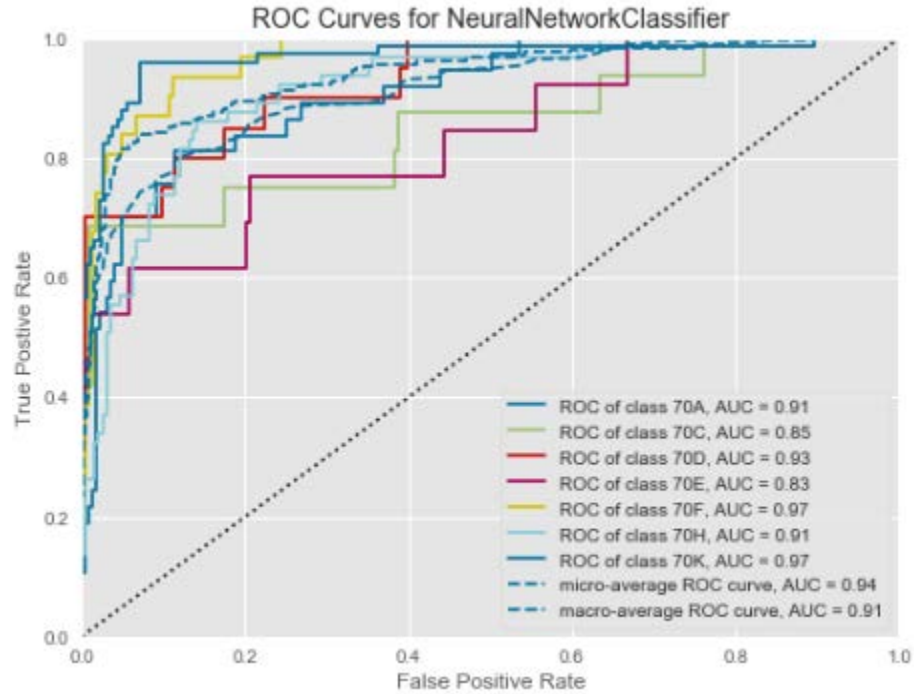


Figure 70. Receiver Operator Curves for Fully Connected Neural Network Classifier

Table 25. Sensitivity and Specificity for Fully Connected Neural Network Classifier

Fully Connected Neural Network Classifier Testing per AOC

	70A	70C	70D	70E	70F	70H	70K
Sensitivity	0.676	0.900	0.824	0.727	0.767	0.740	0.859
Specificity	0.945	0.972	0.975	0.980	0.965	0.940	0.961
No Information Rate	0.145	0.062	0.078	0.051	0.121	0.254	0.289

The complexity and flexibility of the FNN seems to provide no significant advantage over the simpler and more interpretable models explored earlier in the chapter. The FNN’s feature importance and misclassification patterns were also analogous to those of less complex models with few exceptions.

F. MODEL RESULTS AND COMPARISON

In this section, we compare the four models against each other in order to identify the best-suited model for guiding AOC-MS phase I recruitment efforts. Our initial comparison of the four classifiers we’ve studied in this chapter does not suggest significant differences between the classifiers. We turn to Cochran’s Q test, a multivariate nonparametric test for identifying differences in classification patterns between classifiers trained and tested using the same data (Raschka 2018).

We let $\mathcal{C} = \{c_{MLR}, c_{RF}, c_{SVM}, c_{NN}\}$ be the set of four machine learning classifiers. Using the test set predictions, we test the null hypothesis $H_0: c_{MLR} = c_{RF} = c_{SVM} = c_{NN}$ against the alternative that at least one of the classifiers differs from the others. The results of the hypothesis test are in Table 26. We fail to reject the null hypothesis at $\alpha = 0.05$, concluding that we lack sufficient evidence to determine whether there are differences between the classifiers.

Table 26. Comparison of Classifier Performance

Cochran Q Test Results	
Q.Statistic	0.900
df	3.000
P.value	0.828

We likewise compare classifier performance against one another pairwise using the pairwise McNemar test (Brownlee 2018), testing the hypothesis $H_0^{(c,c')}: c = c' \forall (c, c') \in \mathcal{C}$, determining whether classifier c performs equally to classifier c' . The results of the six

tests are located in Table 27. We fail to reject all null hypotheses or to establish any evidence that a particular classifier performs differently from any other one.

Table 27. Pairwise Performance Comparisons Between Machine Learning Classifiers

Pairwise McNemar Test for Machine Learning Classifier Effectiveness

Classifier Comparison	P-Values
NN vs SVM	0.814
NN vs MLR	0.803
NN vs RF	0.831
SVM vs RF	0.823
SVM vs MLR	0.773
RF vs MLR	0.502

Failing to establish a significant difference between classifier performance, we can simply pick the best-performing model, the multinomial logistic classifier, as our model of choice for AOC-MS phase I.

However, given that our intent is to assign AOC matching scores to officers who have not yet designated an AOC at their 6th year of commissioned service, we have no way of using a test set to check the classifier’s performance. We therefore assess the classifiers according to how well their predictions follow the multinomial distribution of AOC assignments in the MFA 70 population, which we introduced in Chapter IV. Since officers will eventually be allocated to the general population during AOC designation, we look for models to generate predictions which closely follow the distribution of AOCs in the designated population and which do not favor a particular AOC over the others.

We generate predictions from all four models using $\mathbf{U}$, a matrix containing the records of 79 undesignated officers in their 6th year of commissioned service. Figure 71 displays the distribution of AOC assignments in the population as well as the distribution of AOC predictions from the neural network classifier on $\mathbf{X}_{test}$ and $\mathbf{U}$ records. We can

visually observe that the distribution of predictions on the unlabeled officers is very different from the other two distributions. We also see that the classifier overwhelmingly predicts undesigned officers will serve as 70Hs at the expense of other AOCs. The remaining three classifiers provide similar prediction patterns as the neural network.

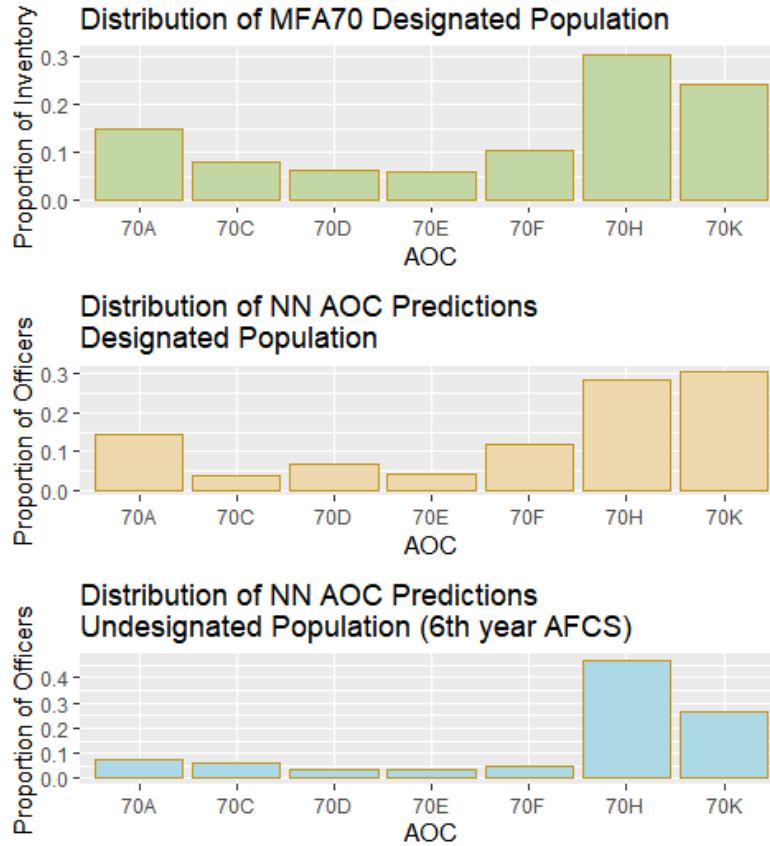


Figure 71. Distribution of AOC Assignments and NN AOC Predictions

In order to establish whether the model predictions fit the same multinomial distribution as the designated population, we use the Chi-squared multinomial goodness of fit (GOF) test (Zelterman 1987). The test uses the Chi-squared statistic to measure how well observed frequencies match expected frequencies under the null hypothesis $H_0^{c \in C}: c \sim \text{Mult}(p_a \in A), \forall c \in C, a \in A$.

We use the Chi-Squared distribution with $|A| - 1 = 7 - 1 = 6$ degrees of freedom. If the probability the test statistic χ^2 comes from the specified distribution is

smaller than $\alpha = 0.05$ we reject the null hypothesis. The results of the GOF tests are located in Table 28. For all classifiers, we reject the null hypothesis and establish that the AOC predictions are not derived from the same distribution as the general population.

Table 28. Pearson GOF Test for Classifier Predictions on Unlabeled Dataset

Multinomial Goodness of Fit Test for Unlabeled Dataset Predictions		
Classifier	Chi Squared Test Stat	Multinomial GOF Test P-Value
MLR	22.6	0.002
RF	32.1	0.001
SVM	18.9	0.004
NN	13.6	0.032

The multinomial GOF test establishes whether the distribution of the predictions comes from the same distribution as the population. We turn to a modification of Cramer's V statistic to analyze the scale of departure from the population distribution (Mangiafico 2016). Cramer's V is a number between 0 and 1 which quantifies the magnitude of the effect.

We use the formula $V_{c \in \mathcal{C}} = \sqrt{\frac{\chi^2}{|U|(|A|-1)}}$ to calculate Cramer's V for every classifier, where χ^2 is the test statistic for the multinomial GOF test, $|U|$ is the number of observations and $|A|$ is the number of AOCs.

A Cramer's V value closer to 1 denotes a larger departure from the distribution of the population. For $\alpha = 7$, a Cramer's V value ≥ 0.217 suggests a large effect size (Cohen 1988). Table 29 displays Cramer's V statistics for $\mathbf{U}$, while Table 30 shows the same information for $\mathbf{X}_{test}$ predictions for comparison. Figure 72 displays the same information graphically.

Table 29. Cramer's V for $\mathbf{U}$ AOC Predictions

Scale of Departure of Unlabeled AOC
Predictions from Population
Distribution

Classifier	Cramer's V	LCI	UCI
MLR	0.219	0.155	0.298
RF	0.260	0.196	0.334
SVM	0.200	0.136	0.287
NN	0.170	0.120	0.263

Table 30. Cramer's V for $\mathbf{X}_{test}$ AOC Predictions

Scale of Departure of Labeled Test Set
Predictions from Population
Distribution

Classifier	Cramer's V	LCI	UCI
MLR	0.091	0.065	0.147
RF	0.105	0.075	0.163
SVM	0.086	0.063	0.144
NN	0.085	0.063	0.146

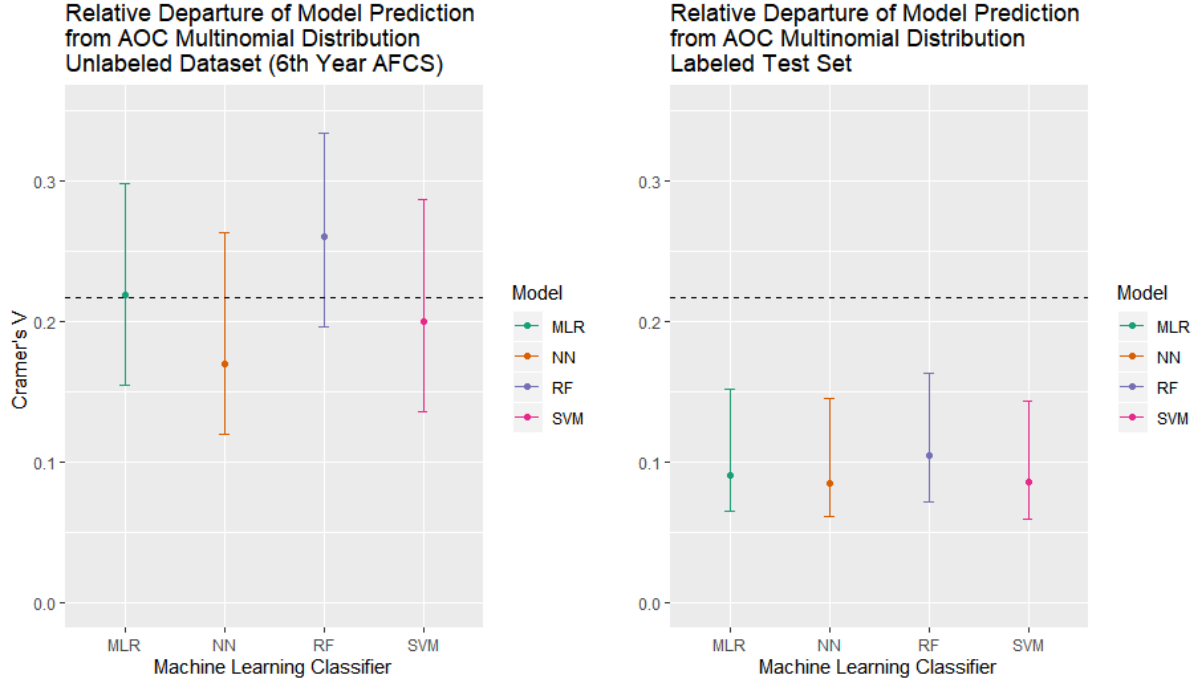


Figure 72. Cramer's V Statistics for $\mathbf{U}$ and $\mathbf{X}_{test}$ Model Predictions

First, we observe that the neural network classifier predictions feature the smallest departure from the population distribution, and that these are closely followed by predictions from the support vector machine. The remaining two classifiers feature large scale departures from the population distribution.

Secondly, model predictions for records from $\mathbf{U}$ feature significantly larger departures from the population distribution than predictions for $\mathbf{X}_{test}$ records, suggesting differences not in the classifiers but in the data itself.

We return to the multidimensional analysis performed on the AOC Predictor dataset in Chapter IV to visualize the records in three dimensions using Euclidean distance. Figure 73 provides a three-dimensional representation of $\mathbf{X}_{test}$, color coded by AOC assignment. We display the same analysis for the records for the undesignated officers in $\mathbf{U}$ in Figure 74. We can observe that the datasets are structured differently, and that for $\mathbf{U}$, the primary separating feature appears to be whether an officer is a 70B or a 67J. The labels on each of the three-dimensional components represent one-dimensional Euclidian distance.

