PERSEREC

OPA-2018-065
PERSEREC-TR-18-14
August 2018

Modeling Insider Threat From the Inside and Outside: Individual and Environmental Factors Examined Using Event History Analysis

Ray Zimmerman
Gregory Friedman
Drumil Munshi
Northrop Grumman Technology Services

David Richmond
Stephanie Jaros
Defense Personnel and Security Research Center
Office of People Analytics

OPA
OFFICE OF PEOPLE ANALYTICS

Approved for Public Distribution
Defense Personnel and Security Research Center
Office of People Analytics

OPA-2018-065 **PERSEREC-TR-18-14** **August 2018**

Modeling Insider Threat From the Inside and Outside: Individual and Environmental Factors Examined Using Event History Analysis

Ray Zimmerman, Gregory Friedman, Drumil Munshi
Northrop Grumman Technology Services

David Richmond, Stephanie Jaros
Defense Personnel and Security Research Center
Office of People Analytics

Released by – Eric L. Lang

Defense Personnel and Security Research Center
Office of People Analytics
400 Gigling Rd.
Seaside, CA 93955

REPORT DOCUMENTATION PAGE

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

1. REPORT DATE: August 2018	2. REPORT TYPE Technical Report	3. DATES COVERED:

4. Modeling Insider Threat from the Inside and Outside: Individual and Environmental Factors Examined Using Event History Analysis	5a. CONTRACT NUMBER: 5b. GRANT NUMBER: 5c. PROGRAM ELEMENT NUMBER:
6. AUTHOR(S): Ray A. Zimmerman, Ph.D., Gregory M. Friedman, Drumil C. Munshi, David Richmond, Ph.D., Stephanie L. Jaros	5d. PROJECT NUMBER: 5e. TASK NUMBER: 5f. WORK UNIT NUMBER:
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Defense Personnel and Security Research Center Office of People Analytics 400 Gigling Road Seaside, CA 93955	8. PERFORMING ORGANIZATION REPORT NUMBER There are two report numbers: PERSEREC-TR-18-14 2018-065
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSORING/MONITOR'S ACRONYM(S) 11. SPONSORING/MONITOR'S REPORT NUMBER(S):

12. DISTRIBUTION/AVAILABILITY STATEMENT: (A)

13. SUPPLEMENTARY NOTES:

ABSTRACT: Insider threat is an ongoing concern for the Department of Defense (DoD). Within the past 8 years, incidents of violence, such as the Fort Hood and Navy Yard shootings, as well as massive unauthorized disclosure of classified information to Wikileaks by Private Manning have caused serious harm to personnel and national security. In addition to policy and procedural changes and implementing programs for improved physical security and continuous evaluation, there have been several attempts at predictive modeling of insider threat events.

For this study, analyses focused on individual factors (e.g., marital status and education) combined with environmental factors (e.g., the property crime rate in an individual's city) in relation to insider threat outcomes that may precede high-impact events. Results demonstrated that event history analysis is a viable approach to examining insider threat behavior using a combination of individual and environmental factors.

14. SUBJECT TERMS:

15. SECURITY CLASSIFICATION OF: Unclassified			16. LIMITATION OF ABSTRACT:	17. NUMBER OF PAGES: 39	19a. NAME OF RESPONSIBLE PERSON: Eric L. Lang, Director
a. REPORT: Unclassified	b. ABSTRACT: Unclassified	c. THIS PAGE: Unclassified			19b. TELEPHONE NUMBER (Include area code): 831-583-2846

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI td. Z39.18

PREFACE

The Department of Defense (DoD) has an ongoing interest in the topic of Insider Threat. Numerous studies have been commissioned by the DoD which focus on developing mathematical models designed to predict insider threat incidents, based on individual characteristics and events.

The Office of People Analytics (OPA) provided internal funding to the Defense Personnel and Security Research Center (PERSEREC), a division of OPA, to further this line of inquiry by conducting a study to determine whether environmental factors should be included in modeling efforts.

Eric L. Lang
Director, PERSEREC

EXECUTIVE SUMMARY

Insider threat is an ongoing concern for the Department of Defense (DoD). Within the past 8 years, incidents of violence, such as the Fort Hood and Navy Yard shootings, as well as massive unauthorized disclosure of classified information to Wikileaks by Private Manning have caused serious harm to personnel and national security.

Much of the Department's response to high-impact insider threat events has been in the form of policy and procedural changes as well as a commitment to programs that improve physical security and continuous evaluation of cleared personnel by leveraging information technology. There have also been attempts to use advanced analytics to build models that predict insider threat events. One such attempt was a prior study by the Defense Personnel and Security Research Center (PERSEREC), a division of the Office of People Analytics (OPA), which utilized individual military personnel data to predict insider threat events. Of course, high-impact insider threat events are statistically rare, making it difficult to develop accurate predictive models. So instead, researchers used four insider threat outcomes that may precede high-impact events. The outcomes included (1) military discharge for unsuitability, (2) being the subject of a criminal investigation, (3) actions resulting in a security incident, and (4) revocation of clearance. A few of the predictive modeling approaches in that study performed moderately well, but not well enough to suggest that they be implemented.

A change in purpose and a new approach guided the current study. Instead of trying to build predictive models for operational use in predicting and preventing high-impact insider threat events, the aim was to explore the use of environmental factors in combination with individual ones to gain a clearer picture of insider threat. Such environmental factors may exist at the level of the work unit, organization, neighborhood, or city. Researchers merged economic and crime statistics for the various places where subjects had lived with individual data from military personnel files. The new approach was the use of discrete-time event history analysis.

Eight models were developed in total. Pseudo R^2 values were at or above .33 for six of the eight models. Goodness of fit measures demonstrated that four of these six models, however, did not provide a good fit, even though they yielded larger pseudo R^2 values.

The predictor variable included most frequently across models was the environmental factor Property Crime. It was statistically significant in five of the eight models. In each of these models an increase in property crime rates corresponded to a decrease in the occurrence of the criterion measure.

