INSTITUTE FOR DEFENSE ANALYSES

Explaining Differences in Predicted O-5 Promotion Outcomes by Race/Ethnicity and Sex Among Navy Officers

Julie Lockwood
Joe King
Rachel Augustine

December 2020
Approved for public release;
distribution is unlimited.
IDA Paper P-20452
Log: H 20-000476

INSTITUTE FOR DEFENSE ANALYSES
4850 Mark Center Drive
Alexandria, Virginia 22311-1882

About This Publication
This work was conducted by the Institute for Defense Analyses under contract HQ0034-14-D-0001, project BE-6-4311, "Assisting implementation of and analysis within DOD's Enterprise Data to Decisions Information Environment," for the Office of the Under Secretary of Defense for Personnel and Readiness (OUSD (P&R)). The views, opinions, and findings should not be construed as representing the official position of either the Department of Defense or the sponsoring organization.

Acknowledgments
The authors sincerely appreciate review and feedback of this publication provided by Matthew Goldberg, Shelley Cazares, and Peter Levine.

For More Information:
Dr. Julie A. Lockwood, Project Leader
jlockwood@ida.org, 703-578-2858
ADM John C. Harvey, Jr., USN (Ret), Director, SFRD
jharvey@ida.org, 703-575-4530

INSTITUTE FOR DEFENSE ANALYSES

IDA Paper P-20452

Explaining Differences in Predicted O-5 Promotion Outcomes by Race/Ethnicity and Sex Among Navy Officers

Julie Lockwood
Joe King
Rachel Augustine

This page is intentionally blank.

Executive Summary

The Department of the Navy (DoN) desires to understand better the drivers of differences in military officer retention and promotion across demographic groups. Among the first major social institutions to begin racial integration, the U.S. Department of Defense and the military services recognize that the process of building a force representative of the U.S. populace remains incomplete. Across the services, incorporation of females and racial/ethnic minorities remains a challenge. There has been more progress among enlisted personnel, with less proportional representation of females and minorities among the officer corps compared to their population share. Among males, retention is generally higher for minority officers than Whites, though they are less likely to promote to the next rank. With the exception of Black females, female officers are generally less likely to remain in service than White males. Minority males and female officers of all racial/ethnic groups are less likely to promote to the next grade than White males. Consistent with these differential retention and promotion rates, females and racial/ethnic minorities rarely attain the most senior military ranks.

To assist efforts to improve representation of females and minorities among Navy senior officers, the DoN asked the Institute for Defense Analyses (IDA) to identify differences in what data features predict retention and O-5 promotion outcomes across race/ethnicity and sex groups of Navy officers. The rank of O-5 (Commander) is pivotal in many ways: it is the first promotion to require a highly selective promotion board, the first rank to be considered a senior officer, and the first rank with the command potential. Attaining the rank of O-5 is the gateway to senior Navy ranks. This research aims to further the Navy’s understanding of factors driving racial/ethnic and sex-based differences in retention and promotion outcomes pertaining to this career milestone.

Data and Methodology

To complete this research on a compressed timeline, we leverage administrative data on military personnel provided by the Defense Manpower Data Center (DMDC) and maintained in IDA’s Personally Identifiable Information Enclave (PII Enclave). Because the DMDC data contain no direct information regarding Navy promotions, we infer promotion based on observing someone attaining the rank of O-5 or not.

We analyze restricted and unrestricted Navy officers of the line that commissioned as O-1s between 2001 and 2018. We include all regular Active Duty officers, as well as Navy Reservists who have been activated (mobilized) for more than 180 days. Our analytic set comprises 45,006 unique officers who collectively served 338,702 person-years between December 2001 and December 2019. We employ a tree-based, discrete-time survival machine learning (ML) model (IDA's Finite-Interval Forecasting Model (FIFE) version 1.3.4). Although FIFE produces retention and promotion forecasts for officers in all years of service and for all future time horizons up to 20 years, we focus on officers in their tenth year of service in the Navy, with retention and promotion forecast four years into the future.

To avoid immediately attributing differences in retention or promotion probability to race, ethnicity, or sex directly, we do not include information on these demographic characteristics when training the models. One implication of this analytic choice is that to the extent that other features in the data strongly correlate with these excluded demographics, systemic differences in retention or promotion associated with these demographic characteristics may be proxied by other features. Although beyond the scope of this paper, follow-on efforts might apply additional analytic tools (currently in a prototype stage at IDA) to identify where relationships discovered by the ML model strongly correlate with various protected class attributes.

To measure the effect of each feature provided to the ML model on an individual's predicted promotion or retention outcome, we use the SHapley Additive exPlanations (SHAP) attribution algorithm. We then calculate and compare feature effects across six demographic groups: White non-Hispanic males, Black non-Hispanic males, Hispanic males (of any race), Other non-Hispanic males (i.e., American Indian and Alaska Native (AIAN), Native Hawaiian and Pacific Islander (NHPI), mixed-race, and other), White non-Hispanic females, and non-White females. This method illuminates differences across demographic groups in which and how much features matter for the outcome under consideration. After identifying which features are most consequential for each demographic group, we then assess the degree to which this importance differs across demographics. Because the majority of officers exit military service prior to fulfilling the minimum eligibility requirements for promotion to O-5, we examine feature effects from two distinct ML models: one predicting retention, and a second predicting promotion to O-5.

Findings

For all demographic groups, we find that many of the most consequential features predicting retention are also the most important predictors of promotion: *officer primary designator*, *officer subspecialty*, *billet designator code*, and *additional officer qualifier designations*. The significance of these career features may intersect with restricted vs. unrestricted line status, and requires further investigation. In addition to career features, family and personal attributes (e.g., *number of dependents*, *citizenship origin*, and *religious* *denomination*) are highly salient for retention outcomes, while the key features predicting O-5 promotions all relate to Navy service regardless of demographic group.

Comparing feature importance of each demographic group to White males, *officer subspecialty*, *citizenship origin*, and to some extent, *qualification designations* matter more for females and racial/ethnic minorities than for White males. Conversely, *officer primary designator* consistently matters more for retention among White males, compared to all other groups. This suggests that retention of females and racial/ethnic minorities is affected by a greater range of factors than those of White males, and that these factors are structurally different in nature. Occupation features such as Navy community and primary designator matter more for retention among White males, while specialized knowledge and training (i.e., officer subspecialty and additional officer qualifier designations) appear to matter more for females and racial/ethnic minorities. We also find variation across demographics in which features are especially predictive of retention relative to White males. For Black male officers, *assigned unit identification code* is particularly significant; for Hispanic males and non-White females, *religious denomination* is especially influential; for Other males, the nature of *citizenship origin* matters most; for White females, officer *subspecialty* assumes foremost importance. Notably, among officers in our analysis set, *number of dependents* is no more consequential for female retention than for White male retention. Prior research finds that females in the military are less likely to be married, less likely to have children, and more likely to be divorced. Our findings suggest that childbearing may not be the driving force behind female attrition, as some have postulated.

Results from our promotion model strongly indicate that *officer subspecialty* is the most consequential predictor of O-5 promotion outcomes for all demographic groups. This feature is especially predictive for females and racial/ethnic minorities; for all but Hispanic males, subspecialty is among the top two most meaningful features. In other words, officer subspecialty matters disproportionately in O-5 promotion outcomes for these groups. Different from our retention model, *officer primary designator* has mixed importance in the promotion model. While this feature is a strong predictor of promotion outcomes for some demographic groups, it matters less for other groups. As a result, officer primary designator is relatively less important than officer subspecialty in predicting O-5 promotion outcomes. These differences in relative importance of occupation and specialty features across demographic groups might be attributable to differential demographic representation in across restricted vs. unrestricted line occupations, and require further investigation.

We then investigate which particular subspecialty codes may account for the outsized role of officer subspecialty. Aggregating subspecialties to the 2-digit level (23 unique categories), we find that approximately half of subspecialty codes are populated almost exclusively by White males, while the other codes are demographically integrated to various degrees. The root cause of this separation is beyond the scope of this project. Some hypotheses include officers' personal preferences and institutional barriers (e.g., historical obstacles to females in various occupations). The confluence of cultural expectations for what constitutes a promotion-enabling career trajectory across restricted vs. unrestricted line status may also explain the importance of integrated subspecialties in these models. Among integrated subspecialties, *regional security studies*, *information sciences*, *oceanography sciences*, and, to some extent, *engineering* disciplines (e.g., aeronautical, mechanical) increase the predicted likelihood of promotion for females and racial/ethnic minorities compared to White males. Other integrated subspecialties (e.g., systems engineering, nuclear engineering) are especially beneficial for promotion outcomes among White males. As a result, the relative importance of officer subspecialty for females and racial/ethnic minorities is driven entirely by a handful of select subspecialties.

Interpretation Caveats

Several important caveats apply to these findings. First and foremost, the relationships we describe are correlational, not causal. Machine learning is a powerful tool that can unearth complex correlations in data, but causality can only be identified when a defensible causal framework exists. Despite the quality and breadth of the administrative data used in this research, this analysis lacks a causal framework and thus cannot measure or substantiate cause:effect relationships. Absent a causal framework, predictive models like those used here should be viewed as forecast and hypothesis generators. Second, feature effects on the predicted outcome depend on the service year and forecast lead length under consideration. Throughout this paper, we focus on Navy officers in their tenth year of service, forecasting retention and promotion four years into the future. Correspondingly, feature explanations pertain to mid-career officers who are weighing the cost and benefits of completing a full Navy career—including personal expectations of potential promotion to O-5. In supplemental analyses, we find the set of most important features differs somewhat at earlier points in the career path, suggesting an evolution in what characteristics influence retention and promotion outcomes over the career.

Next Steps

This project raises many questions for future investigation; we describe some here.

Care should be taken to better understand the role of restricted vs. unrestricted line occupation status and of transfer to non-line occupations in influencing retention and promotion outcomes. This analysis treats those exiting line occupations (e.g., to a staff occupation) as leaving the analysis set. On the one hand, this restriction allows us to specifically examine differences in promotion and retention outcomes among combat command specialties; on the other, by treating changes to non-line occupations as "exits," this may heighten differences (as measured) in these outcomes across demographic groups.

Further research could apply a competing risks approach (like that now available in FIFE 1.3.4) to provide greater insight around these transitions and their retention impact. Additional effort is also needed to understand the relatively greater importance of certain subspecialty and additional officer qualifier designations in minority retention. What is unique about the particular subspecialties driving this effect?

The importance of *Assigned Unit Identification Code* for the retention of Black males is particularly interesting, and suggests a need to understand whether and how time-invariant, unit-specific characteristics affect retention and promotion likelihoods for this group. These unit effects may be positive or negative. Further research is needed to identify and understand the nature and root causes of trends observed here.

Much attention has been given to the home life of female service members under the presumption that family attributes such as marriage and childbearing greatly influence their career choices and outcomes. Our finding that number of dependents does not impact female retention more than White male retention suggests that childbearing may not be the driving force behind female attrition, as some have postulated. More research is clearly needed to better understand the impacts of family life on military service choices and outcomes for both male and female service members, especially as service and societal norms around parenting roles, career aspirations, and occupational and workforce participation continue to evolve. It is worth reemphasizing that these findings pertain to officers in our analysis set at their tenth year of service and may not apply to earlier service years. Follow-on research should explore the importance of childbearing on retention outcomes across the service career.

Further work also is needed to understand what drives the importance of other personal and lifestyle characteristics such as citizenship origin and religious denomination. This may illuminate cultural trends that could aid efforts to improve retention among service members who do not affiliate with these cultural subgroups or personal origins, or have implications for how representative military service members are of American society along these dimensions.

The analyses and findings presented here only consider a subset of Navy personnel, and only part the career path. How might the experience of enlisted members differ? What features matter most for officer retention and promotion at higher grades? How has what matters for retention and promotion evolved over successive generations of Navy service members?

Finally, all findings presented here are correlational, not causal. Moving beyond hypothesis generation and identifying the cause:effect relationships undergirding our results, careful research must identify and exploit experimental or quasi-experimental variation. Many trends identified here are worthy of this level of exploration.

This page is intentionally blank.

Contents

This page is intentionally blank.

1. Introduction

The Department of the Navy (DoN) desires to better understand the drivers of differences in military officer retention and promotion across demographic groups. Among the first major social institutions to begin racial integration, the U.S. Department of Defense (DoD) and the military services recognize that the process of building a force representative of the U.S. populace remains incomplete. Across the services, incorporation of females and minorities remains a challenge.[1,2] This transition has been more complete among enlisted personnel, with less proportional representation of females and racial/ethnic minorities among the officer corps relative to their population share. Among male officers, retention is generally higher for racial/ethnic minorities than White males, though they are less likely to promote to the next rank. With the notable exception of Black females, at the time of this writing, female officers are generally less likely to remain in service than White males. Minority males and female officers of all racial/ethnic groups are less likely to promote to the next grade than White males. Consistent with these differential retention and promotion rates, females and racial minorities rarely attain the most senior military ranks.[3]

Diversity has long concerned the DoD.[4] In June 2020, the Secretary of Defense announced new initiatives aimed at addressing diversity and inclusion in the military. To assist efforts to improve representation of females and racial/ethnic minorities among Navy senior officers, the DoN asked the Institute for Defense Analyses (IDA) to identify differences in what data features predict retention and O-5 promotion outcomes across race/ethnicity and sex groups of Navy officers.[5] The rank of O-5 (Commander) is pivotal in many ways: it is the first promotion to require a highly selective promotion board, the first rank to be considered a senior officer, and the first

1 We define racial and ethnic minorities following official Office of Management and Budget (1997) categories. 2 Karin de Angelis and David R. Segal, “Minorities in the Military,” T*he Oxford Handbook of Military Psychology*, ed. Janice H. Laurence and Michael D. Matthews (Oxford University Press, 2012), 331–32; James Burk and Evelyn Espinoza. “Race Relations Within the US Military,” *Annual Review of Sociology* 38*,* no. 1 (2012): 401–22. https://doi.org/10.1146/annurev-soc-071811-145501.

3 U.S. Department of Homeland Security, “From Representation to Inclusion: Diversity Leadership for the 21st-Century Military (Final Report),” Military Leadership Diversity Commission, 2011. https://www.hsdl.org/?view&did=11390

4 Burk, James, and Evelyn Espinoza, “Race Relations Within the US Military,” *Annual Review of Sociology*, 38, no. 1 (2012): 401–22. https://doi.org/10.1146/annurev-soc-071811-145501

5 This project is part of a pilot program sponsored by the Under Secretary of Defense for Personnel and Readiness (OUSD(P&R)) to enable military Service use and development of the Finite Interval Forecasting Engine (FIFE), a discrete-time survival machine learning (ML) toolkit built by the Institute for Defense Analyses (IDA) and sponsored by OUSD(P&R).

rank with the command potential. Attaining the rank of O-5 is the gateway to senior Navy ranks. This paper aims to further the Navy’s understanding of factors driving racial/ethnic and sex-based differences in retention and promotion outcomes pertaining to this career milestone.

2. Data

A. Data Source

Our research uses administrative military personnel data from the Defense Manpower Data Center (DMDC), which IDA receives on a regular basis as part of an institutional data sharing agreement and maintains in IDA's Personally Identifiable Information Enclave. IDA's DMDC holdings span January 2000 to (presently) June 2020, most of which are measured monthly. We build our analytic set of active duty Navy officers using the DMDC Master, Pay, Family,[6] and Deployment files.

B. Measuring Race/Ethnicity and Sex[7]

We adopt an intersectional approach to race/ethnicity and sex. Intersectionality in the context of race/ethnicity and sex refers to the idea that neither gender nor race alone shape the lived experiences of individuals. Rather, the combination or intersection of race/ethnicity and sex correlates with distinct forms of (dis)advantage, which are fluid, historical, and situationally dependent.[8,9] We create a combined "race-sex" variable, which we make invariant within persons,[10] to account for this multiplicative effect. DMDC data contains two mutually exclusive categories of sex: male and female. We combine race and ethnicity information in the DMDC Master file to create four mutually exclusive categories: non-Hispanic White, non-Hispanic Black, Hispanic (any race), and non-Hispanic Other.[11] For simplicity, from this point forward in this paper, we refer to this dimension as "race," and use "White" to indicate non-Hispanic Whites. Although these racial groups are adequately populated for males, due to the small number of non-

6 We augment the historic DMDC Family files with information from the Defense Enrollment Eligibility Reporting System (DEERS).

7 Throughout this paper, we refer to biological sex, rather than the social construct of gender, though we recognize gendered social beliefs are an important driver of sex-based differences in society.

8 Irene Browne and Joya Misra, "The Intersection of Gender and Race in the Labor Market." *Annual Review of Sociology* 29, no. 1 (2003): 487–513.

9 Kimberle Crenshaw, "Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics." University of Chicago Legal Forum 1989, no. 1 (1989): 139–67. https://chicagounbound.uchicago.edu/uclf/vol1989/iss1/8.

10 For this research, we assign individuals their modal sex and race/ethnicity categories.

11 Non-Hispanic Other includes Asian, Native Hawaiian or Other Pacific Islander (NHPI), American Indian or Alaska Native (AIAN), mixed race, and some other race.

White females in our analytic set, we combine all Black, Hispanic, and Other females into a single non-White female group.

C. Scoping

We restrict the analysis to active duty Navy officers who commissioned as O-1s in 2001-2018 and who entered as restricted or unrestricted officers of the line. While focusing on newly commissioned line officers reduces the universe of Navy officers considered, these restrictions help mitigate confounding influences on promotion outcomes, and support estimates targeted to this subpopulation.

Because the DMDC data contain no direct information regarding Navy promotions, we infer promotion based on observing someone to have attained the rank of O-5 or not.[12] We thus model promotion outcomes indirectly and potentially with some delay. Our retention model measures whether or not someone is present in our administrative data year-to-year. Our population restrictions help mitigate the lack of direct promotion and evaluation metrics in the DMDC data.

The analysis population includes regular Active Duty (AD) line officers, as well as Reservists who have been activated (mobilized) for more than 180 days. Annual counts of the universe of these officers contained in IDA's DMDC data is displayed in Figure 1, alongside publicly available DMDC counts for this same universe. We measure officers' service duration based on the length of time we observe each person in the data. Because we are unable to identify service duration among left-truncated[13] individuals, we exclude those who were present in the first month of our DMDC data (January 2000).[14] We additionally exclude officers who commissioned in calendar year 2000 from all analyses as DMDC began transferring data to a new database during this time, resulting in an impartial year of data.[15] As the duration to promote to O-5 depends critically on officers' commissioning rank, we restrict the analysis to officers who commissioned as O-1s.[16] We also restrict to combat command specialties – restricted or unrestricted line officers – as the promotion and retention opportunities available to these occupations differ markedly from that of

[12] This project was commissioned with an ambitious 4-month timeline that restricted the data available.

[13] Left-truncation is a statistical term referring to when an observation (e.g., a person) is at risk of an event (e.g., death, leaving the military, promoting to O-5) before the start of a study. In our case, these are officers whose Navy careers begin sometime before the beginning of our data in January 2000 and remain Navy officers after this time. Importantly, we completely miss individuals who exited service prior to January 2000 (i.e., left-*censored* individuals). Individuals who we *do* observe in our data that commissioned prior to January 2000 are therefore a non-representative sample, whose inclusion would bias our retention and promotion predictions.

[14] An alternative service duration measure would be to use the start date of individuals' credible active federal military service. However, we would not observe potential breaks in individuals' active duty military careers, nor would we observe cohort attrition prior to the start of our data.

[15] One possible consequence of using this impartial year of data would be that retention and promotion estimates for the individuals in this cohort may be inflated, as early exiters might not have been migrated to the new database prior to their exit.

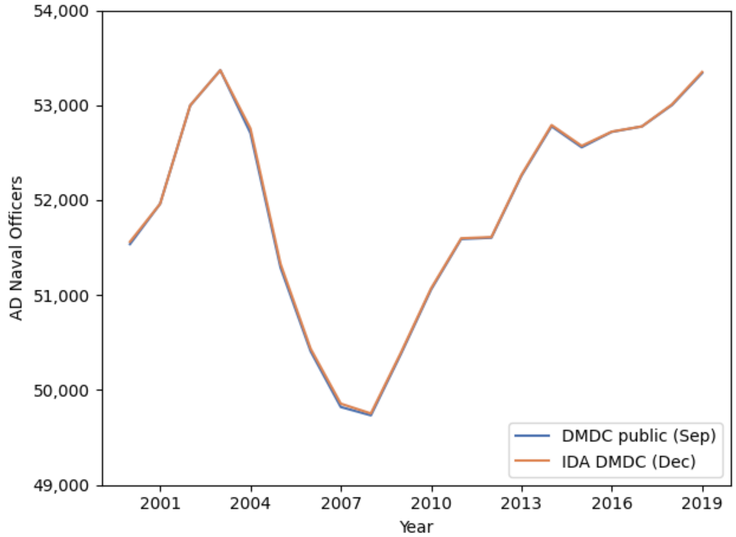
[16] Restricting to O-1s further mitigates the risk of mistakenly attributing left-censored individuals as newly commissioned Navy officers that might otherwise re-enter at a higher rank in later years.

staff and limited duty officer occupations. We keep only officers who begin their career as an officer of the line (according to primary designator), dropping observations after any switch to a non-line designator. Finally, because survival and promotion probabilities for individuals observed only once across the period are undefined in our ML model, prior to training our model we drop these individuals, keeping only officers observed for two or more periods. Notably, this excludes all officers who commissioned in the latest year of our data (2019). To facilitate modeling using the entire set of individuals in our analytic set, we reduce our monthly data to the annual level.