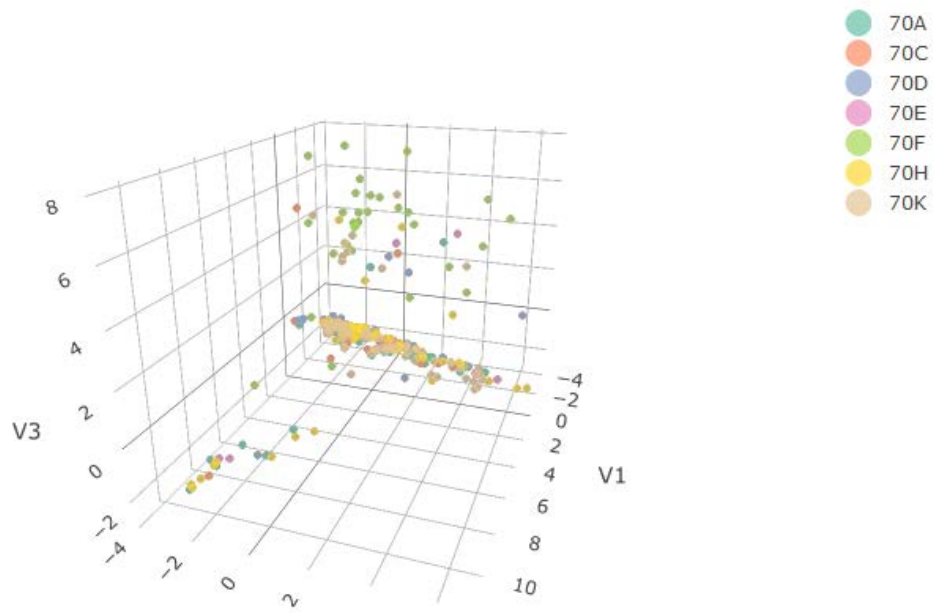


Figure 73. Three-Dimensional Mapping of X_{test}

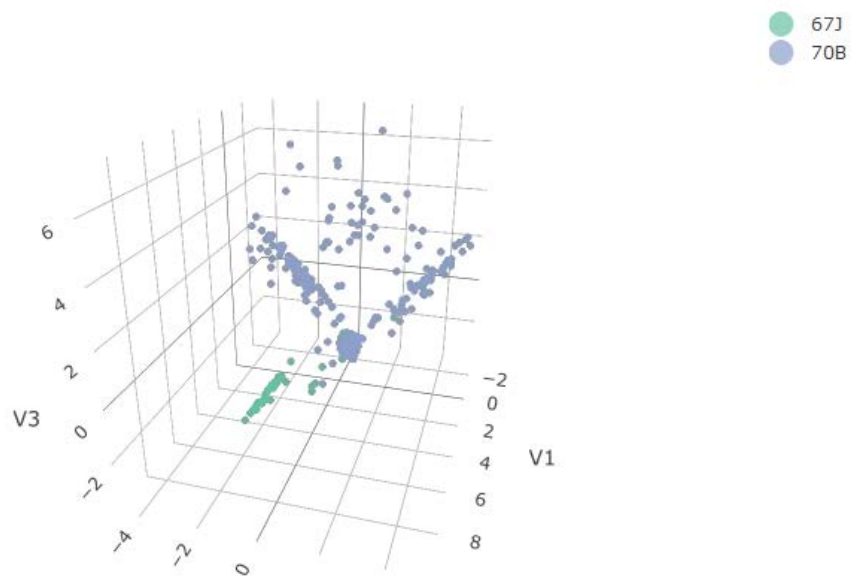


Figure 74. Three-Dimensional Mapping of U

We theorize that the two years of experience separating officers in $\mathbf{U}$ from those in $\mathbf{X}_{test}$ provide significant development experiences which help separate the groups into the specialized AOCs while the more junior cohort is still dominated by the entry-level AOC experience. Since 70H is the logical continuation of 70B and 67J work, it is likely that the classifiers are using 70H as a proxy for their more junior status, therefore predicting a higher proportion of officers to be matches for 70H.

However, all officers in $\mathbf{U}$ will be designated into specialized AOCs within two years and will likely look much like the officers in $\mathbf{X}_{train}$. The skills, knowledge and attributes which will make these officers successful in their AOCs are likely already present in the officers.

AOC-MS phase I needs classifiers which can recognize these differences earlier in officers' careers in order to better recommend AOC matches. This will require improved data collection that relies on self-professed officer talents and desires as opposed to only relying on assignment and education data.

As part of the Army talent alignment process introduced in Chapter II, officers provide HRC with an electronic resume using the Assignment Interactive Module 2.0 (AIM2). As part of this resume, officers self-profess skills, knowledge and attributes. We theorize that training classifiers using assignment and education data in addition to AIM2 resume data will yield a much richer picture of officer aptitude and will help separate officers from one another earlier in their careers better than educational and assignment data alone. AIM2 resume data was not available for completion of this study; therefore, this theory cannot yet be tested.

In the absence of improved training data, we pick the multinomial logistic regression classifier as the model of choice for AOC-MS phase I. It performed best of all four models when predicting against $\mathbf{X}_{test}$ records. While its predictions departed from the population distribution more than those of the neural network and the support vector machine, the departure barely meets the threshold for being considered large. Lastly, the model is simple and easily explainable to officers and decision-makers.

VI. AREA OF CONCENTRATION DESIGNATION ALGORITHMS

In this chapter we study algorithmic approaches for solving the AOC designation problem during phase II of AOC-MS and compare them to the results of the FY2020 AOC designation process. To do so, we report on the results of a pilot program conducted in coordination with HRC, APPD and the office of the MSC chief during May and June of 2020.

In Section A we introduce the AOC-MS pilot program. In Section B we investigate the process for determining the correct AOC-MS talent matching algorithm (TMA) variant based on the multinomial distribution of officer preferences and OFM targets. In Section C of this chapter we examine the stable marriage variant of the AOC-MS TMA and in Section D we address the ILP variant. Lastly, in Section E we compare both variants of the AOC-MS TMA.

A. THE AREA OF CONCENTRATION MATCHING SYSTEM PILOT PROGRAM

In order to validate the AOC-MS phase II concept, HRC MSC branch and the office of the MSC chief, provided AOC preferences for 30 PYG2012 70Bs in the mandatory window of designation to fill 40 OFM-produced slots. The designation problem was infeasible, with 10 slots going unfilled and no direction from APPD or OCMSC for panel members regarding how the shortages should be distributed between AOCs.

HRC sanitized preference records of all PII and demographic data, assigning letter codes to officers as unique identifiers, so that officer $o \in O = \{A, B, C \dots AA, AB, AC, AD\}$. These letter-based unique identifiers do not match the number-based unique identifiers previously provided by HRC, preventing the matching of pilot program preference records with assignment data in the AOC Designation Records dataset. Twenty-seven officers ranked all 7 AOCs in order of preference while 3 of the 30 officers provided no AOC preferences.

Four out of seven panel members provided officer preference data and word picture scores for the designating officers. For officer preferences, panel members ranked officers into four tiers, with Tier 1 representing the most preferred tier. Tier 1 rankings were limited to under 25% of designating officers, and combined Tier 1 and Tier 2 rankings were limited to under 50% of officers. Tier 3 and 4 rankings were unrestricted. Word picture scores ranged from 1 to 6.5. For further information regarding how word picture scores are developed, refer to Figure 7 and Chapter I of this study.

Panel members also provided their AOC designation votes to be tallied by HRC using the current AOC designation methodology. Due to the previously discussed force structure imbalances caused by the current AOC designation process, the OFM prescribed large designation targets for short AOCs like 70D and 70F. However, officer preferences were heavily skewed towards overstrength AOCs like 70A and 70C, creating a friction between officer preferences and force structure needs that the current designation process has repeatedly failed to address.

The panel's proclivity for granting officers their top preference once again missed OFM targets for understrength AOCs and exceeded them for overstrength AOCs, deepening the already-existing AOC inventory disparities. Table 31 introduces the designation targets for PYG2012 along with the final designation provided by the FY2020 AOC designation panel. The bottom row of the table features the proportion of target fill with panel underfills highlighted in yellow and overfills highlighted in red. We can observe that the panel designated 150% more officers than necessary to 70A and 33% more than necessary to 70C while filling AOC 70D at 25% of its OFM target and 70F at 0% of its OFM target.

Table 31. FY2020 AOC Designation Process Results

	70A	70C	70D	70E	70F	70H	70K	Total
FY2020 AOC Designation Panel Results								
Minimum OFM Target	1	2	8	4	8	13	4	40
FY2020 Panel Designation	3	3	2	4	0	13	5	30
Panel Departure From Target	2	1	-6	0	-8	0	1	-10
Panel FY2020 Proportion Fill	3	1.5	0.25	1	0	1	1.25	0.75

Lastly, HRC provided a list of officers with prior AOC experience or LTHET acceptances to accompany the preference data. Figure 75 visualizes the AOC-MS Pilot Program dataset using the missing no data completeness visualization package for Python (Bilogur 2018).

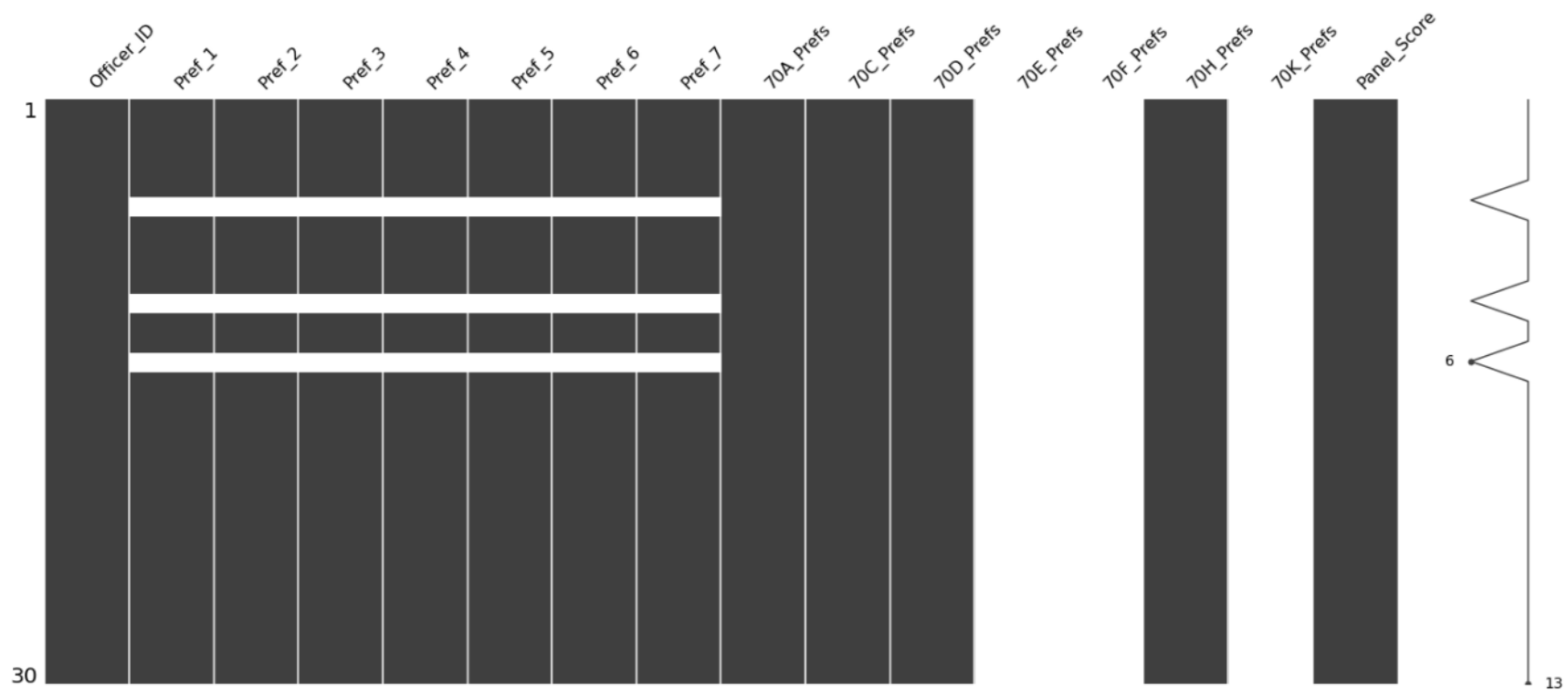


Figure 75. AOC-MS Pilot Program Missing Data

Given the missing preference data from both officers and AOCs, we address the need for the AOC-MS phase II to be robust to missing preference data. Since we cannot make machine learning classifier model predictions for the designating officers in the pilot program due to the incompatibility of unique officer identifiers, the imputation methodology is different between the pilot program and a full-scale implementation of AOC-MS. In the full implementation we impute missing preference data using the multinomial logistic classifier discussed in Chapter V to predict the probability of AOC assignment for each officer. For imputed officer preferences, we preference the AOCs in the order of magnitude for model-predicted probabilities.

For imputed AOC preferences, we rank order officers according to the model-predicted probabilities for the AOC with missing preferences. We then split the data into quartiles, assigning Tier 1 preferences to officers in the first quartile of model predicted probabilities of AOC assignment and Tier 4 to the officers in the lowest quartile.

For the pilot program only, we impute missing officer preference data by making officers indifferent to AOC assignment. For AOCs, we impute missing preference data using the following logic: We assign a Tier 1 preference for officers with prior AOC experience or LTHET acceptance and Tier 2 otherwise, relaxing the tier ranking constraints, making these AOCs largely indifferent to officer assignments. Table 32 displays sample officer preference data records and Table 33 shows AOC panel member preference records. We will use these records in the remainder of this chapter to find the best matching of officers and AOCs.

