Annual Report for Cooperative Agreement W911NF2020027

Developing Predictive Models of U.S. Army Career Pathways through the Integration of Multiple Army Administrative and Survey Data and Other Non-DOD Data Sources

Josh Goldstein (PI), Joel Thurston, Joanna Schroeder, Zhengyuan Zhu, and Stephanie Shipp

Social and Decision Analytics Division
Biocomplexity Institute
University of Virginia

January 6, 2023

DEVELOPING PREDICTIVE MODELS OF U.S. ARMY CAREER PATHWAYS THROUGH THE INTEGRATION OF MULTIPLE ARMY ADMINISTRATIVE AND SURVEY DATA AND OTHER NON-DOD DATA SOURCES

Abstract

Responding to a critical gap in its talent management capabilities, the U.S. Army seeks new means to develop and assess Soldiers' activities, training, promotions, and other experiences across Soldiers' careers. This collaborative research project between the University of Virginia's Social and Decision Analytics Division within the Biocomplexity Institute and the United States Army Research Institute for the Behavioral and Social Sciences extends our previous work examining first term Soldier attrition. Here we develop new models for the career pathways of Soldiers integrating administrative records accessed through the Army Analytics Group's Person-event Data Environment (AAG PDE) along with social, economic, and environmental data external to the Department of Defense. During this collaboration, we are developing profiles of career pathways based on individual Soldier and group characteristics (e.g., socioeconomic status, training, job experiences, occupation type) and reviewing private-sector talent management tools and techniques. We conducted a qualitative analysis of data from subject matter expert interviews and document analysis. We developed discrete-time Markov models of Soldier career pathways applied to Army data in the PDE to identify the factors that drive career pathways.

Contents

1 Introduction

The U.S. Army collects significant administrative and survey data about Soldiers, including training, positions held, deployments, command climate, accessions, pay, waivers, demographics, health, global assessments, and family characteristics. Our research leverages and integrates these data sources to develop predictive models and metrics describing Soldiers' career pathways. The goal is to develop a longitudinal characterization of the nature and acquisition of Soldiers' knowledge, skills, and behaviors and their relation to Areas of Concentration (AOC) for Commissioned Officers and Military Occupation Specialty (MOS) for Enlisted Soldier assignments.

This research leverages the previous collaborative research between the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) and the University of Virginia's Social and Decision Analytics Division within the Biocomplexity Institute examining Soldier attrition (Keller et al., 2020, 2021). A unique feature of this research is the data used. These are data provided in the Army Analytics Group's Person-event Data Environment (AAG PDE). AGG PDE is a business intelligence platform that provides a secure repository for data sources on U.S. military personnel. We have overcome many hurdles in accessing U.S. Department of Defense (DOD) data within the PDE by creating a data science infrastructure to develop analytic models within the enclave. Extending our attrition model research to address Army career progression can lead to new quantitative approaches to personnel planning and talent management utilizing data collected routinely for the administration of the Army. Another unique feature of this research is the inclusion of qualitative methods to inform model development and provide context for the analytical results.

2 Qualitative and Descriptive Analysis

2.1 Document Analysis

We leveraged qualitative analysis, specifically document analysis, in this project to provide rich description that informs and aids in the formation, interpretation, and validation of research questions and statistical models. We developed a standard operating procedure for creating a document analysis research design. The procedure includes background information on qualitative research and document analysis, as well as the steps necessary to complete the analysis. The document analysis of Army administrative materials on Office career progression serves as an example model of document analysis. This procedure has been used to train research staff and students on conducting a document analysis.

The document discovery is a corpus of documents and metadata used to identify documents appropriate for research questions. We expanded the document discovery by tracking additional document metadata (keywords and years). Additionally, we reviewed three additional documents, bring the total number of documents in the corpus to 55 (not including duplicates for years).

We also expanded our use of the Pamphlet on Commissioned Officer Professional Development and Career Management (PAM 600-3). The PAM 600-3, as the primary Army doctrine describing career progression for Officers, is key to contextualizing findings from our quantitative models. We acquired ten versions of the PAM 600-3, dating from 1998 to 2019. We analyzed these documents over time to understand thematic changes in Army Officer promotion. Additionally, we acquired the long-form version of the PAM 600-3, version 2014, which includes the officer development models for each branch. The officer development models are detailed descriptions of lists of broadening and key developmental assignments for each branch and functional areas, organized lists of assignments across prototypical career over years of publication for the document. These descriptions can be used to understand a prototypical or ideal Army Officer career.