Figure 2 displays the cumulative effect of each of these analytic restrictions on the number of officers remaining in the analysis. The bottom-most line represents the fully restricted analytic set used to train our ML models. Our resulting analytic set comprises 45,006 officers and 338,702 person-years spanning December 2001 to December 2019. Table 1 reconciles how these various scoping restrictions affect the number of unique persons remaining by race-sex group (see also Figure A-1 in Appendix A for each demographic's share across analytic restrictions). More detailed analyses of the resulting composition of individuals by race-sex are displayed in Table 2. The analytic set is overwhelmingly male (83.2%), of whom the vast majority are White (82.7%).

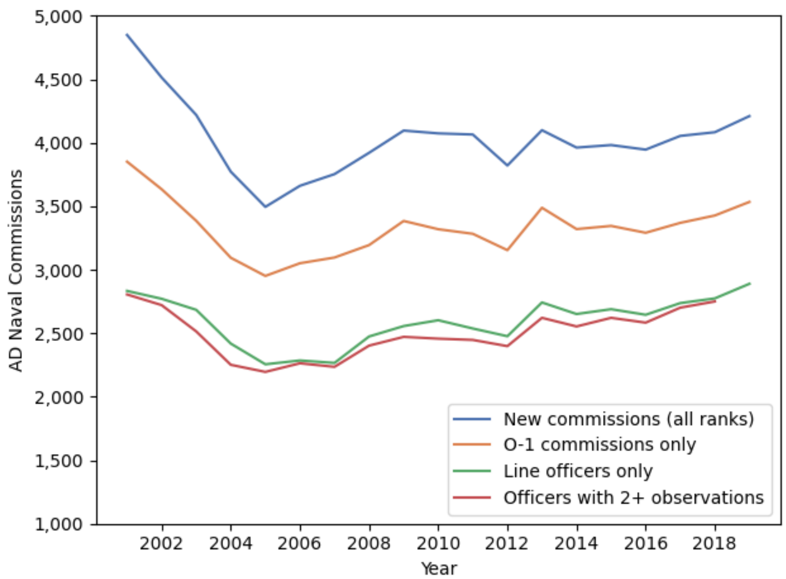
Table 3 presents O-5 promotion outcomes among the 2001-2004 commissioning cohorts,[17] the only cohorts in our data for whom O-5 promotions are well observed (see Appendix A, Figure A-3; Appendix A, Figure A-4 shows the distribution of promotion duration for these cohorts). For males, about 22% of officers that commissioned in these years were observed to have promoted to O-5, compared to about 14% of females. Conditional on serving at least 14 years in the Navy (i.e., the minimum observed service length to promote to O-5 in our analytic set), 70% of males and 83% of females were observed to promote to O-5. Lower female retention is therefore the primary driver of female underrepresentation among O-5s at the population level, because conditional on serving 14 or more years, females are actually more likely to promote to O-5. Among males and females, Whites were the most likely to promote to O-5, relative to other racial groups.

[17] We define officers' commissioning cohorts as the first year they appear in the DMDC Master file. This method aligns with other publicly available counts of new Navy commissions; see Appendix A, Figure A-2.



Note: Counts derived from IDA's Active Duty DMDC Master (December) file, 2000-2019. Publicly available DMDC counts from https://www.dmdc.osd.mil/appj/dwp/dwp_reports.jsp

Figure 1. Universe of AD Navy officers: IDA DMDC Data and Public DMDC Counts



Note: IDA DMDC annual data, 2001-2019. Sample restrictions are cumulative, e.g., "Line officers only" includes only O-1 line officers. The bottom line is the number in our analytic set.

Figure 2. Number of Newly-Commissioned AD Navy officers Net of Various Restrictions

Table 1. Population Reconciliation

		Unique Officers in Period	Commissioned During Period (all ranks)	Commissioned as O-1	Enter As Line Officers	Observed 2+ Periods	Analysis Set
Male	**All**	103,580	59,863 (57.8%)	50,609 (48.9%)	40,907 (39.5%)	37,436 (36.1%)	37,436 (36.1%)
	White	85,020	47,425 (55.8%)	40,135 (47.2%)	33,643 (39.6%)	30,954 (36.4%)	30,954 (36.4%)
	Black	6,613	3,978 (60.2%)	3,387 (51.2%)	2,016 (30.5%)	1,822 (27.6%)	1,822 (27.6%)
	Hispanic	6,745	4,628 (68.6%)	4,006 (59.4%)	3,009 (44.6%)	2,663 (39.5%)	2,663 (39.5%)
	Other	5,202	3,832 (73.7%)	3,081 (59.2%)	2,239 (43.0%)	1,997 (38.4%)	1,997 (38.4%)
Female	**All**	24,473	16,732 (68.4%)	12,580 (51.4%)	8,396 (34.3%)	7,570 (30.9%)	7,570 (30.9%)
	White	18,358	12,208 (66.5%)	9,186 (50.0%)	6,399 (34.9%)	5,787 (31.5%)	5,787 (31.5%)
	Non-White	6,115	4,524 (74.0%)	3,394 (55.5%)	1,997 (32.7%)	1,783 (29.2%)	1,783 (29.2%)

Note: DMDC annual (December) data, 2001-2019. Percentages in parentheses denote the share of individuals remaining from the universe of Navy AD and activated Reserve officers in the first column, by demographic. The far right column represents the number of individuals in the analytic set.

Table 2. Demographics of Analytic Set

	Total	Male				Total Males	Female		Total Females
		White	**Black**	**Hispanic**	**Other**		**White**	**Non-White**	
Officers	45,006	30,954	1,822	2,663	1,997	37,436	5,787	1,783	7,570
Share of Total Officers	100.0%	68.8%	4.0%	5.9%	4.4%	83.2%	12.9%	4.0%	16.8%
Share of Sex		82.7%	4.9%	7.1%	5.3%	100.0%	76.4%	23.6%	100.0%

Note: DMDC annual (December) data, 2001-2019.

Table 3. Promotion among 2001-2004 Commissioning Cohorts

		Officers in 2001-2004 Cohorts	Officers who Serve 14+ Years *(% of cohort)*	Officers who Promote to O-5 *(% of cohort)* *(% of those serving 14+ yrs.)*
Male	**All**	9,155	2,904 (31.7%)	2,023 (22.1%) (69.7%)
	White	7,691	2,437 (31.7%)	1,729 (22.5%) (70.9%)
	Black	519	160 (30.8%)	101 (19.5%) (63.1%)
	Hispanic	574	192 (33.4%)	120 (20.9%) (62.5%)
	Other	371	115 (31.0%)	73 (19.7%) (63.5%)
Female	**All**	1,555	253 (16.3%)	212 (13.6%) (83.8%)
	White	1,242	187 (15.1%)	163 (13.1%) (87.2%)
	Non-White	313	66 (21.1%)	49 (15.7%) (74.2%)

Note: DMDC annual (December) data, 2001-2019. Restricted to officers in analytic set that commissioned in 2001-2004.

D. Data Construction

In addition to the analytic scoping above, we performed a series of data cleaning and data engineering operations prior to training our ML models. This included dropping features that were more than 99.9% missing or invariant, ensuring correct data typing, forward/backward filling missing values in a given feature per scrambled Social Security Number,[18] min-max normalize numeric features, and engineering a host of features from the DMDC data. Appendix Table C1 displays the full list of all 306 features used to train our machine learning model.[19] Finally, prior to training our model, we use the PanelDataProcessor method in FIFE 1.3.4[20] to create our duration measure and identify right-censored[21] individuals.

[18] Forward/backward filling is a commonly used imputation technique used in longitudinal data. Missing values are imputed from the nearest non-missing value of the same person. We backward-fill a selection of features presumed to be time-invariant (e.g., date of birth, state, and country of birth). All other features are forward-filled.

[19] At training time, we exclude race, ethnic affiliation, and sex from the model. If our goal was pure prediction it would make sense to include these features; however, because this is not our goal and because including these features would likely have the unwanted effect of absorbing covariance from other features, we exclude them.

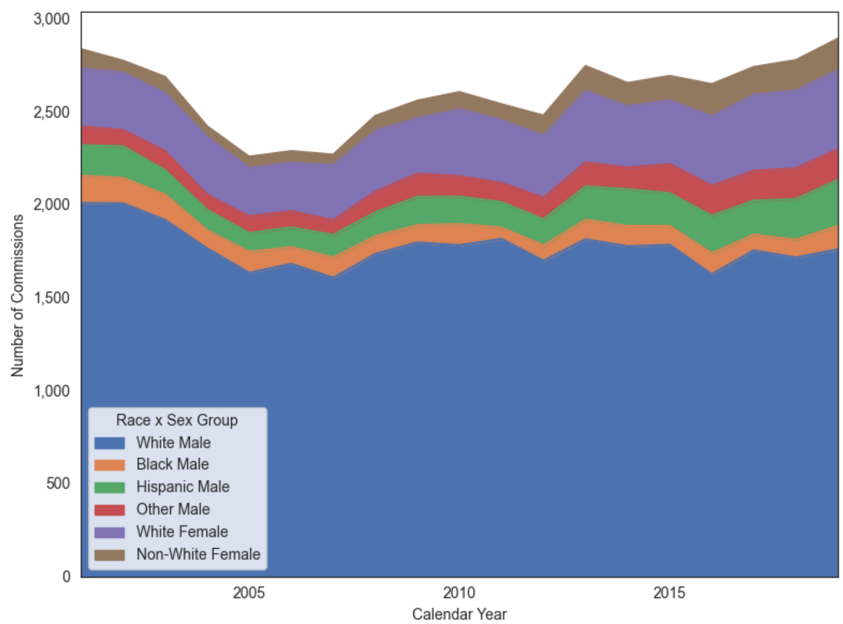
[20] Details are described at https://fife.readthedocs.io/en/latest.

[21] Right-censoring is a statistical term meaning the event under consideration (i.e., exit from service or promotion to O-5) is not observed during the study for a given individual. Importantly, our survival models assume the rates of exits and O-5 promotions are similar for censored individuals as for those we observe to have experienced these outcomes.

E. Descriptive Statistics

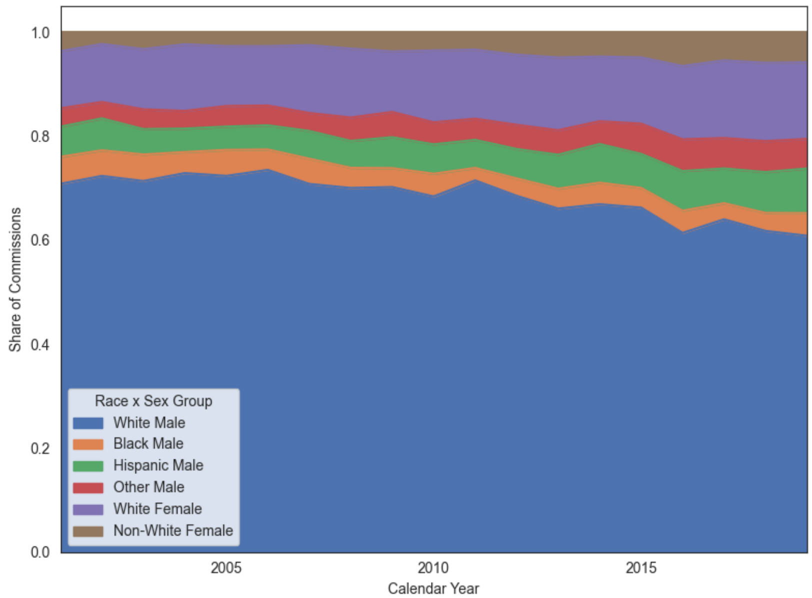
1. Annual commissions by demographic group

Figure 3 displays commissions by demographic group per calendar year among Navy officers in our analytic set. Over the period of analysis, the number of newly commissioned officers in our analytic set fluctuates, decreasing from about 2,800 to 2,200, before rebounding in later years. Figure 4 normalizes the raw counts in Figure 3, illustrating the change in the relative share of each demographic group. Across time, the proportion of White males decline, accelerating in 2010 amidst a rising share of female officers.



Note: Active Duty Navy officers commissioned as O-1s in 2001 or later who began their career as an officer of the line. DMDC annual data, 2001-2019.

Figure 3. O-1 Commissions by Demographic per Calendar Year



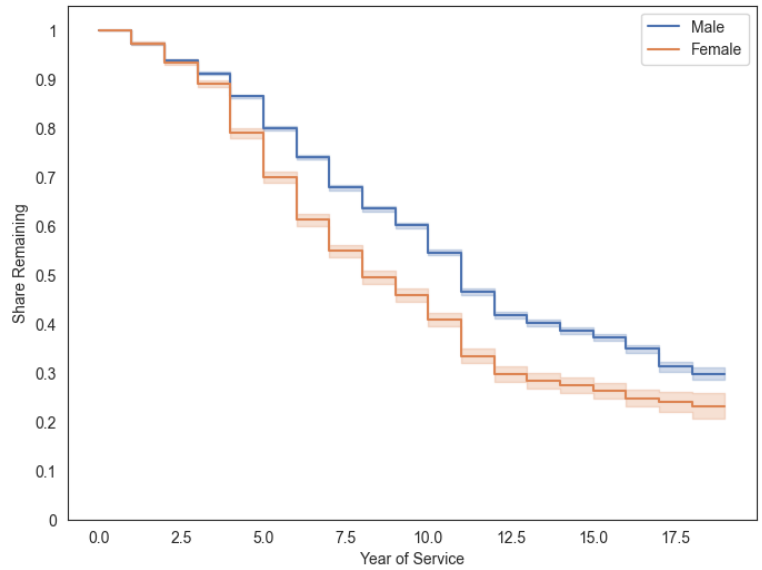
Note: Active Duty Navy officers commissioned as O-1s in 2001 or later who began their career as an officer of the line. DMDC annual data, 2001-2019.

Figure 4. Share of O-1 Commissions by Demographic per Calendar Year

2. Retention by demographic group

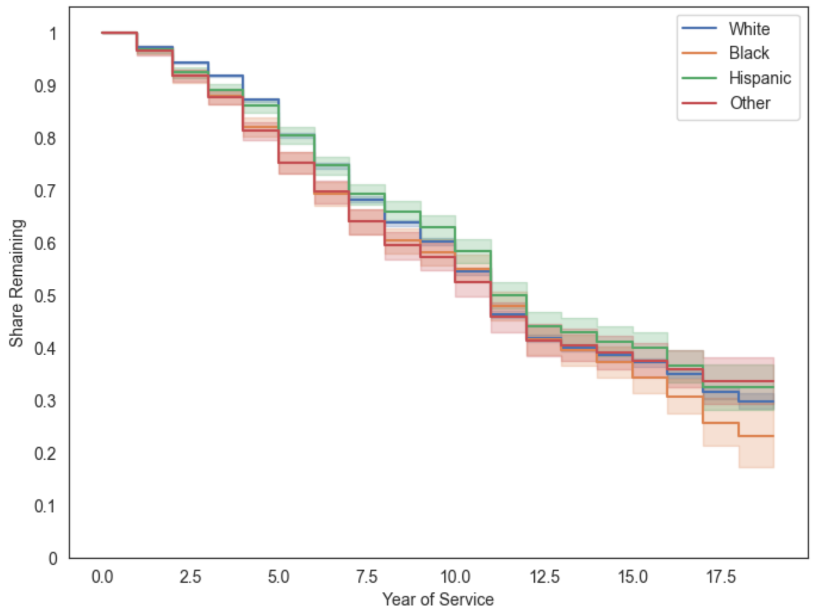
Figure 5 displays retention trends by year of service among males and females for all commissioning cohorts in our data. Although initial retention rates are similar among males and females, they diverge beginning in the fourth year of Navy service as female retention drops relative to males and remains lower thereafter. Not only do fewer females commission as O-1s in the Navy (see Figure 4), but relative to their original cohort size, even fewer females serve long enough to be considered eligible to promote to O-5. This further illustrates that retention differences are a primary driver of female underrepresentation among O-5s.

Figure 6 and Figure 7 plot retention rates by year of service and race among males and females, respectively. Among males, Whites have a marginally higher retention rate in years 3-8 compared to Blacks and Others; Hispanics have a slightly higher retention rate in years 5-8 relative to Blacks and Others. Conditional on remaining in service for 9 years, retention rates for subsequent service years are statistically indistinguishable between racial groups. Among females, retention rates for Whites and non-Whites move together across the early years of service, with non-White female retention statistically higher in years 11-15 than that of White females.



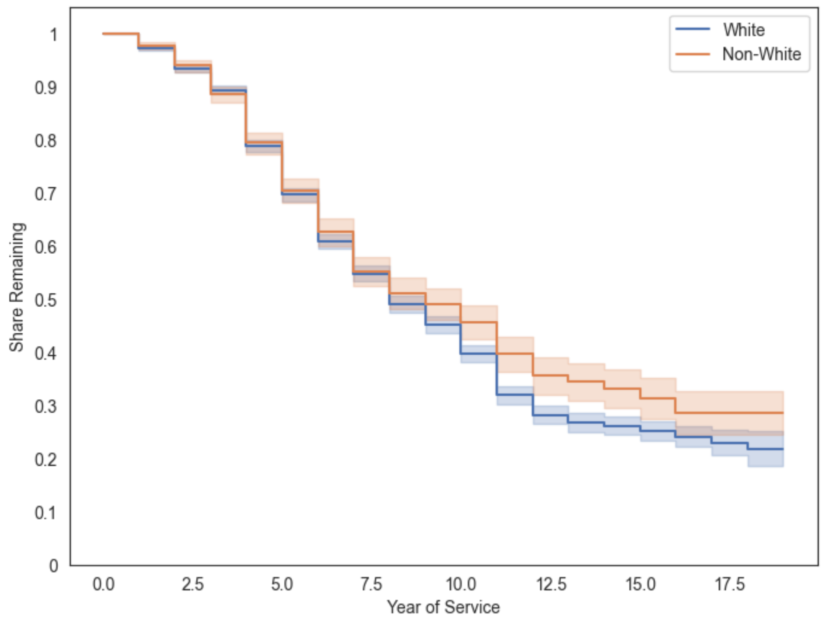
Note: Active Duty Navy officers commissioned as O-1s in 2001 or later who began their career and remain as an officer of the line. DMDC annual data, 2001-2019. Kaplan-Meier estimator shown, 95% confidence intervals shaded.

Figure 5. Retention by Year of Service and Sex



Note: Active Duty Navy officers commissioned as O-1s in 2001 or later who began their career and remain as an officer of the line. DMDC annual data, 2001-2019. Kaplan-Meier estimator shown, 95% confidence intervals shaded.

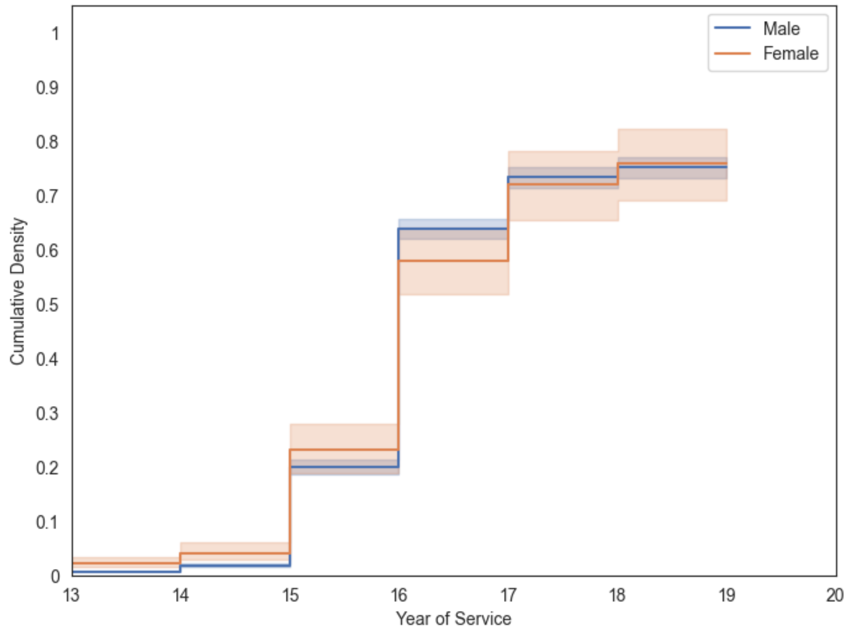
Figure 6. Retention by Year of Service and Race, Males



Note: Active Duty Navy officers commissioned as O-1s in 2001 or later who began their career and remain as an officer of the line. DMDC annual data, 2001-2019. Kaplan-Meier estimator shown, 95% confidence intervals shaded..

Figure 7. Retention by Year of Service and Race, Females

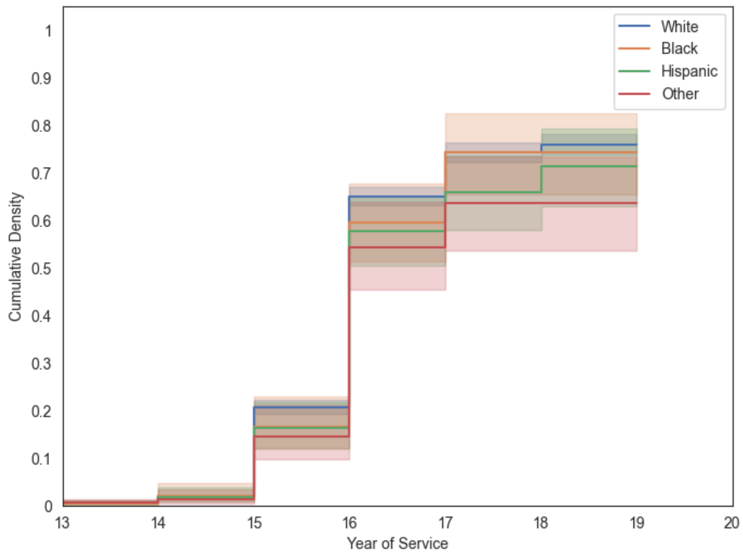
3. Promotion by demographic group



Note: Active Duty Navy officers commissioned as O-1s in 2001 or later who began their career and remain as an officer of the line. DMDC annual data, 2001-2019. Cumulative density function from Kaplan-Meier estimate, 95% confidence intervals shaded. Observed attritions removed prior to plotting. Censored observations (i.e., the remaining share) could still promote or exit service in the future. Also see Appendix A, Figure A-5 for a kernel density estimate of promotion durations for males and females.