Table 32. Sample Officer Preferences for FY2020 AOC-MS Pilot Program

	Pref_1	Pref_2	Pref_3	Pref_4	Pref_5	Pref_6	Pref_7
Officer_ID							
A	70K	70C	70A	70H	70E	70F	70D
AA	70C	70A	70H	70K	70D	70E	70F
AB	70E	70H	70K	70A	70F	70C	70D
AC	70H	70A	70K	70C	70F	70E	70D
AD	70A	70D	70C	70H	70K	70F	70E

Table 33. Sample AOC Preferences for FY2020 AOC-MS Pilot Program

	70A_Prefs	70C_Prefs	70D_Prefs	70E_Prefs	70F_Prefs	70H_Prefs	70K_Prefs	Panel_Score
Officer_ID								
A	4	2	3	2	2	2	1	3.0
AA	1	1	3	2	2	3	2	5.5
AB	4	3	2	1	2	1	2	4.0
AC	2	3	2	2	1	2	2	5.0
AD	1	3	1	2	2	3	2	5.0

B. TALENT MATCHING ALGORITHM (TMA) VARIANT SELECTION

In Chapter IV we introduced the infeasibility of the AOC designation problem in years when the number of available officers for designation was insufficient to meet the minimum OFM targets. We also introduced the conflict between satisfying officer preferences and meeting OFM targets, acknowledging the current process's tendency to disregard designation targets in favor of preference satisfaction, creating deepening imbalances over the past five PYGs. The logical consequence of this conflict is that meeting OFM targets and rebalancing AOC strength will likely result in lower officer preference satisfaction unless the distribution of officer preferences is very similar to the distribution of OFM targets.

Quickly rectifying the force structure disparities created by the current process will likely result in PYGs with large proportions of dissatisfied officers who may be less likely to remain in the service, while continuing to satisfy preferences at the expense of filling manpower targets will likely result in an unhealthy force unable to fulfill all of its assigned missions. AOC-MS addresses this conflict by providing the MSC with two variants of the AOC-MS TMA.

The first variant uses a stable marriage algorithm to match officers and AOCs and requires strict AOC designation targets. This variant is appropriate for PYGs where officer

preferences are closely aligned with OFM targets. The second variant uses an ILP to simultaneously set targets and make officer/AOC matches. This variant is adequate for PYGs with significant differences between the distribution of officer preferences and OFM targets as it gives the MSC the flexibility to balance the conflict between preferences and OFM targets.

In order to determine the appropriate AOC-MS TMA variant to utilize, we analyze the multinomial distribution of officer preferences and compare it against the distribution of OFM targets. We utilize the multinomial goodness of fit test and Cramer's V statistic, using the same procedure introduced in Chapter V for comparing the distribution of machine learning classifier predictions against the distribution of AOC assignments. For the selection of AOC-MS TMA variant, a Cramer's $V \geq 0.217$ indicates a large-scale difference between the distributions (Cohen 1988) and requires the use of ILP variant of the algorithm. Cramer's $V < 0.217$ indicates that use of the stable marriage variant is warranted.

Figure 76 displays the distribution of OFM targets, officer first preferences and panel designations for the FY2020 AOC designation panel. Visually, we can observe the discrepancy between officer preferences and OFM targets, with officers over-preferring AOCs like 70A and 70C and under-preferring AOCs like 70D and 70F. We can also observe that the panel again opted for satisfying officer preferences instead of OFM targets.

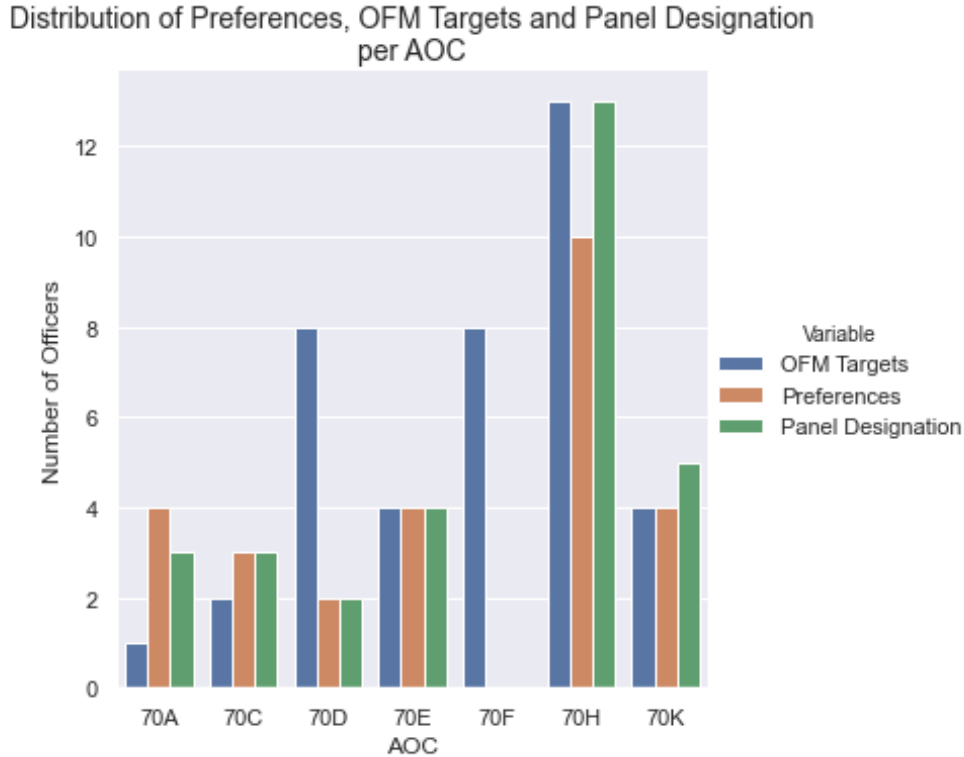


Figure 76. Discrepancy Between Preferences, Targets and Designations for AOC-MS Pilot Program

For PYG 2012, Cramer's $V = 0.35$, indicating a large-scale difference between the distributions and the need to use the ILP variant of the AOC-MS TMA. For completeness, we apply both variants of the TMA to the pilot program preferences and OFM targets in order to analyze them and compare them to the current designation process.

C. TALENT MATCHING ALGORITHM (TMA) STABLE MARRIAGE VARIANT

In Chapter II we introduced the Gale-Shapley matching algorithm as an approach for solving the AOC designation problem. In Gale-Shapley's version of the algorithm, matching sets are equally sized, all preferences are strictly ranked and there are no missing preferences. None of these requirements are met by the AOC designation data. In the first section of this chapter, we discussed our approach for handling missing preferences.

In part 1 of this section we discuss the use of an ILP to create the equally sized sets required by the Gale-Shapley algorithm. In part 2 we introduce the adaptations required for creating strictly ranked preference lists. In part 3 we explore the stable marriage variant of the AOC-MS TMA and in part 4 we analyze the algorithm results and compare them against the current process.

1. The AOC-MS Target-Setting Algorithm (TSA)

In general, the number of officers up for designation will not strictly equal the number of designations required by the OFM output. In order to create equally sized sets for the matching algorithm to utilize, we turn to an ILP. The AOC-MS target-setting algorithm (TSA) maximizes the proportion of AOC fills across one or multiple PYGs while discouraging disparities between AOCs.

The AOC-MS TSA considers the proportion of fill for multiple PYGs, providing a more holistic view of AOC strength than a single-PYG view. If an AOC is overfilled across several PYGs, a shortage during the current designation cycle may not be as harmful as it would be for AOCs with a history of shortages. The decision about how many PYGs to consider is a policy decision; however, we recommend considering no more than 5 PYGs, as the difference in experience beyond this threshold makes the most senior officers in the considered group substantially different from designating officers.

In order to assess the AOC-MS target-setting algorithm's performance against the FY2020 AOC designation panel results, we allocate the 10 70B shortages across the seven AOCs. In the current designation process, the panel member votes decide which AOCs receive shortages and which receive overages. These choices are not coordinated across panel members and do not consider whether AOCs receiving shortages are already staffed below OFM targets in previous year groups. Therefore, there is no method for ensuring these shortages are distributed optimally. The AOC-MS TSA rectifies this shortcoming.

a. AOC-MS TSA Formulation

We introduce the formulation for AOC-MS TSA utilizing NPS format (Dell 2019).

Indices

Y Index for officer's promotion year groups $y \in Y = \{y_D, y_2 \dots y_\psi\}$. For the PYG2012 AOC designation problem, $Y = \{2012, 2011, 2010, 2009, 2008\}$.

A Index for officer AOCs $a \in A$, where $A = \{70A, 70C, 70D, 70E, 70F, 70H, 70K\}$.

Data Parameters

i_{70B, y_D} Total 70B inventory from the designating PYG available for designation [Officers]

$i_{a \in A, y \in Y}$ Total on-hand inventory per AOC a for the PYG $y \forall a \in A, y \in Y$ [Officers]

$o_{a \in A, y_1}$ OFM-produced minimum targets per AOC a for the designating PYG $y_1 \forall a \in A$ [Authorizations]

$o_{a \in A, y \in Y}$ OFM-produced authorizations per AOC a for the PYG $y \forall a \in A, y \in Y$ [Authorizations]

User-Defined Parameters

ψ Number of PYGs considered for assessing AOC fill rates. $\psi = |Y|$. [Number of PYGs]

ρ Scalar between 0 and 1 designating the minimum proportion of officers that must be designated to all AOCs. [Scalar]

π^+ Positive scalar weighing penalty variables $P_{a \in A}^+$ [Scalar]

π^- Positive scalar weighing penalty variables $P_{a \in A}^-$ [Scalar]

Decision Variables

$D_{a \in A}$	Integer variable reflecting positive or negative departures from OFM-produced designation targets for designating officers from PYG y_1 to AOC a . $\forall a \in A$ [officers]
$P_{a \in A}^+$	Nonnegative real-valued variable penalizing the objective function for filling AOCs below 100% of OFM authorizations $\forall a \in A$ [Underfill penalty]
$P_{a \in A}^-$	Nonpositive real-valued variable penalizing the objective function for filling AOCs above 100% of OFM authorizations $\forall a \in A$ [Overfill penalty]

Objective Function

$$\max_{D_a, P_a^+, P_a^-} \frac{1}{|A|} \sum_{a \in A} \left(\frac{D_a + o_{a,y_1} + \sum_{y \in Y} i_{a,y}}{\sum_{y \in Y} o_{a,y}} - \pi^+ P_a^+ + \pi^- P_a^- \right)$$

Constraints

$$\sum_{a \in A} D_a + o_{a,y_1} = i_{70B,y_1} \quad (1)$$

$$D_a + o_{a,y_1} \geq \rho * o_{a,y_1} \quad \forall a \in A \quad (2)$$

$$\frac{1}{\sum_{y \in Y} o_{a,y}} \sum_{y \in Y} (i_{a,y} + D_a + o_{a,y_1}) + P_a^+ \geq 1 \quad \forall a \in A \quad (3)$$

$$\frac{1}{\sum_{y \in Y} o_{a,y}} \sum_{y \in Y} (i_{a,y} + D_a + o_{a,y_1}) + P_a^- \leq 1 \quad \forall a \in A \quad (4)$$

$$D_a \in \mathbb{Z} \quad \forall a \in A \quad (5)$$

$$P_a^+ \geq 0 \quad \forall a \in A \quad (6)$$

$$P_a^- \leq 0 \quad \forall a \in A \quad (7)$$

Model Description

The objective function maximizes the sum of the AOC fill proportions across all PYGs while penalizing both overfills and underfills. The goal of the objective function is to keep all AOCs as filled at near 100% of OFM authorizations across the ψ -year horizon. The penalty variables P_a^+ and P_a^- are weighed using scalars π^+ and π^- , allowing analysts to penalize the objective function differently for overfills and underfills.

Constraint (1) ensures the sum of AOC designation targets is equal to the number of officers available for designation. Constraint (2) creates a designation floor per AOC utilizing the ρ scalar to scale down the OFM-produced target. This constraint ensures at least $\rho * o_{a,y_1}$ officers designate on a given FY. Constraint (3) ensures that each AOC is filled above its OFM authorizations and in cases where such a proportion fill is infeasible, activates the penalty variable P_a^+ . Constraint (4) ensures AOCs are not filled above OFM targets, activating the penalty variable P_a^- when this is infeasible. Together, constraints (3) and (4) drive the AOC fill percentage towards 100%, encouraging overfills or shortages according to the weight of the scalars π^+ and π^- . Constraint (5) ensures that the departure variables D_a are integer-valued. Constraint (6) ensures P_a^+ is nonnegative, while constraint (7) makes P_a^- nonpositive.

b. AOC-MS TSA Implementation

In order to reallocate the infeasible OFM targets for PYG2012, we implement the AOC target-setting program using Python's Pyomo optimization package (Hart et al. 2017) and the Coin-OR Branch and Bound (CBC) solver (Forrest et al. 2020). We set the user-defined parameters to $\psi = 5$, $\rho = \frac{1}{4}$, $\pi^+ = 2$, and $\pi^- = 1.5$, penalizing underfills more than we penalized overfills and ensuring that the AOC designation process fills at least a quarter of OFM-produced targets per AOC.

The AOC-MS TSA is a relatively simple ILP with 35 equations and 21 variables. CBC solved the problem in 0.01 CPU seconds on a 64-bit Samsung PC with 8GB RAM and an Intel core i7-8550U CPU processor, reaching an optimal objective value of 1.27.

c. AOC-MS TSA and Current Process Comparison

In the aggregate, there are more than enough available officers to fill all OFM slots in PYGs 2008–2012. Despite this, the FY2020 AOC designation panel continued to overfill AOCs already overfilled by the previous designation processes while keeping AOCs 70D and 70F short. By contrast, the AOC target-setting algorithm ensures that all AOCs were staffed at a minimum of 100% of OFM authorizations. Figure 77 compares the 5-year fill rates between the FY2020 AOC designation panel and the AOC-MS TSA against the pre-designation 5-year fill rate.