2.2 Descriptive Analysis of Officer KSBs

We conducted a literature review on Soldier knowledge, skill, and behavior (KSB) acquisition, with a focus on skills. To date, the research on Soldier skill acquisition explores mainly their relationship to civilian skills, or skill equivalencies (Solutions for Information Design, 2014; Wegner, 2017). Skill equivalencies also includes the idea of transferable skills, which are skills that Soldiers can easily translate to the civilian workforce (Chu, 2017; Curry Hall, 2015; Hayden et al., 2014; Kirchner and Akdere, 2019; Zoli et al., 2015). Tools like crosswalks are used both by the research community and soldiers themselves to understand skill equivalences. Also relevant is the idea of soft skills versus technical skills acquisition (Hayes and Hogan, 2021; Haynie, 2021; Schulker, 2017). The acquisition of soft skills is particularly emphasized and imparted through military culture, while acquisition of technical skills is emphasized in Officer development models (Hayes and Hogan, 2021; Haynie, 2021; Schulker, 2017).

We began a descriptive analysis of Officer skill acquisition. We used the Pamphlet on Commissioned Officer Professional Development and Career Management (PAM 600-3), version 2014 to build a dataset of trainings and certifications for every rank in the officer development models for each Branch [1]. Figure 1 shows the career development model for the Infantry branch from the PAM 600-3. Information from the skills training section, as well as any available information on certifications, was used to build a dataset of skill acquisition by rank for each branch.

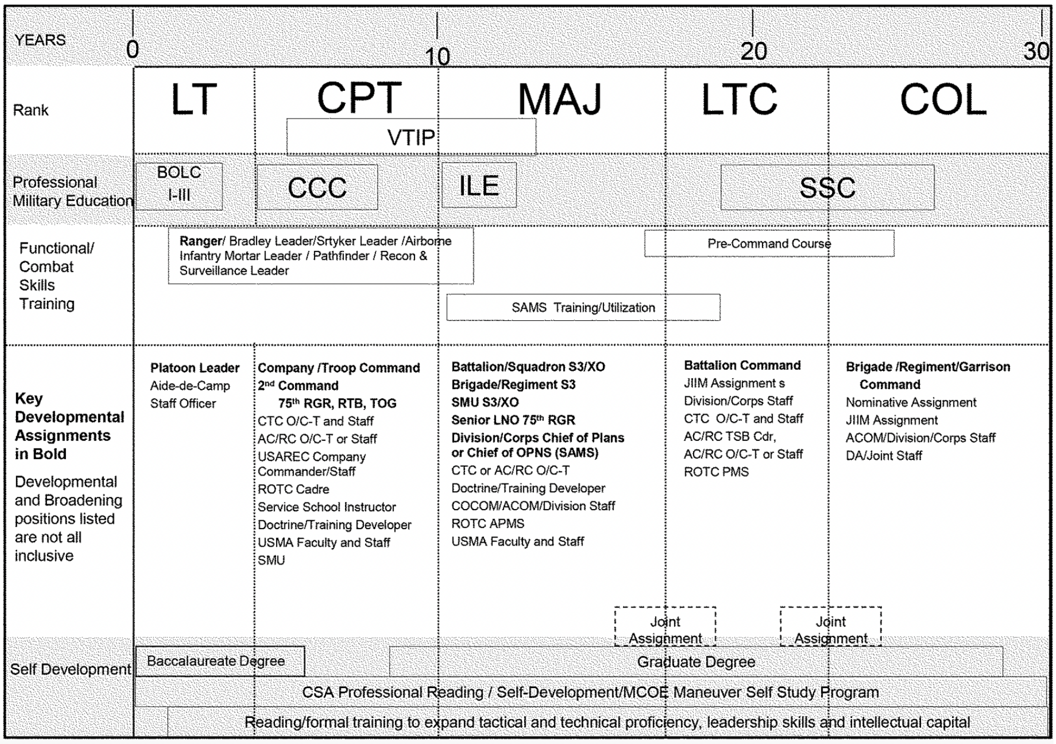

Figure 1: Example active component officer development model for the Infantry branch. Information from the Functional/Combat Skills Training section was extracted to built a dataset of Officer skills acquisition by rank.
From pg. 62 of the Pamphlet on Commissioned Officer Professional Development and Career Management (PAM 600-3), version 2014