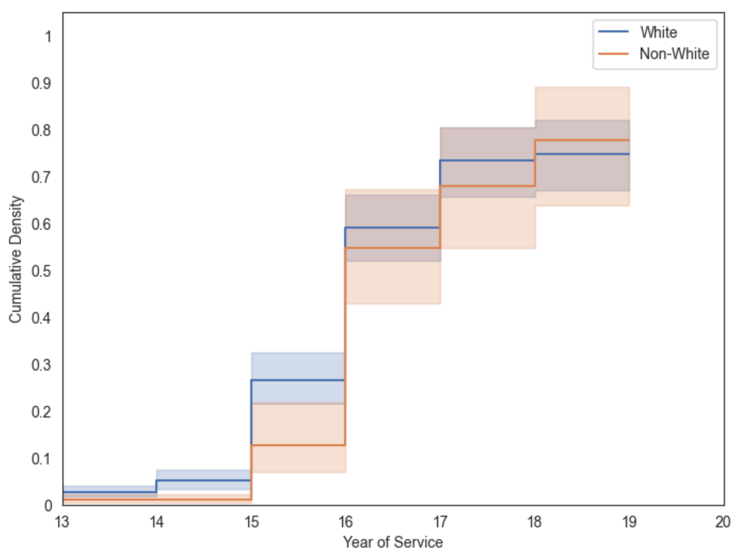
Figure 8. Share Promoted to O-5 by Year of Service among Males and Females

Figure 8 plots promotion rates by year of service among males and females. Importantly, prior attritions have been removed from this plot (as well as Figure 9 and Figure 10), thereby already accounting for a major driver of sex-based representation disparities among mid-career Navy officers. Conditional on not already exiting Navy service, promotion rates for all comparisons in Figures 8-10 are statistically indistinguishable from each other, meaning promotion rates are approximately similar in service years 13-19.



Note: Active Duty Navy officers commissioned as O-1s in 2001 or later who began their career and remain as an officer of the line. DMDC annual data, 2001-2019. Cumulative density function from Kaplan-Meier estimate, 95% confidence interval shaded. Observed attritions removed prior to plotting. Censored observations (i.e., the remaining share) could still promote or exit service in the future. Also see Appendix A, Figure A-6 for a kernel density estimate of promotion durations by race among males.

Figure 9. Share Promoted to O-5 by Year of Service and Race, Males



Note: Active Duty Navy officers commissioned as O-1s in 2001 or later who began their career and remain as an officer of the line. DMDC annual data, 2001-2019. Cumulative density function from Kaplan-Meier estimate, 95% confidence interval shaded. Observed attritions removed prior to plotting. Censored observations (i.e., the remaining share) could still promote or exit service in the future. Also see Appendix A, Figure A-7 for a kernel density estimate of promotion durations by race among females.

Figure 10. Share Promoted to O-5 by Year of Service and Race, Females

This page is intentionally blank.

3. Methodology

A. Machine Learning Model

We model officer retention and O-5 promotion using the Finite Interval Forecasting Engine (FIFE) version 1.3.4, a free and open source machine learning package developed by the IDA and sponsored by OUSD (P&R).[22] FIFE offers an array of machine learning and other models for discrete-time survival analysis, a statistical method focused on modeling the duration of time until one or more events occur and where time is measured discretely (e.g., annually), rather than continuously (e.g., nanoseconds). For our retention model, we use FIFE's LGBSurvivalModeler, a tree-based ML model that is based on a classic single-risk survival model, whereby the outcome at a given time horizon is binary (remain in service[23] vs. exit service). For each time horizon, the LGBSurvivalModeler fits a LightGBM binary classifier model. Each model produces a probability of remaining in service through the last period of the time horizon, conditional on remaining in service through the periods before. The cumulative product of the predictions from these models form an estimated survival function. The survival probabilities at the time horizon *t* periods into the future are defined as

$$Pr(T_{i\tau} \geq t | X_{i\tau}) \tag{1}$$

where $X_{i\tau}$ is a vector of feature values for individual *i* at time τ, and $T_{i\tau}$ is the number of consecutive future periods the individual remains after time τ.

For our promotion model, we use FIFE's LGBStateModeler, a tree-based ML model, to forecast the future value of a feature, or "state," conditional on survival to that point. In our case, we model the binary outcome of having achieved the rank of O-5 or higher (vs. remain in service at any lower rank) at the given time horizon. The probabilities of being in state *d* at a given time horizon, conditional on survival to that time horizon, are

$$Pr(D_{it} = d \mid T_{i\tau} \geq t, X_{i\tau}) \tag{2}$$

where D_{it} is the state of individual *i, t* periods into the future.

[22] FIFE is written in Python and available via the Python Package Index (PyPI), or via GitHub at https://github.com/IDA-HumanCapital/fife.

[23] Individuals remaining in service in the last period of data are considered right-censored.

We train retention and promotion models separately using all data from our analytic set. Further, because we are chiefly interested in the correlates of in-sample performance, rather than out-of-sample performance, we do not specify a validation or holdout set, which is different from typical practice with machine learning.[24] All reported results are based on the training set.

B. Shapley Additive exPlanations (SHAP) Values

Machine learning models are designed for prediction, rather than coefficient estimates classically used in regression analysis. Nonetheless, a family of algorithms have been developed to explain or "credit" the prediction to the features used, thereby not only demystifying the traditional "black-box" nature of ML, but also offering an analogous concept to a regression coefficient.[25] We use a prominent explanation method: SHapley Additive exPlanations (SHAP).[26,27] Grounded in game-theoretic Shapley values, SHAP computes local explanations that quantify each feature's contribution to a given observation's prediction. For example, SHAP measures how much a given officer's primary designator code contributed to their predicted probability of retention in a particular year. This compares to a global explanation algorithm, which simply quantifies the impact of features on the model as a whole.[28] SHAP's local explanations thereby offer a considerably more nuanced understanding about which features influence a person's prediction. Further, by quantifying each feature's contribution towards an individual prediction, we can calculate separate explanations for various data subsets (e.g., Black males from the 2002 cohort in their tenth year of Navy service), as well as calculate statistical moments across individuals' SHAP values.

---

24 Validation sets are commonly used to test how out-of-sample performance varies with hyperparameters (e.g., number of boosting iterations). Holdout sets are commonly used to estimate the out-of-sample performance of a chosen model.

25 Fellow analysts might question why we do not simply estimate a regression model. In this application, ML is advantageous, as it does not require specifying the functional relationship between the 300+ input features and the output *a-priori*. This means we do not codify (potentially incorrect) functional form assumptions, but instead allow the algorithm to learn the contours of the data.

26 Scott M Lundberg, Gabriel G. Erion, and Su-In Lee, "Consistent Individualized Feature Attribution for Tree Ensembles," University of Washington, February 11, 2018. https://arxiv.org/pdf/1802.03888. Scott Lundberg, et al. "From Local Explanations to Global Understanding with Explainable AI for Trees," *Nature Machine Intelligence 2*, no. 1 (2020): 56–67. https://doi.org/10.1038/s42256-019-0138-9.

27 SHAP is a Python module (available on PyPI and GitHub), with functionality for both tree-based and deep-learning models. Because we use a tree-based model, we use SHAP's TreeExplainer method.

28 An example of a global explanation would be XGBoost's (another tree-based algorithm) get_scores method, which counts the number of times a feature was used to split the data as leaves are grown.

Intuitively, Shapley values are calculated by comparing a model prediction with and without each feature. However, the contributions of the features depend on the order in which they are omitted from the model. One method to address this issue would be to compute predictions for all possible orders of omitted features. In general, however, this method is computationally intractable. SHAP's TreeExplainer method calculates exact Shapley values by utilizing the internal structure of tree-based models, measuring the effects of features based on a set of calculations for each leaf in a tree.[29]

By the term "contribution," we mean that SHAP values quantify each feature's effect on the change in model prediction for each observation (i.e., a partial effect from the mean of the prediction in the training set). Averaging individuals' SHAP values by feature, our reported SHAP values represent average partial effects (i.e., the average contribution of a feature on the outcome, conditional on the other features in the model); these are not average marginal effects (i.e., the average instantaneous change in the outcome in response to an instantaneous change to the value of the feature, conditional on other features in the model). In our case, SHAP values capture a change in log odds of retention or promotion relative to the baseline. Summing SHAP values across features for a given individual equals that person's prediction. Individuals missing data for a given feature have a SHAP value of zero.[30]

We estimate SHAP values separately by race-sex group for all individuals in our training set at their tenth year of service, forecasting four years into the future (see Appendix B for figures based on individuals' first year of service, forecasting six years into the future). Conditional on having served ten years in the Navy, officers at this time are typically mid-career and must decide whether to remain on their path towards a full Navy service career, or exit. We use this same service year and forecast horizon to produce SHAP values from our promotion model, except that we restrict to cohorts that commissioned in 2001-2004, since promotion outcomes are only well-observed for these individuals (see Appendix A3). Conditional on having served ten years, SHAP values from the promotion model illuminate feature effects on the predicted probability of advancement at a time when officers in our analytic set are first promoting to O-5. Figure 11 and Figure 12 show the distribution of feature-wise SHAP values according to the mean and standard deviation, respectively, from the retention model. These distributions are very similar to those from the promotion model. SHAP values may be either positive or negative, depending on whether a feature increases or decreases the log odds of retention or promotion for a set of individuals. The importance of a feature is the average magnitude of its contribution (i.e.

[29] Scott Lundberg, et al. "From Local Explanations to Global Understanding with Explainable AI for Trees."

[30] Ibid.

averaged across all officers in a particular subset), whether positive or negative. Correspondingly, we report feature importance plots in absolute terms.

We examine differences by race-sex group among the most important features predicting retention or promotion, comparing each demographic group to White males as the reference group. We calculate this difference in three steps. First, we identify the union of most important features for the race-sex comparison demographic group and White males.[31] Second, we normalize the feature-wise vector of mean SHAP values by subtracting the grand mean across all features:

$$\ddot{X} = X - \mu \quad (3)$$

where X is a vector of mean SHAP values for a given demographic and μ is mean of that vector (i.e. the demographic-specific SHAP value grand mean). We do this to account for the possibility that some demographic groups may have higher average SHAP values than others. Third, restricting to the union of most important features from (3), we subtract White males' mean absolute SHAP value per feature from the mean absolute SHAP value of the corresponding feature in the race-sex comparison demographic. That is,

$$\Delta = \left|\ddot{X}_{ComparisonDemo}\right| - \left|\ddot{X}_{WhiteMales}\right| \quad (4)$$

where Δ is a feature-wise vector of average differences in normalized absolute SHAP values.

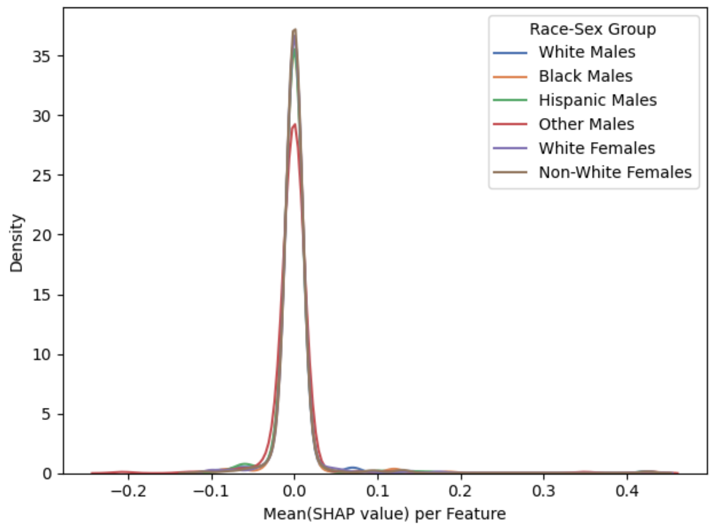
Conceptually, differences in a feature's effect on the prediction between demographic groups can be decomposed into two parts: first, differences in the effect itself, including the direction (i.e., increase/decrease the probability of retention or promotion) and magnitude (i.e., the size of the contribution, in absolute terms); and second, differences in the distribution of demographic groups across features. For example, relative to other Navy designators, females are disproportionately present in health care designators while males are disproportionately present in tactical designators.[32] If the average partial effect of a feature is different for officers in a health care occupation than for officers in a tactical occupation, we can expect the average partial effect of that feature to differ between males and females. We focus especially on differences in the magnitude of feature effects,

---

[31] For both the retention and promotion models, we define the most important features for a given service year, forecast lead length, and race-sex group combination by z-score normalizing the distribution of mean absolute SHAP values for that set and using all of the features above 60th percentile. We additionally exclude 37 features that are linear combinations of our outcome (e.g., dates, duration since particular dates, person age, pay amounts indexed to inflation) returned from this algorithm from all figures. These excluded features enhance the precision of our predictions, but offer little meaningful information otherwise.

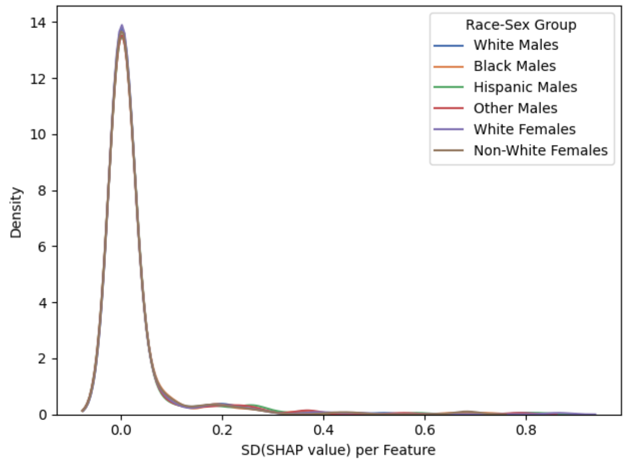
[32] U.S. Department of Defense, Office of the Assistant Secretary of Defense (Force Management Policy)(Accession Policy), "Population Representation Reports," Appendix B, Table B-28, various years.

illustrated by the most important features. Notably, this tells us about disparities in which and how much features matter across groups. For example, if a particular feature matters considerably more for White males than another race-sex comparison demographic, this indicates that this feature has a stronger association with the outcome for White males, possibly warranting follow-up investigation. Where possible – among individual categories of a categorical feature (officer subspecialty) – we also examine the effect direction. While categories generally have the same effect direction on the prediction across demographic groups, they occasionally diverge, possibly warranting follow-up examination.



Note: Active Duty Navy officers commissioned as O-1s in 2001 or later who began their career and remain as an officer of the line. DMDC annual data, 2001-2019. Results from retention model; distribution of mean absolute SHAP values similar for promotion model.

Figure 11. Kernel Density of Feature-Wise Mean SHAP Values by Race-Sex Group, Service Year 10, Forecast Lead 4 Years



Note: Active Duty Navy officers commissioned as O-1s in 2001 or later who began their career and remain as an officer of the line. DMDC annual data, 2001-2019. Results from retention model; distribution of standard deviations of absolute SHAP values similar for promotion model.

Figure 12. Kernel Density of Feature-Wise SHAP Value Standard Deviations by Race-Sex Group, Service Year 10, Forecast Lead 4 Years

4. Results

A. Retention Model

Figures 13-17 display the most important features for each demographic pair. The left and center subplots in these figures show the feature-wise mean absolute SHAP values for the given comparison demographic and White males, respectively. These represent the feature's average absolute partial effect on the mean of the prediction (in log odds), relative to the mean of the outcome variable in the training set.[33] The horizontal axis of these two subplots is equal in scale, while the horizontal axis scale in the right-most plot is magnified. The length of orange (comparison demographic) and blue (White males) bars in the left and center subplots indicate the magnitude of the feature's impact on the retention prediction. Longer bars denote a greater effect, in absolute terms. The right-most subplot in these figures shows the difference between the normalized mean absolute SHAP values across groups, which we calculate from Equation 4.[34] This difference captures the degree to which a feature explains relatively more of the prediction for the comparison demographic versus White males. Positive (orange) bars indicate a larger effect in absolute terms for the comparison demographic, whereas negative (blue) bars denote a larger effect in absolute terms for White males.

Overall, the union of most important predictors of retention across all demographic groups include *officer primary designator code*,[35] *officer subspecialty*,[36] *billet designator*

[33] The training set comprises all individuals across all cohorts in our analysis set. However, SHAP values are estimated for a subset of these individuals: we restrict to individuals at their tenth year of service in the Navy. For the promotion model, the SHAP subset is restricted even further to officers in their tenth year of service *and* who commissioned in 2001-2004.

[34] The x-axis for the right subplot is not on the same scale as the left and center, as differences are often small and might be otherwise missed.

[35] An officer's primary specialty (designator) code captures the Navy specialty education and training an officer possesses. This is specific to a particular community (e.g., line [unrestricted, restricted, and restricted line special duty], staff, limited duty, or warrant). This code describes the type of billet for which an officer is qualified. For more information see: https://www.public.navy.mil/bupers-npc/reference/noc/NOOCSVOL1/Pages/default.aspx.

[36] This refers to an officer's professional educational discipline that is secondary to their primary specialty designator. Subspecialties have degree requirements that are specific to a given discipline and also require a master's degree or higher from an accredited educational institution. For more information, see: https://www.public.navy.mil/bupers-npc/reference/noc/NOOCSVOL1/Pages/default.aspx.

code,[37] *additional officer qualifier designations*,[38] *number of dependents*, *home residence type*,[39] *citizenship origin*, and *religious denomination*.[40] It is important to bear in mind that while these are partial effects (i.e., conditional on other features in the model), they are not causal. Compared to White males, the relative importance of some of these features in predicting retention vary somewhat across demographics: *officer subspecialty*, *citizenship origin*, and, to some extent, *qualification designations* matter more for comparison demographics than for White males, while *officer primary designator* matters consistently more for retention among White males than for comparison demographics. This suggest that comparison demographic retention is strongly influenced by a wider array of factors than that of White males, and that these factors are structurally different in nature. What matters more for retention among White males is the type of job (i.e., Navy community and specific occupation) they are qualified for, while for members of comparison demographics specialized knowledge and training matters more. This may be driven by the relative status of integrated subspecialties after accounting for membership in a restricted vs. unrestricted line occupation.

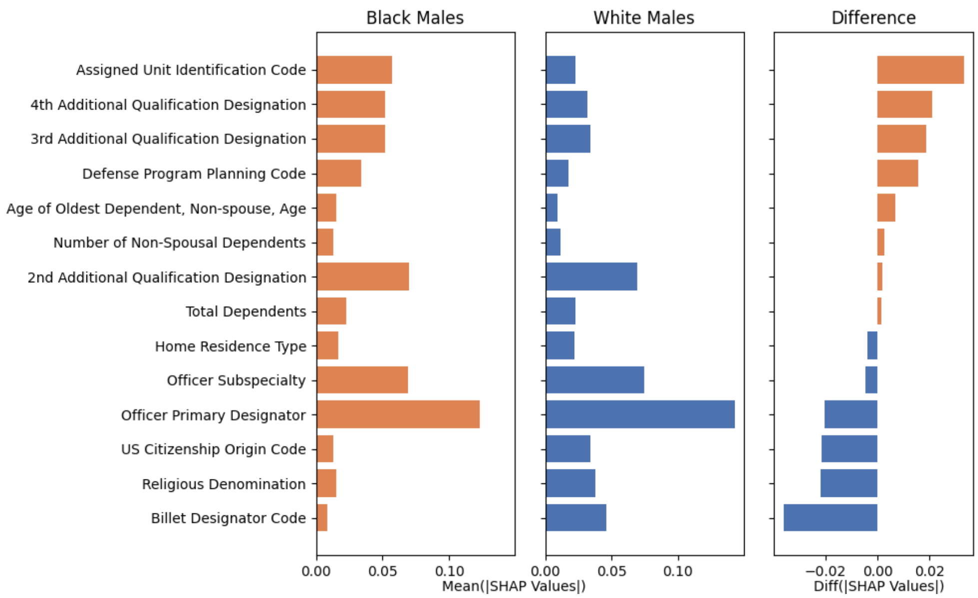
Drilling down into specific demographic comparisons, *assigned unit identification code* (UIC) has the largest effect among Black vs. White males (Figure 13). An officer's assigned UIC matters considerably more for Black male officers than White male officers for retention; for no other group is assigned UIC so consequential to retention. For Hispanic males and non-White females, *religious denomination* has among the largest differential effect (Figures 14 and 17). Religious denomination, therefore, matters a great deal more for retention among Hispanic males and non-White females relative to White males. Among male officers with a race of Other, *citizenship origin* is of particular salience (Figure 15). Perhaps surprisingly, family factors, including *number of dependents* and *home residence type*, are not disproportionately more consequential for White or non-White females (Figures 16 and 17). Instead, career-related features, including qualification designation and officer subspecialty, matters considerably more for females than White male officers in retention.