Some predictors occurred less frequently but had significant effects on their outcome measure. The predictor with the largest measured effect on any model was College Graduate. Although it was significant in only three of the models, the odds ratios indicated that a college graduate was up to 14 times less likely to experience an unsuitability discharge than members of the reference group.

Results demonstrated the ability to successfully create models of insider threat behavior using a combination of individual and environmental factors. The inclusion of environmental factors in model development provides valuable contextual information and may contribute significantly to the development of effective predictive models in the future. Consideration of external data sources for future research should be expanded to other sources of environmental data where available.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES IN APPENDICES

INTRODUCTION

Insider threat is an ongoing concern for the Department of Defense (DoD). Within the past 8 years, incidents of violence, such as the Fort Hood and Navy Yard shootings, as well as massive unauthorized disclosure of classified information to Wikileaks by Private Manning have caused serious harm to personnel and national security.

U.S. Army Major Nidal Hasan killed 13 people and wounded 32 others at Fort Hood on 5 November 2009. Soon after the incident, the press revealed that the Joint Terrorism Task Force (JTTF) had been aware of Major Hassan's communications with a Muslim cleric who had ties to al-Qaeda. In December of that year, Federal Bureau of Investigation (FBI) Director Robert Mueller appointed former FBI Director William Webster to lead a commission looking into the handling of this information. The Webster Commission determined that the FBI had failed to follow up on leads and to share information with DoD. DoD conducted its own internal investigation and concluded that the Department was not prepared to identify or deal with the risk of violence by trusted insiders.

Aaron Alexis, a DoD contractor, killed 12 people and wounded three others at the Washington Navy Yard on 16 September 2013. Subsequently DoD conducted an internal review of programs, policies, and procedures related to physical security and military installations as well as vetting and continuous evaluation of DoD personnel and contractors with security clearances. Recommendations based on this review included implementation of programs for continuous evaluation as well improved identity management services at military installations.

In 2010, then-Private Bradley Manning leaked classified information to Wikileaks. In October 2011, President Obama issued Executive Order 13587, *Structural Reforms to Improve the Security of Classified Networks and the Responsible Sharing and Safeguarding of Classified Information*. Thirteen months later, the White House issued the corresponding *National Insider Threat Policy and Minimum Standards for Executive Branch Insider Threat Programs* (hereinafter, Minimum Standards). Together, these documents provided the authority and established the minimum requirements for agencies' mandatory insider threat programs intended to protect classified assets both on- and off-line.

Prompted by the federal mandate, the Department of Defense (DoD) issued Directive 5205.16, *The DoD Insider Threat Program*, on September 30, 2014. Like EO 13587 and the Minimum Standards, DoD Directive 5205.16 limited the scope of the enterprise-level DoD Insider Threat Program to only those individuals who could misuse their authorized access; that is, their access to classified information. In accordance with language from the National Defense Authorization Act of FY17, however, DoD revised 5205.16 in January 2017, and expanded the definition of an insider to include any "person who has or had been granted eligibility for access to classified information or eligibility to hold a sensitive position." Without accounting for former clearance holders, this expansion broadened the scope of the DoD Insider Threat Program from 2.9 million civilian, Service, and contractor personnel to more than 4.2 million people (Office of the Director of National Intelligence, 2015).

BACKGROUND

Malicious insider threat behavior may manifest in a number of ways, including espionage, workplace violence, sabotage, and fraud. It also may be motivated by a number of factors, such as ideology, profit, and revenge. This complex interplay among individual, interpersonal, and organizational factors poses significant challenges to those who are committed to prevention, detection, and mitigation efforts. Efforts to *predict* these events are complicated further by the fact that high-impact insider threat incidents are statistically rare.

By definition, rare events provide few data points that can be used for predictive research. One potential solution is to assume that these incidents “are just the high-magnitude tail of some underlying distribution of events” and as such, are correlated with far more common events that can be observed, modeled, and validated (JASON, The MITRE Corporation, 2009, p. 22). For example, professional misconduct and on-the-job issues preceded all known cases of DoD workplace homicide that occurred between 2009 and 2015, and the majority of the perpetrators had a documented history of aggression (Smith, Jaros, & Chandler, 2016).

These pathway models (Shaw & Sellers, 2015) provide both a useful theoretical and methodological framework by focusing attention on predicating events that are both observable and more common than high-impact insider threat incidents. They also highlight the importance of the environmental context, and its effects on the likelihood and type of insider threat behavior. For instance, in his macro-level study of espionage, Lillbacka (2016) concluded that ideologically motivated spies were more likely to emerge from socio-culturally coherent groups that supported a foreign country or entity. Similarly, Moore, et al. (2016) found that cyber-related counterproductive workplace behaviors were less likely in organizations rated by their employees as high in organizational justice.

A previous study by the Defense Personnel and Security Research Center (PERSEREC), a division of the Office of People Analytics (OPA), explored how to apply advanced analytics to internally held DoD personnel data in order to identify predictors of four insider threat outcomes that may precede high-impact events (Zimmerman, et al., in progress). The outcomes for the previous study included (1) military discharge for unsuitability, (2) being the subject of a criminal investigation, (3) actions resulting in a security incident, and (4) revocation of clearance. The methodology for that study focused on predictor data accumulated over a person’s career up to the time of an outcome event, for example, rate of promotion or number of marriages. Although a few of the predictive modeling approaches in that study performed moderately well, the effort was not an unqualified success. In the end, advanced analytics were not sufficient for predicting such outcomes, given the predictors used in that study.

It should be noted that PERSEREC researchers are aware of other efforts applying advanced analytics to look at insider threat. However, documentation of these efforts has not been broadly disseminated and cannot be summarized here.

PURPOSE

Like the previous OPA/PERSEREC modeling effort, the current study was internally funded. The purpose of this effort was not to build predictive models for operational use in predicting and preventing high-impact insider threat events. Its purpose was rather to explore the use of environmental factors in combination with individual ones to gain a clearer picture of insider threat.