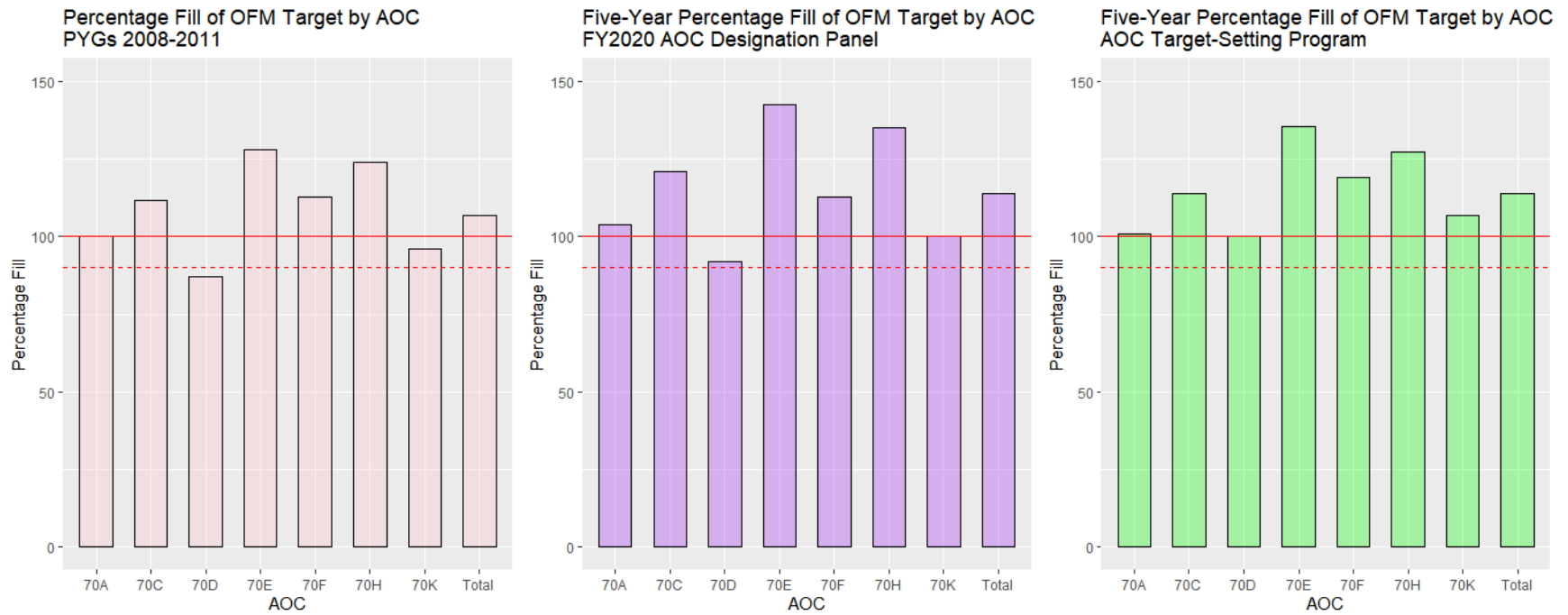


Figure 77. Percentage Fill of OFM Targets for FY2020 AOC Designation

Table 34 displays the raw AOC designation targets produced by the AOC-MS TSA and by the FY2020 AOC designation panel using ψ -PYG horizons for calculating fill percentages and holding all other user-defined parameters constant.

Table 34. FY2020 AOC Designation Panel and AOC-MS TSA Raw Targets

	70A	70C	70D	70E	70F	70H	70K	Total
Targets								
5 year	1	1	5	2	3	4	14	30
4 year	1	1	7	2	3	4	12	30
3 year	4	1	6	2	3	4	10	30
2 year	3	1	5	2	3	4	12	30
1 year	1	1	6	3	4	13	2	30
Panel	3	3	2	4	0	13	5	30

Table 35 displays the ψ -PYG horizon fill rates using the FY2020 AOC designation panel allocations while Table 36 displays the same information using the AOC target-setting program allocations. In the tables, we highlight in yellow instances where the designation process filled an AOC below 100% of OFM targets. The current process underfills at least one AOC at every level of ψ while the AOC-MS target-setting algorithm fills all AOCs at 100% of OFM targets at every ψ -year horizon.

Table 35. FY2020 AOC Designation Panel Allocation Fill Percentages

	70A	70C	70D	70E	70F	70H	70K	Total
Current Process Fill								
5 year	1.05	1.21	0.921	1.43	1.13	1.35	1	1.14
4 year	1.08	1.15	0.839	1.43	1.05	1.38	0.99	1.13
3 year	0.974	1.25	0.826	1.47	1	1.5	1.08	1.18
2 year	1	1.38	0.8	1.27	0.895	1.42	1.19	1.18
1 year	1.18	1.5	0.429	1.2	0.889	1.5	1.45	1.27

Table 36. AOC-MS TMA Fill Percentages

	70A	70C	70D	70E	70F	70H	70K	Total
AOC-MS Fill								
5 year	1.015	1.147	1	1.357	1.191	1.274	1.070	1.143
4 year	1.038	1.074	1	1.348	1.128	1.280	1.057	1.132
3 year	1.000	1.150	1	1.353	1.103	1.368	1.147	1.177
2 year	1.000	1.231	1	1.091	1.053	1.222	1.333	1.176
1 year	1.000	1.167	1	1.000	1.333	1.500	1.318	1.268

After using the AOC-MS TSA to generate designation targets which match the number of officers available, we create a preference list of length $|O|$ for every AOC $a \in A$, and each officer has a preference list of length $|A|$. We transform the matching problem from a many-to-one problem, where we match several officers to a single AOC to a one-to-one problem where we match officers to AOC slots $a_s \in A_s$ where $|A_s| = |O|$.

2. Strictly Ordered Preference Lists

By having consultants rank order officers in preference tiers, we introduce the problem of ties. This is a matter of practicality, as asking panel members to provide strict rankings of close to 100 officers in large year groups is unrealistic. While the first five or ten rankings might be informative, the panel member is likely indifferent between officers in the middle of the preference list. We manage the problem of preference ties by using the word picture scores to rank order officers within preference tiers, generating a strict ranking between all designating officers.

For example, if the target-setting algorithm generates three 70A slots, three officers will be matched to three different AOC slots 70A, 70A2 and 70A3, with all three slots being exact clones of each other and featuring the exact same preference list. Officer preference lists expand from length 7 to length $|A_s|$ by having officers express the same preferences for each AOC slot.

In the case of a theoretical officer whose first two AOC preferences are 70A and 70C, with three available slots for 70A and two for 70C, the officer's preference list would

transform from $\{70A, 70C, \dots, opref_7\}$ to $\{70A, 70A1, 70A2, 70C, 70C1, \dots, opref_{s_{|A_s|}}\}$

where officer rank the slots for their preferred AOCs in order. However, in practice officers are indifferent to slots within a particular AOC and AOCs are indifferent to officers within their slots.

The result of the adaptations are two square preference tables of dimension $|O| \times |O|$, with table **OP** _{$|O| \times |O|$} holding officer preferences for AOC slots and the table **AP** _{$|O| \times |O|$} holding ordinal AOC slot preferences for officers in the form $\{A, B, C, \dots, cpref_{|O|}\}$. Sample rows and columns from table **OP** are displayed on Table 37, while Table 38 features sample entries from **AP**.

Table 37. Sample Records from officer Preference Table **OP**.

Officer	A	AA	AB	AC	AD	B	C	D	E	F
Preference										
1	70K	70C	70E	70H	70A		70K	70H	70D	70E
2	70K2	70C2	70E2	70H2	70A2		70K2	70H2	70D2	70E2
3	70K3	70C3	70E3	70H3	70A3		70K3	70H3	70K	70E3
4	70K4	70A	70E4	70H4	70D		70K4	70H4	70K2	70E4
5	70K5	70A2	70H	70H5	70D2		70K5	70H5	70K3	70C
6	70C	70A3	70H2	70H6	70C		70C	70H6	70K4	70C2
7	70C2	70H	70H3	70H7	70C2		70C2	70H7	70K5	70C3
8	70C3	70H2	70H4	70H8	70C3		70C3	70H8	70H	70H
9	70A	70H3	70H5	70H9	70H		70E	70H9	70H2	70H2
10	70A2	70H4	70H6	70H10	70H2		70E2	70H10	70H3	70H3

Table 38. Sample Records from AOC Preference Table **AP**.

AOC Slot	70A	70A2	70A3	70C	70C2	70C3	70D	70D2	70E	70E2
Preference										
1	AA	AA	AA	AA	AA	AA	M	M	F	F
2	T	T	T	T	T	T	T	T	AB	AB
3	L	L	L	U	U	U	AD	AD	AA	AA
4	AD	AD	AD	S	S	S	O	O	S	S
5	H	H	H	D	D	D	H	H	D	D
6	S	S	S	O	O	O	L	L	O	O
7	AC	AC	AC	N	N	N	AC	AC	AC	AC
8	I	I	I	M	M	M	I	I	AD	AD
9	Q	Q	Q	L	L	L	F	F	L	L
10	C	C	C	W	W	W	P	P	M	M

3. AOC-MS TMA Stable Marriage Variant Implementation

After adapting the AOC designation problem to meet the one-to-one structure of the Gale-Shapley marriage problem, we execute the AOC-MS TMA stable marriage variant with officers proposing assignments to AOCs in accordance to the logic laid out in Algorithm 1, located in Chapter II.

Algorithm 2 provides pseudocode for the AOC-MS TMA stable marriage variant. This algorithm was implemented in Python; however, it is language-agnostic and can be implemented in R or any other programming language. The inputs to the implemented function are the preference tables for officers and AOCs.

Algorithm 2. AOC-MS Talent Matching Algorithm Pseudocode –
Stable Matching Variant. Adapted from Stable Marriage Problem
(2013)

AOCMatching (officer Preferences, AOC Preferences)

#1. Boolean Check to see whether matrices are square and dimensions are equal
and preference values are unique

if Boolean True:

 continue

else:

 error message

 break

#2. Initialize data structure settings all officers and all AOCs to free

officer Status per officer = Free

AOC Slot Status per AOC Slot = Free

#3. Initialize data structure for storing matches

Matchings = Empty data structure

#4 .Extract indices

officer List = Indices for officer Preferences

AOC Slot List = Indices for AOC Slot Preferences

#5. Randomly shuffle officer list to avoid ordinal impacts

Shuffled officers = Sample (officer List, Without Replacement, for length of
officer List)

#6. Impute Preferences

Predict probabilities from Multinomial Model for all officer records

Store probabilities in a matrix

#6a. Find and impute officer missing preferences

Iterate through Shuffled officers:

 if officer AOC Preferences are Blank:

 find officer AOC probability list

 sort officer AOC probability list in descending order

 store preferences in order of probability in officer Preference
 Matrix

#6b. Find and impute AOC preferences

Predict probabilities from Multinomial Model for officer record

Iterate through AOC Slots:

 if AOC Preferences are Blank:

 find vector with AOC probabilities for all officers

 sort list in order and separate into quartiles

 assign officers into Tiers 1 through 4

 sort officers by panel score within tier

```

        Iterate through AOC slots:
            Store AOC preference list

#7 Initiate Main Matching Loop
initiate freeCount to number of officers

#7a. Initial Matching Check
while freeCount > 0:

    #7b. Make Tentative Matches
    Iterate through Shuffled officer List:
        if officer Status == Free:
            Iterate through AOC Slots:
                if AOC Slot Status == Free:
                    Match officer and AOC
                    Store match in Matching Data Structure
                    Update officer Status = Matched
                    Update AOC Status = Matched
                freeCount -= 1

    #7c. Check to see whether existing matches should be broken

    else:
        If AOC Slot Status != Free:
            Retrieve Assigned officer
            #7d. Does the AOC prefer officer to Assigned officer?
            Iterate through AOC Slot Preferences:
                If officer before Assigned officer in
                Preferences:
                    return True
                else:
                    return False

            If AOC Prefers officer to Assigned officer:
                Break Match Between Assigned officer and
                AOC
                Match officer and AOC
                Store match in Matching Data Structure
                Update officer Status = Matched
                Update Assigned officer Status = Free
                Update AOC Status = Matched

    #7e. Finished tentative matching for officer

#Rematch officers until all officers are finally matched

return (Matching)

```

In summary, the algorithm iterates through officers, who propose assignments to AOCs in their preference list. If the slot is free, the officer is tentatively assigned to it. If

the slot is not free, the algorithm checks to see whether the AOC prefers the officer in question to its currently assigned officers. If so, the currently assigned officer is replaced by the officer in question. The newly released officer is then matched with a new slot. This process continues until all officers are matched.

We use the AOC-MS TMA stable marriage variant to designate officers according to the six sets of targets outlined on Table 34, including targets derived from the AOC-MS TSA using $\psi = 1$ through $\psi = 5$ in addition to the AOC designation panel's designation allocations. For all six matching problems, the AOC-MS TMA stable marriage variant successfully assign all 30 officers to AOCs in an average of 0.004 seconds on a 64-bit Samsung PC with 8GB RAM and an Intel core i7-8550U CPU processor.

For each set of matchings, we ensure that the matches produced by the algorithm are stable. We present the AOC-MS stability checking algorithm to verify the stability of matches. As described in Chapter II, matchings are stable if there exists no officer-AOC pair that prefer each other to the model-assigned match. We check stability for the six sets of matches generated using the AOC-MS TMA stable marriage variant and contrast them with the panel designations derived using the current process. Pseudocode for the AOC-MS Stability-Checking algorithm is located in Algorithm 3.

Algorithm 3. AOC-MS Stability Checking Algorithm

```
CheckStability (officer Preferences, consultant Preferences, Matching)

#1. Check for officer stability first
#1a. Initialize officer Matching Stability variable to True

officer Stable = True

#1b. Check every slot per officer to see if there is a slot they like more than
the current one

Iterate through officers:
    Retrieve officer AOC designation as Current Slot

    Iterate through officer Preference List:
    #1c. If officer prefers another slot better

        If Another Slot is before Current Slot and both slots not in same AOC:
            Iterate through AOC Preference List:

                #1d. Check AOC preference list to see if AOC prefers officer
                to its currently assigned officers

                If officer before Assigned officer in AOC Slot Preferences:
                    #1e. Match is unstable.
                    Update officer Stable = False
                    Store Unstable Pair

                else:
                    pass

#1e. Report officer stability
if officer Stable == True:
    report Matching is officer Stable
else:
    report Matching is Unstable

#2. Check for AOC stability
#2a. Initialize AOC Matching Stability variable to True
AOC Stable = True

#2b. Check every AOC Matching to see if there is an officer they like more than
the currently assigned
Iterate through AOC Slots:
    Retrieve AOC designation as Current officer

    Iterate through AOC Preference List:
    #2c. If AOC prefers another officer better
        If Another officer is before Current officer and both officers not
        in same AOC:
            Iterate through officer Preference List:
```

```

#2d. Check officer preference list to see if officer prefers
#officer to its current AOC

    If AOC before Current AOC in officer Preferences:
        #2e. Match is unstable.
        Update AOC Stable = False
        Store Unstable Pair
    else:
        pass

#2e. Report AOC stability
if AOC Stable == True:
    report Matching is AOC Stable

else:
    report Matching is Unstable

return (Unstable Pairs)

```

The stability-checking algorithm iterates through officers and checks their AOC assignment. It then searches the officer's preference list for better-preferred assignments. When it finds a more preferred AOC, the algorithm checks the AOC's preference list and determines whether the AOC prefers the officer over its currently assigned officers. The algorithm repeats the process, checking AOC preference lists first. Finally, the algorithm returns a list of unstable matches or reports that the matching scheme is stable.