[1]For the Cyber branch, information from the PAM 600-3, version 2018 was used

We used network analysis to describe the skill acquisition dataset. Figure 2 shows a birds-eye view of the constructed network. The network includes 166 nodes, 51 of which are unique Branch/Rank combinations and 115 of which are unique Trainings/Certifications. These nodes are connected by 417 unique Training/Certification - Branch/Rank edges. The training with the highest degree centrality, or connection to the most Branch/Rank combinations, was Ranger (18), followed by Air Assault (15) and Airborne (15). Only three branches list certifications in the Officer development model: Cyber, Engineer, and Military Police. Figure 3 shows a close-up of the network, zooming in on the Cyber Branch. We see that Cyber Ranks require both unique trainings and certifications and trainings highly central to the network overall.

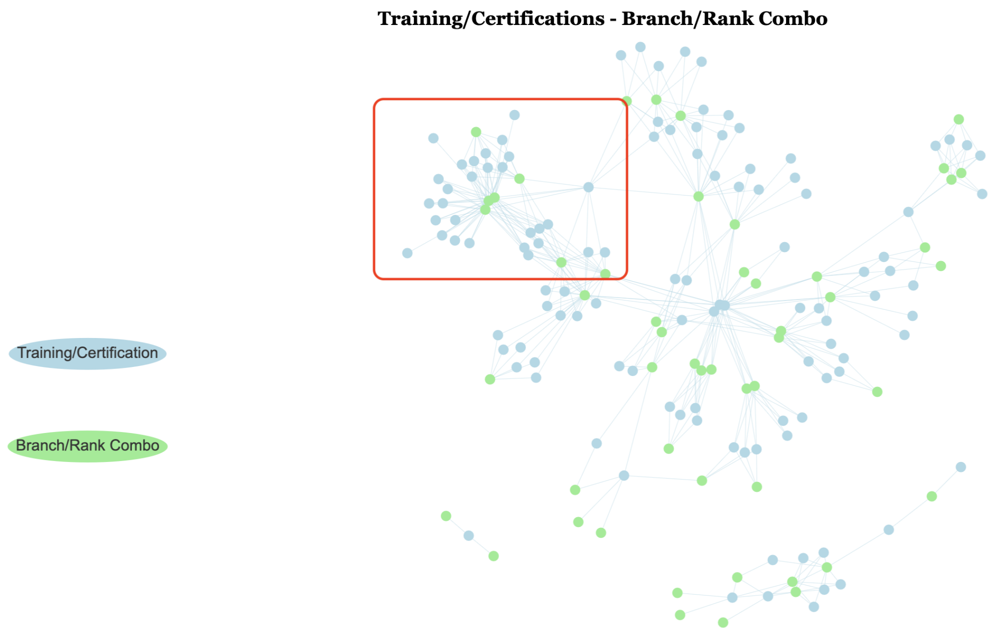

Figure 2: This figure shows the network of connections between Branch/Rank combinations and their associated Trainings/Certifications. We see a mix of trainings that are connected to several positions and trainings that are isolated.
Data Source: SDAD computations and PAM 600-3 active component officer development models.

We plan to use clustering methods to classify a typology of skills using the skill acquisition dataset. This typology, paired with the Digital Training Management System data tables in the PDE, will be used to build skill profiles for individuals and understand the relationship between career development models (ideal skill acquisition) and observed patterns of skill acquisition.