The approach for the current study differs from the previous work in three ways. First, analysts accounted for contextual factors in people's environment that can influence their behavior. Such environmental factors may exist at the level of their work unit, organization, neighborhood, or city. To do this, analysts integrated economic and crime statistics for the various places where subjects had lived. Second, rather than analyzing a single snapshot of the variables, annual snapshots were taken for both the individual and environmental variables up to the year prior to the outcome event. This was an important modification to the methodology because it provides a greater level of precision in measuring relevant predictor data. Thirdly, rather than focusing on building predictive models, the goal of this study was to identify the individual and environmental factors having a nexus with negative outcome personnel events.

METHOD

PERSEREC used data from manpower and personnel files maintained by the Defense Manpower Data Center (DMDC) and from open sources maintained by Department of Commerce, Internal Revenue Service, and Department of Justice. The DMDC files provided information on individuals while the aggregated data from open sources provided macro-level, environmental information.

DMDC maintains Department of Defense (DoD) historical manpower and personnel records as flat files on a mainframe. Other DMDC data came from two in-house information systems: the Defense Central Index of Investigations (DCII) and the Joint Personnel Adjudication System (JPAS). DCII is a common index of investigation dossiers for all of the military criminal investigative organizations, while JPAS is the system of record for DoD clearance holders and contains current and past security clearance levels as well as security incident records. In addition to these individual-level data sources, macro-level (i.e., aggregated) environmental data from open sources included economic and crime statistics.

Subjects included all Service members in the regular components who served between 1999 and 2017. Reservists on active duty were not included in the records extracted from the mainframe.

PREDICTORS

Individual predictors drawn from active duty personnel, pay, and family files included demographics, marital status, occupation, pay grade, bonuses, awards and special pay, and ages of dependents. Dates and locations of deployments were retrieved from the Contingency Tracking System, while dates and circumstances of injuries came from the Casualty File.

Environmental predictors were selected to represent measurable indicators of common factors such as regional crime rates, economic conditions, and job availability. Additional consideration for predictor inclusion was the availability of data for the time periods of interest (see STATISTICAL APPROACH) as well as regional completeness of data to include all locations in the United States. Finally, zip code or county identifiers were required in order to link environmental predictors to each subject's individual predictor data.

Environmental data were drawn from open source Internal Revenue Service (IRS) files. The IRS data are aggregated (at the source) by zip code and include tax filing information, such as household size, and rates of various tax-filing statuses (e.g., pension and unemployment benefits, home ownership, and mortgage deductions). Unfortunately, it was only possible to use data from 2008 onwards since there were too few predictors in the earlier IRS data.

Data from other open sources were aggregated (at the source) by county, based on the Federal Information Processing Standards (FIPS) codes. Examples of economic and labor statistics from the U.S. Department of Commerce are county job density (number of jobs per person), average earnings per capita, and average earnings per job. Finally, the United States Department of Justice annual crime rates for reported property and violent crime were also included.

In addition to individual and contextual variables, the passage of time was also considered as a possible predictor variable, and was represented by length of time serving in the military.

In total, 42 predictors were considered. Appendix A contains a description of the predictors.

CRITERION MEASURES

Since actual insider threat incidents are too rare to be used with inferential statistics, four surrogate measures from DMDC personnel files were used in this study. These measures included 1) unsuitability attrition, 2) being the subject of a criminal investigation, 3) having a security incident, and 4) losing access to classified information.

Unsuitability attrition came from the Active Duty Transactions File which contains discharge dates and reasons for discharge. Discharge records that indicated the individual was discharged due to a behavior issue were indicative of a negative outcome. People in the non-negative outcome group were those who were still serving or who had been discharged for other reasons.

The negative outcome group for being a subject of a criminal investigation consisted of service members with one or more records in DCII, in which the individual was flagged as the subject of the investigation. Individuals without such records made up the non-negative group.

Individuals with a record of a security incident in JPAS constituted the negative outcome group for the third outcome measure, having a security incident. Individuals in JPAS without a security incident constituted the non-negative group.

Lastly, individuals who lost access to classified information, as reported in JPAS, were placed in the fourth group, losing access to classified information. The non-negative outcome group contained individuals in JPAS without loss of access.

More detailed information regarding criterion measure coding can be found in Appendix B.

STATISTICAL APPROACH

Event history analysis is frequently employed in social science to study factors leading up to a particular event. For this study, the events of interest are the outcomes discussed in the previous section. The rational for using this approach is perhaps best explained in the following:

> Although event histories are ideal for studying the causes of events, they typically possess two features—censoring and time-varying explanatory variables—that create major problems for standard statistical procedures such as multiple regression. In fact, the attempt to apply standard methods can lead to severe bias or loss of information. (Allison, 2014).

Rather than developing statistical models focused on a single time period—e.g., just before the occurrence of a particular outcome—PERSEREC used an event history analysis to examine the series of events leading up to the outcome. While predictor variables like deployments are measured on a time continuum, variables like property crimes are measured as an annual rate, so analysts were not able to use continuous time proportional hazards regression. For discrete time measurement, binomial regression is the appropriate alternative. Binomial regression using the complimentary-log-log link would have produced a discrete time proportional hazards model. However, the resulting regression coefficients could only be converted to hazard ratios, which are difficult to interpret. Logistic regression (i.e., binomial regression employing a logit link function) was chosen instead due to ease of interpretation and the judgment of researchers that the proportional hazards assumption was not relevant, given the exploratory nature of the current study.

The dataset was prepared as a set of annual records for each person. Continuous time variables were converted to dichotomous variables for use in discrete time modeling. For example, if a person was deployed for any portion of a given year, the dichotomous variable was coded as 1 for that year. If they were not deployed in that year, then the value was coded as 0. In addition, time was treated as a categorical variable and converted to a set of dichotomous variables, one for each year in the study period, except for the last year of the period.

The model building was exploratory. That is, there were no hypotheses generated prior to the study with regard to which predictor variables would be determined to be important in the models. Exploratory modeling is an iterative process that involves adding or removing predictor variables at successive steps until a model is found that best fits the data. The danger is that the final model will suffer from overfitting and not be generalizable. To address this concern, the dataset was randomly partitioned into an exploratory sample and a confirmatory sample so that the predictor variables chosen during the modeling phase could be used to create final models in the confirmatory sample. Exploratory modeling is focused on selecting predictor variables, rather than creating a final model, and generates hypotheses as to which variables should go into the model. The models developed for the confirmatory sample are then used for hypothesis testing.