The AOC-MS Stability Checking algorithm successfully checked officer and AOC stability for all six matching sets on an average of 0.002 seconds in matching schemes 100% stable matches and in 0.009 seconds when checking the panel's AOC designations under the current process, where the algorithm identified one officer-unstable pair and six AOC-unstable pairs.

4. AOC-MS TMA Stable Marriage Variant Results

We evaluate the AOC-MS TMA stable marriage variant using the measures of effectiveness outlined in Chapter IV, comparing the talent matching algorithm's results to the panel's AOC designations. Second, we perform sensitivity analysis by comparing AOC designations using the targets generated by the AOC-MS TSA using values of ψ ranging from 1 to 5.

Understanding the conflict between officer preferences and proportion fill introduced earlier in the chapter and knowing that the panel ‘cheats’ by disregarding force structure targets to favor honoring officer preferences, we aim to create a fair comparison between the two processes by holding the AOC designation targets constant. Since we have no way of knowing how the panel would have designated officers if it were held accountable for meeting strict designation targets, we use the panel’s designations as targets for the AOC-MS TMA stable marriage variant. Using these targets causes both processes to perform equally poorly in filling OFM targets while allowing for better comparisons of the remaining measures of effectiveness. Table 39 displays the results of the comparison. Better performance is highlighted in green, worse performance in yellow and bad performance in orange.

Table 39. MOE Comparison Between AOC-MS TMA and Panel Designation

	Panel Targets (AOC-MS)	Panel Targets (Panel Designation)
AOC-MS TMA MOE		
Officer 1st Preference Satisfaction	0.8333	0.7667
Officer Top Two Pref Satisfaction	0.8667	0.8667
Consultant Preference Satisfaction***	0.8333	0.8181
OFM Target Percentage Fill	0.7715	0.7715
Panel Score Distribution Kruskal P-Value	0.094	0.0743
Stable	1	0

*** consultant Preference Satisfaction may be inflated due to imputed preferences

AOC-MS outperformed the current process in three of the six measures of effectiveness and tied in the remaining three. As expected, both processes failed to balance the AOC inventory fairly, and the Kruskal-Wallis tests for both processes yield no evidence that the distribution of panel scores is different across AOCs. Both processes matched officers with their top two choices at the same rate; however, AOC-MS managed to improve the rate of officer first preference satisfaction and the rate of consultant preference satisfaction. It is important to note that the consultant preference satisfaction may be

inflated due to imputed AOC preferences. AOC-MS also produced 100% officer-stable and AOC-stable matchings while the designation panel did not.

We now assess the algorithm's sensitivity to targets set by the AOC-MS TSA algorithm. The results of the assessment are listed on Table 40, where better performance is highlighted in green, worse performance in yellow and bad performance in orange.

Table 40. MOE Comparison for AOC-MS TMA Across PYG Fill Horizons

	5 Year	4 Year	3 Year	2 Year	1 Year
AOC-MS TMA MOE					
Officer 1st Preference Satisfaction	0.467	0.533	0.5	0.5	0.633
Officer Top Two Pref Satisfaction	0.7	0.7	0.767	0.733	0.8
Consultant Preference Satisfaction***	0.833	0.767	0.733	0.8	0.767
OFM Target Percentage Fill	1	1.01	1	1	0.988
Panel Score Distribution Kruskal P-Value	0.073	0.069	0.008	0.016	0.2
Stable	1	1	1	1	1

*** consultant Preference Satisfaction may be inflated due to imputed preferences

As expected, we can observe a negative correlation between the OFM target percentage fill and 1st preference satisfaction. When the AOC-MS TSA enforces AOC designations which fill the gaps in understrength AOCs, the officer preference satisfaction rate plummets. A high rate of officer preference satisfaction using the panel's designations resulted in an abysmal fill rate.

Performance across the remaining measures of effectiveness was comparable to the first example and relatively consistent with the exception of the 1-PYG horizon matching (using only PYG2012 targets to calculate fill rates). The better fill rate performance for the $\psi = 1$ problem relates directly to the fact that the AOC-MS target-setting algorithm set a lower target for 70K (2 as opposed to 14), more closely matching the distribution of officer preferences.

In two occasions we reject the Kruskal-Wallis test’s null hypothesis that the panel scores are equally distributed across AOCs; however, given the small population designating to the majority of AOCs, $n_a \leq 3$, the Kruskal-Wallis test may not be as reliable as it could be in PYGs with larger 70B inventories.

Because of the requirement for equal-sized matching sets and strict targets, the stable marriage of the AOC-MS TMA works best for PYGs where the needs of the MSC are closely aligned with the desires of officers, delivering stable matches and better MOE performance than the current designation process.

D. TALENT MANAGEMENT ALGORITHM INTEGER LINEAR PROGRAM (ILP)

We now turn to the approach for solving the AOC designation problem in PYGs with substantial differences between officer preferences and designation targets. In Chapter II.a we presented a linear programming method for solving assignment problems. This section provides two integer linear program (ILP) formulations and examines their implementations in order to solve the AOC designation problem.

Section 1 explores adaptations required by the ILP approach. The second section examines a comparison formulation which exists only for the sake of comparison with the current process. The third section explores a second formulation, the AOC-MS self-target-setting and designation algorithm (STSDA), the second variant of the AOC-MS TMA, which attempts to balance both preference satisfaction and AOC fill rates, creating its own AOC targets during the designation process.

1. Integer Linear Program Adaptations

Unlike the AOC-MS TMA stable marriage variant, which required ordinal preferences from both officers and AOCs to find stable matches, the ILP requires a numerical representation of officer ordinal preferences and AOC tiered preferences. Instead of requiring two $|O| \times |O|$ square preference tables, ILPs can manage $|O| \times |A|$ matrices to capture the preference relationship between officers and AOCs numerically.

In order to mirror in the generic assignment problem ILP we presented in Equation 1, we construct a matrix $\mathbf{S}$ to capture the satisfaction coefficient $s_{o,a}$, expressing the mutual satisfaction between officer $o \in O$ for AOC $a \in A$. We use the satisfaction coefficients $s_{o,a} \in \mathbf{S}$ as scalars modifying the ILP's binary designation variable $X_{o,a}$, indicating whether officer o is designated to AOC a .

We construct $s_{o,a}$ as a linear combination of two coefficients. The first, $os_{o,a} \in \mathbf{OS}_{|O| \times |A|}$ captures whether officer o 's first or second AOC preferences are satisfied, while $as_{o,a} \in \mathbf{AS}_{|O| \times |A|}$ shows whether officers designated to AOC a were in the AOC's top two tiers. The choice of what level of preference to credit for officers and AOCs is a policy choice. For the purposes of this study, we credit only the top two preferences for officers and the top two preference tiers for AOCs. In both matrices, top preferences are depicted with a 1, second preferences with a 0.5 and all other preferences with a 0. Table 41 displays ten sample records from the officer Satisfaction matrix $\mathbf{OS}$ while Table 42 does the same for the AOC Satisfaction matrix $\mathbf{AS}$.

Table 41. Officer Satisfaction Matrix $\mathbf{OS}$ Sample Records

	70A	70C	70D	70E	70F	70H	70K
A	0.0	0.5	0.0	0	0.0	0.0	1.0
AA	0.5	1.0	0.0	0	0.0	0.0	0.0
AB	0.0	0.0	0.0	1	0.0	0.5	0.0
AC	0.5	0.0	0.0	0	0.0	1.0	0.0
AD	1.0	0.0	0.5	0	0.0	0.0	0.0
B	0.0	0.0	0.0	0	0.0	0.0	0.0
C	0.0	0.5	0.0	0	0.0	0.0	1.0
D	0.5	0.0	0.0	0	0.0	1.0	0.0
E	0.0	0.0	1.0	0	0.0	0.0	0.5
F	0.0	0.5	0.0	1	0.0	0.0	0.0

Table 42. AOC Satisfaction Matrix **AS** Sample Records

	70A	70C	70D	70E	70F	70H	70K
A	0.0	0.5	0.0	0.5	0.5	0.5	1.0
AA	1.0	1.0	0.0	0.5	0.5	0.0	0.5
AB	0.0	0.0	0.5	1.0	0.5	1.0	0.5
AC	0.5	0.0	0.5	0.5	1.0	0.5	0.5
AD	1.0	0.0	1.0	0.5	0.5	0.0	0.5
B	0.0	0.0	0.0	0.0	0.0	0.0	0.0
C	0.5	0.0	0.5	0.5	0.5	0.5	0.5
D	0.0	0.5	0.0	0.5	0.5	0.0	0.5
E	0.0	0.0	0.5	0.5	0.5	0.0	0.5
F	0.0	0.0	0.5	1.0	0.5	0.0	0.5

The satisfaction matrix $\mathbf{S} = \frac{1}{\omega + \alpha} (\omega \mathbf{OS} + \alpha \mathbf{AS})$ provides a normalized numerical representation of the mutual satisfaction between officers and AOCs. The scalars ω and α are user-defined scalars representing how to weigh officer and AOC preferences relative to one another. Table 43 displays ten sample records from the satisfaction matrix $\mathbf{S}$ when $\omega = \alpha = 1$, highlighting mutual preferences in dark green and mutual indifference in light green.

Table 43. Satisfaction Matrix $\mathbf{S}$ Sample Records

	70A	70C	70D	70E	70F	70H	70K
A	0	0.5	0	0.25	0.25	0.25	1
AA	0.75	1	0	0.25	0.25	0	0.25
AB	0	0	0.25	1	0.25	0.75	0.25
AC	0.5	0	0.25	0.25	0.5	0.75	0.25
AD	1	0	0.75	0.25	0.25	0	0.25
B	0	0	0	0	0	0	0
C	0.25	0.25	0.25	0.25	0.25	0.25	0.75
D	0.25	0.25	0	0.25	0.25	0.5	0.25
E	0	0	0.75	0.25	0.25	0	0.5
F	0	0.25	0.25	1	0.25	0	0.25

We observe that officer A listed AOC 70K as their top preference and that AOC 70K ranked officer A in its top tier. Therefore, the satisfaction coefficient for officer $o = A$ and AOC $a = 70K$ $s_{A,70K} = \frac{1}{\omega + \alpha} (\omega * os_{A,70K} + \alpha * as_{A,70K}) = \frac{1}{2} (1 * 1 + 1 * 1) = 1$. We should expect that unless the number of available 70K slots is low relative to the number of one-to-one matches like this one, officer A will be designated as a 70K. We likewise expect officer AA to be designated as a 70C and would be surprised if officer AB was assigned as a 70A. Figure 78 displays a histogram of the coefficients $s_{o,a}$ for the 30 designating 70Bs.

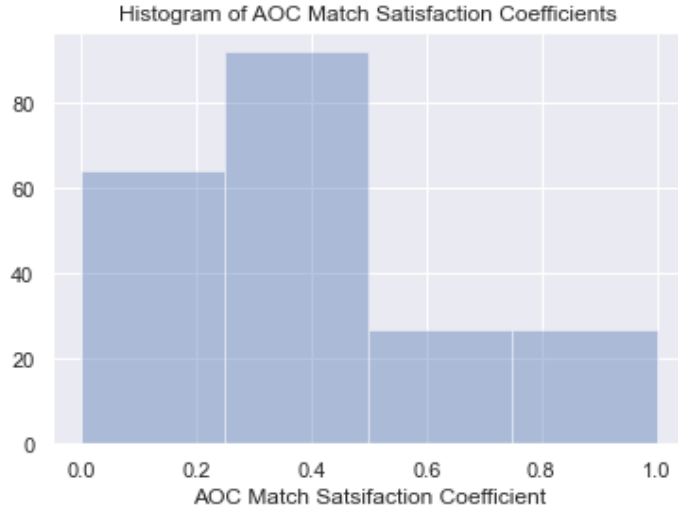


Figure 78. Histogram of AOC Satisfaction Coefficients

The distribution is right-skewed with the majority of satisfaction coefficients clustered near zero. However, when we examine the distribution of the maximum match coefficient for each officer, $\max(s_{o,a \in A})$, we can see that there are 27 coefficients where $\max(s_{o,a \in A}) \geq 0.5$, meaning that relaxing all force structure constraints, we could create 27 matches where both officers and AOCs receive one of their top two preferences. Figure 79 provides a histogram depicting the distribution of maximum satisfaction coefficients.

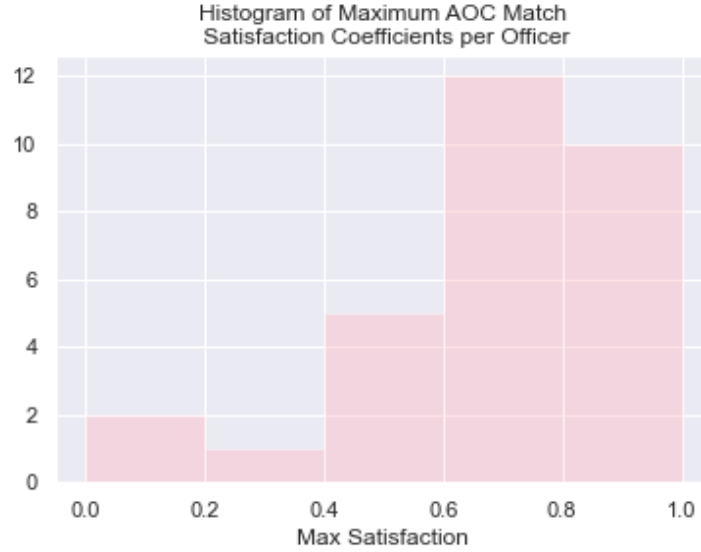


Figure 79. Histogram of Maximum AOC Satisfaction Coefficients per Officer

2. Comparing the AOC-MS ILP with the Current Process

We provide the formulation for the comparison ILP and further expounds on the interplay between mutual preferences and AOC designation targets to create the best outcomes for officers and the MSC.

a. AOC-MS Comparison ILP Formulation

We introduce the formulation for the AOC-MS comparison ILP utilizing NPS format.