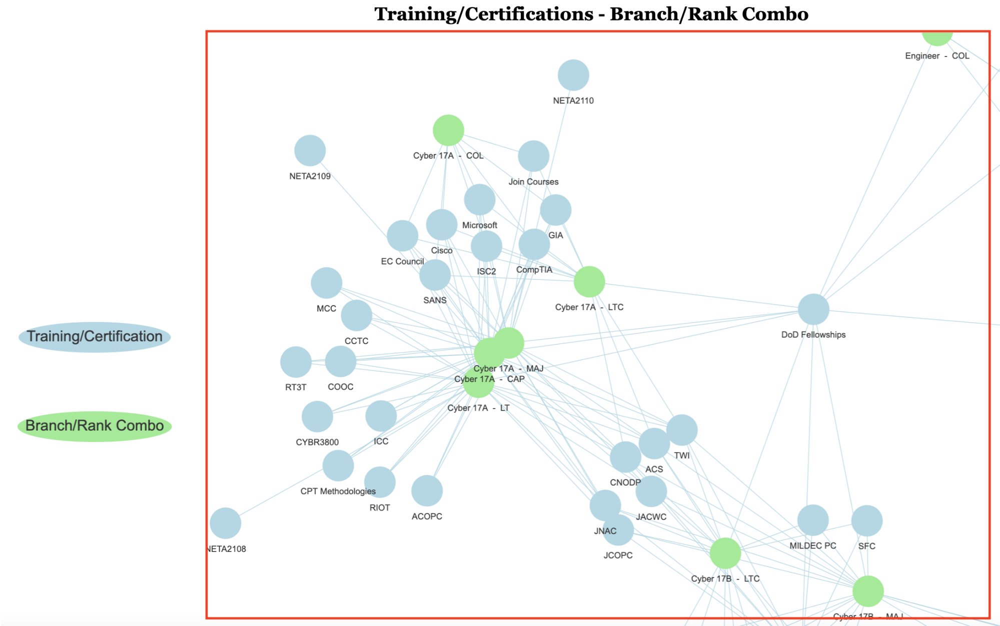

Figure 3: This figure shows a zoomed view of network of connections between Branch/Rank combinations and their associated Trainings/Certifications. Cyber is one of the only branches that lists certifications in their career development model. Cyber positions require several unique trainings, but also trainings with several connections in the wider network.
Data Source: SDAD computations and PAM 600-3 active component officer development models.

3 Model Results for Officer Career Pathways

We developed a discrete-time Markov model for the career pathways of soldiers in the Army. The model consists of a finite number of states and contains a finite number of time epochs at which individuals can transition between states. It considers the number of individuals in a state at a given time to be random variables. The states in our model are represented by all feasible combinations of Soldier Military Occupational Specialty (MOS) and rank. Associated with this model is a transition probability matrix(TPM) that represents the probability that an individual in state j moves to state k (e.g., probability of promotion in rank) with the constraint that the rows must sum to 1.

We built a Bayesian multinomial logistic regression model following Albert and Chib (1993) that processes historical individual-level information to classify an individual to job states. Such probabilities face competing risks from various sources, most commonly spatial and temporal uncertainty, where resorting to a latent modeling framework is beneficial (Silverman et al., 2019). Furthermore, the historical information is sufficiently sparse to necessitate explicit accounting for such sparsity within the modeling framework (Krishnapuram et al., 2005).

3.1 Generating Career Pathways

We describe the following data-generating processes to produce Soldier career pathways. We denote y_{ijk} as the polychotomous variable describing an individual i at job type j transitioning to k, where $i = 1, 2, \ldots, N$ and $j, k = 1, 2, \ldots, J$.

1. Multinomial logistic regression with intercept only:

$$\Pr(y_{ijk} = 1|\boldsymbol{\beta}_{j0}) = \frac{\exp(\beta_{0jk})}{\sum_k \exp(\beta_{0jk})}, \quad \boldsymbol{\beta}_{j0} \sim \mathcal{N}_J(\mu_{\boldsymbol{\beta}}, \Sigma_{\boldsymbol{\beta}}), \quad j = 1, 2, \ldots, J,$$

   where $\boldsymbol{\beta}_{j0}$ is a $J \times 1$ vector of intercepts.

2. Multinomial logistic regression with intercept and covariates:

$$\Pr(y_{ijk} = 1|\mathbf{X}_j, \boldsymbol{\beta}_j) = \frac{\exp(\mathbf{X}_j\boldsymbol{\beta}_{jk})}{\sum_k \exp(\mathbf{X}_j\boldsymbol{\beta}_{jk})}, \qquad \boldsymbol{\beta}_j \sim \mathcal{MN}_{J,p}(\mathbf{M}_{\boldsymbol{\beta}}, \mathbf{U}_{\boldsymbol{\beta}}, \mathbf{V}_{\boldsymbol{\beta}}),$$

where $j = 1, 2, \ldots, J$, $\boldsymbol{\beta}_j$ is a $J \times p$ matrix of coefficients and $\mathcal{MN}_{J,p}$ is the matrix-normal density with location parameter $\mathbf{M}$ (of order $J \times p$) and scale parameters $\mathbf{U}$ and $\mathbf{V}$ of orders $J \times J$ and $p \times p$ respectively. While considering all possible states, we impose priors on tensor normal densities (Xu and Zhang, 2019).