Finally, all of the criterion measures are based on rare events. Logistic regression is most efficient with equally balanced criterion data and has difficulty in modeling rare events. The rare event problem is further compounded by the fact that the data consist of person-year records and the negative outcome can only occur in one of those years, if at all. For example, consider fifty people with 20 years of records each and one of them had a security incident in the 20th year. One record with a negative event out of a thousand poses a problem for logistic regression.

To deal with the rare event problem, researchers decided to employ narrowly defined time periods. Specifically, two 5-year time periods were selected. The first time period was 2002 – 2006 and the second was 2008 – 2012. These time periods correspond to the early years of the War on Terror and to the Great Recession and its aftermath, respectively. It was considered that both of these time periods resulted in increased hardship and stress which might be identified through measures available during those periods (e.g., deployment activity, unemployment rates, etc.)

SAMPLING

Random sampling was used to partition cases into two samples of relatively equal size. The first sample was used as an exploratory group (n=5,570,414) to investigate the data and to select individual predictor variables based on their relationship to each of the outcome measures. The second sample was used as a confirmatory group (n=5,567,555) – i.e., to verify the exploratory models in a separate sample. Demographic characteristics of the two samples are displayed in Table 1.

Table 1
Demographic Characteristics by Sample

Demographic Characteristic	Exploratory Sample		Confirmatory Sample	
	n	%	n	%
Service				
Army	2,756,462	49.5	2,756,732	49.5
Navy	396,198	7.1	394,343	7.1
Air Force	1,413,963	25.4	1,411,552	25.4
Marine Corps	1,003,791	18.0	1,004,928	18.0
Rank				
E1-E4	2,987,995	53.6	2,990,778	53.7
E5-E9	1,774,076	31.8	1,770,066	31.8
W1-W5	51,035	0.9	49,640	0.9
O1-O3	425,390	7.6	426,447	7.7
O4-O9	331,918	6.0	330,624	5.9
Marital Status				
Never married	2,483,719	44.6	2,484,419	44.6
Married	2,843,031	51.0	2,840,889	51.0
Separated	5,021	0.1	4,615	0.1
Annulled/Divorced	229,925	4.1	229,131	4.1
Widowed	3,008	0.1	2,736	<0.1
UNKNOWN[1]	5,710	0.1	5,765	0.1
Education				
Less than High School	224,494	4.0	222,728	4.0
High School Graduate	4,328,989	77.7	4,325,768	77.7
College Graduate	579,230	10.4	583,686	10.5
Advanced Degree	356,345	6.4	354,035	6.4
UNKNOWN	81,356	1.5	81,338	1.5

[1] UNKNOWN is a legitimate value provided in the source data. It does not refer to null values in the data set.

RESULTS

For model development, analysts generated a single model for each outcome measure (i.e., Unsuitability Discharge, DCII Investigation, Security Incident, and Access Suspended) by time period (2002 to 2006 and 2008 to 2012). This resulted in eight models:

(1) Model 1: Unsuitability Discharge (2002 – 2006)

(2) Model 2: Unsuitability Discharge (2008 – 2012)

(3) Model 3: DCII Investigation (2002 – 2006)

(4) Model 4: DCII Investigation (2008 – 2012)

(5) Model 5: Security Incident (2002 – 2006)

(6) Model 6: Security Incident (2008 – 2012)

(7) Model 7: Access Suspended (2002 – 2006)

(8) Model 8: Access Suspended (2008 – 2012)

Predictor variables found to be statistically significant in the exploratory phase were included in the confirmatory model development while non-significant predictors were omitted from the confirmatory phase. The results for confirmatory models are displayed below.

Pseudo $R^2$and Goodness of Fit statistics (including both Osius-Rojek Z and Stukel χ^2) were generated to measure the effectiveness of each model. The results of these analyses can be found in Table 2.

Table 2
Effect Size and Goodness of Fit for Confirmatory Models

			Goodness of Fit			
Criterion Measure	**Time Period**	**Pseudo R^2**	**Osius-Rojek Z**	**p**	**Stukel x^2 (df=2)**	**p**
Unsuitability Discharge	2002-2006	.61	1.38	0.17	25.97	<0.01
Unsuitability Discharge	2008-2012	.61	2.29	0.02	28.39	<0.01
DCII Investigation	2002-2006	.49	16.92	<0.01	0.64	0.73
DCII Investigation	2008-2012	.64	14.94	<0.01	27.94	<0.01
Security Incident	2002-2006	.06	0.51	0.61	3.89	0.14
Security Incident	2008-2012	.63	2.88	<0.01	5.61	0.06
Access Suspended	2002-2006	.06	-0.47	0.64	8.23	0.02
Access Suspended	2008-2012	.33	0.08	0.94	14.67	<0.01

The pseudo R^2 value reflects the variability in the criterion measure that is accounted for by the model. These values were at or above .33 for six of the eight models created: Unsuitability Discharge, all years; DCII Investigation, all years; Security Incident, 2008 – 2012; and Access Suspended, 2008 – 2012.

Note that for the goodness of fit tests in Table 2, the null hypothesis is that the model is a good fit to the data; the alternative hypothesis is that the model does not fit the data. Thus, small *p* values indicate a poor fit. Thus, most of the models did not provide a good fit, even though they yielded larger pseudo R^2 values. The exceptions were the models for Unsuitability Discharge, 2002 – 2006 (Z=1.38, p=0.17) and DCII Investigation, 2002 – 2006 (χ^2= 0.64, p=0.73). The remaining two models (Security Incident, 2002 – 2006 and Access Suspended, 2002 – 2006) displayed small R^2 (each with R^2= .06). However, both of these models displayed non-significant goodness of fit measures (Security Incident, 2002 - 2006 [Z=0.51, p=0.61 and χ^2=3.89, p=0.14] and Access Suspended, 2002 – 2006 [Z=-0.47, p=0.64]).