[37] Billet designator codes are an occupation code entail the primary Navy specialty qualifications required of an incumbent. This is specific to a particular community. For more information see: https://www.public.navy.mil/bupers-npc/reference/noc/NOOCSVOL1/Pages/default.aspx

[38] Additional qualification designations are unique qualifications awarded an officer, indicating recognition of specific skills and knowledge. These augment officer designator codes. For more information see: https://www.public.navy.mil/bupers-npc/reference/noc/NOOCSVOL1/Pages/default.aspx

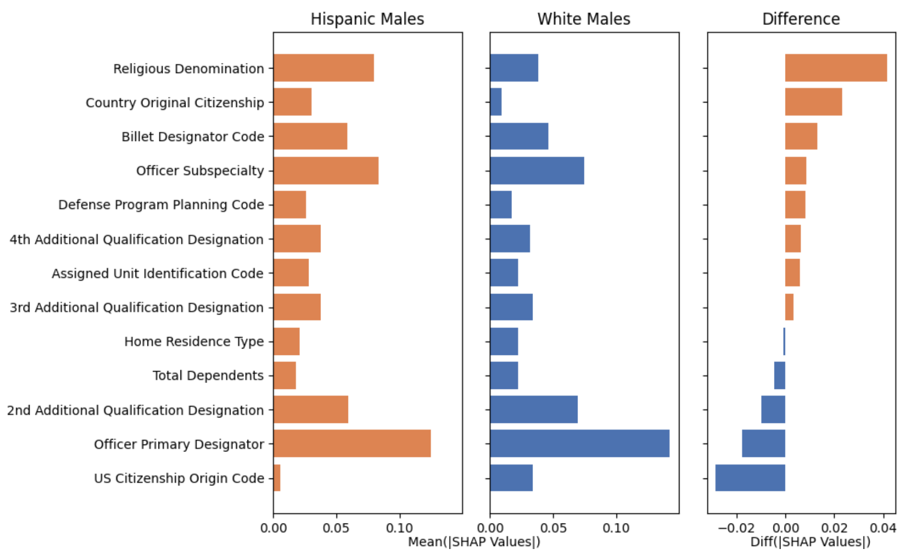
[39] This feature denotes whether the individual receives a housing allowance for living in a duty or residence location, with or without dependents.

[40] Feature descriptions in all plots are limited to 40 characters. Longer descriptions therefore may be cut short, see Appendix C for full feature descriptions.



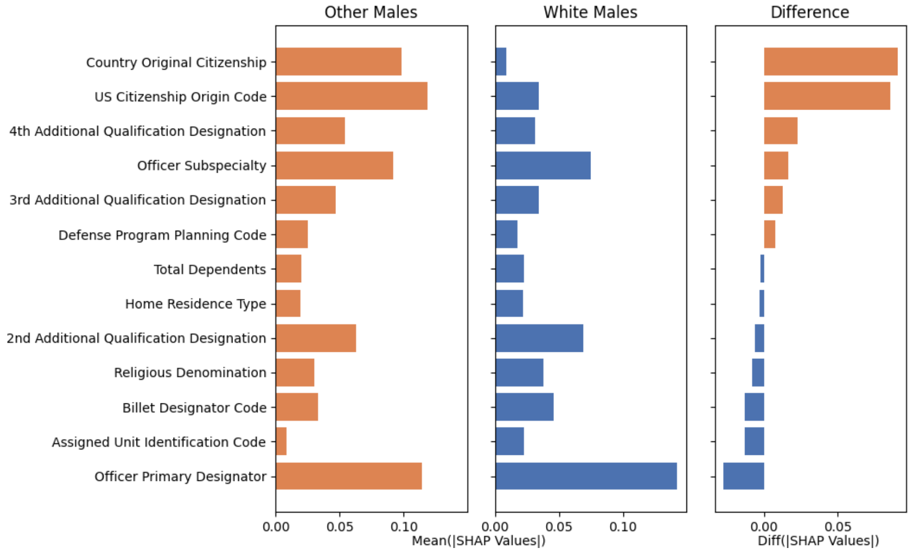
Note: Active Duty Navy officers commissioned as O-1s in 2001-2018 who began their career and remain as an officer of the line. DMDC annual data, 2001-2019. Top features according to mean of absolute SHAP values and their absolute difference. Orange bars in right subplot represent greater absolute feature importance for the comparison demographic group; blue bars indicate greater absolute feature importance for White males.

Figure 13. Top Features Predicting Retention among Black Males and White Males, Service Year 10, Forecast Lead 4 Years



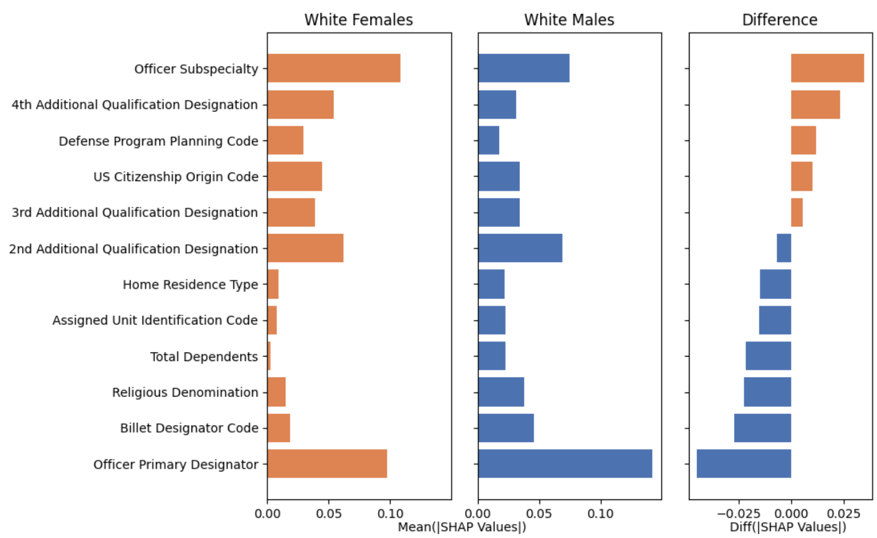
Note: Active Duty Navy officers commissioned as O-1s in 2001-2018 who began their career and remain as an officer of the line. DMDC annual data, 2001-2019. Top features according to mean of absolute SHAP values and their absolute difference. Orange bars in right subplot represent greater absolute feature importance for the comparison demographic group; blue bars indicate greater absolute feature importance for White males.

Figure 14. Top Features Predicting Retention among Hispanic Males and White Males, Service Year 10, Forecast Lead 4 Years



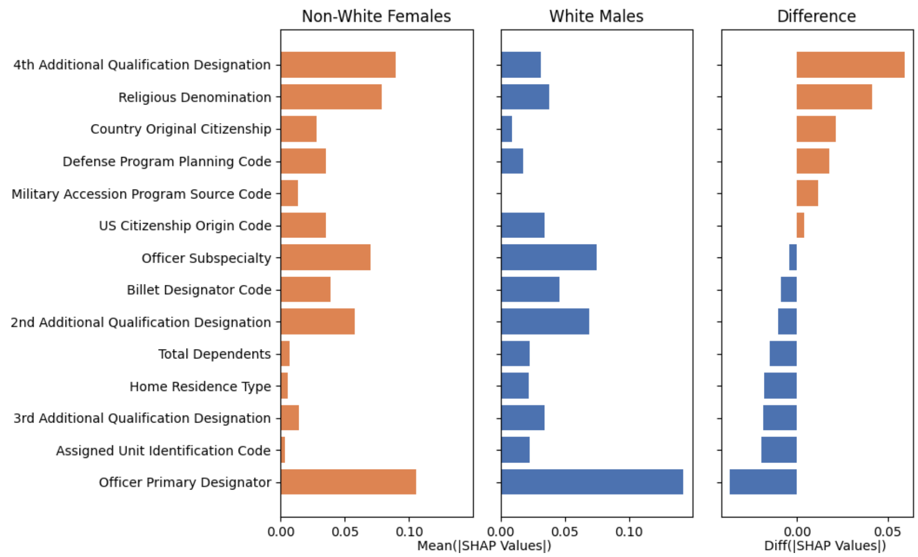
Note: Active Duty Navy officers commissioned as O-1s in 2001-2018 who began their career and remain as an officer of the line. DMDC annual data, 2001-2019. Top features according to mean of absolute SHAP values and their absolute difference. Orange bars in right subplot represent greater absolute feature importance for the comparison demographic group; blue bars indicate greater absolute feature importance for White males.

Figure 15. Top Features Predicting Retention among Other Males and White Males, Service Year 10, Forecast Lead 4 Years



Note: Active Duty Navy officers commissioned as O-1s in 2001-2018 who began their career and remain as an officer of the line. DMDC annual data, 2001-2019. Top features according to mean of absolute SHAP values and their absolute difference. Orange bars in right subplot represent greater absolute feature importance for the comparison demographic group; blue bars indicate greater absolute feature importance for White males.

Figure 16. Top Features Predicting Retention among White Females and White Males, Service Year 10, Forecast Lead 4 Years



Note: Active Duty Navy officers commissioned as O-1s in 2001-2018 who began their career and remain as an officer of the line. DMDC annual data, 2001-2019. Top features according to mean of absolute SHAP values and their absolute difference. Orange bars in right subplot represent greater absolute feature importance for the comparison demographic group; blue bars indicate greater absolute feature importance for White males.

Figure 17. Top Features Predicting Retention among Non-White Females and White Males, Service Year 10, Forecast Lead 4 Years

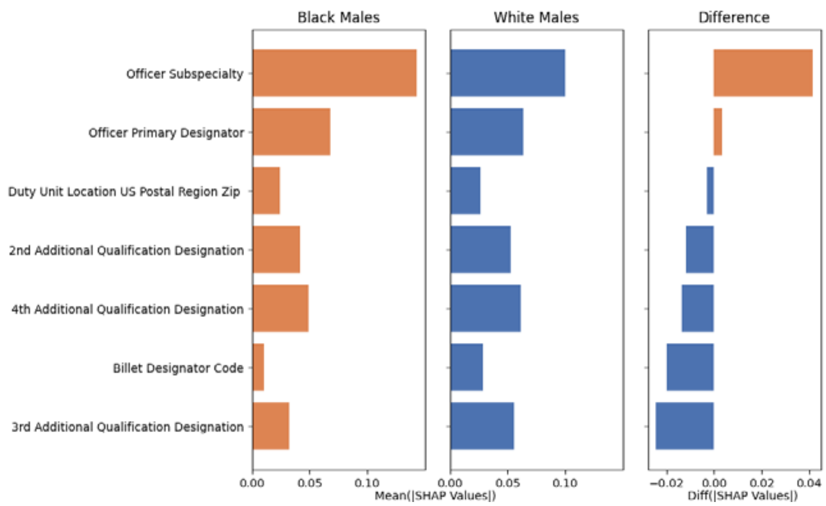
B. Promotion model

Figures 18, 20, 22, 24, and 26 examine the most important features predicting O-5 promotion at officers' tenth year of Navy service, forecasting four years into the future. Although we restrict to the same service duration and forecast lead length as our previous retention model, the SHAP results depicted in Figures 18-22 were generated using only officers that commissioned in 2001-2004. As with the retention model plots, the left and center subplots capture the feature's average absolute partial effect of the mean of the prediction (in log odds) and are scaled equally, while the right-most subplot is the absolute difference between groups and has a magnified scale. Compared to the previous retention figures, nearly all highly predictive features for all demographic groups relate to officers' Navy service: *officer subspecialty*, *officer primary designator code*, *billet designator code*, *additional officer qualifier designations*, and *duty unit location zip code*. We no longer observe a strong influence of family, religious denomination, or citizenship origin on predicted O-5 promotion outcomes. Still, some of these effects may be proxied by duty unit location zip, which is an important promotion feature.

For all demographic groups, officer subspecialty is the most consequential feature predicting O-5 promotion outcomes. Moreover, compared to White males, this feature is even more meaningful for promotion outcomes among comparison demographic groups. With the exception of Hispanic males, subspecialty is among the top two most significant features, relative to White males. Officer subspecialty is therefore of notable consequence in predicting both O-5 promotion outcomes and retention outcomes among females and racial/ethnic minorities. Different from our retention model, *officer primary designator* has mixed importance. Among Black males, Other males, and White females this feature is relatively more consequential, while for Hispanic males and non-White females it is less. We also see that billet designator code more greatly influences promotion outcomes among White males, in general. The effect of additional qualification designators is mixed across demographics.

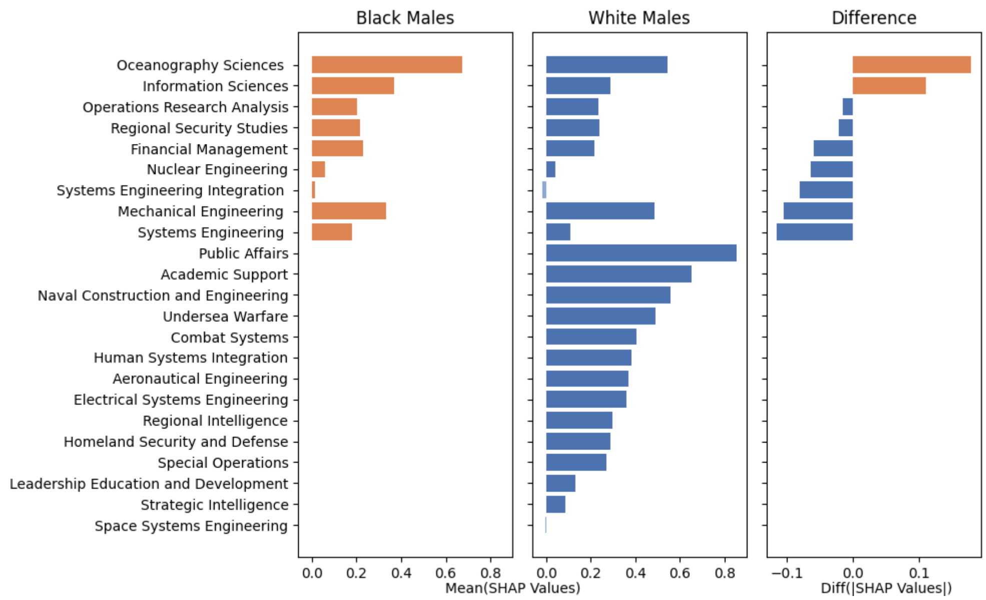
To begin to unpack the particular importance of officer subspecialty in O-5 promotion outcomes for each demographic group, Figures 19, 21, 23, 25, and 27 examine the average partial effect of individual officer subspecialties. Although our promotion model is trained using 4-digit subspecialties, for parsimony we aggregate to the 2-digit level when constructing these figures. Positive values (darker orange/blue bars) in the left and center subplots indicate the subspecialty increases the probability of promotion, whereas lighter bars represent the opposite effect. In the right subplot of these figures, orange bars signify greater absolute feature importance for the comparison demographic; blue bars indicate greater absolute feature importance for White males. Missing bars in the left and right subplot indicate no officers were present in that particular aggregated subspecialty code. On the one hand, the absence of comparison demographic members in a given subspecialty is partly attributable to females and ethnic/racial minorities also being numeric minorities in the Navy, meaning fewer comparison demographic officers are available to populate various subspecialties. On the other, their absence may also indicate segregation, whether driven by personal choices, institutional barriers (e.g., females on submarines), or other influences. Observed differences might also be attributable to differential demographic representation among restricted and unrestricted line occupations. Overwhelmingly, the effects of individual subspecialties point in the same direction for all demographic groups, indicating they change the model's predicted probability of promotion in the same manner (i.e., positively or negatively). Still, approximately half of aggregated subspecialties are only populated by White males. These findings require further investigation.

Across most comparison demographic groups, *regional security studies*, *information sciences*, *oceanography sciences*, and, to some extent, *engineering* disciplines (aeronautical, mechanical) disproportionately increase the predicted likelihood of promotion for females and racial/ethnic minorities relative to White males. For example, conditional on remaining in service for ten years, membership in oceanography sciences increases the predicted likelihood of promotion 96.2% ($1.962 = \exp^{0.673919}$) on average for Black male officers, holding all other factors constant. If their baseline promotion likelihood at this same year of service was for example 30%, this would correspond to a promotion probability of 58.9% ($0.589 = 0.3 \times 1.962$). For White male officers, the average effect is smaller, raising their predicted promotion prospects 71.9%, on average. After accounting for differences in the distribution of category-specific mean SHAP values between groups (see methods section), specializing in oceanography sciences raises Black male's promotion likelihood 19.7% more than White males (Figure 18). Still, the number of consequential subspecialties for comparison demographics is few. As a result, the overall greater feature importance of officer subspecialty for comparison demographics is driven entirely by a handful of select subspecialties (e.g., regional security studies, information sciences, and oceanography sciences). The confluence of cultural expectations for what constitutes a promotion-enabling career trajectory with restricted vs. unrestricted line status may also explain the importance of integrated subspecialties in these models.



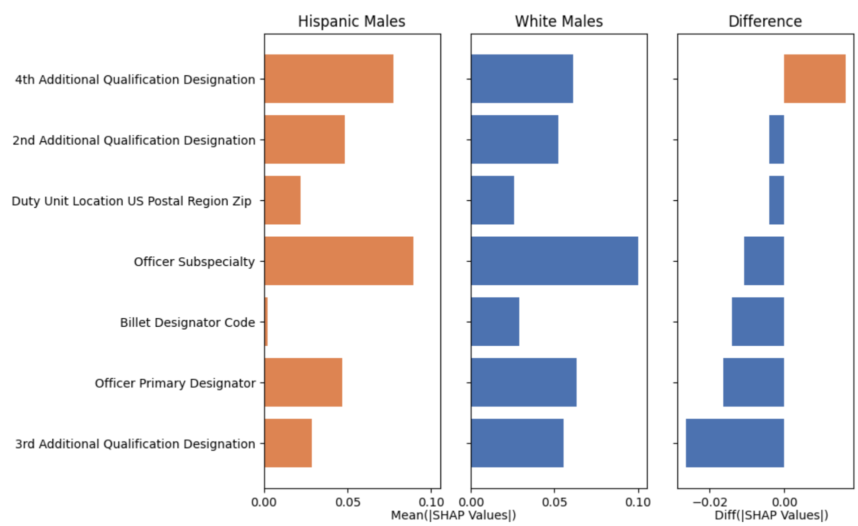
Note: Active Duty Navy officers commissioned as O-1s in 2001-2004 who began their career and remain as an officer of the line. DMDC annual data, 2001-2019. Top features according to mean absolute SHAP values and their difference. Orange bars in right subplot represent greater absolute feature importance for the comparison demographic group; blue bars indicate greater absolute feature importance for White males.

Figure 18. Top Features Predicting O-5 Promotion among Black Males and White Males, Service Year 10, Forecast Lead 4 Years



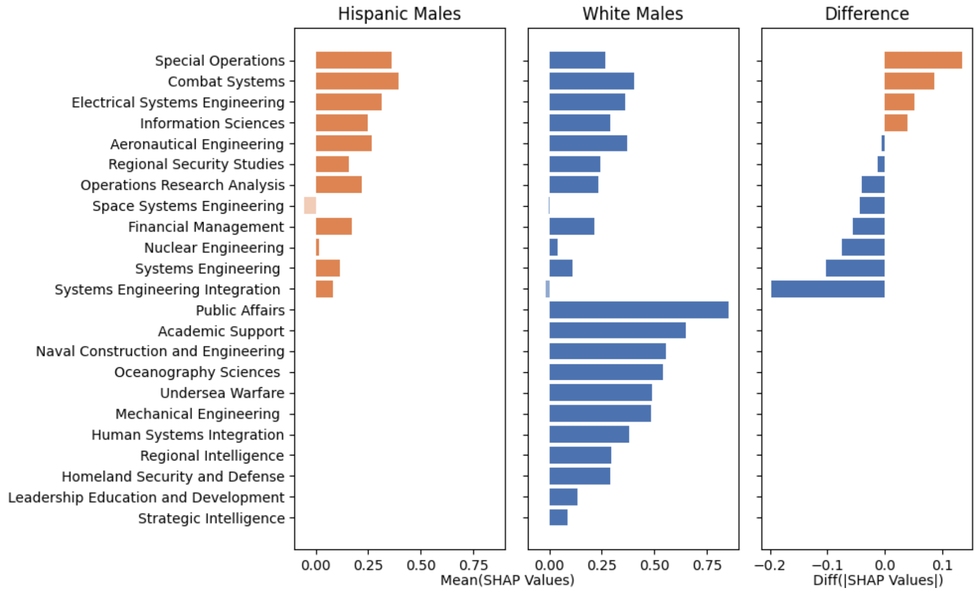
Note: Active Duty Navy officers commissioned as O-1s in 2001-2004 who began their career and remain as an officer of the line. DMDC annual data, 2001-2019. Top categories of feature according to mean of SHAP values and SHAP value absolute difference. Positive values (darker orange or blue bars) in left and center plots indicate higher probability of promotion. Orange bars in right subplot represent greater absolute feature importance for the comparison demographic group; blue bars in this same subplot indicate greater absolute feature importance for White males. Empty bars indicate no comparison demographic officers were present in that subspecialty.

Figure 19. Officer Subspecialty Code, Top Categories Predicting O-5 Promotion among Black Males and White Males, Service Year 10, Forecast Lead 4 Years



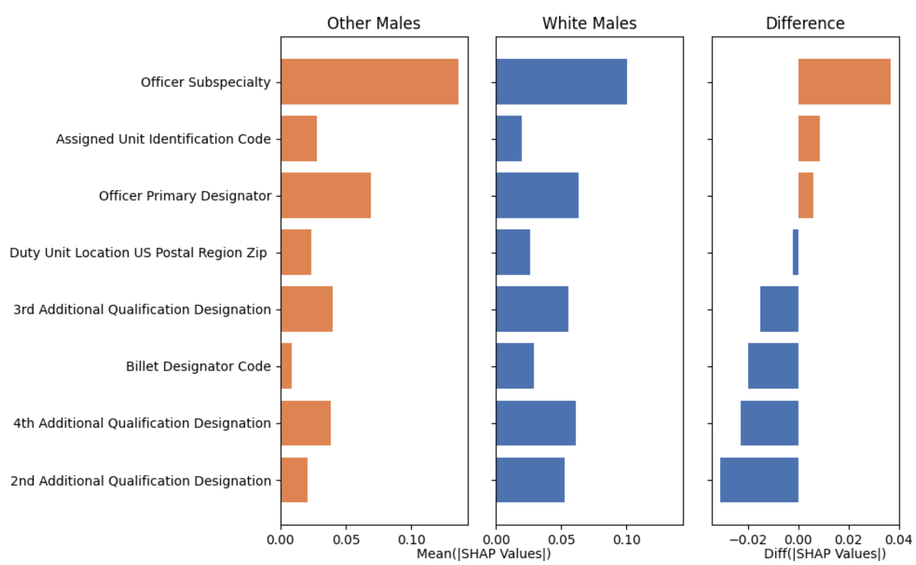
Note: Active Duty Navy officers commissioned as O-1s in 2001-2004 who began their career and remain as an officer of the line. DMDC annual data, 2001-2019. Top features according to mean absolute SHAP values and their difference. Orange bars in right subplot represent greater absolute feature importance for the comparison demographic group; blue bars indicate greater absolute feature importance for White males.