3. Latent multinomial logistic regression with intercept and covariates: The underlying latent reference domain is denoted as $\mathcal{S}$. We extend the modeling framework in (2.) to include,

$$\Pr(y_{ijk} = 1|\mathbf{X}_j, \boldsymbol{\beta}_j, \boldsymbol{\alpha}_{\mathcal{S}}) = \frac{\exp(\mathbf{X}_j\boldsymbol{\beta}_{jk} + \boldsymbol{\alpha}_{ijk})}{\sum_k \exp(\mathbf{X}_j\boldsymbol{\beta}_{jk} + \boldsymbol{\alpha}_{ijk})}, \quad \boldsymbol{\beta}_j \sim \mathcal{MN}_{J,p}(\mathbf{M}_{\boldsymbol{\beta}}, \mathbf{U}_{\boldsymbol{\beta}}, \mathbf{V}_{\boldsymbol{\beta}}),$$

$$\boldsymbol{\alpha}_{ijk} \sim \mathcal{N}(\mathbf{0}, \Sigma_{\boldsymbol{\alpha}}), \qquad\qquad j, k = 1, 2, \ldots, K.$$

Based on the type of latent effect desired, we have the following cases:

(a) *Spatial:* Denoting Δ as the spatial inter-site separation and $||\cdot||$ as the Euclidean norm, the structure is imposed via the following formats.

(i) Parametric Kernels of the form:

$$\Sigma_{\boldsymbol{\alpha}} = \begin{cases} \tilde{K}(||\Delta||) = \sigma^2 \exp(-\phi_s||\Delta||^{\nu}), \nu \in [0, 2], \text{(Power exponential)} \\ \tilde{K}(||\Delta||) = \sigma^2(\phi_s||\Delta||)^{\nu} K_{\nu}(\phi_s||\Delta||), \qquad \text{(Matérn)} \end{cases}$$

where K_ν is the modified Bessel function of order ν, $\phi_s \sim U(0.1, 30)$, and $\sigma^2 \sim IG(2, 1)$, which is the Inverse-Gamma probability distribution. We could have an additional individual-level risk component (commonly termed as the *nugget*) $\epsilon \sim N(0, \tau^2)$, $\tau^2 \sim IG(2, 0.1)$

(ii) $\Sigma_{\boldsymbol{\alpha}}$ could correspond to a sparse Gaussian Markov Random field (GMRf) (Rue and Held, 2005) or transformed Gaussian Markov Random field (tGMRf) (Prates et al., 2015).

While generating synthetic data, we use 10 placeholder job titles, with the 10th job title indicating *attrition* or the event an individual leaves the organization, the norm being that once individuals leave, they cannot rejoin the organization. An upward line indicates promotion, while a downward line indicates demotion. A snapshot of 9 individuals from a simulated Markov Decision process using synthetic probabilities for 1000 individuals is shown in Figure (4). The 9 individuals started from job title 1. This framework capturing transitions between career states gives us a generative model for career pathways of Army officers, and is applied to Army data in the PDE in the following section.