The remaining tables in this section provide details of the confirmatory logistic regression. Asterisks in the significance (***p***) columns indicate predictors that were statistically significant in the final confirmatory models. Predictors without an asterisk are included in the tables to reflect that they were part of the final confirmatory models though they did not demonstrate significance.

The regression statistics for the predictor variables used within each model include coefficient estimates and odds ratios. Combined, these statistics demonstrate the effect each variable had on the model. Positive estimate values indicate an increasing likelihood of the event, while negative estimates indicate a decreasing likelihood. The rate of the increase or decrease in likelihood is reflected in the size of the odds ratio. When an estimate has a negative value (reflecting a decrease in the likelihood of the outcome event), an inverted ratio was calculated to represent the size of that decrease.

MODELS 1 & 2: UNSUITABILITY DISCHARGE

As shown in Table 3, seven predictor variables contributed to Model 1. The predictor with the strongest relationship to the criterion was College Graduate with an odds ratio of 8.69. This indicates that a person with a college education was nearly nine times less likely to experience an unsuitability discharge than a person in the reference group, which for this predictor is a person with an UNKNOWN level of education (See Appendix C for a description of all reference groups for categorical predictors.) The predictor Deployed, with an odds ratio of 2.03, indicates that a person who was deployed at some point between 2002 and 2006 was twice as likely not to experience an unsuitability discharge the following year when compared to a person who was not deployed.

Table 3
Model 1: Unsuitability Discharge (2002 – 2006)

Predictor	I/E[1]	Estimate	Error	*z*	*p*	Odds Ratio[2]
Number of Children 5-14 Years Old	I	-0.85	0.10	-8.36	<0.01*	[2.34]
College Graduate	I	-2.16	0.38	-5.69	<0.01*	[8.69]
Deployed	I	-0.71	0.11	-6.66	<0.01*	[2.03]
Property Crime Rate	E	-0.20	0.03	-5.83	<0.01*	[1.22]
Per Capita Dividends, Interest, & Rent	E	0.53	0.04	12.14	<0.01*	1.70
Employer Contributions for Pension & Insurance	E	0.68	0.09	7.96	<0.01*	1.97
Average Earnings Per Job	E	-0.64	0.09	-7.47	<0.01*	[1.89]

[1]Individual/Environmental
[2]Numbers in brackets denote inverted odds ratios

Results from Model 2, which also considers unsuitability discharges but for a different time period than Model 1, are shown in Table 4. As in Model 1, College Graduate was the strongest predictor of an unsuitability discharge with an odds ratio of 13.89. This suggests a person who had a college degree was nearly 14 times less likely to experience an unsuitability discharge than those with an educational level of UNKNOWN, when considering the years 2008 to 2012. The variable Violent Crime is a measure of the violent crime rate for the county in which a person resides. The odds ratio of 2.62 for the variable Violent Crime indicates that for each standard deviation increase in the crime rate in a person's county of residence, that person is 2.62 times more likely to experience an unsuitability discharge.

Table 4
Model 2: Unsuitability Discharge (2008 – 2012)

Predictor	I/E[1]	Estimate	Error	*Z*	*P*	Odds Ratio[2]
Number of Children 19-25 Years Old	I	0.30	0.08	3.93	<0.01*	1.36
Serving in Navy	I	-1.27	0.27	-4.65	<0.01*	[3.55]
Serving in Air Force	I	-0.96	0.12	-8.00	<0.01*	[2.60]
Divorced	I	-0.53	0.15	-3.59	<0.01*	[1.70]
Married	I	-0.80	0.07	-11.74	<0.01*	[2.23]
College Graduate	I	-2.63	0.32	-8.28	<0.01*	[13.89]
Received Special Pay	I	-1.17	0.07	-17.12	<0.01*	[3.23]
Percent Joint Tax Returns	E	-0.14	0.10	-1.50	0.13	[1.15]
Percent Returns with Pension or Annuity Income	E	-0.45	0.06	-7.32	<0.01*	[1.56]
Percent Returns with Child Care Credit	E	0.41	0.12	3.37	<0.01*	1.51
Percent Returns with Earned Income Credit	E	0.31	0.07	4.25	<0.01*	1.36
Percent Prepared Returns	E	0.26	0.05	4.86	<0.01*	1.30
Violent Crime Rate	E	0.96	0.20	4.94	<0.01*	2.62
Property Crime Rate	E	-0.25	0.05	-5.05	<0.01*	[1.28]
Per Capita Personal Current Transfer Receipts	E	-0.56	0.07	-8.16	<0.01*	[1.76]
Employer contributions for pension & insurance	E	-0.34	0.07	-4.61	<0.01*	[1.40]
Average Earnings Per Job	E	1.03	0.18	5.83	<0.01*	2.79
Average Wages & Salaries	E	-1.52	0.16	-9.31	<0.01*	[4.59]

[1] Individual/Environmental
[2]Numbers in brackets denote inverted odds ratios

MODELS 3 & 4: DCII INVESTIGATION

Table 5 shows results of the Model 3 analysis for DCII Investigation (2002 – 2006). This model included counts of dependent children in three different age groups as significant predictors of a DCII Investigation outcome. In all three groups, an increase in the number of children corresponded to an increase in the likelihood of the DCII Investigation result. The largest odds ratio of any predictor was for O1 – O3 (i.e., indicating a person with rank between O1 and O3 inclusive). Here, a person with rank O1 – O3 was approximately 3.5 times less likely to experience a DCII Investigation than those in the reference group, which was personnel with a rank of Warrant Officer (i.e., W01 – W05), when considering the years from 2002 to 2006.