Figure 20. Top Features Predicting O-5 Promotion among Hispanic Males and White Males, Service Year 10, Forecast Lead 4 Years



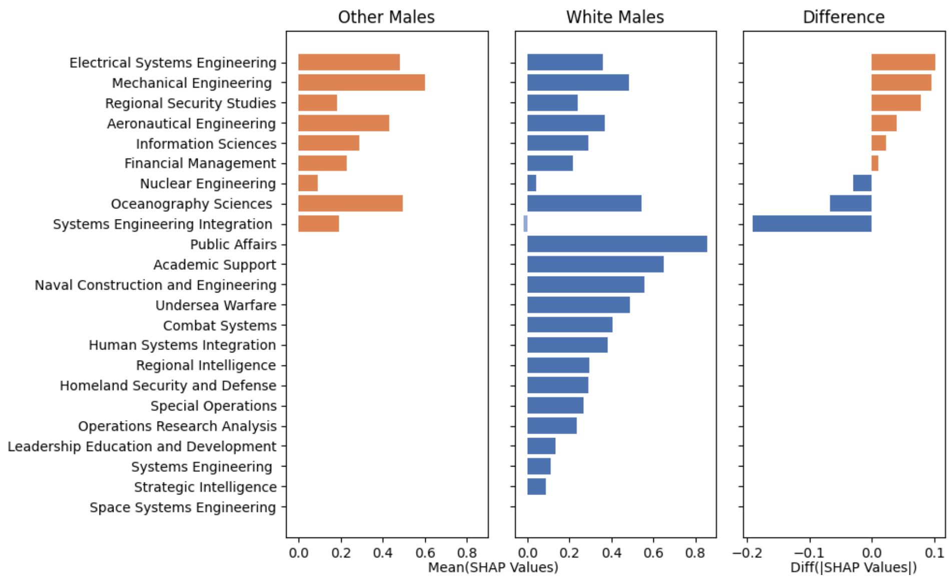
Note: Active Duty Navy officers commissioned as O-1s in 2001-2004 who began their career and remain as an officer of the line. DMDC annual data, 2001-2019. Top categories of feature according to mean of SHAP values and SHAP value absolute difference. Positive values (darker orange or blue bars) in left and center plots indicate higher probability of promotion. Orange bars in right subplot represent greater absolute feature importance for the comparison demographic group; blue bars in this same subplot indicate greater absolute feature importance for White males. Empty bars indicate no comparison demographic officers were present in that subspecialty.

Figure 21. Officer Subspecialty Code, Top Categories Predicting O-5 Promotion among Hispanic Males and White Males, Service Year 10, Forecast Lead 4 Years



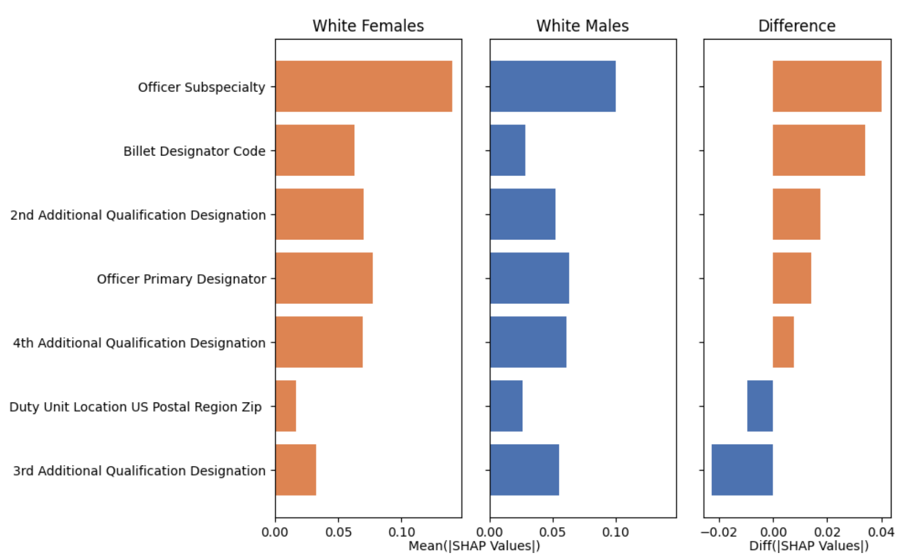
Note: Active Duty Navy officers commissioned as O-1s in 2001-2004 who began their career and remain as an officer of the line. DMDC annual data, 2001-2019. Top features according to mean absolute SHAP values and their difference. Orange bars in right subplot represent greater absolute feature importance for the comparison demographic group; blue bars indicate greater absolute feature importance for White males.

Figure 22. Top Features Predicting O-5 Promotion among Other Males and White Males, Service Year 10, Forecast Lead 4 Years



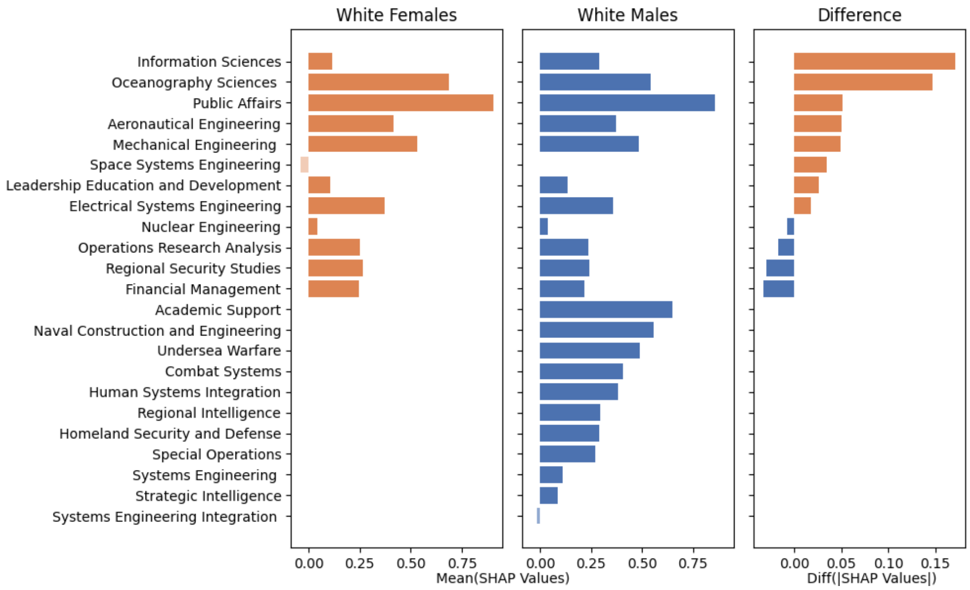
Note: Active Duty Navy officers commissioned as O-1s in 2001-2004 who began their career and remain as an officer of the line. DMDC annual data, 2001-2019. Top categories of feature according to mean of SHAP values and SHAP value absolute difference. Positive values (darker orange or blue bars) in left and center plots indicate higher probability of promotion. Orange bars in right subplot represent greater absolute feature importance for the comparison demographic group; blue bars in this same subplot indicate greater absolute feature importance for White males. Empty bars indicate no comparison demographic officers were present in that subspecialty.

Figure 23. Officer Subspecialty Code, Top Categories Predicting O-5 Promotion among Other Males and White Males, Service Year 10, Forecast Lead 4 Years



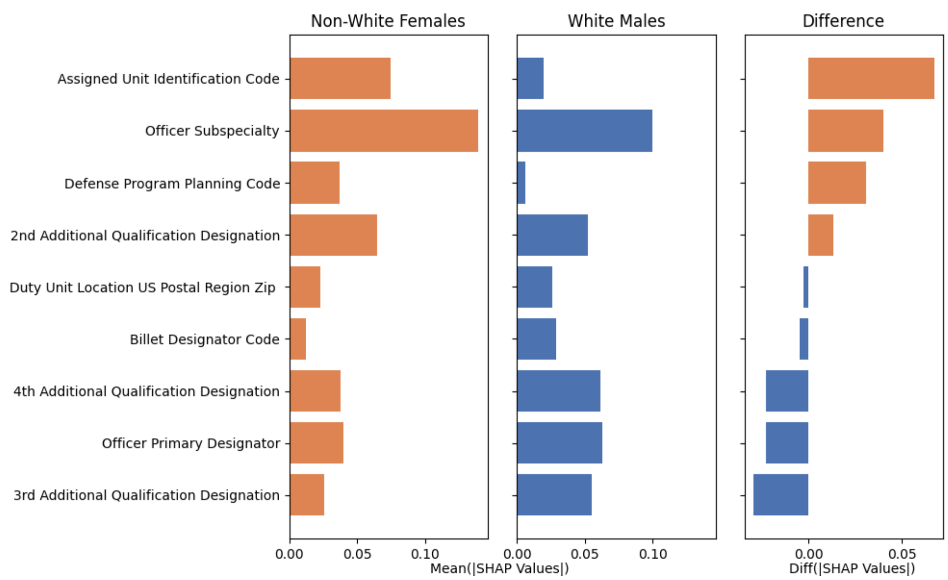
Note: Active Duty Navy officers commissioned as O-1s in 2001-2004 who began their career and remain as an officer of the line. DMDC annual data, 2001-2019. Top features according to mean absolute SHAP values and their difference. Orange bars in right subplot represent greater absolute feature importance for the comparison demographic group; blue bars indicate greater absolute feature importance for White males.

Figure 24. Top Features Predicting O-5 Promotion among White Females and White Males, Service Year 10, Forecast Lead 4 Years



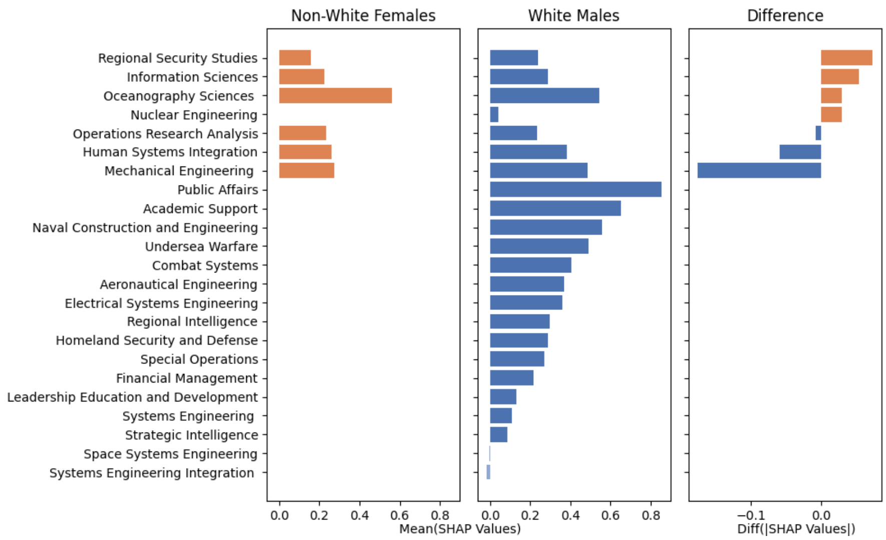
Note: Active Duty Navy officers commissioned as O-1s in 2001-2004 who began their career and remain as an officer of the line. DMDC annual data, 2001-2019. Top categories of feature according to mean of SHAP values and SHAP value absolute difference. Positive values (darker orange or blue bars) in left and center plots indicate higher probability of promotion. Orange bars in right subplot represent greater absolute feature importance for the comparison demographic group; blue bars in this same subplot indicate greater absolute feature importance for White males. Empty bars indicate no comparison demographic officers were present in that subspecialty.

Figure 25. Officer Subspecialty Code, Top Categories Predicting O-5 Promotion among White Females and White Males, Service Year 10, Forecast Lead 4 Years



Note: Active Duty Navy officers commissioned as O-1s in 2001-2004 who began their career and remain as an officer of the line. DMDC annual data, 2001-2019. Top features according to mean absolute SHAP values and their difference. Orange bars in right subplot represent greater absolute feature importance for the comparison demographic group; blue bars indicate greater absolute feature importance for White males.

Figure 26. Top Features Predicting O-5 Promotion among Non-White Females and White Males, Service Year 10, Forecast Lead 4 Years



Note: Active Duty Navy officers commissioned as O-1s in 2001-2004 who began their career and remain as an officer of the line. DMDC annual data, 2001-2019. Top categories of feature according to mean of SHAP values and SHAP value absolute difference. Positive values (darker orange or blue bars) in left and center plots indicate higher probability of promotion. Orange bars in right subplot represent greater absolute feature importance for the comparison demographic group; blue bars in this same subplot indicate greater absolute feature importance for White males. Empty bars indicate no comparison demographic officers were present in that subspecialty.

Figure 27. Officer Subspecialty Code, Top Categories Predicting O-5 Promotion among Non-White Females and White Males, Service Year 10, Forecast Lead 4 Years

5. Conclusion

This paper examines the correlates of race and sex-based differences in retention and O-5 promotional outcomes for restricted and unrestricted Navy officers of the line who commissioned as O-1s between 2001 and 2018. Using administrative military personnel data from DMDC, we employ a tree-based discrete-time survival ML model to produce retention and promotion forecasts. Because the majority of officers exit military service prior to fulfilling the minimum eligibility requirements for promotion to O-5, we examine feature effects from two distinct ML models: one predicting retention, and a second predicting promotion to O-5. We focus on four-year retention and promotion forecasts for officers in their tenth year of service. To quantify the effect of each feature provided to the ML model on each person's prediction, we apply the SHAP explanation algorithm. Averaging local effects by feature, we compare feature effects across six demographic groups: White non-Hispanic males, Black non-Hispanic males, Hispanic males (of any race), Other males, White non-Hispanic females, and non-White females. This method illuminates differences across demographic groups in which and how much features matter for the outcome under consideration. After identifying which features are most consequential for each demographic group, we then assess the degree to which this importance differs across demographics.

A. Synopsis of Findings

For all demographic groups, we find that many of the most consequential features predicting retention are also the most important predictors of promotion: *officer primary designator*, *officer subspecialty*, *billet designator code*, and *additional officer qualifier designations*. The significance of these career features may intersect with restricted vs. unrestricted line status, and requires further investigation. In addition to career features, family and personal attributes (e.g., *number of dependents*, *citizenship origin*, and *religious denomination*) are highly salient for retention outcomes, while the key features predicting O-5 promotions all relate to Navy service regardless of demographic group.

Comparing feature importance of each demographic group to White males, *officer subspecialty*, *citizenship origin*, and to some extent, *qualification designations* matter more for females and racial/ethnic minorities than for White males. Conversely, *officer primary designator* consistently matters more for retention among White males, compared to all other groups. This suggests that retention of females and racial/ethnic minorities is affected by a greater range of factors than that of White males, and that these factors are structurally different in nature. Occupation features such as Navy community and primary designator matter more for retention among White males, while specialized knowledge and training (i.e., officer subspecialty and additional officer qualifier designations) appear to matter more for females and racial/ethnic minorities. We also find variation across demographics in which features are especially predictive of retention relative to White males. For Black male officers, *assigned unit identification code* is particularly significant; for Hispanic males and non-White females, *religious denomination* is especially influential; for Other males, the nature of *citizenship origin* matters most; for White females, officer *subspecialty* assumes foremost importance. Notably, among officers in our analysis set, *number of dependents* is no more consequential for female retention than for White male retention. Prior research has found that females in the military are less likely to be married, less likely to have children, and more likely to be divorced. Our findings suggest that childbearing may not be the driving force behind female attrition, as some have postulated.

Results from our promotion model strongly indicate that *officer subspecialty* is the most consequential predictor of O-5 promotion outcomes for all demographic groups. This feature is especially predictive for females and racial/ethnic minorities: for all but Hispanic males, subspecialty is among the top two most meaningful features. In other words, officer subspecialty matters disproportionately in O-5 promotion outcomes for these groups. Different from our retention model, *officer primary designator* has mixed importance in the promotion model. While this feature is a strong predictor of promotion outcomes for some demographic groups, it matters less for others. As a result, officer primary designator is relatively less important than officer subspecialty in predicting O-5 promotion outcomes. These differences in relative importance of occupation and specialty features across demographic groups might be attributable to differential demographic representation in across restricted vs. unrestricted line occupations, and require further investigation.

We then investigate which particular subspecialty codes may account for the outsized role of officer subspecialty. Aggregating subspecialties to the 2-digit level (23 unique categories), we find that approximately half of subspecialty codes are populated almost exclusively by White males, while the other codes are demographically integrated to various degrees. The root cause of this separation is beyond the scope of this project. Some hypotheses include officers' personal preferences, and institutional barriers (e.g., historical obstacles to females in various occupations). The confluence of cultural expectations for what constitutes a promotion-enabling career trajectory with restricted vs. unrestricted line status may also explain the importance of integrated subspecialties in these models. Among integrated subspecialties, *regional security studies*, *information sciences*, *oceanography sciences*, and, to some extent, *engineering* disciplines (e.g., aeronautical, mechanical) increase the predicted likelihood of promotion for females and racial/ethnic minorities compared to White males. Other integrated subspecialties (e.g., systems engineering, nuclear engineering) are especially beneficial for promotion outcomes among White males. As a result, the relative importance of officer subspecialty for females and racial/ethnic minorities is driven entirely by a handful of select subspecialties.

B. Interpretation Caveats

Several important caveats apply to these findings. First and foremost, the relationships we describe are correlational, not causal. Machine learning is a powerful tool that can unearth complex correlations in data, but causality can only be identified when a defensible causal framework exists. Despite the quality and breadth of the administrative data used in this research, this analysis lacks a causal framework and thus cannot measure or substantiate cause:effect relationships. Absent a causal framework, predictive models like those used here should be viewed as forecast and hypothesis generators. Second, feature effects on the predicted outcome depend on the service year and forecast lead length under consideration. Throughout this paper, we focus on Navy officers in their tenth year of service, forecasting retention and promotion four years into the future. Correspondingly, feature explanations pertain to mid-career officers who are weighing the cost and benefits of completing a full Navy career—including personal expectations of potential promotion to O-5. In other research, we find the set of most important features differ somewhat at earlier points in the career path, suggesting an evolution in what characteristics influence retention and promotion outcomes over the career.

C. Avenues for Future Investigation

This project raises many questions for future investigation; we describe some here.

Care should be taken to better understand the role of restricted vs. unrestricted line occupation status and of transfer to non-line occupations in influencing retention and promotion outcomes. This analysis treated those exiting line occupations as leaving the analysis set. Further research could apply a competing risks approach (like that now available in FIFE 1.3.4) to provide greater insight around these transitions and their retention impact. Additional effort also is needed to understand the relatively greater importance of certain subspecialty and additional officer qualifier designations in minority retention. What is unique about the particular subspecialties driving this effect?

The importance of *Assigned Unit Identification Code* for the retention of Black males is particularly interesting, and suggests a need to understand whether and how time-invariant, unit-specific characteristics affect retention and promotion likelihoods for this group. These unit effects may be positive or negative. Further research is needed to identify and understand the nature and root causes of trends observed here.

Much attention has been given to the home life of female service members under the presumption that family choices such as marriage and childbearing greatly influence their career choices and outcomes. Our finding that number of dependents does not affect female retention more than White male retention among the officers studied suggests that childbearing may not be the driving force behind female attrition, as some have postulated. More research is clearly needed to better understand the impacts of family life on military service choices and outcomes for both male and female service members, especially as service and societal norms around parenting roles, career aspirations, and occupational and workforce participation continue to evolve.

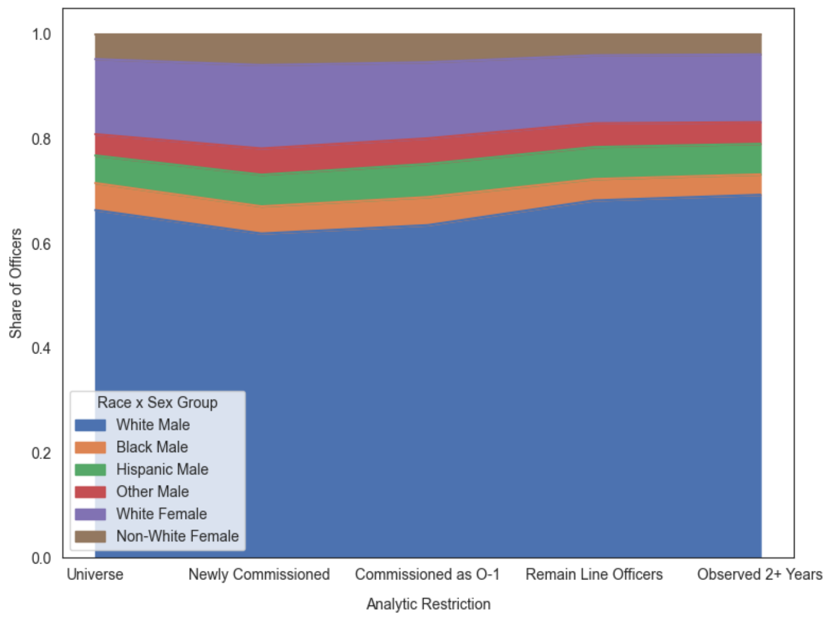
Further work also is needed to understand what drives the importance of other personal and lifestyle characteristics like citizenship origin and religious denomination. This may illuminate cultural trends that could aid efforts to improve retention among service members who do not affiliate with these cultural subgroups or personal origins, or have implications for the broader matter of how representative military service members are of American society along these dimensions.