Indices

O Index for designating officers $o \in O = \{o_1, o_2 \dots o_{|O|}\}$. For the PYG2012 AOC designation $O = \{o_A, o_B \dots o_{AD}\}$ and $|O| = 30$

A Index for officer AOCs $a \in A$, where $A = \{70A, 70C, 70D, 70E, 70F, 70H, 70K\}$.

Data Parameters

$t_{a \in A}$	AOC-MS TSA minimum targets per AOC $a \forall a \in A$ [Authorizations]
$os_{o \in O, a \in A}$	Satisfaction coefficient expressing preference of officer o for AOC $a \forall a \in A, o \in O$ $os_{o,a} = \begin{cases} 1 & \text{if AOC } a \text{ is officer } o\text{'s first preference} \\ 0.5 & \text{if AOC } a \text{ is officer } o\text{'s second preference} \\ 0 & \text{otherwise} \end{cases}$ [Satisfaction score]
$as_{o \in O, a \in A}$	Satisfaction coefficient expressing preference of AOC a for officer $o \forall a \in A, o \in O$ $as_{o,a} = \begin{cases} 1 & \text{if AOC } a \text{ listed officer } o \text{ in first tier} \\ 0.5 & \text{if AOC } a \text{ listed officer } o \text{ in second tier} \\ 0 & \text{otherwise} \end{cases}$ [Satisfaction score]
$s_{o \in O, a \in A}$	Normalized satisfaction coefficient expressing mutual assignment satisfaction between officer o and AOC $a, s_{o,a} = \frac{1}{\omega + \alpha} (\omega * os_{o,a} + \alpha * as_{o,a}) \forall o \in O, a \in A$ [Satisfaction score]

User-Defined Parameters

ω	Positive scalar for weighing officer preference coefficients $os_{a,o}$ when constructing $s_{o,a}$ [Scalar]
α	Positive scalar for weighing AOC preference coefficients $as_{a,o}$ when constructing $s_{o,a}$ [Scalar]

Decision Variable

$X_{o \in O, a \in A}$	Binary variable reflecting whether officer o is designated to AOC $a . \forall o \in O, a \in A$ [Officers]
------------------------	---

Objective Function

$$\max_{X_{o,a}} \frac{1}{|O|} \sum_{a \in A} \sum_{o \in O} s_{o,a} X_{o,a}$$

Constraints

$$\sum_{o \in O} X_{o,a} = 1 \quad \forall a \in A \quad (1)$$

$$\sum_{a \in A} X_{o,a} = t_a \quad \forall o \in O \quad (2)$$

$$X_{o,a} \in \{0,1\} \quad \forall o \in O, a \in A \quad (3)$$

Model Description

The objective function maximizes the sum of satisfaction coefficients across all AOC designations. Constraint (1) ensures all officers are designated to an AOC. Constraint (2) ensures the AOC target is satisfied by the designation process. Constraint (3) ensures the decision variable is binary.

b. AOC-MS Comparison ILP Implementation

We again implemented the AOC-MS comparison ILP using Pyomo and the CBC solver, iterating through different values of ω and α as well as through the ψ -PYG horizon targets created by the AOC-MS target-setting program. The program features 210 binary variables and 37 equations. CBC solved the problem on an average of 0.01 seconds.

The choice of ω and α seemed to influence the results little, while the choice of target horizon greatly influenced the designation results and their performance against AOC-MS MOE. We set $\omega = 2$ and $\alpha = 1$.

c. AOC-MS Comparison ILP Results

We test the AOC-MS comparison ILP using the same methodology we used for testing the AOC-MS TMA stable matching variant. We compare the ILP results with those

of the current process. Table 44 displays the results of the comparison, highlighting better performance in green, worse performance in yellow and bad performance in orange.

Table 44. MOE Comparison Between AOC-MS Comparison ILP and Panel Designation

	Panel Targets (AOC-MS)	Panel Targets (Panel Designation)
AOC-MS ILP MOE		
Officer 1st Preference Satisfaction	0.8667	0.7667
Officer Top Two Pref Satisfaction	0.9	0.8667
Consultant Preference Satisfaction***	0.8333	0.8181
OFM Target Percentage Fill	0.7715	0.7715
Panel Score Distribution Kruskal P-Value	0.171	0.0743
Stable	0	0

*** consultant Preference Satisfaction may be inflated due to imputed preferences

We observe that the ILP outperformed the current process in three MOEs and tied in the remaining three. While the ILP failed to deliver stable matches, the proportion of officers who received their top preference is higher than the current process and slightly higher than the stable matching variant reached using the same set of targets.

The AOC-MS comparison ILP shows that an ILP can outperform the current designation system and provide comparable results to the stable marriage variant of the TMA given the same targets.

3. The AOC-MS Self-Target Setting and Designation Algorithm (STSDA)

In order to provide the MSC with tools to balance the conflict between preferences and OFM targets, we formulate an ILP that can set its own targets during the designation process, in essence combining the work of the AOC-MS TSA and the AOC-MS comparison ILP into a single program, the AOC-MS self-target setting and designation algorithm (STSDA)

a. AOC-MS STSDA Formulation

We introduce the formulation for the AOC-MS STSDA utilizing NPS format.

Indices

- Y Index for officer promotion year groups $y \in Y = \{y_D, y_2 \dots y_\psi\}$. For the PYG2012 AOC designation problem, $Y = \{2012, 2011, 2010, 2009, 2008\}$.
- A Index for officer AOCs $a \in A$, where $A = \{70A, 70C, 70D, 70E, 70F, 70H, 70K\}$.
- O Index for designating officers $o \in O = \{o_1, o_2 \dots o_{|O|}\}$. For the PYG2012 AOC designation $O = \{o_A, o_B \dots o_{AD}\}$ and $|O| = 30$

Data Parameters

- i_{70B, y_D} Total 70B inventory from the designating PYG available for designation [Officers]
- $i_{a \in A, y \in Y}$ Total on-hand inventory per AOC a for the PYG $y \forall a \in A, y \in Y$ [Officers]
- $o_{a \in A, y_1}$ OFM-produced minimum targets per AOC a for the designating PYG $y_1 \forall a \in A$ [Authorizations]
- $o_{a \in A, y \in Y}$ OFM-produced authorizations per AOC a for the PYG $y \forall a \in A, y \in Y$ [Authorizations]
- $os_{o \in O, a \in A}$ Satisfaction coefficient expressing preference of officer o for AOC $a \forall a \in A, o \in O$
- $$os_{o,a} = \begin{cases} 1 & \text{if AOC } a \text{ is officer } o\text{'s first preference} \\ 0.5 & \text{if AOC } a \text{ is officer } o\text{'s second preference} \\ 0 & \text{otherwise} \end{cases}$$
- [Satisfaction score]
- $as_{o \in O, a \in A}$ Satisfaction coefficient expressing preference of AOC a for officer $o \forall a \in A, o \in O$

$$as_{o,a} = \begin{cases} 1 & \text{if AOC } a \text{ listed officer } o \text{ in first tier} \\ 0.5 & \text{if AOC } a \text{ listed officer } o \text{ in second tier} \\ 0 & \text{otherwise} \end{cases}$$

[Satisfaction score]

$s_{o \in O, a \in A}$ Normalized satisfaction coefficient expressing mutual assignment satisfaction between officer o and AOC a , $s_{o,a} = \frac{1}{\omega + \alpha} (\omega * os_{o,a} + \alpha * as_{o,a}) \forall o \in O, a \in A$ [Satisfaction score]

User-Defined Parameters

ψ	Number of PYGs considered for assessing AOC fill rates. $\psi = Y $. [Number of PYGs]
ρ^-	Scalar between 0 and 1 designating the minimum proportion of OFM targets that must be filled for all AOCs. [Scalar]
ρ^+	Scalar greater than 1 designating the maximum proportion of OFM targets that must be filled for all AOCs. [Scalar]
π^+	Positive scalar weighing penalty variables $P_{a \in A}^+$ [Scalar]
π^-	Positive scalar weighing penalty variables $P_{a \in A}^-$ [Scalar]
ω	Positive scalar for weighing officer preference coefficients $os_{a,o}$ when constructing $s_{o,a}$ [Scalar]
α	Positive scalar for weighing AOC preference coefficients $as_{a,o}$ when constructing $s_{o,a}$ [Scalar]
σ	Positive scalar for weighing mutual preference satisfaction portion of objective function [Scalar]
ϕ	Positive scalar for weighing force structure portion of objective function [Scalar]

Decision Variables

$X_{o \in O, a \in A}$	Binary variable reflecting whether officer o is designated to AOC a . $\forall o \in O, a \in A$ [Officers]
$P_{a \in A}^+$	Nonnegative real-valued variable penalizing the objective function for filling AOCs below 100% of OFM authorizations $\forall a \in A$ [Underfill penalty]
$P_{a \in A}^-$	Nonpositive real-valued variable penalizing the objective function for filling AOCs above 100% of OFM authorizations $\forall a \in A$ [Overfill penalty]

Objective Function

$$\max_{X_{o,a}, P_a^+, P_a^-} \phi \left[\frac{1}{|A|} \sum_{a \in A} \left(\frac{\sum_{o \in O} X_{o,a} + \sum_{y \in Y} i_{a,y}}{\sum_{y \in Y} o_{a,y}} - \pi^+ P_a^+ + \pi^- P_a^- \right) \right] \\ + \sigma \left[\frac{1}{|O|} \sum_{a \in A} \sum_{o \in O} s_{o,a} X_{o,a} \right]$$

Constraints

$$\sum_{a \in A} \sum_{o \in O} X_{o,a} = i_{70B,y_1} \quad (1)$$

$$\sum_{o \in O} X_{o,a} = 1 \quad \forall o \in O \quad (2)$$

$$\sum_{a \in A} X_{o,a} \leq \rho^+ * o_{a,y_1} \quad \forall a \in A \quad (3)$$

$$\sum_{a \in A} X_{o,a} \geq \rho^- * o_{a,y_1} \quad \forall a \in A \quad (4)$$

$$\frac{1}{\sum_{y \in Y} o_{a,y}} \left(\sum_{y \in Y} i_{a,y} + \sum_{o \in O} X_{o,a} \right) + P_a^+ \geq 1 \quad \forall a \in A \quad (5)$$

$$\frac{1}{\sum_{y \in Y} o_{a,y}} \left(\sum_{y \in Y} i_{a,y} + \sum_{o \in O} X_{o,a} \right) + P_a^- \leq 1 \quad \forall a \in A \quad (6)$$

$$X_{o,a} \in \{0,1\} \quad \forall o \in O, a \in A \quad (7)$$

$$P_a^+ \geq 0 \quad \forall a \in A \quad (8)$$

$$P_a^- \leq 0 \quad \forall a \in A \quad (9)$$

Model Description

The objective function has two parts. The first part is similar to the objective function for the AOC-MS TSA. Weighed by the scalar ϕ , it aims to keep all AOCs as filled at near 100% of OFM authorizations across the ψ -year horizon. The penalty variables P_a^+ and P_a^- are weighed using scalars π^+ and π^- , allowing the user to penalize the objective function differently for overfills and underfills.

The second part of the objective function is similar to the objective function for the AOC-MS comparison ILP. It is weighed by the scalar σ and maximizes the mutual satisfaction between officers and AOCs by summing across the binary designation variable $X_{o,a}$, weighed by the satisfaction coefficient $s_{o,a}$. For any particular designation, the MSC can determine how to weigh both parts of the objective function relative to one another in order to balance the need to fill force structure gaps with the desire to honor officer and consultant preferences.

Constraint (1) ensures the number of designations is equal to the number of officers available for designation. Constraint (2) ensures each officer is designated to only one AOC. Constraint (3) creates a designation ceiling per AOC utilizing the ρ^+ scalar to set at a proportion of OFM targets greater than 1. Constraint (4) provides a designation floor per AOC using the scalar ρ^- . Constraint (5) drives AOCs towards being filled above OFM authorizations and in cases where such a proportion fill is infeasible, activates the penalty variable P_a^+ . Constraint (6) drives AOCs to be filled at 100% of OFM authorization or below, activating the penalty variable P_a^- when this is infeasible. Together, constraints (5) and (6) drive the AOC fill percentage towards 100%, encouraging overfills or shortages according to the weight of the scalars π^+ and π^- . Constraint (7) ensures that the designation variable $X_{o,a}$ is integer-valued. Constraint (8) ensures P_a^+ is nonnegative, while constraint (9) makes P_a^- nonpositive.

b. AOC-MS STSDA Implementation

When implemented in Pyomo to solve the FY2020 AOC designation problem, the AOC-MS STSDA has a total of 224 variables, with 210 binary, 7 non-negative real variables, 7 non-positive real variables and 59 constraints. We solved the problem with multiple values for user-set parameters. CBC solved the designation problem in an average of 0.03 seconds.

c. AOC-MS STSDA Results

AOC-MS STSDA has nine user-defined parameters. To understand the relative effects of valuing preference satisfaction against force structure concerns, we choose to vary only the meta-parameters ϕ and σ , which assign weight to satisfying OFM requirements and honoring officer and AOC preferences. We hold constant the remaining user-defined parameters at the values outlined in Table 45.

Table 45. AOC-MS STSDA Fixed Parameter Values

	Parameter	Value
Upper AOC Designation Bound Prop.	$\rho +$	2
Lower AOC Designation Bound Prop.	$\rho -$	0.2
Underfill Penalty Scalar	$\pi +$	1
Overfill Penalty Scalar	$\pi -$	1
Officer Satisfaction Weight	ω	2
AOC Satisfaction Weight	α	1
PYGs Considered for Fill Proportion	ψ	1

We compare the results of three AOC-MS parameter configurations: First, we set ϕ to a high value and σ to a low value, then we set the parameters equal to each other and lastly, we set ϕ to a low value and σ to a high value. Our assessment of AOC-MS STSDA performance against AOC-MS MOE is located on Table 46.