Table 5
Model 3: DCII Investigation (2002 – 2006)

Predictor	I/E[1]	Estimate	Error	*z*	*p*	Odds Ratio[2]
Number of Children 1-4 Years Old	I	0.08	0.02	5.50	<0.01*	1.09
Number of Children 5-14 Years Old	I	0.05	0.01	4.56	<0.01*	1.05
Number of Children 15-18 Years Old	I	0.11	0.02	4.99	<0.01*	1.12
O1-O3	I	-1.27	0.05	-23.30	<0.01*	[3.56]
Injured in Hostile Situation	I	0.43	0.12	3.66	<0.01*	1.53
Received Special Pay	I	0.32	0.05	6.93	<0.01*	1.37
Property Crime Rate	E	-0.04	0.01	-6.19	<0.01*	[1.04]
Per Capita Income Maintenance Benefits	E	0.07	0.01	6.10	<0.01*	1.07
Per Capita Unemployment Compensation	E	0.28	0.02	15.23	<0.01*	1.32
Average Proprietors' Income	E	-0.04	0.01	-4.68	<0.01*	[1.04]

[1]Individual/Environmental
[2]Numbers in brackets denote inverted odds ratios

Model 4, shown in Table 6, contained the largest number of predictor variables of statistical significance (18). Three of the predictors (Number of Children 19 – 25 Years Old, Injured in Non-Hostile Situation, and Percent Returns with Self Employment Income) survived the initial exploratory phase but were determined non-significant in the final confirmatory analysis. A large number of economic measures contributed to the final model, although with mixed results. For example, an increase in the predictor Average Earnings Per Job (i.e., suggesting a person's county of residence had more high-paying jobs) increased the likelihood of a DCII investigation. Conversely, an increase in the predictor Average Wages & Salaries (i.e., suggesting a person's county of residence had more highly paid residents) corresponded with a decrease in DCII Investigation occurrences. Of note, Model 4 was the only model to have both types of injury designation (Injured in Hostile Situation and Injured in Non-Hostile situation) as significant predictors. Both types of injury designation increased the likelihood of a DCII Investigation outcome.

Table 6
Model 4: DCII Investigation (2008 – 2012)

Predictor	I/E[1]	Estimate	Error	*z*	*p*	Odds Ratio[2]
Number of Children 1-4 Years Old	I	0.10	0.02	6.29	<0.01*	1.10
Number of Children 19-25 Years Old	I	0.01	0.02	0.52	0.60	1.01
Serving in Army	I	0.79	0.04	19.09	<0.01*	2.20
Serving in Air Force	I	-0.83	0.05	-18.22	<0.01*	[2.29]
O1-O3	I	-1.10	0.05	-21.94	<0.01*	[3.01]
College Graduate	I	-0.27	0.04	-7.15	<0.01*	[1.31]
Injured in Hostile Situation	I	0.45	0.10	4.34	<0.01*	1.57
Injured in Non-Hostile Situation	I	0.29	0.21	1.35	0.18	1.33
Percent Joint Tax Returns	E	-0.26	0.03	-10.43	<0.01*	[1.30]
Percent Returns with Pension or Annuity Income	E	0.12	0.01	9.15	<0.01*	1.13
Percent Returns with Self Employment Income	E	-0.02	0.02	-0.66	0.51	[1.02]
Percent Returns with Child Care Credit	E	0.09	0.03	2.67	<0.01*	1.09
Percent Returns with Earned Income Credit	E	0.17	0.02	9.57	<0.01*	1.19
Percent Prepared Returns	E	-0.04	0.02	-2.73	<0.01*	[1.04]
Violent Crime Rate	E	0.24	0.05	4.35	<0.01*	1.27
Property Crime Rate	E	-0.11	0.01	-9.18	<0.01*	[1.12]
Per Capita Income Maintenance Benefits	E	-0.11	0.02	-6.84	<0.01*	[1.12]
Per Capita Unemployment Compensation	E	0.07	0.01	5.30	<0.01*	1.07
Number of Jobs Per Capita	E	-0.07	0.01	-6.30	<0.01*	[1.08]
Average Earnings Per Job	E	0.26	0.04	6.10	<0.01*	1.29
Average Wages & Salaries	E	-0.19	0.04	-4.61	<0.01*	[1.21]

[1]Individual/Environmental
[2]Numbers in brackets denote inverted odds ratios

MODELS 5 & 6: SECURITY INCIDENT

In Table 7, two economic measures stand out for Model 5, which measured security incidents in the years 2002 to 2006. Considering the odds ratios, higher Per Capita Personal Income increased the likelihood of experiencing a security incident. Conversely, higher Per Capita Unemployment Compensation, which suggests living in areas of higher unemployment, had the opposite effect, with higher unemployment rates actually decreasing the likelihood of a security incident.

Table 7
Model 5: Security Incident (2002 – 2006)

Predictor	I/E[1]	Estimate	Error	*z*	*P*	Odds Ratio[2]
Number of Children 1-4 Years Old	I	0.06	0.04	1.58	0.11	1.07
Serving in Army	I	1.23	0.05	23.66	<0.01*	3.43
E1-E4	I	0.35	0.05	7.01	<0.01*	1.41
High School Graduate	I	0.38	0.07	5.66	<0.01*	1.47
Per Capita Personal Income	E	0.22	0.03	8.10	<0.01*	1.24
Per Capita Unemployment Compensation	E	-0.45	0.06	-7.29	<0.01*	[1.56]

[1]Individual/Environmental
[2]Numbers in brackets denote inverted odds ratios

Model 6 results are shown in Table 8. Per Capita Personal Income had the strongest relationship to the outcome measure with an odds ratio of 2.06. This same predictor was also influential in Model 5, the first of the two models which considered security incidents as the outcome measure. In both models, higher Per Capita Personal Income increased the likelihood of experiencing a security incident. Similarly, in both Model 5 and Model 6, predictors concerning unemployment benefits showed that for those living in an area with higher unemployment compensation there was a decrease in the likelihood of a security incident.