The analyses and findings presented here only consider a subset of Navy personnel, and only part the career path. Many questions relate to other individuals and conditions not studied here. How might the experience of enlisted members differ? What features matter most for officer retention and promotion at higher grades? How has what matters for retention and promotion evolved over successive generations of Navy service members?

Finally, all findings presented here are correlational, not causal. Moving beyond hypothesis generation and identifying the cause:effect relationships undergirding trends identified here, careful research must identify and exploit experimental or quasi-experimental variation. Many trends identified here are worthy of this level of exploration.

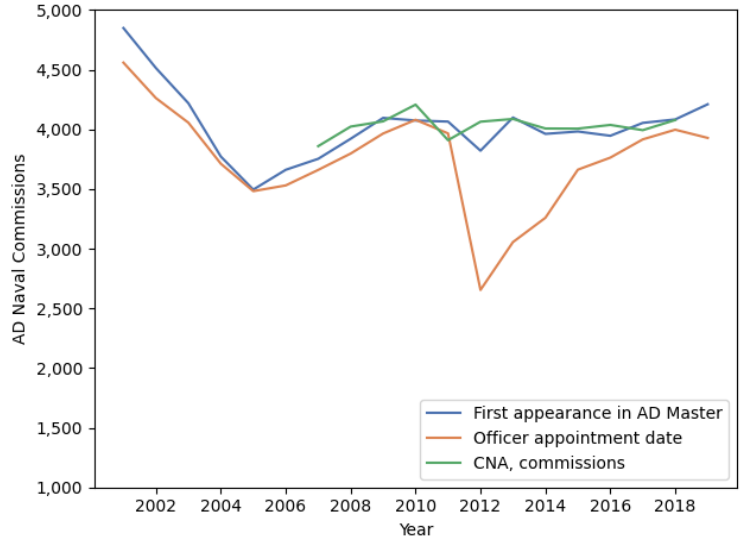
Appendix A.
Supplementary Figures

Figures in this section augment the primary results described in the main text.



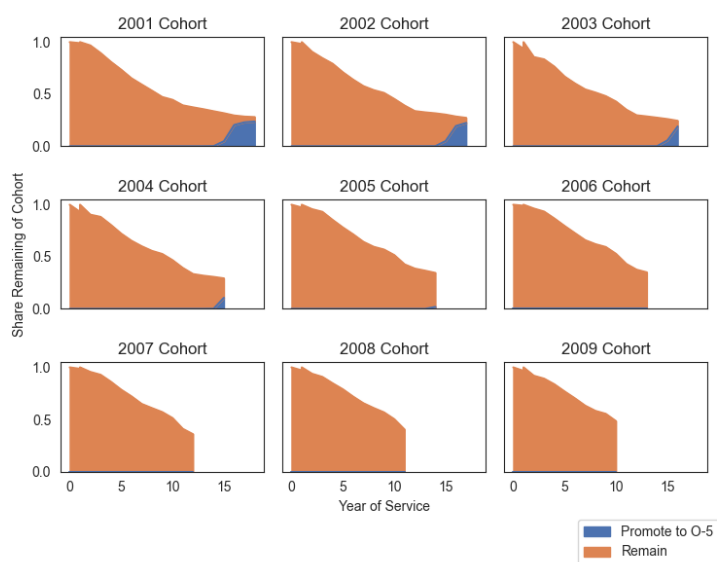
Note: DMDC annual data, 2001-2019.

Figure A-1. Relative Share of each Demographic Across Analytic Restrictions



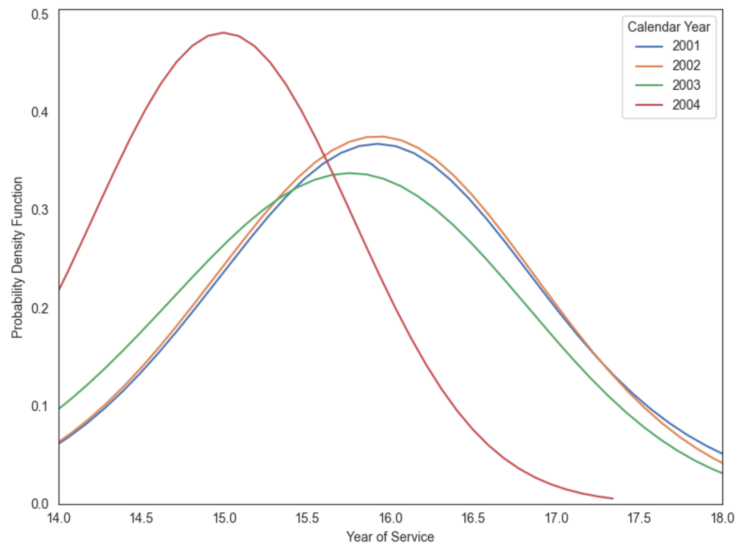
Note: DMDC annual (December) data, 2001-2019. Officer appointment date measured according to "OFF_APNT_DT", a variable contained in the Active Duty Master file. CNA commissions from Population Representation Reports, Appendix B, Table B-23 various years.

Figure A-2. Identifying Commissioning Year among AD Navy Officers, Comparing Methods



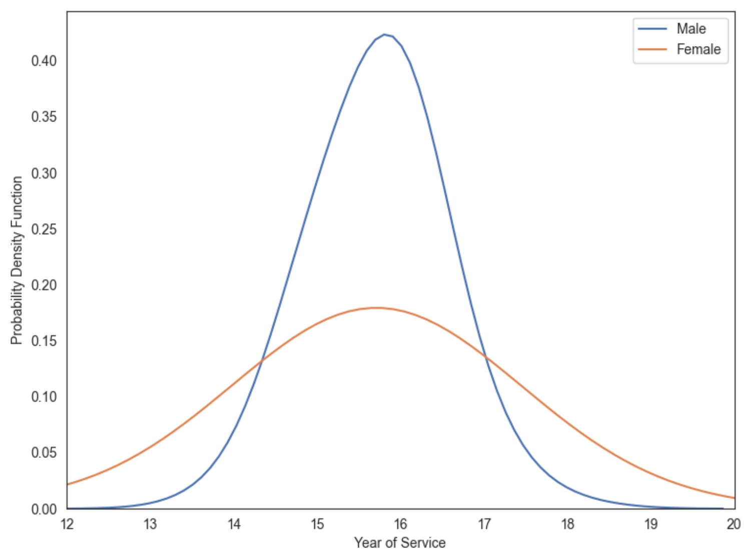
Note: Active Duty Navy officers commissioned as O-1s in 2001-2009 who began their career and remain as an officer of the line. DMDC annual data, 2001-2019. Commissioning cohort based on first year of appearance in AD Master file.

Figure A-3. Share Promoted to O-5 by Commissioning Cohort



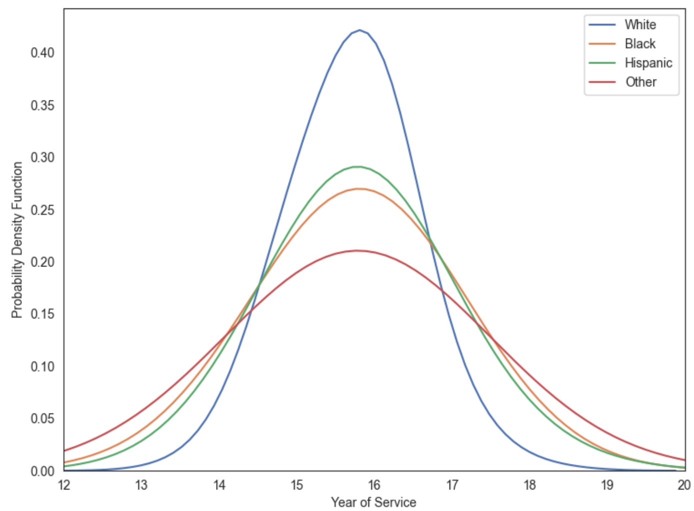
Note: Active Duty Navy officers commissioned as O-1s in 2001 or later who began their career and remain as an officer of the line. DMDC annual data, 2001-2019. Restricted to individuals observed to have promoted to O-5.

Figure A-4. Kernel Density Estimate of Observed Promotion Duration to O-5 among CY01-04 Commissioning Cohorts



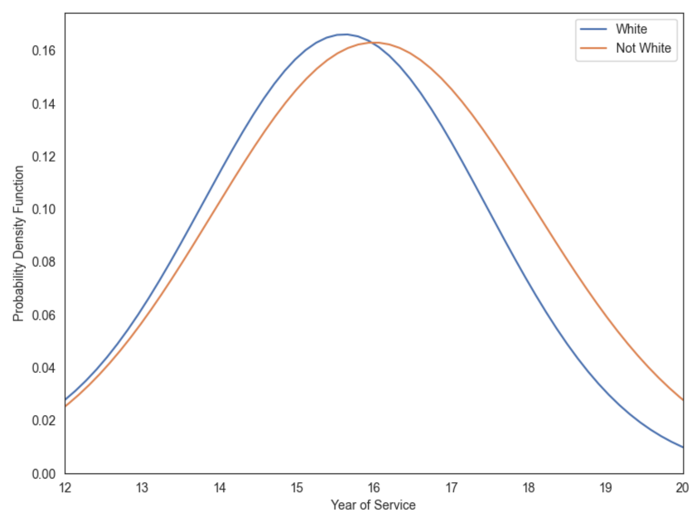
Note: Active Duty Navy officers commissioned as O-1s in 2001 or later who began their career and remain as an officer of the line. DMDC annual data, 2001-2019. Restricted to individuals observed to have promoted to O-5.

Figure A-5. Kernel Density Estimate of Observed Promotion Duration to O-5, Males vs. Females



Note: Active Duty Navy officers commissioned as O-1s in 2001 or later who began their career and remain as an officer of the line. DMDC annual data, 2001-2019. Restricted to individuals observed to have promoted to O-5.

Figure A-6. Kernel Density Estimate of Observed Promotion Duration to O-5 by Race, Males



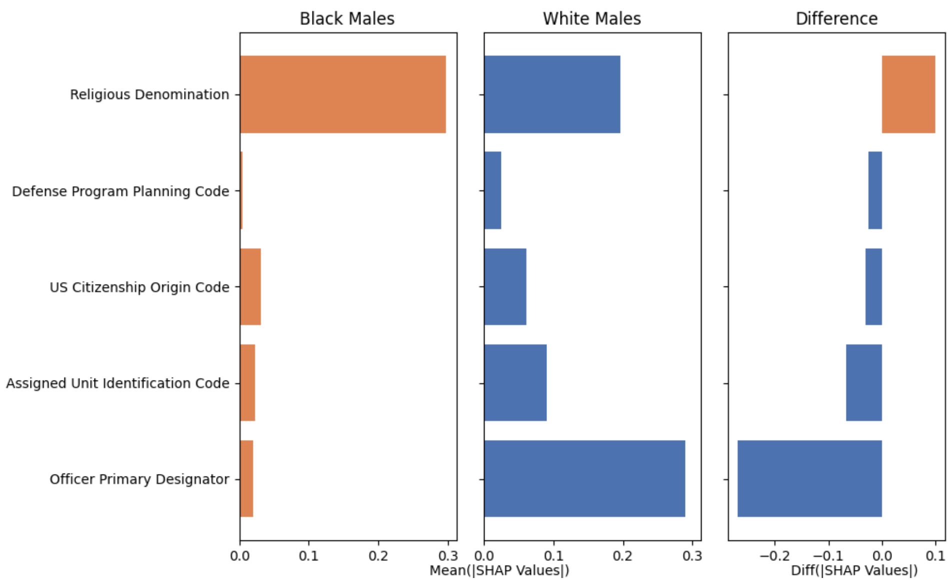
Note: Active Duty Navy officers commissioned as O-1s in 2001 or later who began their career and remain as an officer of the line. DMDC annual data, 2001-2019. Restricted to individuals observed to have promoted to O-5.

Figure A-7. Kernel Density Estimate of Observed Promotion Duration to O-5 by Race, Females

Appendix B.
Most Consequential Features for Retention at the Beginning of Navy Service

The following figures depict the most consequential predictors of retention six years into the future from the start of newly-commissioned O-1 line officers' Navy service careers.

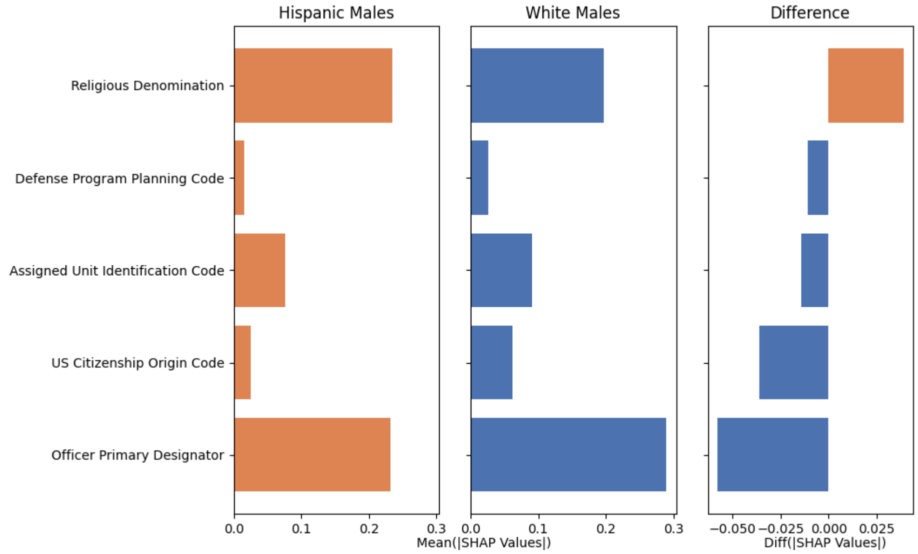
Black males



Note: Active Duty Navy officers commissioned as O-1s in 2001-2018 who began their career and remain as an officer of the line. DMDC annual data, 2001-2019. Top features according to mean of absolute SHAP values and their absolute difference. Orange bars in right subplot represent greater absolute feature importance for the minority group; blue bars indicate greater absolute feature importance for White males.

Figure B-1. Top Features Predicting Retention among Black Males and White Males, Service Year 0, Forecast Lead 6 Years

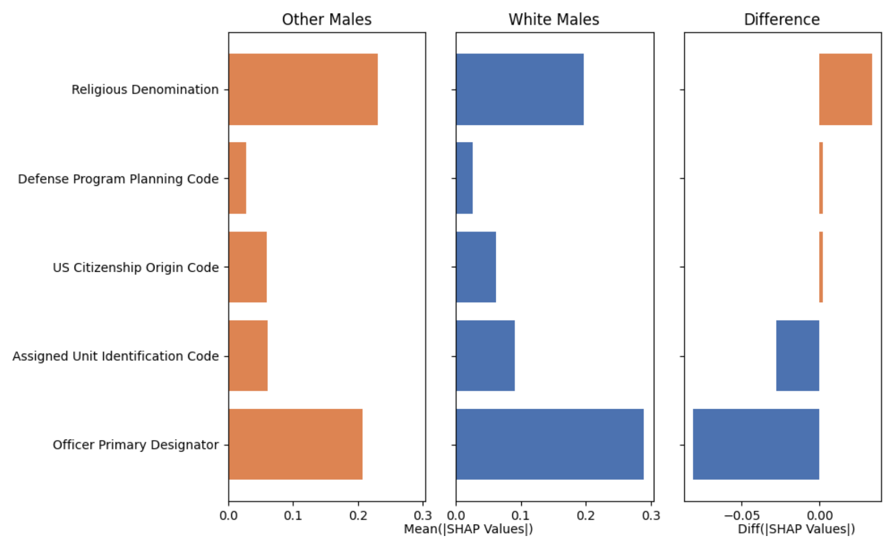
Hispanic males



Note: Active Duty Navy officers commissioned as O-1s in 2001-2018 that began their career and remain as an officer of the line. DMDC annual data, 2001-2019. Top features according to mean of absolute SHAP values and their absolute difference. Orange bars in right subplot represent greater absolute feature importance for the minority group; blue bars indicate greater absolute feature importance for White males.

Figure B-2. Top Features Predicting Retention among Hispanic Males and White Males, Service Year 0, Forecast Lead 6 Years

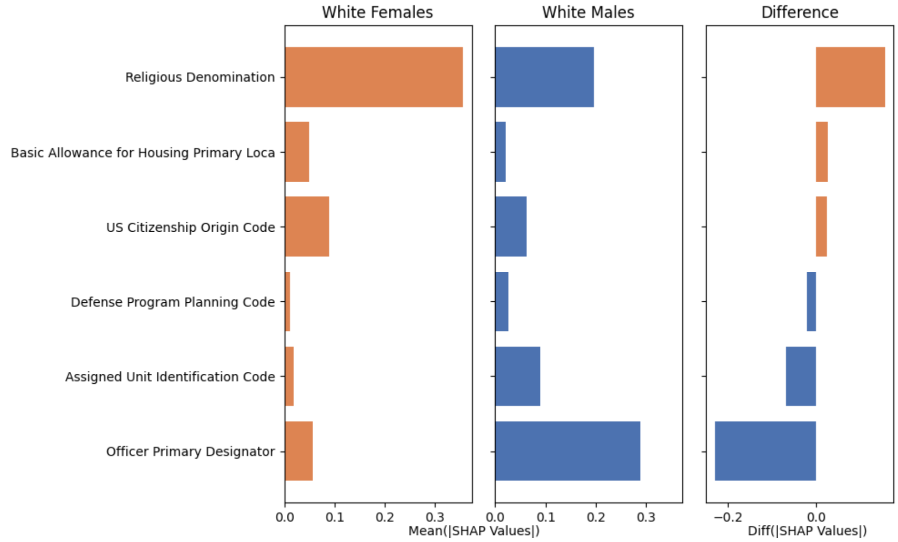
Other males



Note: Active Duty Navy officers commissioned as O-1s in 2001-2018 who began their career and remain as an officer of the line. DMDC annual data, 2001-2019. Top features according to mean of absolute SHAP values and their absolute difference. Orange bars in right subplot represent greater absolute feature importance for the minority group; blue bars indicate greater absolute feature importance for White males.

Figure B-3. Top Features Predicting Retention among Other Males and White Males Service Year 0, Forecast Lead 6 Years

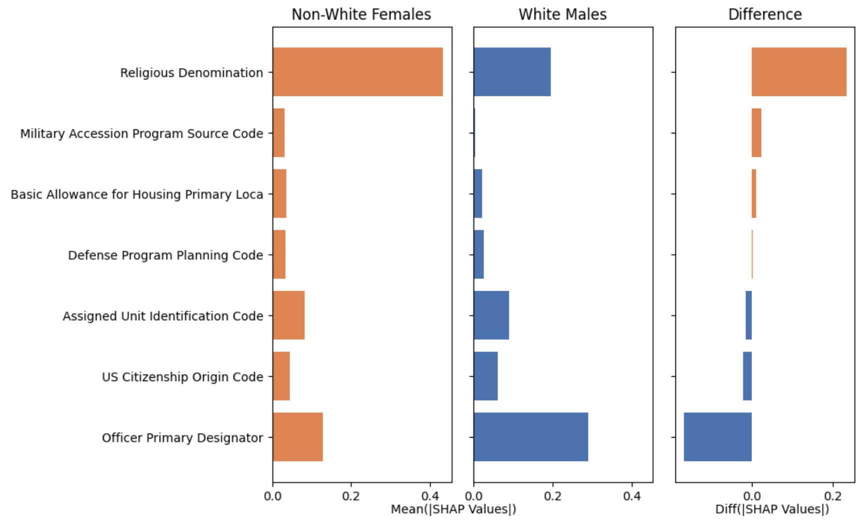
White females



Note: Active Duty Navy officers commissioned as O-1s in 2001-2018 who began their career and remain as an officer of the line. DMDC annual data, 2001-2019. Top features according to mean of absolute SHAP values and their absolute difference. Orange bars in right subplot represent greater absolute feature importance for the minority group; blue bars indicate greater absolute feature importance for White males.

Figure B-4. Top Features Predicting Retention among White Females and White Males, Service Year 0, Forecast Lead 6 Years

Non-White females



Note: Active Duty Navy officers commissioned as O-1s in 2001-2018 who began their career and remain as an officer of the line. DMDC annual data, 2001-2019. Top features according to mean of absolute SHAP values and their absolute difference. Orange bars in right subplot represent greater absolute feature importance for the minority group; blue bars indicate greater absolute feature importance for White males.

Figure B-5. Top Features Predicting Retention among Non-White Females and White Males, Service Year 0, Forecast Lead 6 Years

This page is intentionally blank.