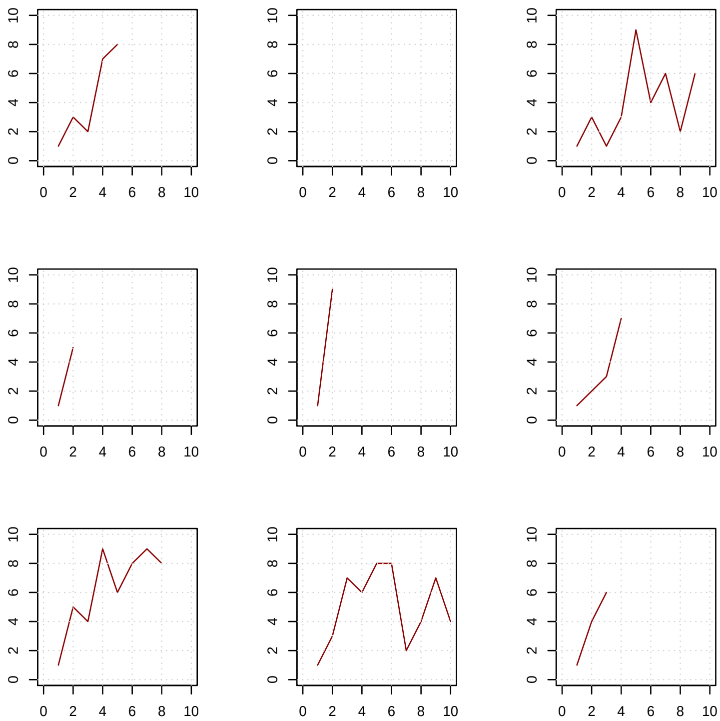

Figure 4: Synthetic career pathways for 9 individuals generated from the intercept-only Markov Decision process model. Each individual begins in the first of 10 placeholder job states; the empty panel is an immediate simulated attrition.

3.2 Modeling Results for Army Officers

For example, we implement the multinomial logistic regression model described above on Army officer career pathways to estimate covariates that drive transitions between officer ranks. Model covariates include demographics (gender, race, age at accession), marital status, education level, selection source, counterproductive behaviors (article 15 proceedings, letters of reprimand, court martial), benefits at accession, and deployment history. The overall year-to-year transition probabilities are given in Figure 5 and selected model effects are shown in Figures 6, 7.

	2nd Lieutenant	1st Lieutenant	Captain	Major	Senior Officer	Discharged
2nd Lieutenant	0.27	0.70	<0.01	<0.01	–	0.01
1st Lieutenant	<0.01	0.54	0.40	<0.01	–	0.06
Captain	<0.01	<0.01	0.81	0.10	<0.01	0.09
Major	–	<0.01	<0.01	0.83	0.11	0.06
Senior Officer	–	–	–	–	0.87	0.13
Discharged	–	–	–	–	–	1.00

Figure 5: Total year-to-year (2001-2020) estimated rank transition probabilities. The large probabilities on the diagonal represent the likelihood of staying in rank from one year to the next. In contrast, the upper-right part of the matrix represents the yearly chance of promotion.

Age at Accession

	2LT	1LT	CPT	MAJ	LTC	Senior	Discharged
2LT	0.20	0.36	0.01	-0.01	-0.01	-0.00	0.03
1LT	-0.01	0.42	0.37	-0.00	-0.00	0.02	0.06
CPT	-0.01	0.01	0.66	0.37	0.00	0.03	0.19
MAJ	-0.01	-0.00	0.02	0.77	-0.00	-0.00	0.15
LTC	-0.00	0.00	0.02	-0.00	0.48	0.02	0.16
Senior	-0.02	-0.00	0.01	-0.00	-0.01	0.13	0.02

Figure 6: Estimated effect of age at accession on year-to-year rank transition probabilities for Army officers. Green highlighted effects are significantly positive, indicating that older age at accession is associated with a larger probability of promotions.

The Bayesian multinomial logistic regression modeling framework allows us to identify the factors that drive career pathways in the PDE. We implement this model by MOS grouping to obtain detailed results for Infantry, Armor, and Cyber MOS branches.

Our next steps for modeling include building out the state space of officer careers. In addition to rank and MOS, we will define career state using Officer position descriptions obtained from Officer Evaluation Reports. Because this data includes Personal Identifiable Information, we have worked with RFL to set up access in a separate cloud computing environment. We are performing text analysis to clean and standardize Officer positions and scrub PII before linking data back to the rest of our Army administrative data on the PDE. Additionally, we will gain an understanding of how career

Article 15							
	2LT	1LT	CPT	MAJ	LTC	Senior	Discharged
2LT	-0.66	-0.94	0.00	-0.01	0.01	0.02	-0.05
1LT	0.01	-1.17	-1.03	-0.01	0.03	0.01	-0.37
CPT	0.01	0.02	-1.49	-0.58	0.03	0.03	-0.81
MAJ	0.00	-0.01	-0.00	-1.43	-0.02	-0.01	-0.21
LTC	0.00	-0.02	-0.02	-0.03	-0.78	0.08	0.32
Senior	0.00	-0.01	0.01	-0.00	0.02	-0.29	-0.04

Figure 7: Estimated effect of an Article 15 proceeding in an officer's career on year-to-year rank transition probabilities. Purple highlighted effects are significantly negative, indicating that Article 15 is associated with a lower probability of promotion, e.g., from Captain to Major.

pathways vary for different groups by using the model to perform subgroup analysis by MOS, accession pathway, and career outcomes (attrition/cause of separation, character of service).