Table 8
Model 6: Security Incident (2008 – 2012)

Predictor	I/E[1]	Estimate	Error	z	P	Odds Ratio[2]
Number of Children 1-4 Years Old	I	0.13	0.04	3.50	<0.01*	1.14
Number of Children 19-25 Years Old	I	0.16	0.05	3.30	<0.01*	1.18
Serving in Navy	I	-0.41	0.11	-3.91	<0.01*	[1.51]
E5-E9	I	0.09	0.05	1.91	0.06	1.10
Never Married	I	0.22	0.05	4.47	<0.01*	1.24
High School Graduate	I	0.53	0.05	9.65	<0.01*	1.69
Percent Joint Tax Returns	E	-0.30	0.04	-8.06	<0.01*	[1.35]
Percent Returns with Retirement Distributions	E	0.42	0.04	11.61	<0.01*	1.53
Percent Returns with Unemployment Income	E	-0.28	0.04	-7.16	<0.01*	[1.33]
Percent Returns with Earned Income Credit	E	0.49	0.04	13.47	<0.01*	1.63
Property Crime Rate	E	-0.05	0.03	-2.10	0.04*	[1.05]
Per Capita Personal Income	E	0.72	0.09	7.86	<0.01*	2.06
Per Capita Net Earnings	E	-0.40	0.08	-4.98	<0.01*	[1.49]
Per Capita Income Maintenance Benefits	E	0.20	0.04	5.46	<0.01*	1.22
Employer contributions for pension & insurance	E	0.45	0.04	12.94	<0.01*	1.57
Number of Jobs Per Capita	E	-0.50	0.03	-17.84	<0.01*	[1.65]
Average Wages & Salaries	E	0.14	0.05	2.64	0.01*	1.15

[1]Individual/Environmental
[2]Numbers in brackets denote inverted odds ratios

MODELS 7 & 8: ACCESS SUSPENDED

Model 7, shown in Table 9, contained the fewest predictors of any of the eight models, with only two variables retained in the final model. Of the two, Serving in the Army resulted in the largest odds ratio (5.38), indicating that an active duty member serving in the army was 5.38 times more likely to experience a suspended access than someone in the reference group (Serving in the Marines), when considering the years 2002 to 2006.

Table 9
Model 7: Access Suspended (2002 – 2006)

Predictor	I/E[1]	Estimate	SE	z	P	Odds Ratio
Serving in Army	I	1.68	0.18	9.33	<0.01*	5.38
Deployed	I	0.01	0.14	0.04	0.97	1.01
Per Capita Retirement & Other	E	0.47	0.12	3.86	<0.01*	1.60

[1]Individual/Environmental

Table 10 provides the results for Model 8. The predictor with the largest odds ratio in Model 8 was Serving in the Army, with a value of 7.40. This predictor also had the largest odds ratio in the previous model, Model 7, which also measured access suspended as the outcome.

Table 10
Model 8: Access Suspended (2008 – 2012)

Predictor	I/E[1]	Estimate	SE	z	P	Odds Ratio[2]
Number of Children 5-14 Years Old	I	0.14	0.04	3.77	<0.01*	1.15
Serving in Army	I	2.00	0.12	16.58	<0.01*	7.40
E5-E9	I	0.35	0.07	4.78	<0.01*	1.42
Percent Joint Tax Returns	E	-0.02	0.04	-0.40	0.69	[1.02]

[1]Individual/Environmental
[2]Numbers in brackets denote inverted odds ratios

ALL MODELS

The predictor variable with statistical significance in most of the models (n=5) was Property Crime. In each of the five models in which Property Crime was significant, an increase in property crime actually decreased the odds of a negative outcome. Violent Crime, on the other hand, was significant in only two of the models (Models 2 and 4) but showed an increase in the likelihood of a negative outcome for an increase in the rate of violent crime value.

One predictor of interest is Percent Joint Tax Returns which was only available in the models representing the years 2008 to 2012 due to limitations of IRS data availability prior to 2008. This predictor occurred in all the models for which it was eligible during the exploratory phase. In the confirmatory phase it was significant in two models (Models 4 and 6) where the effect was to decrease the negative outcome likelihood for any increase in the number of joint returns filed.

The predictor with the largest measured effect on any model was College Graduate. Although this predictor was significant in only three of the models, it suggested a College Graduate was nearly nine times less likely to experience an unsuitability discharge in the time period 2002 – 2006 and nearly 14 times less likely to experience an unsuitability discharge in the years 2008 – 2012.

When considering the Department of Commerce economic and labor data aggregated by county, Average Wages and Salaries had the highest odds ratio of any predictor from that data source. The odds ratio of 4.59 suggests that for each standard deviation increase in average wage a person was nearly five times less likely to receive an unsuitability discharge in the years 2008 to 2012. Several other variables obtained from the Department of Commerce, including Average Earnings Per Job, Average Wages and Salaries, Per Capita Income Maintenance Benefits, and Per Capita Unemployment Compensation, were significant in three of the models. This suggests these predictors may have some generalizability across multiple types of negative outcome events.

CONCLUSIONS

This study demonstrated the effectiveness of using external environmental and contextual data as a supplement to individual personnel measures as predictive modeling indicators of insider threat behavior. Although there are limitations to the availability of some external data sources, these data can provide valuable contextual information and may contribute significantly to the development of effective predictive modeling efforts.

Results were mixed for the direction of the relationship between a predictor and a criterion measure. Some predictors, such as living in an area with higher violent crime rates, demonstrated an increase in likelihood of a negative security event. Other predictors, such as possessing a college degree, showed that those with a college degree are less likely to commit a security incident or to have their access suspended when compared to those with an educational level coded as UNKNOWN.

In general, the results demonstrate an ability to model a relationship between certain predictor variables and potential insider threat indicators such as Unsuitability Discharge or DCII Investigation. Exactly which variables will best serve this purpose can be determined through further modeling studies. However, the predictor variables which showed statistical significance here should certainly be considered as candidates for future modeling efforts.

LIMITATIONS

The crime data used in this study was provided through the Uniform Crime Report maintained by the Department of Justice, Federal Bureau of Investigation. Although considered a comprehensive measure of crime statistics in the United States, the local law enforcement agencies providing the crime data participate on a voluntary basis. Therefore, it is possible that some agencies may have underreported incidents of crime in their area.

There was also an issue with non-existent or unusable data for some sources for the first half of the time period. In particular, usable economic data aggregated by zip code was not available prior to 2008.