Appendix C.
Features Included in ML Models

Table C-1. Features Included in ML Models

Feature Name	Feature Description	Feature Type	Categorical Type	Source Dataset
BONUS_1_ORIG_ENTITLEMENT_AMOUNT	Bonus 1 Original Entitlement Amount	continuous		Active Duty Pay
BONUS_1_ORIG_START_DATE	Bonus 1 Original Start Date	date		Active Duty Pay
BONUS_1_PAID_CURR_MONTH_AMOUNT	Bonus 1 Paid Current Month Amount	continuous		Active Duty Pay
BONUS_1_PAID_TO_DATE_AMOUNT	Bonus 1 Paid to Date Amount	continuous		Active Duty Pay
BONUS_1_PAYMENT_ELECTION_CODE	Bonus 1 Payment Election Code	categorical	nominal	Active Duty Pay
BONUS_1_TYPE_CODE	Bonus 1 Type Code	categorical	nominal	Active Duty Pay
BONUS_2_ORIG_ENTITLEMENT_AMOUNT	Bonus 2 Original Entitlement Amount	continuous		Active Duty Pay
BONUS_2_ORIG_START_DATE	Bonus 2 Original Start Date	date		Active Duty Pay
BONUS_2_PAID_CURR_MONTH_AMOUNT	Bonus 2 Paid Current Month Amount	continuous		Active Duty Pay
BONUS_2_PAID_TO_DATE_AMOUNT	Bonus 2 Paid to Date Amount	continuous		Active Duty Pay
BONUS_2_PAYMENT_ELECTION_CODE	Bonus 2 Payment Election Code	categorical	nominal	Active Duty Pay
BONUS_2_TYPE_CODE	Bonus 2 Type Code	categorical	nominal	Active Duty Pay
BONUS_3_ORIG_ENTITLEMENT_AMOUNT	Bonus 3 Original Entitlement Amount	continuous		Active Duty Pay
BONUS_3_PAID_CURR_MONTH_AMOUNT	Bonus 3 Paid Current Month Amount	continuous		Active Duty Pay
BONUS_3_PAID_TO_DATE_AMOUNT	Bonus 3 Paid to Date Amount	continuous		Active Duty Pay
CZTE_ENTITLEMENT_DURING_HOSP	Combat Zone Tax Exclusion Entitlement During Hospitalization Indicator Code	categorical	nominal	Active Duty Pay
EITC_CURR_MO_AMT	Earned Income Tax Credit Current Month Amount	continuous		Active Duty Pay
FSGLI_COVERAGE_AMOUNT	Family Service Members Group Life Insurance Coverage Amount	continuous		Active Duty Pay

Feature Name	Feature Description	Feature Type	Categorical Type	Source Dataset
FSGLI_COVERAGE_DEDUCATION_AMOUNT	Family Service members Group Life Insurance Coverage Deduction Amount	continuous		Active Duty Pay
FSGLI_COVERAGE_EFFECTIVE_DATE	Family Service members Group Life Insurance Coverage Effective Date	date		Active Duty Pay
FSGLI_DECLINATION_IND_CODE	Family Service members Group Life Insurance Declination Indicator Code	categorical	nominal	Active Duty Pay
FSSA_PD_CURR_MO_AMT	Family Supplemental Subsistence Allowance Paid Current Month Amount	continuous		Active Duty Pay
FICA_TAX_WITHHELD_CURR_MO_AMT	Federal Insurance Compensation Act Tax Withheld Current Month Amount	continuous		Active Duty Pay
FICA_TAX_WITHHELD_YTD_AMOUNT	Federal Insurance Compensation Act Tax Withheld Year To Date Amount	continuous		Active Duty Pay
FICA_WAGES_PAID_CURR_MONTH_AMT	Federal Insurance Compensation Act Wages Paid Current Month Amount	continuous		Active Duty Pay
FICA_WAGES_YTD_AMOUNT	Federal Insurance Compensation Act Wages Paid Year To Date Amount	continuous		Active Duty Pay
HIGH_DEPLOYMENT_DAY_QUANTITY	High Deployment Day Quantity	continuous		Active Duty Pay
HIGH_DEPLOYMENT_PAY_AMOUNT	High Deployment Pay Amount	continuous		Active Duty Pay
MEDICARE_TAX_WITHH_CURR_MO_AMT	Medicare Tax Withheld Current Month Amount	continuous		Active Duty Pay
MEDICARE_TAX_WITHH_YTD_AMOUNT	Medicare Tax Withheld Year To Date Amount	continuous		Active Duty Pay
MEDICARE_TAX_WAGESPD_CURR_MO_AMT	Medicare Taxable Wages Paid Current Month Amount	continuous		Active Duty Pay
MEDICARE_TAX_WAGESPD_YTD_AMOUNT	Medicare Taxable Wages Paid Year To Date Amount	continuous		Active Duty Pay
MGIB_ADDTL_CONTRIB_PD_CURRMO_AMT	Montgomery GI Bill Additional Contribution Paid Current Month Amount	continuous		Active Duty Pay

Feature Name	Feature Description	Feature Type	Categorical Type	Source Dataset
MGIB_CONTRIB_CUMULATIVE_AMOUNT	Montgomery GI Bill Contribution Cumulative Amount	continuous		Active Duty Pay
MGIB_CONTRIB_PD_CURR_MONTH_AMT	Montgomery GI Bill Contribution Paid Current Month Amount	continuous		Active Duty Pay
OCOLA_2_AMOUNT	Overseas Cost of Living Allowance 2 Amount	continuous		Active Duty Pay
OCOLA_2_BARRACKS_PD_CURR_MO_AMT	Overseas Cost of Living Allowance 2 Barracks Paid Current Month Amount	continuous		Active Duty Pay
OCOLA_BARRACKS_PD_CURR_MO_AMT	Overseas Cost of Living Allowance Barracks Paid Current Month Amount	continuous		Active Duty Pay
OCOLA_DEPENDENT_QTY	Overseas Cost of Living Allowance Dependent Quantity	continuous		Active Duty Pay
OCOLA_LOCATION_CODE	Overseas Cost of Living Allowance Location Code	categorical	nominal	Active Duty Pay
OCOLA_UNIQ_LUMP_SUM_PD_CURR_AMT	Overseas Cost of Living Allowance Unique Lump Sum Paid Current Amount	continuous		Active Duty Pay
ROTC_BOOKS_FEES_PD_CURRMO_AMT	Reserve Officer Training Corps Books and Fees Paid Current Month Amount	continuous		Active Duty Pay
ROTC_STIPEND_PAY_AMOUNT	Reserve Officer Training Corps Stipend Payment Amount	continuous		Active Duty Pay
ROTC_SUBSISTENCE_PD_CURR_MO_AMT	Reserve Officer Training Corps Subsistence Paid Current Month Amount	continuous		Active Duty Pay
ROTC_SUMMER_TRAIN_ENCAMP_PAY_AMT	Reserve Officer Training Corps Summer Training Encampment Payment Amount	continuous		Active Duty Pay
ROTC_UNIFORM_COMMUTATION_AMOUNT	Reserve Officer Training Corps Uniform Commutation Amount	continuous		Active Duty Pay
SPECIAL_PAY_1_ORIGINAL_START_DT	Special Pay 1 Original Start Date	date		Active Duty Pay
SPECIAL_PAY_1_PAID_CURR_MO_AMT	Special Pay 1 Paid Current Month Amount	continuous		Active Duty Pay
SPECIAL_PAY_1_STOP_DATE	Special Pay 1 Stop Date	date		Active Duty Pay
SPECIAL_PAY_1_TYPE_CODE	Special Pay 1 Type Code	categorical	nominal	Active Duty Pay
SPECIAL_PAY_2_ORIGINAL_START_DT	Special Pay 2 Original Start Date	date		Active Duty Pay

Feature Name	Feature Description	Feature Type	Categorical Type	Source Dataset
SPECIAL_PAY_2_PAID_CURR_MO_AMT	Special Pay 2 Paid Current Month Amount	continuous		Active Duty Pay
SPECIAL_PAY_2_STOP_DATE	Special Pay 2 Stop Date	date		Active Duty Pay
SPECIAL_PAY_2_TYPE_CODE	Special Pay 2 Type Code	categorical	nominal	Active Duty Pay
SPECIAL_PAY_3_ORIGINAL_START_DT	Special Pay 3 Original Start Date	date		Active Duty Pay
SPECIAL_PAY_3_PAID_CURR_MO_AMT	Special Pay 3 Paid Current Month Amount	continuous		Active Duty Pay
SPECIAL_PAY_3_TYPE_CODE	Special Pay 3 Type Code	categorical	nominal	Active Duty Pay
SPECIAL_PAY_4_PAID_CURR_MO_AMT	Special Pay 4 Paid Current Month Amount	continuous		Active Duty Pay
TSP_BONUS_PAY_CONTRIB_CURRMO_AMT	Thrift Savings Plan Bonus Pay Contribution Current Month Amount	continuous		Active Duty Pay
TSP_BONUS_PAY_CONTRIB_YTD_AMOUNT	Thrift Savings Plan Bonus Pay Contribution Year To Date Amount	continuous		Active Duty Pay
TSP_CATCHUP_CONTRIB_AMOUNT	Thrift Savings Plan Catch-up Contribution Amount	continuous		Active Duty Pay
TSP_CATCHUP_CONTRIB_CURR_MO_AMT	Thrift Savings Plan Catch-up Contribution Current Month Amount	continuous		Active Duty Pay
TSP_CONTRIB_AMOUNT	Thrift Savings Plan Contribution Amount	continuous		Active Duty Pay
TSP_CONTRIB_YTD_AMOUNT	Thrift Savings Plan Contribution Year To Date Amount	continuous		Active Duty Pay
TSP_GOVT_MATCH_CONTRIB_PPD_AMT	Thrift Savings Plan Government Match Contribution Pay Period Amount	continuous		Active Duty Pay
TSP_INCNT_PAY_CONTRIB_CURRMO_AMT	Thrift Savings Plan Incentive Pay Contribution Current Month Amount	continuous		Active Duty Pay
TSP_INCNT_PAY_CONTRIB_YTD_AMOUNT	Thrift Savings Plan Incentive Pay Contribution Year To Date Amount	continuous		Active Duty Pay
TSP_SPEC_PAY_CONTRIB_CURR_MO_AMT	Thrift Savings Plan Special Pay Contribution Current Month Amount	continuous		Active Duty Pay
TSP_SPEC_PAY_CONTRIB_YTD_AMOUNT	Thrift Savings Plan Special Pay Contribution Year To Date Amount	continuous		Active Duty Pay
TOTAL_ENTITLEMENTS_PD_CURRMO_AMT	Total Entitlements Paid Current Month Amount	continuous		Active Duty Pay

Feature Name	Feature Description	Feature Type	Categorical Type	Source Dataset
TOTAL_ENTITLEMENTS_PD_YTD_AMT	Total Entitlements Paid Year To Date Amount	continuous		Active Duty Pay
TOTAL_FLPB_CURR_AMOUNT	Total Foreign Language Proficiency Bonus Current Amount	continuous		Active Duty Pay
TOTAL_NEG_ALLOWANCE_CURR_AMT	Total Negative Allowance Current Amount	continuous		Active Duty Pay
TOTAL_NEG_BASIC_DRILLPAY_CURRAMT	Total Negative Basic and Drill Pay Current Amount	continuous		Active Duty Pay
TOTAL_NEG_BONUS_PAY_CURR_AMT	Total Negative Bonus Pay Current Amount	continuous		Active Duty Pay
TOTAL_NEGATIVE_COMP_CURR_MO_AMT	Total Negative Compensation Current Month Amount	continuous		Active Duty Pay
TOTAL_NEG_SPEC_PAY_CURR_MO_AMT	Total Negative Special Pay Current Amount	continuous		Active Duty Pay
TOTAL_POS_ALLOWANCE_CURR_MO_AMT	Total Positive Allowance Current Month Amount	continuous		Active Duty Pay
TOTAL_POS_BASIC_DRILLPAY_CURRAMT	Total Positive Basic and Drill Pay Current Amount	continuous		Active Duty Pay
TOTAL_POS_BONUS_PAY_CURR_MO_AMT	Total Positive Bonus Pay Current Month Amount	continuous		Active Duty Pay
TOTAL_POS_COMP_CURR_MO_AMOUNT	Total Positive Compensation Current Month Amount	continuous		Active Duty Pay
TOTAL_POS_SPEC_PAY_CURR_MO_AMT	Total Positive Special Pay Current Month Amount	continuous		Active Duty Pay
ACCRUED_LEAVE_PAY_AMOUNT	Accrued Leave Pay Amount	continuous		Active Duty Pay
ACT_DUTY_SEPARATION_DATE	Active Duty Separation Date	date		Active Duty Pay
ACT_DUTY_SVC_PROJ_END_DATE	Active Duty Service Projected End Date	date		Active Duty Pay
STR_ACCT_CD	Active Duty Strength Accounting Code	categorical	nominal	Active Duty Master
TAFMS_DT	Active Federal Military Service Base Calendar Date	date		Active Duty Master
TAFMS_MN_QY	Active Federal Military Service Months Quantity	continuous		Active Duty Master
TAFMS_YR_QY	Active Federal Military Service Years Quantity	continuous		Active Duty Master

Feature Name	Feature Description	Feature Type	Categorical Type	Source Dataset
depn_nonspouse_oldest_0_26	Age of Oldest Dependent, Non-spouse, Ages 0-26	continuous		Created
depn_nonspouse_youngest_0_26	Age of Youngest Dependent, Non-spouse, Ages 0-26	continuous		Created
ASSGN_UIC_BASE_ID	Assigned Unit Base Id	categorical	nominal	Active Duty Master
ASSGN_UIC_CD	Assigned Unit Identification Code	categorical	nominal	Active Duty Master
ASGND_UNIT_LOC_CNTRY	Assigned Unit Location Country Code	categorical	nominal	Active Duty Pay
ASGND_UNIT_LOC_ZIP_CODE	Assigned Unit Location US Postal Region Zip Code	categorical	nominal	Active Duty Pay
ASULOC_ZIPX_ID	Assigned Unit Location US Postal Region Zip Extension Code	categorical	nominal	Active Duty Master
ASULOC_ST_CD	Assigned Unit Location US State Alpha Code	categorical	nominal	Active Duty Master
ASULOC_CNGS_DIST1_CD	Assigned Unit Location US State Congressional District Code	categorical	nominal	Active Duty Master
ASULOC_CNTY_CD	Assigned Unit Location US State County Code	categorical	nominal	Active Duty Master
ASSGN_UIC_MJR_CMD_CD	Assigned Unit Major Command Code	categorical	nominal	Active Duty Master
AVG_36_MN_BSC_PAYAMT	Average of 36 Highest Month Basic Pay Amount	continuous		Active Duty Pay
AVIATION_CAREER_INCN_PAY_AMT	Aviation Career Incentive Pay Amount	continuous		Active Duty Pay
AVIATION_OFF_CONT_PAY_OE_AMT	Aviation Officer Continuation Pay Original Entitlement Amount	continuous		Active Duty Pay
AVIATION_OFF_CONT_PAY_PTD_AMT	Aviation Officer Continuation Pay Paid to Date Amount	continuous		Active Duty Pay
AVIATION_OFF_CONT_PAY_PD_AMT	Aviation Officer Continuation Pay Period Amount	continuous		Active Duty Pay
AVIATION_SVC_BASE_DATE	Aviation Service Base Date	date		Active Duty Pay
BRKS_COLA_ALLOW_IND_1	Barracks Cost of Living Allowance Indicator Code	categorical	nominal	Active Duty Pay

Feature Name	Feature Description	Feature Type	Categorical Type	Source Dataset
BAH_2_AMOUNT	Basic Allowance for Housing 2 Amount	continuous		Active Duty Pay
BAH_AMOUNT	Basic Allowance for Housing Amount	continuous		Active Duty Pay
BAH_DEPENDENT_TYPE	Basic Allowance for Housing Dependent Type	categorical	nominal	Active Duty Pay
BAQ_AMT	Basic Allowance for Housing Partial Amount	continuous		Active Duty Pay
BAH_ELIGIBILITY_STATUS	Basic Allowance for Housing Primary Location Eligibility Status Code	categorical	nominal	Active Duty Pay
BAH_ZIP_CODE	Basic Allowance for Housing Primary Location US Postal Region Zip Code	categorical	nominal	Active Duty Pay
BAS_AMOUNT	Basic Allowance for Subsistence Amount	continuous		Active Duty Pay
BASIC_PAY_AMOUNT	Basic Pay Amount	continuous		Active Duty Pay
CAREER_SEA_PAYAMT	Career Sea Pay Amount	continuous		Active Duty Pay
CAREER_SEA_PAY_PREMIUM_AMT	Career Sea Pay Premium Amount	continuous		Active Duty Pay
CSBP_DT	Career Status Bonus Program Category Calendar Date	date		Active Duty Master
CSBP_CAT_CD	Career Status Bonus Program Category Code	categorical	nominal	Active Duty Master
CSBP_PTCN_DT	Career Status Bonus Program Participation Calendar Date	date		Active Duty Master
CSBP_TYP_CD	Career Status Bonus Program Type Code	categorical	nominal	Active Duty Master
CLOTH_MONEY_UNIF_EQUIP_ALLOW_AMT	Clothing Monetary Uniformed Equipment Allowance Amount	continuous		Active Duty Pay
COLOC_DEP_QY	Collocated Dependents Quantity	continuous		Active Duty Master
COLOC_DEP_TYP_CD	Collocated Dependents Type Code	categorical	nominal	Active Duty Master
CZTE_COUNTRY_CODE	Combat Zone Tax Exclusion Country Code	categorical	nominal	Active Duty Pay
CZTE_INDICATOR_CODE	Combat Zone Tax Exclusion Indicator Code	categorical	nominal	Active Duty Pay
CMD_SPND_DEP_QY	Command Sponsored Dependents Quantity	continuous		Active Duty Master

Feature Name	Feature Description	Feature Type	Categorical Type	Source Dataset
CMD_STAT_CD	Command Status Code	categorical	nominal	Active Duty Master
CONUS_COLA_AMOUNT	Continental United States Cost of Living Allowance Amount	continuous		Active Duty Pay
CONUS_COLA_BASIS_CODE	Continental United States Cost of Living Allowance Basis Code	categorical	nominal	Active Duty Pay
CONUS_COLA_ZIP_CODE	Continental United States Cost of Living Allowance US Postal Region Zip Code	categorical	nominal	Active Duty Pay
COLA_DEP_QTY_2	Cost of Living Allowance 2 Dependent Quantity	continuous		Active Duty Pay
COLA_DEP_QTY_1	Cost of Living Allowance Dependent Quantity	continuous		Active Duty Pay
US_CITZ_CTRY_ORIG_CD	Country Original Citizenship	categorical	nominal	Active Duty Master
cumsum_dplyd_days	Cumulative Sum of Deployed Days	continuous		Created
cumsum_dplyd_mo	Cumulative Sum of Deployed Months	continuous		Created
CURRENT_MONTH_ALLOWANCE_IND_CD	Current Month Allowance Indicator Code	categorical	nominal	Active Duty Pay
CURRENT_MONTH_PAY_IND_CODE	Current Month Pay Indicator Code	categorical	nominal	Active Duty Pay
unmarried_but_previously_married	Currently Unmarried but Have Been Married	categorical	nominal	Created
days_dep	Days Deployed	continuous		Created
days_af	Days Deployed in Afghanistan	continuous		Created
days_iz	Days Deployed in Iraq	continuous		Created
days_ku	Days Deployed in Kuwait	continuous		Created
days_kg	Days Deployed in Kyrgyzstan	continuous		Created
days_ou	Days Deployed in Missing or Unknown	continuous		Created
days_me	Days Deployed in Other Middle East Countries	continuous		Created
days_qa	Days Deployed in Qatar	continuous		Created
days_fe	Days Deployed in the Far East	continuous		Created
days_us	Days Deployed in US	continuous		Created
PEC_CD	Defense Program Planning Code	categorical	nominal	Active Duty Master

Feature Name	Feature Description	Feature Type	Categorical Type	Source Dataset
DENTAL_PREMIUM_DEDUCTION_AMOUNT	Dental Premium Deduction Amount	continuous		Active Duty Pay
DENTAL_PREMIUM_EFFECTIVE_DATE	Dental Premium Effective Date	date		Active Duty Pay
DEPENDENTS_QY	Total Dependents	continuous		Active Duty Master
deplyd_mo	Deployed in a Month	continuous		Created
DIVING_DUTY_PAYAMT	Diving Duty Pay Amount	continuous		Active Duty Pay
DTY_DOD_OCC_CD	Duty DoD Occupation Code	categorical	nominal	Active Duty Master
DTY_SVC_OCC_CD	Billet Designator Code	categorical	nominal	Active Duty Master
DTY_UIC_BASE_ID	Duty Unit Base Id	categorical	nominal	Active Duty Master
DTY_UIC_CD	Duty Unit Identification Code	categorical	nominal	Active Duty Master
DTULOC_CTRY_CD	Duty Unit Location Country Code	categorical	nominal	Active Duty Master
DTULOC_ZIP_ID	Duty Unit Location US Postal Region Zip Code	categorical	nominal	Active Duty Master
DTULOC_ZIPX_ID	Duty Unit Location US Postal Region Zip Extension Code	categorical	nominal	Active Duty Master
DTULOC_ST_CD	Duty Unit Location US State Alpha Code	categorical	nominal	Active Duty Master
DTULOC_CNTY_CD	Duty Unit Location US State Congressional District Code	categorical	nominal	Active Duty Master
DTULOC_CNGS_DIST1_CD	Duty Unit Location US State County Code	categorical	nominal	Active Duty Master
DTY_UIC_MJR_CMD_CD	Duty Unit Major Command Code	categorical	nominal	Active Duty Master
ASSGN_UIC_NV_ASHR_AFLT_CD	Duty Unit Navy Ashore Afloat Code	categorical	nominal	Active Duty Master
EITC_YTD_AMOUNT	Earned Income Tax Credit Year To Date Amount	continuous		Active Duty Pay
EDU_LVL_CD	Education Level Code	categorical	ordinal	Active Duty Master