Table 46. MOE Comparison for AOC-MS STSDA Across Three Parameter Configurations

	$\Phi = 5 \sigma = 1$	$\Phi = 1 \sigma = 1$	$\Phi = 1 - \sigma = 5$
AOC-MS MOE			
Officer 1st Preference Satisfaction	0.467	0.667	0.867
Officer Top Two Pref Satisfaction	0.767	0.9	0.9
Consultant Preference Satisfaction***	0.8	0.767	0.733
OFM Target Percentage Fill	1.17	0.84	0.782
Panel Score Distribution Kruskal P-Value	0.038	0.676	0.117
Stable	0	0	0

*** consultant Preference Satisfaction may be inflated due to imputed preferences

When $\phi = 5\sigma$, the STSDA favors force structure concerns over preference satisfaction concerns, yielding designation results comparable to those yielded by the designation models using AOC-MS TSA targets, where the designation filled AOCs to OFM targets at the expense of fulfilling mutual preferences. When $5\phi = \sigma$, we observe results slightly better than those yielded by the current designation process where officers get their top choices but AOCs are filled unevenly, failing to satisfy OFM targets. However, when $\phi = \sigma$, we find a convergence of MOE performance centered roughly between 0.7 and 0.8. In this configuration, AOC-MS STSDA attempts to strike the best possible balance between OFM fills and preference satisfaction. We observe that consultant preferences are satisfied at roughly the same rate across all configurations. However, for reasons we have already discussed, analysis of AOC preference fill rates is unreliable.

The AOC-MS STSDA provides the MSC chief with the flexibility to steer the designation process towards a desired policy outcome or to balance performance across conflicting MOEs in PYGs

E. TALENT MATCHING ALGORITHM VARIANT COMPARISON

Both AOC-MS TMA variants delivered similar results when tested on the same set of targets. AOC-MS matches the strengths of each approach to the specifics of the

designation problem, providing the MSC with more flexibility and control over designation outcomes than the current process provides.

The stable marriage variant of the algorithm has two salient advantages: First, it is a simple algorithm which can be easily understood by officers and MSC leaders without extensive mathematics, statistics or computational training. Secondly, the ability to guarantee stable matches ensures that the entire designating officer cohort is as satisfied as possible within the constraints of targets and preference distributions.

However, the STSDA ILP variant features compelling advantages over the stable marriage variant. Its flexibility accommodates the addition of policy constraints and allows the MSC to steer the outcomes of the designation process to balance force structure and officer satisfaction concerns. Together, both variants provide the MSC with a range of designation options with which to optimize AOC designation to meet the needs of officers and of the MSC.

THIS PAGE INTENTIONALLY LEFT BLANK

VII. RECOMMENDATIONS

In this study we have examined at length the current AOC designation process for the U.S. Army MSC and investigated mathematical approaches for solving the AOC designation problem. We explored three AOC-designation related datasets extensively, identifying MSC-wide trends governing AOC structure, assignments, retention and preference satisfaction. We recognized salient problems with the current AOC designation process including its failure to balance officer preference satisfaction against the force structure needs of the MSC, a lack of meaningful input by AOC consultants in shaping their own AOCs and its subjectivity for human bias and error.

We proposed the area of concentration matching system (AOC-MS) as a technology-enabled solution to the AOC designation problem, managing MSC officers over a 6-year life cycle in three distinct phases: talent development (phase I), talent matching (phase II) and talent utilization (phase III). We presented measures for evaluating the effectiveness of each phase of AOC-MS in managing, allocating and utilizing MSC officer talent.

In support of AOC-MS phase I recruitment and talent development efforts, we introduced four different machine learning classifiers which used officer assignment and educational data for predicting AOC aptitude and for recommending possible matches between officers and AOCs. We selected the multinomial logistic classifier as the AOC-MS phase I model of choice due to its 79% correct classification rate on an independent test set and ease of interpretability. However, we cautioned that predictions from AOC-MS machine learning classifiers skewed heavily towards a single AOC when predicting on officers earlier in their careers. In order to address this shortfall, we recommended retraining the models using self-professed skills derived from officer's official resumes in order to match them with the skills and attributes required by each AOC.

Finally, we detailed algorithmic approaches for rectifying the current process's shortfalls. We formulated two approaches for matching officers and AOCs during phase II of AOC-MS. The first used an adaptation of the Gale-Shapley stable marriage algorithm,

while the second used an integer linear program to decide designations based on mutual preferences between officers and AOCs as well as manpower target fill rates. When evaluated against the AOC-MS measures of effectiveness, both designation algorithms outperformed the current process while performing similarly to each other.

We have provided the MSC with valuable insights regarding the state of its force and the results of its talent and force management efforts to date. We presented a compelling mathematical framework for approaching the AOC designation problem. This framework is applicable to other large-scale military and civilian talent management problems and contributes to the body of scholarly work regarding human resource and talent management applications of operations research techniques. Particularly, we add to the small but growing field of literature regarding the use of machine learning algorithms to predict job aptitude.

A. RECOMMENDATIONS FOR THE MEDICAL SERVICE CORPS

At the time of writing, the Office of the Chief, MSC (OCMSC) has scheduled a decision brief for the corps chief, MG LeMaster, regarding the possibility of implementing AOC-MS. In that brief we will recommend the MSC adopt AOC-MS as its AOC designation method and update MSC and Army policies according to the recommendations outlined in this section and in Chapter III. We also recommend OCSMSC negotiate agreements with APPD and HRC Health Services Division regarding responsibilities for planning and executing AOC-MS.

In order to retrain the AOC-MS phase I machine learning classifiers to deliver better AOC predictions, we recommend HRC provide access to AIM 2.0 officer resume data and that AOC consultants provide write-ups regarding the ideal knowledge, skills and attributes required for their AOCs. We recommend that responsibility for maintaining and running the AOC-MS phase I algorithms be given to APPD operations research analysts with support from OCMSC, AOC consultants and HRC MSC branch. Under this system, APPD would generate AOC match scores for officers in their 5th, 6th and 7th year of AFCS and provide AOCs with lists of officers to consider as recruitment prospects.

We also recommend the development of an online portal for providing officers access to their AOC match scores. This online portal could be developed by HRC utilizing the AIM-2.0 architecture. It would also be used as an AOC designation marketplace where officers and consultants express mutual designation preferences.

At the start of phase II of AOC-MS, APPD analysts should execute the AOC-MS target setting algorithm and provide designation targets to HRC. We also recommend operations research analysts at HRC Health Services Division be charged with collecting the data from the AIM 2.0 portal and running the AOC-MS phase II talent management algorithms. We recommend HRC and APPD analysts conduct post-designation analysis and MOE assessments in order to brief the MSC chief regarding designation courses of action.

Finally, we recommend APPD and HRC track officers through phase III of AOC-MS in order to gauge the long-term effectiveness of AOC-MS and recommend adjustments to the designation process and policies as required.

B. FUTURE WORK

There remains considerable work to be done in order to operationalize AOC-MS. The majority of the required work exists in retraining the AOC-MS phase I machine learning classifiers to process text data in officer's record briefs as well as self-professed skills and attributes in AIM 2.0 resumes in order to render more precise AOC aptitude predictions. This is a complex endeavor which can easily underpin an operations research graduate thesis project.

Computationally, additional work is needed to automate the collection and manipulation of officer records and preferences as well as consultant inputs in order to minimize the manual work performed by HRC personnel during the AOC designation process. This work on its own is not sufficient for a graduate thesis; however, it can make an excellent final project for an intermediate or advanced computational methods course.

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

- Abadi, M, Agarwal, A, Barham, P et al. (2015) *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, version 2.3.0. <http://tensorflow.org/>
- Ahuja, BK, Magnati, TL, Orlin, JB (1993) *Network Flows: Theory, Algorithms, and Applications*. (Upper Saddle River, NJ 07458)
- Akaike, H (2011) Akaike's information criterion. *International Encyclopedia of Statistical Science*. https://link.springer.com/referenceworkentry/10.1007%2F978-3-642-04898-2_110
- Alfaro, C, Perez, S, Valencia, C et al. (2018) The equivalence between two classic algorithms for the assignment problem. ArXiv:1810.03562 [Cs, Math], October. <http://arxiv.org/abs/1810.03562>
- AMEDD Personnel Proponent Directorate (2019) Medical Service Corps Objective Force Model brief. Presentation, Conference Call with Author, Fort Sam Houston, TX
- AMEDD Personnel Proponent Directorate (2020) Medical Service Corps histograms. Presentation, Conference Call with Author, Fort Sam Houston, TX
- Armacost, AP, Lowe, JK (2005) Decision support for the career field selection process at the us air force academy. *European Journal of Operational Research* 160 (3): 839–50. <https://doi.org/10.1016/j.ejor.2002.11.001>
- Bengfort, B, Bilbro, R (2017) *Yellowbrick*, version 0.5. <http://www.scikit-yb.org/>
- Bilogur, A (2018) Missingno: a missing data visualization suite. *Journal of Open Source Software* 3 (22): 547. <https://doi.org/10.21105/joss.00547>
- Brown, D (2019) All about the Gale-Shapley algorithm. duncanjbrown.com. Accessed June 10, 2019 <https://duncanjbrown.com/all-about-the-gale-shapley-algorithm/>
- Brownlee, J (2018a) How to calculate McNemar's test to compare two machine learning classifiers. *Machine Learning Mastery*. Accessed July 24, 2018 <https://machinelearningmastery.com/mcnemars-test-for-machine-learning/>
- Brownlee, J (2018b) A gentle introduction to dropout for regularizing deep neural networks. *Machine Learning Mastery*. Accessed December 2, 2018 <https://machinelearningmastery.com/dropout-for-regularizing-deep-neural-networks/>
- Burkard, R, Dell'Amico, M, Martello, S (2012) *Assignment Problems*. Other Titles in Applied Mathematics. <https://doi.org/10.1137/1.9781611972238>

- Chatterjee, S, Simonoff, JS (2013) Multinomial regression. *Wiley Handbooks in Applied Statistics*. (Hoboken, NJ), 177–90
- Chollet, F (2015) Keras, version 2.3.0. <https://github.com/fchollet/keras>
- Cohen, J (1988) *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. (Hillsdale, N.J)
- Comtois, D (2020) *Summarytools: Tools to Quickly and Neatly Summarize Data*, version 0.9.6. <https://CRAN.R-project.org/package=summarytools>
- Cui, Y (2015) Solving the stable marriage problem in Erlang. theburningmonk.com. Accessed March 13, 2015 <https://theburningmonk.com/2015/03/solving-the-stable-marriage-problem-in-erlang/>
- Dell, RF (2019) NPS format: a supplement to optimization class notes. Class notes, Linear Programming, Naval Postgraduate School Department of Operations Research, Monterey, CA
- Department of the Army Secretariat for Selection Boards (2014) AMEDD officer promotion selection boards officer professional development. Presentation, MSC Junior officer Week, Fort Sam Houston, TX
- Devore, JL (2014) *Probability and Statistics for Engineering and the Sciences*. 9th Ed. (Boston, MA)
- Domingos, P (2012) A few useful things to know about machine learning. *Communications of the ACM* 55 (10): 78. <https://doi.org/10.1145/2347736.2347755>
- Erel, I, Stern, LH, Tan, C et al. (2019) Selecting directors using machine learning. *National Bureau of Economic Research Working Paper Series*, June, 55. <https://www.nber.org/papers/w24435>
- Faraway, JJ (2016) *Extending the Linear Model with R : Generalized Linear, Mixed Effects and Nonparametric Regression Models*. 2nd Ed. Chapman & Hall Texts in Statistical Science Series. (Boca Raton, FL)
- Forrest, J, Ralphs, T, Vigerske, S et al. (2020) *Coin-or Branch and Cut (CBC)*, version 2.10.5. C++. <https://github.com/coin-or/Cbc>
- Fox, J, Hong, J (2009) Effect displays in R for multinomial and proportional-odds logit models: extensions to the effects package. *Journal of Statistical Software* 32 (1): 1–24. <https://www.jstatsoft.org/article/view/v032i01>
- Gale, D., Shapley, L (1962) College admissions and the stability of marriage. *The American Mathematical Monthly* 69 (1): 9. <https://doi.org/10.2307/2312726>

- Garrett, D, Vannucci, J, Silva, R et al. (2005) Genetic algorithms for the sailor assignment problem. Proceedings of the 2005 Conference on Genetic and Evolutionary Computation - GECCO '05. (Washington, DC). <https://doi.org/10.1145/1068009.1068333>
- Gass, SI, Collins, RW, Meinhardt, CW et al. (1988) The army manpower long-range planning system. *Operations Research* 36 (1): 5. <https://doi.org/10.1287/opre.36.1.5>
- Ginn, R (2008) *The History of the U.S. Army Medical Service Corps*. (Washington, DC). <https://history.army.mil/catalog/pubs/30/30-19.html>
- Google (2019) Assignment as a min cost flow problem. Google Developers. Accessed September 29, 2019 https://developers.google.com/optimization/flow/assignment_min_cost_flow
- Greedy Algorithms (2011). Brilliant Math & Science Wiki. Accessed 25 February 2020 <https://brilliant.org/wiki/greedy-algorithm/>
- Gupta, S, Iwama, K, Miyazaki, S (2015) Stable Nash equilibria in the Gale-Shapley matching game. ArXiv:1509.04344 [Cs], September. <http://arxiv.org/abs/1509.04344>
- Hall, P. (1935) On representatives of subsets. *Journal of the London Mathematical Society* 10 (1): 26–30. <https://doi.org/10.1112/jlms/s1-10.37.26>
- Hart, WE, Watson, JP, Woodruff, DL (2011) Pyomo: modeling and solving mathematical programs in python. *Mathematical Programming Computation* 3 (3): 219–260. <http://www.pyomo.org/documentation>
- Headquarters, Department of the Army (2007) *Professional Education and Training Programs of the Army Medical Department*. AR 351-3, Washington, DC. https://armypubs.army.mil/epubs/DR_pubs/DR_a/pdf/web/AR%20351-3.pdf
- Headquarters Department of the Army (2014) *Army Military Human Resource Records Management*. AR 600-8-104, Washington, DC. https://armypubs.army.mil/epubs/DR_pubs/DR_a/pdf/web/r600_8_104.pdf
- Headquarters Department of the Army (2015) *The Army Force Modernization Proponent System*. AR 5-22, Washington, DC. https://armypubs.army.mil/epubs/DR_pubs/DR_a/pdf/web/r5_22.pdf
- Headquarters Department of the Army (2017) *Army Commands, Army Service Component Commands, and Direct Reporting Units*. AR 10-87, Washington, DC. https://armypubs.army.mil/epubs/DR_pubs/DR_a/pdf/web/ARN2541_AR10-87_WEB_Final.pdf