4 Summary

In this report, we summarized our collaborative research with ARI to advance the Army's understanding and capabilities to manage talent. Over the last year, we completed the following activities:

- Expanded document review to provide context and inform statistical model development and create realistic career profiles
- Literature review and a descriptive analysis of Officer KSA acquisition
- Developed discrete-time Markov models of Soldier career pathways applied to Army data in the PDE
- Built a Bayesian multinomial logistic regression model that captures key features driving transitions in an officer's career

References

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American statistical Association*, 88(422):669–679.

Chu, D. (2017). The military pathway to skills. Presentation, Building America's Skilled Technical Workforce.

Curry Hall, K. (2015). Veteran employment: Lessons from the 100,000 jobs mission. Technical report, RAND Corporation.

Hayden, S., Ledwith, K., Shengli, D., and Buzzetta, M. (2014). Assessing the career-development needs of student veterans: A proposal for career intervention. *The Professional Counselor*, 4:129–138.

Hayes, T. and Hogan, R. (2021). The soft skills veterans bring to the workforce. In Ainspan, N. and Saboe, K., editors, *Military Veteran Employment: A Guide for the Data Driven Leader*, pages 117–133. Oxford University Press.

Haynie, M. (2021). Why hire veterans? In Ainspan, N. and Saboe, K., editors, *Military Veteran Employment: A Guide for the Data Driven Leader*, pages 117–133. Oxford University Press.

Keller, S., Goldstein, J., Higdon, D., and Shipp, S. (2020). Using Administrative and External Data Sources to Model First Term Attrition of Army Enlisted Soldiers. Deliverable to US. Army Research Institute. Technical report, Social & Decision Analytics Laboratory, Bicomplexity Institute of Virginia Tech.

Keller, S., Goldstein, J., Higdon, D., and Shipp, S. (2021). Using Administrative and External Data Sources to Model First Term Attrition of Army Enlisted Soldiers. Deliverable to US. Army Research Institute. Technical report, Social & Decision Analytics Laboratory, Bicomplexity Institute of Virginia Tech.

Kirchner, M. and Akdere, M. (2019). An empirical investigation of the acquisition of leadership ksas in the us army: Implications for veterans' career transitions. *Journal of Veterans Studies*, 4:110–127.

Krishnapuram, B., Carin, L., Figueiredo, M. A., and Hartemink, A. J. (2005). Sparse multinomial logistic regression: Fast algorithms and generalization bounds. *IEEE transactions on pattern analysis and machine intelligence*, 27(6):957–968.

Prates, M. O., Dey, D. K., Willig, M. R., and Yan, J. (2015). Transformed gaussian markov random fields and spatial modeling of species abundance. *Spatial Statistics*, 14:382–399.

Rue, H. and Held, L. (2005). *Gaussian Markov random fields: theory and applications*. CRC press.

Schulker, D. (2017). The recent occupation and industry employment patterns of american veterans. *Armed Forces and Society*, 43:695–710.

Silverman, J. D., Roche, K., Holmes, Z. C., David, L. A., and Mukherjee, S. (2019). Bayesian multinomial logistic normal models through marginally latent matrix-t processes. *arXiv preprint arXiv:1903.11695*.

Solutions for Information Design (2014). Pilot study translating military skills to civilian employment. Technical report, Solutions for Information Design, LLC.

Wegner, J. (2017). Helping soldiers leverage army knowledge, skills, and abilities in civilian jobs. Technical report, RAND Corporation.

Xu, C. and Zhang, Z. (2019). Random tensors and their normal distributions.

Zoli, C., Maury, R., and Fay, D. (2015). Missing perspectives: Servicemembers' transition from service to civilian life data-driven research to enact the promise of the post-9/11 gi bill. Technical report, Syracuse University Institute for Veterans and Military Families.