Feature Name	Feature Description	Feature Type	Categorical Type	Source Dataset
EDU_DSCPL_CD	Educational Discipline Code	categorical	nominal	Active Duty Master
EDU_TIER_CD	Educational Tier Code	categorical	ordinal	Active Duty Master
ENL_AGMT_DRTN_YR_QY	Enlisted Active Service Agreement Duration Years Quantity	continuous		Active Duty Master
FAITH_GRP_CD	Religious Denomination	categorical	nominal	Active Duty Master
FSA_AMOUNT	Family Separation Allowance Amount	continuous		Active Duty Pay
FED_TAX_WITHHELD_CURR_MO_AMOUNT	Federal Tax Withheld Current Month Amount	continuous		Active Duty Pay
FED_TAX_WITHHELD_YTD_AMOUNT	Federal Tax Withheld Year To Date Amount	continuous		Active Duty Pay
FED_TAX_WITHHOLD_ALLOWANCE_QTY	Federal Tax Withholding Allowance Quantity	continuous		Active Duty Pay
FED_TAX_WITHHOLD_MARITAL_STAT	Federal Tax Withholding Marital Status Code	categorical	nominal	Active Duty Pay
FED_TAX_WAGES_PAID_CURR_MO_AMT	Federal Taxable Wages Paid Current Month Amount	continuous		Active Duty Pay
FED_TAX_WAGES_PAID_YTD_AMT	Federal Taxable Wages Paid Year To Date Amount	continuous		Active Duty Pay
FILE_DATE	Year (continuous, linear)	date		Active Duty Master
FGN_LANGUAGE_1_PROF_PAY_EFF_DT	Foreign Language 1 Proficiency Pay Effective Date	date		Active Duty Pay
HAZ_DUTY_INC_PAY_1_AMT	Hazardous Duty Incentive Pay 1 Amount	continuous		Active Duty Pay
HAZ_DUTY_INC_PAY_1_TYPE	Hazardous Duty Incentive Pay 1 Type Code	categorical	nominal	Active Duty Pay
HAZ_DUTY_INC_PAY_2_AMT	Hazardous Duty Incentive Pay 2 Amount	continuous		Active Duty Pay
HAZ_DUTY_INC_PAY_2_TYPE	Hazardous Duty Incentive Pay 2 Type Code	categorical	nominal	Active Duty Pay
HAZ_DUTY_INC_PAY_3_AMT	Hazardous Duty Incentive Pay 3 Amount	continuous		Active Duty Pay
HAZ_DUTY_INC_PAY_3_TYPE	Hazardous Duty Incentive Pay 3 Type Code	categorical	nominal	Active Duty Pay
HLTH_PROF_BOARD_CERT_SPAY_AMT	Health Professional Board Certified Special Pay Amount	continuous		Active Duty Pay
HLTH_PROF_SAVED_PAY_AMT	Health Professional Saved Pay Amount	continuous		Active Duty Pay

Feature Name	Feature Description	Feature Type	Categorical Type	Source Dataset
HOR_CTRY_CD	Home of Record Country Code	categorical	nominal	Active Duty Master
HOR_ST_CD	Home of Record US State Alpha Code	categorical	nominal	Active Duty Master
HOSTILE_FIRE_IMM_DANGR_PAYAMT	Hostile Fire Imminent Danger Pay Amount	continuous		Active Duty Pay
INADEQUATE_GOVT_QTRS_RENTAL_AMT	Inadequate Government Quarters Rental Amount	continuous		Active Duty Pay
JPME_CMPL_DT	Joint Professional Military Education Effective Calendar Date	date		Active Duty Master
JPME_LVL_CD	Joint Professional Military Education Level Code	categorical	nominal	Active Duty Master
LEGAL_RESIDENCE_STATE_CODE	Legal Residence US State Code	categorical	nominal	Active Duty Pay
LOST_TIME_DAYS_QTY	Lost Time Day Quantity	continuous		Active Duty Pay
MA_CTRY_CD	Mailing Address Country Code	categorical	nominal	Active Duty Master
MA_ST_CD	Mailing Address US Postal Region State Code	categorical	nominal	Active Duty Master
MA_ZIPX_ID	Mailing Address US Postal Region ZIP Extension Identifier	categorical	nominal	Active Duty Master
MA_ZIP_ID	Mailing Address US Postal Region ZIP Code	categorical	nominal	Active Duty Master
MA_CNTY_CD	Mailing Address US State County Code	categorical	nominal	Active Duty Master
married	Marital Status Code	categorical	nominal	Created
married_with_children	Married with Children	categorical	nominal	Created
ACC_SRC_CD	Military Accession Program Source Code	categorical	nominal	Active Duty Master
MIL_AERO_RTG_CD	Military Aeronautical Rating Code	categorical	nominal	Active Duty Master
OCC_CRER_GRP_CD	Military Career Category Code	categorical	nominal	Active Duty Master
PEBD_DT	Military Longevity Pay Service Base Calendar Date	date		Active Duty Master

Feature Name	Feature Description	Feature Type	Categorical Type	Source Dataset
PEBD_YR_QY	Military Longevity Pay Service Years Quantity	continuous		Active Duty Master
MILITARY_PAY_STATUS_CODE	Military Pay Status Code	categorical	nominal	Active Duty Pay
MGIB_ACT_DUTY_CONTRIB_STATUS_CD	Montgomery GI Bill Active Duty Contribution Status Code	categorical	nominal	Active Duty Pay
MGIB_ADDTL_CONTRIB_CUM_AMT	Montgomery GI Bill Additional Contribution Cumulative Amount	continuous		Active Duty Pay
months_since_new_nonspouse_nonchild	Months Since Having a New Non-spouse, Non-Child Dependent	continuous		Created
months_since_first_child	Months Since the Birth of an Individual's First Child	continuous		Created
months_since_new_child	Months Since the Birth of an Individual's Most Recent Child	continuous		Created
months_since_unmarried_to_married	Months since the Individual's Marital Status Changed to Married	continuous		Created
months_since_married_to_unmarried	Months since the Individual's Marital Status Changed to Unmarried	continuous		Created
NR_AQD4_CD	4th Additional Qualification Designation	categorical	ordinal	Active Duty Master
NR_AQD2_CD	2nd Additional Qualification Designation	categorical	ordinal	Active Duty Master
NR_AQD3_CD	3rd Additional Qualification Designation	categorical	ordinal	Active Duty Master
never_married	Never Married Status	categorical	nominal	Created
unmarried_with_children	Never Married with Children	categorical	nominal	Created
NUCL_OFF_ACC_BONUS_AMT	Nuclear Officer Accession Bonus Amount	continuous		Active Duty Pay
marital_status_changes	Number of Changes to Marital Status	continuous		Created
depn_nonspouse_num	Number of Non-Spousal Dependents	continuous		Created
depn_nonspouse_num_0_2	Number of Dependents, Non-spouse, Ages 0-2	continuous		Created
depn_nonspouse_num_12_15	Number of Dependents, Non-spouse, Ages 12-15	continuous		Created

Feature Name	Feature Description	Feature Type	Categorical Type	Source Dataset
depn_nonspouse_num_16_18	Number of Dependents, Non-spouse, Ages 16-18	continuous		Created
depn_nonspouse_num_19_26	Number of Dependents, Non-spouse, Ages 19-26	continuous		Created
depn_nonspouse_num_27_plus	Number of Dependents, Non-spouse, Ages 27 and up	continuous		Created
depn_nonspouse_num_3_5	Number of Dependents, Non-spouse, Ages 3-5	continuous		Created
depn_nonspouse_num_6_11	Number of Dependents, Non-spouse, Ages 6-11	continuous		Created
OFF_ASVC_OBLG_END_DT	Date Eligible for Separation or Transfer	date		Active Duty Master
OFF_ACT_STAT_PE_DT	Maximum Date of Eligible Active Duty Status	date		Active Duty Master
OFF_APNT_DT	Officer Appointment Date	date		Active Duty Master
cohort	Officer Appointment Date Calendar Year	categorical	nominal	Created
OFF_HARDSHIP_DUTY_PAY_AMT	Officer Hardship Duty Pay Amount	continuous		Active Duty Pay
OVERSEAS_ALLOWANCE_BASIS_CODE	Overseas Allowance Basis Code	categorical	nominal	Active Duty Pay
OCOLA_ALLOWANCE_AMOUNT	Overseas Cost of Living Allowance Amount	continuous		Active Duty Pay
OHA_2_AIR_COND_UTIL_IND_CODE	Overseas Housing Allowance 2 Air Conditioning Utilities Indicator Code	categorical	nominal	Active Duty Pay
OHA_2_ELEC_UTILITIES_IND_CODE	Overseas Housing Allowance 2 Electric Utilities Indicator Code	categorical	nominal	Active Duty Pay
OHA_2_HEATING_UTIL_IND_CODE	Overseas Housing Allowance 2 Heating Utilities Indicator Code	categorical	nominal	Active Duty Pay
OHA_2_MONTHLY_HOUSING_AMOUNT	Overseas Housing Allowance 2 Monthly Housing Amount	continuous		Active Duty Pay
OHA_2_MIHA_MISC_AMOUNT	Overseas Housing Allowance 2 Move In Housing Allowance Miscellaneous Amount	continuous		Active Duty Pay
OHA_2_MIHA_RENT_AMOUNT	Overseas Housing Allowance 2 Move In Housing Allowance Rent Amount	continuous		Active Duty Pay

Feature Name	Feature Description	Feature Type	Categorical Type	Source Dataset
OHA_2_MIHA_SECURITY_AMOUNT	Overseas Housing Allowance 2 Move In Housing Allowance Security Amount	continuous		Active Duty Pay
OHA_2_TRASH_DISPOSAL_IND_CODE	Overseas Housing Allowance 2 Trash Disposal Indicator Code	categorical	nominal	Active Duty Pay
OHA_2_WATER_OR_SEWER_IND_CODE	Overseas Housing Allowance 2 Water or Sewer Indicator Code	categorical	nominal	Active Duty Pay
OHA_AIR_COND_UTIL_IND_CODE	Overseas Housing Allowance Air Conditioning Utilities Indicator Code	categorical	nominal	Active Duty Pay
OHA_ALLOWANCE_AMOUNT	Overseas Housing Allowance Amount	continuous		Active Duty Pay
OHA_CURRECY_CODE	Overseas Housing Allowance Currency Code	categorical	nominal	Active Duty Pay
OHA_ELEC_UTILITIES_IND_CODE	Overseas Housing Allowance Electric Utilities Indicator Code	categorical	nominal	Active Duty Pay
OHA_HEATING_UTIL_IND_CODE	Overseas Housing Allowance Heating Utilities Indicator Code	categorical	nominal	Active Duty Pay
OHA_HOUSING_STATUS_CODE	Overseas Housing Allowance Housing Status Code	categorical	nominal	Active Duty Pay
OHA_IND_SHARNG_OSEAS_HOUS_QTY	Overseas Housing Allowance Individual Sharing Overseas Housing Quantity	categorical	nominal	Active Duty Pay
OHA_LOCATION_CODE	Overseas Housing Allowance Location Code	categorical	nominal	Active Duty Pay
OHA_MONTHLY_HOUSING_AMOUNT	Overseas Housing Allowance Monthly Housing Amount	continuous		Active Duty Pay
OHA_MIHA_MISC_AMOUNT	Overseas Housing Allowance Move In Housing Allowance Miscellaneous Amount	continuous		Active Duty Pay
OHA_MIHA_RENT_AMOUNT	Overseas Housing Allowance Move In Housing Allowance Rent Amount	continuous		Active Duty Pay
OHA_MIHA_SECURITY_AMOUNT	Overseas Housing Allowance Move In Housing Allowance Security Amount	continuous		Active Duty Pay
OHA_1_SPECIAL_STATUS	Overseas Housing Allowance Special Status Code	categorical	nominal	Active Duty Pay
OHA_TRASH_DISPOSAL_IND_CODE	Overseas Housing Allowance Trash Disposal Indicator Code	categorical	nominal	Active Duty Pay

Feature Name	Feature Description	Feature Type	Categorical Type	Source Dataset
OHA_WATER_OR_SEWER_IND_CODE	Overseas Housing Allowance Water or Sewer Indicator Code	categorical	nominal	Active Duty Pay
PAY_CONT_DURING_HOSP_REHB_IND	Pay Continuation During Hospitalization and Rehabilitation Indicator Code	categorical	nominal	Active Duty Pay
PG_MOD_CD	Pay Grade Modifier Code	categorical	nominal	Active Duty Master
PAY_PLAN_CODE	Pay Plan Code	categorical	ordinal	Active Duty Pay
PG_CD	Pay Plan Grade Identifier	categorical	nominal	Active Duty Master
PAY_PLAN_PAY_GRADE_EFF_DATE	Pay Plan Pay Grade Effective Date	date		Active Duty Pay
PRM_DTY_STN_ARRV_DT	Permanent Duty Station Arrival Calendar Date	date		Active Duty Master
PRM_DTY_STN_DPRT_DT	Permanent Duty Station Departure Calendar Date	date		Active Duty Master
PERM_DUTY_STATION_GOVT_QTRS	Permanent Duty Station Government Quarters Assignment or Adequacy Code	categorical	nominal	Active Duty Pay
AGE_QY	Person Age Quantity	continuous		Active Duty Master
DOB_DT	Person Birth Date	date		Active Duty Master
POB_CNTRY_CD	Person Birth Place Country Code	categorical	nominal	Active Duty Master
POB_ST_CD	Person Birth Place US State Code	categorical	nominal	Active Duty Master
MA_CNGS_DIST1_CD	Person Mailing Address US State Congressional District Code	categorical	nominal	Active Duty Master
PRI_DOD_OCC_CD	Primary DoD Occupation Code	categorical	nominal	Active Duty Master
PRI_SVC_OCC_CD	Officer Primary Designator	categorical	nominal	Active Duty Master
PME_LVL_CD	Professional Military Education Level Code	categorical	nominal	Active Duty Master

Feature Name	Feature Description	Feature Type	Categorical Type	Source Dataset
SEC_DOD_OCC_CD	Secondary DoD Occupation Code	categorical	nominal	Active Duty Master
SEC_SVC_OCC_CD	Officer Subspecialty	categorical	nominal	Active Duty Master
SEPARATION_PAY_AMOUNT	Separation Pay Amount	continuous		Active Duty Pay
SERVICE_COMPONENT_CODE	Service Component Code	categorical	nominal	Active Duty Pay
SGLI_COVERAGE_CURRENT_AMT_EFF_DT	Service members Group Life Insurance Coverage Current Amount Effective Date	date		Active Duty Pay
SGLI_COVERAGE_ELECTED_AMOUNT	Service members Group Life Insurance Coverage Elected Amount	continuous		Active Duty Pay
SGLI_FULL_TIME_DEDUCTION_AMOUNT	Service members Group Life Insurance Full Time Deduction Amount	continuous		Active Duty Pay
SGLI_TRAUMATIC_DEDUCTION_AMOUNT	Service members Group Life Insurance Traumatic Deduction Amount	continuous		Active Duty Pay
SPECIAL_PAY_10_PAID_CURR_MO_AMT	Special Pay 10 Paid Current Month Amount	continuous		Active Duty Pay
SPECIAL_PAY_5_PAID_CURR_MO_AMT	Special Pay 5 Paid Current Month Amount	continuous		Active Duty Pay
SPECIAL_PAY_6_PAID_CURR_MO_AMT	Special Pay 6 Paid Current Month Amount	continuous		Active Duty Pay
SPECIAL_PAY_7_PAID_CURR_MO_AMT	Special Pay 7 Paid Current Month Amount	continuous		Active Duty Pay
SPECIAL_PAY_8_PAID_CURR_MO_AMT	Special Pay 8 Paid Current Month Amount	continuous		Active Duty Pay
SPECIAL_PAY_9_PAID_CURR_MO_AMT	Special Pay 9 Paid Current Month Amount	continuous		Active Duty Pay
spouse_age	Spouse Age	continuous		Created
spouse_educ	Spouse Education	categorical	ordinal	Created
STATE_TAX_WITHHELD_YTD_AMOUNT	State Tax Withheld Year To Date Amount	continuous		Active Duty Pay
STATE_TAX_WAGES_PAID_CURR_MO_AMT	State Taxable Wages Paid Current Month Amount	continuous		Active Duty Pay
STATE_TAX_WAGES_PAID_YTD_AMOUNT	State Taxable Wages Paid Year To Date Amount	continuous		Active Duty Pay
STATE_TAX_WITHHELD_CURR_MO_AMT	State Taxes Withheld Current Month Amount	continuous		Active Duty Pay
COMP_CD	Uniformed Service Organization Component Code	categorical	nominal	Active Duty Master
RANK_DT	Uniformed Service Rank Effective Calendar Date	date		Active Duty Master

Feature Name	Feature Description	Feature Type	Categorical Type	Source Dataset
RANK_MN_QY	Uniformed Service Rank Months Quantity	continuous		Active Duty Master
RANK_YR_QY	Uniformed Service Rank Years Quantity	continuous		Active Duty Master
CITIZ_ORIG_CD	US Citizenship Origin Code	categorical	nominal	Active Duty Master
CITIZ_STATUS_CD	US Citizenship Status Code	categorical	nominal	Active Duty Master
VHA_BASIS_ID_1	Home Residence Type	categorical	nominal	Active Duty Pay
off_apt_yr_month	Officer Appointment Date Year and Month	categorical	ordinal	Created

This page is intentionally blank.

Appendix D.
Illustrations

Tables

Figures

This page is intentionally blank.

Appendix E.
Bibliography

Angelis, Karin de, and David R. Segal. "Minorities in the Military." *The Oxford Handbook of Military Psychology*. Edited by Janice H. Laurence and Michael D. Matthews. Oxford University Press, 2012.

Asch, Beth J., Trey Miller, and Andrea Malchiodi. "A New Look at Gender and Minority Differences in Officer Career Progression in the Military." RAND Corporation, National Defense Research Institute (U.S.), Santa Monica, CA, 2012.

Burk, James, and Evelyn Espinoza. "Race Relations Within the US Military." *Annual Review of Sociology* 38, no. 1 (2012): 401–22. https://doi.org/10.1146/annurev-soc-071811-145501.

Browne, Irene, and Joya Misra. "The Intersection of Gender and Race in the Labor Market." *Annual Review of Sociology* 29, no. 1 (2003): 487–513. https://doi.org/10.1146/annurev.soc.29.010202.100016.

Crenshaw, Kimberle. "Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics." *University of Chicago Legal Forum* 1989, no. 1 (1989): 139–67. https://chicagounbound.uchicago.edu/uclf/vol1989/iss1/8.

"From Representation to Inclusion: Diversity Leadership for the 21st-Century Military (Final Report)." Military Leadership Diversity Commission, 2011.

Hosek, Susan D., Peter Tiemeyer, M. Rebecca Kilburn, Debra A. Strong, Selika Ducksworth, and Reginald Ray. *Minority and Gender Differences in Officer Career Progression.* Monograph Reports. RAND Corporation, 2001. https://www.rand.org/pubs/monograph_reports/MR1184.html.

Lundberg, Scott M., Gabriel G. Erion, and Su-In Lee. "Consistent Individualized Feature Attribution for Tree Ensembles." February 11, 2018. https://arxiv.org/pdf/1802.03888.

Lundberg, Scott M., Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M. Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. "From Local Explanations to Global Understanding with Explainable AI for Trees." *Nature Machine Intelligence* 2, no. 1 (2020): 56–67. https://doi.org/10.1038/s42256-019-0138-9.

This page is intentionally blank.

Appendix F.
Abbreviations

AD	Active Duty
AIAN	American Indian and Alaska Native
DMDC	Defense Manpower Data Center
DoD	Department of Defense
DoN	Department of the Navy
FIFE	Finite-Interval Forecasting Model
IDA	Institute for Defense Analyses
ML	Machine Learning
NHPI	Native Hawaiian and Pacific Islander
OUSD (P&R)	Under Secretary of Defense for Personnel and Readiness
PII Enclave	Personally Identifiable Information Enclave
SHAP	SHapley Additive exPlanations
UIC	Unit Identification Code

This page is intentionally blank.

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

1. REPORT DATE (DD-MM-YY)	2. REPORT TYPE	3. DATES COVERED (From – To)
XX-12-2020	Final	

4. TITLE AND SUBTITLE Explaining Differences in Predicted O-5 Promotion Outcomes by Race/Ethnicity and Sex Among Navy Officers	5a. CONTRACT NO. HQ0034-14-D-0001
	5b. GRANT NO.
	5c. PROGRAM ELEMENT NO(S).
6. AUTHOR(S) Julie Lockwood Joe King Rachel Augustine	5d. PROJECT NO.
	5e. TASK NO. BE-6-4311
	5f. WORK UNIT NO.
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Institute for Defense Analyses 4850 Mark Center Drive Alexandria, VA 22311-1882	8. PERFORMING ORGANIZATION REPORT NO. IDA Paper P-20452 Log: H 20-000476
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) USD(P&R) 1400 Defense Pentagon, Arlington, VA 22202	10. SPONSOR'S / MONITOR'S ACRONYM(S) USD(P&R)
	11. SPONSOR'S / MONITOR'S REPORT NO(S).

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The U.S. Department of the Navy desires to better understand the drivers of differences in officer retention and promotion across demographic groups. This Institute for Defense Analyses research identifies the differences in predicted O-5 promotion outcomes across race/ethnicity and sex groups of Navy line officers commissioned as O-1s in 2001-2018. Using administrative military personnel data, we train two machine learning (ML) models: one predicts retention, the other predicts promotion. We calculate the effect of each feature on an individual's likelihood of retention or promotion, and compare across demographics to identify differences in which and how much features matter. We find that officer subspecialty and, to some extent, qualification designations matter more for female and minority retention than for White males, while officer primary designator consistently matters more for retention among White males. In the promotion model, officer subspecialty is especially influential for female and minority promotion outcomes, while officer primary designator has mixed importance across demographics, suggesting primary designator is less important than subspecialty to promotion. Notably, conditional on serving ten years in the Navy, number of dependents is no more consequential for female retention than for White males, suggesting that childbearing may not drive differential female attrition as some have postulated.

15. SUBJECT TERMS

Race, sex, gender, gender differences, racial differences, attrition, retention, attrition prediction, retention predication, FIFE, RPM, officer retention, Navy, military officers, naval officers, machine learning, predictive analytics.

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NO. OF PAGES	19a. NAME OF RESPONSIBLE PERSON Lernes Hebert
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER (Include Area Code)
U	U	U	U	86	(703) 571-0114

This page is intentionally blank.