- Headquarters Department of the Army (2018) *Army Medical Department Officer Professional Development And Career Management*. DA Pam 600–4, Washington, DC. https://armypubs.army.mil/epubs/DR_pubs/DR_a/pdf/web/ARN13783_P600_4_admin_FINAL.pdf
- Headquarters Department of the Army (2019) *Smartbook DA PAM 600-4 Medical Service Corps*. DA Pam 600–4 SB, Washington, DC. <https://www.milsuite.mil/book/groups/smartbook-600-4>
- Headquarters Department of the Army (2019a) *Composition, Mission, and Functions of the Army Medical Department*. AR 40-1, Washington, DC. https://armypubs.army.mil/epubs/DR_pubs/DR_a/pdf/web/ARN11363_AR40_1_FINAL.pdf
- Headquarters Department of the Army (2019b) *evaluation reporting System*. AR 623-3, Washington, DC. https://armypubs.army.mil/epubs/DR_pubs/DR_a/pdf/web/ARN14342_AR623-3_FINAL.pdf
- Headquarters Department of the Army (2019c) *Smartbook DA Pam 611–21, Military Occupational Classification and Structure*. DA Pam 611–21 SB, Washington, DC. <https://www.milsuite.mil/book/servlet/JiveServlet/previewBody/424619-102-3-1168444/DA%20Pam%20611-21%20SB%20Dec%202019%20v2.pdf>
- Holz, BW, Wroth, JM (1980) Improving strength forecasts: support for army manpower management. *Interfaces* 10 (6): 37–52. <https://doi.org/10.1287/inte.10.6.37>
- Hothorn, T, Zeileis, A (2015) Partykit: a modular toolkit for recursive partytioning in r. *Journal of Machine Learning Research* 16 (December): 5 <http://www.jmlr.org/papers/volume16/hothorn15a/hothorn15a.pdf>
- Irving, R, Manlove, D, O’Malley, G (2009) Stable marriage with ties and bounded length preference lists. *Journal of Discrete Algorithms, Selected papers from the 2nd Algorithms and Complexity in Durham Workshop ACiD 2006*, 7 (2): 213–19. <https://doi.org/10.1016/j.jda.2008.09.003>
- Iwama, K, Miyazaki, S (2008) A survey of the stable marriage problem and its variants. *International Conference on Informatics Education and Research for Knowledge-Circulating Society (Icks 2008)*, 131–36. <https://doi.org/10.1109/ICKS.2008.7>
- Jain, P (2020) Complete guide of activation functions. Medium. Accessed May 8, 2020 <https://towardsdatascience.com/complete-guide-of-activation-functions-34076e95d044>
- James, G, Witten, D, Hastie, T et al., eds. (2013) *An Introduction to Statistical Learning: With Applications in R*. Springer Texts in Statistics 103. (New York, NY)
- Kelleher, JD (2019) *Deep Learning*. <https://doi.org/10.7551/mitpress/11171.001.0001>

- Kimmons, S (2020) More than half of officers receive top choice in first atap cycle. Army News Service. www.army.mil. Accessed January 28, 2020
https://www.army.mil/article/232041/more_than_half_of_officers_receive_top_choice_in_first_atap_cycle
- Korobov, M, Lopuhin, K (2019) *ELI5* Documentation, version 0.9.0. R Documentation.
<https://readthedocs.org/projects/eli5/downloads/pdf/stable/>
- Krijthe, JH (2015) *Rtsne: T-Distributed Stochastic Neighbor Embedding Using Barnes-Hut Implementation*, version 0.15. <https://github.com/jkrijthe/Rtsne>
- LeNail, A (2019) NN-svg: publication-ready neural network architecture schematics. *Journal of Open Source Software* 4 (33): 747. <https://doi.org/10.21105/joss.00747>
- Liaw, A, Wiener, M (2002) Classification and regression by randomforest. *R News* 2 (3): 18–22. https://www.researchgate.net/profile/Andy_Liaw/publication/228451484_Classification_and_Regression_by_RandomForest/links/53fb24cc0cf20a45497047ab/Classification-and-Regression-by-RandomForest.pdf
- Mangiafico, S (2016) R handbook: tests for paired nominal data. *R Handbook*. Accessed 2016 https://rcompanion.org/handbook/H_05.html
- McGee, J.P. (2019) The Army’s New Approach to People - War on the Rocks Interview by R Evans. <https://warontherocks.com/2019/12/the-armys-new-approach-to-people/>
- McKinney, W, Walt, SVD, Millman, J (2010) Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference*, 51–56.
<https://conference.scipy.org/proceedings/scipy2010/>
- McMurry, P (2019) OFM output information provided via personal communication, October 10, 2019
- McMurry, P, Fulton, L, Brooks, M et al. (2010) Optimizing army medical department officer accessions. *The Journal of Defense Modeling and Simulation* 7 (3): 133–43. <https://doi.org/10.1177/1548512910368527>
- Meyer, D, Dimitriadou, E, Hornik, K et al. (2019) *E1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien*. <https://CRAN.R-project.org/package=e1071>
- Meyer, D, Zeileis, A, Hornik, K (2006) The strucplot framework: visualizing multi-way contingency tables with VCD. *Journal of Statistical Software* 17 (3): 1–48.
<http://www.jstatsoft.org/v17/i03/>

- National Resident Matching Program (2020) How the matching algorithm works. The Match, National Resident Matching Program. Accessed February 4, 2020 <http://www.nrmp.org/matching-algorithm/>
- O'Connor, Brandon (2019) Assigning branches to cadets takes on new system. [www.army.mil](https://www.army.mil/article/227124/assigning_branches_to_cadets_takes_on_new_system). Accessed September 12, 2019 https://www.army.mil/article/227124/assigning_branches_to_cadets_takes_on_new_system
- Office of the Chief, Medical Service Corps (2019a) *MSC Policy Memo 19–002, 70B Area of Concentration (AOC) Designation Policy*. MSC 19–002, Fort Sam Houston, TX. https://cac.amedd.army.mil/msc/leader_dev/70B_AREA_OF_CONCENTRATION_AOC_Designation_Policy19-002.pdf
- Office of the Chief, Medical Service Corps (2019b) *MSC Policy Memo 19–003, 67J Secondary Area of Concentration (AOC) Designation Policy*. MSC 19–003, Fort Sam Houston, TX. https://cac.amedd.army.mil/msc/leader_dev/67J_Secondary_AOC_Designation_Policy19-003.pdf
- Office of the Chief, Medical Service Corps (2020a) Careers in the Medical Service Corps. U.S. Army Medical Service Corps. Accessed January 2020 <https://medicalsevicecorps.amedd.army.mil/careers.html>
- Office of the Chief, Medical Service Corps (2020b) Medical Service Corps. Medical Service Corps. Accessed January 2020 <https://www.milsuite.mil/book/docs/DOC-491431>
- O'Hara, RB, Kotze, DJ (2010) Do not log-transform count data. *Methods in Ecology and Evolution* 1 (2): 118–22. <https://doi.org/10.1111/j.2041-210X.2010.00021.x>
- Oja, H (2008) Applied nonparametric statistical methods, 4th edition by Peter Sprent, Nigel C. Smeeton. *International Statistical Review* 76 (1): 152–152. https://doi.org/10.1111/j.1751-5823.2007.00039_14.x
- Oliphant, T (2006) *NumPy: A Guide to NumPy*. <http://www.numpy.org/>
- Osipenko, Alexander (2019) Gale–Shapley algorithm simply explained. Medium. Accessed May 26, 2019 <https://towardsdatascience.com/gale-shapley-algorithm-simply-explained-caa344e643c2>
- Pardalos, PM, Hearn, DW, Hager, WW, eds. (1997) *Network Optimization*. Lecture Notes in Economics and Mathematical Systems 450. (Berlin ; New York)
- Pedregosa, F., Varoquaux, G., Gramfort, A. et al. (2011) Scikit-Learn: machine learning in Python. *Journal of Machine Learning Research* 12: 2825–2830. <https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>

- Peters, S (2020) ORB sample. milsuite.mil. Accessed February 13, 2020
<https://www.milsuite.mil/book/docs/DOC-733388>
- Python Core Team (2019) *Python: A Dynamic, Open Source Programming Language*.
<https://www.python.org/>
- R Core Team (2019) *R: A Language and Environment for Statistical Computing*, version 3.6.1. (Vienna, Austria). <https://www.R-project.org/>
- Rakotomamonjy, A (2003) Variable selection using SVM-based criteria. *Journal of Machine Learning Research* 3: 14. <https://www.jmlr.org/papers/volume3/rakotomamonjy03a/rakotomamonjy03a.pdf>
- Raschka, S (2018) Model evaluation, model selection, and algorithm selection in machine learning. ArXiv:1811.12808 [Cs, Stat], Accessed July 5, 2020. <http://arxiv.org/abs/1811.12808>
- Ree, MJ, Earles, JA, Teachout, MS (1994) Predicting job performance: not much more than G. *Journal of Applied Psychology* 79 (4): 518–24. <https://doi.org/10.1037/0021-9010.79.4.518>
- Romero-Medina, A (2001) Sex-equal stable matchings. *Theory and Decision* 50 (3): 197. <https://link.springer.com/article/10.1007/s00453-012-9672-0>
- Roth, AE, Peranson, E (1999) The redesign of the matching market for American physicians: some engineering aspects of economic design. *The American Economic Review* 89 (4): 49. <https://www.aeaweb.org/articles?id=10.1257/aer.89.4.748>
- Roth, AE, Rothblum, UG, Vande Vate, JH (1993) Stable matchings, optimal assignments, and linear programming. *Mathematics of Operations Research* 18 (4): 803–28. <https://doi.org/10.1287/moor.18.4.803>
- Rotich, KJ (2015) History, evolution and development of human resource management: a contemporary perspective. *Global Journal of Human Resource Management* 3 (3): 58–73 <http://www.eajournals.org/wp-content/uploads/History-Evolution-and-Development-of-Human-Resource-Management-A-Contemporary-Perspective.pdf>
- Shrimpton, D, Newman, AM (2005) The U.S. Army uses a network optimization model to designate career fields for officers. *Interfaces* 35 (3): 230–37. <https://doi.org/10.1287/inte.1050.0137>
- Sison, CP, Glaz, J (1995) Simultaneous confidence intervals and sample size determination for multinomial proportions. *Journal of the American Statistical Association* 90 (429): 366–69. <https://doi.org/10.2307/2291162>

- Stable marriage problem (2013). GeeksforGeeks. Accessed July 20, 2013
<https://www.geeksforgeeks.org/stable-marriage-problem/>
- Stefl, M (2008) Common competencies for all healthcare managers: the healthcare leadership alliance model: *Journal of Healthcare Management* 53 (6): 360–73.
<https://doi.org/10.1097/00115514-200811000-00004>
- Stoner, N, Campbell, A (2019) The Navy is strangling its most promising talent management initiatives. USNI Blog. Accessed August 20, 2019
<https://blog.usni.org/posts/2019/08/20/the-navy-is-strangling-its-most-promising-talent-management-initiatives>
- U.S. Army Medical Department Office of Medical History (2009) Milestones in the evolution of the medical service corps. The United States Army Medical Service Corps. Accessed May 15, 2009 <https://history.amedd.army.mil/booksdocs/HistoryofUSArmyMSC/appendixa.html>
- U.S. Army Talent Management Task Force (2019) Army talent based branching. U.S. Army Talent Management. Accessed 5 February 2020 <https://talent.army.mil/project/army-talent-based-branching/>
- U.S. Army Talent Management Task Force (2019) *The officer's Guide to the Army Talent Alignment Process (ATAP)*. (Washington, DC). https://ssilrc.army.mil/wp-content/uploads/2019/12/HR_Tech_WOAC_The_officers_Guide_to_ATAP.pdf
- U.S. Army (n.d.). U.S. Army. www.army.mil. Accessed January 21, 2020 <https://www.army.mil/>
- U.S. Navy (2019) MyNavy assignment smart sheet. Navy Personnel Command. Accessed February 5, 2019 [https://www.public.navy.mil/bupers-npc/enlisted/mna/Documents/MyNavy%20Assignment%20Smart%20Sheet%20\(rev1\)_Nov%202019.pdf](https://www.public.navy.mil/bupers-npc/enlisted/mna/Documents/MyNavy%20Assignment%20Smart%20Sheet%20(rev1)_Nov%202019.pdf)
- Vanderplas, J (2016) In-depth: decision trees and random forests. Python Data Science Handbook. (Sebastopol, CA). <https://jakevdp.github.io/PythonDataScienceHandbook/05.08-random-forests.html>
- Vannucci, J, Garrett, D, Dasgupta, D (2006) Comparative analysis of the sailor assignment problem. Proceedings of the 8th Annual Conference on Genetic and Evolutionary Computation - GECCO '06. (Seattle, WA), 1881. <https://doi.org/10.1145/1143997.1144310>
- Venables, WN, Ripley, BD (2002) *Modern Applied Statistics with S*. Fourth. (New York). <http://www.stats.ox.ac.uk/pub/MASS4>
- Waring, E, Quinn, M, McNamara, A et al. (2020) *Skimr: Compact and Flexible Summaries of Data*. <https://CRAN.R-project.org/package=skimr>

- Waskom, M, Botvinnik, O, O’Kane, D et al. (2017) *Mwaskom/Seaborn: V0.8.1* (September 2017), version v0.8.1. <https://doi.org/10.5281/zenodo.883859>
- Whitaker, L (2020) Fundamentals-Part III. Class notes, Machine and Statistical Learning, Naval Postgraduate School Department of Operations Research, Monterey, CA
- Wickham, H (2016) *Ggplot2: Elegant Graphics for Data Analysis*, version 3.3.2. <https://ggplot2.tidyverse.org>
- Wing, MK, Weston, S, Williams, A et al. (2019) *Caret: Classification and Regression Training*, version 6.0-86, <https://CRAN.R-project.org/package=caret>
- Wright, MN, Ziegler, A (2017) Ranger: a fast implementation of random forests for high dimensional data in c++ and r. *Journal of Statistical Software* 77 (1): 1–17. <https://doi.org/10.18637/jss.v077.i01>
- Zelterman, D (1987) Goodness-of-fit tests for large sparse multinomial distributions. *Journal of the American Statistical Association* 82 (398): 624–29. <https://doi.org/10.1080/01621459.1987.10478475>
- Zimmermann, T, Kotschenreuther, L, Schmidt, K (2016) Data-driven HR - resume analysis based on natural language processing and machine learning. ArXiv:1606.05611 [Cs], Accessed February 4, 2020. <http://arxiv.org/abs/1606.05611>

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