Lastly, data stored in the JPAS Incident table can occasionally contain non-incident data. This may result in a slight increase in false positive cases for the security incident criterion measure. Future efforts should consider analytic methods to minimize these false positive cases.

RECOMMENDATIONS

Several factors should be considered when determining which predictors are good candidates for further investigation and inclusion in future predictive modeling efforts. Frequently occurring predictors such as Average Earnings Per Job and

Property Crime Rate demonstrate they may be useful indicators across multiple types of negative outcomes for security clearance holders.

Alternatively, certain predictors, such as College Graduate, may occur less frequently but have a strong relationship to a specific type of negative outcome. In this case, the strong relationship between College Graduate and Unsuitability Discharge suggest it is a good candidate for further investigation of that particular type of risk.

Additionally, further studies should consider new predictors which measure features such as relative disparity and comparative properties. For instance, combining the environmental predictor Per Capita Income with an individual predictor such as pay (as derived from rank) could demonstrate the degree of income disparity a person experiences relative to where they live. Such features may contribute significantly to further model development.

Finally, consideration of external data sources for future research should be expanded to other sources of environmental data where available. Sites providing open source data generally provide caveats on its use, the data collection methodology and warnings regarding its interpretation to prevent misuse and over-generalizations. These factors should all be taken into account when exploring potential environmental data collection for additional insider threat modelling efforts.

REFERENCES

Allison, P. D. (2014) *Event history analysis: Regression for longitudinal event data.* Thousand Oaks, CA: Sage Publications, Inc.

Department of Defense. (2014, Sep. 30). *The DoD insider threat program* (DoD Directive 5205.16(d)). Revised January 25, 2017.

Executive Order 13587, *Structural Reforms To Improve the Security of Classified Networks and the Responsible Sharing and Safeguarding of Classified Information*, October 7, 2011.

JASON, The MITRE Corporation. (2009). *Rare events* (Publication No. JSR-09-108). Retrieved from https://www.dtic.mil/get-tr-doc/pdf?AD=ADA510224.

Lillbacka, R. (2016). The social context as a predictor of ideological motives for espionage. *International Journal of Intelligence and Counterintelligence, 30*(1), 117-146. Doi:10.1080/08850607.2016.1230704.

Moore, A., Savinda, J., Monaco, E., Moyes, J., Rousseau, D., Perl, S., Cowley, J., & Parshall, A. (2016). *The critical role of positive incentives for reducing insider threats* (Report No. CMU/SEI-2016-TR-014). Retrieved from http://resources.sei.cmu.edu/library/asset-view.cfm?assetid=484917.

Office of the Director of National Intelligence (2012). *National insider threat policy and minimum standards for Executive Branch insider threat programs* [Presidential Memorandum]. Retrieved from https://www.dni.gov/index.php/ic-legal-reference-book/presidential-memorandum-nitp-minimum-standards-for-insider-threat-program.

Office of the Director of National Intelligence. (2015). *2015 Annual Report on Security Clearance Determinations*. Retrieved from https://www.dni.gov/files/documents/Newsroom/Reports%20and%20Pubs/2015-Annual_Report_on_Security_Clearance_Determinations.pdf.

S.2943 Res. 114 Cong. (2017) (enacted). Retrieved from: https://www.congress.gov/bill/114th-congress/senate-bill/2943/text.

Shaw, E. & Sellers, L. (2015) Application of the critical-path method to evaluate insider risks. *Studies in Intelligence, 59*(2), 1-8. Retrieved from https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/csi-studies/studies/vol-59-no-2/pdfs/Shaw-Critical%20Path-June-2015.pdf.

Smith, C., Jaros, S., & Chandler, C. (2016). *Foreground factors of DoD workplace homicide: A comparative case analysis of incidents between 2009-2015* (Report No. TR 17-01). Seaside, CA: Defense Personnel and Security Research Center/Office of People Analytics. (For Official Use Only). DTIC: AD1036694.

Zimmerman, R.A., Seneviratna, G.S., Friedman, G.M., Ciani, D.A., Munshi, D.C., & Richmond, D.A. (In progress). *Modeling insider threat risk using predictive analytics*. Seaside, CA: Defense Personnel and Security Research Center/Office of People Analytics.

APPENDIX A:

PREDICTOR DESCRIPTIONS

APPENDIX A

Table A-1
Predictor Source/Aggregation/Description

Predictor	Source*	Level of Aggregation (based on residence)	Description
Average Earnings Per Job	DOC	County	Average earnings per job is total earnings divided by total full-time and part-time employment. Earnings is the sum of three components of personal income--wages and salaries, supplements to wages and salaries, and proprietors' income. The BEA employment series for states and local areas comprises estimates of the number of jobs, full-time plus part-time, by place of work. Full-time and part-time jobs are counted at equal weight. Both employment for wages and salaries and proprietors employment are included, but the employment of unpaid family workers and volunteers is not included.
Average Proprietors' Income	DOC	County	Proprietors' income with inventory valuation and capital consumption adjustments is the current-production income (including income in kind) of sole proprietorships, partnerships, and tax-exempt cooperatives. Corporate directors' fees are included in proprietors' income. Proprietors' income includes the interest income received by financial partnerships and the net rental real estate income of those partnerships primarily engaged in the real estate business.
Average Wages & Salaries	DOC	County	Average wages and salaries are wages and salaries divided by total wage and salary employment.
College Graduate	DMDC	Individual	Highest educational level achieved is Bachelor's Degree
Deployed	DMDC	Individual	Deployed into operational area
Divorced	DMDC	Individual	Marital Status
E1-E4	DMDC	Individual	Rank between E01 to E04, inclusive
E5-E9	DMDC	Individual	Rank between E05 to E09, inclusive
Employer Contributions for Pension & Insurance	DOC	County	Consists of employer payments to private and government pension plans and to private insurance funds such as for group health and life insurance; workers' compensation; and supplemental unemployment insurance.
Never Married	DMDC	Individual	Marital Status
High School Graduate	DMDC	Individual	Highest educational level achieved is High School Diploma
Injured in Hostile Situation	DMDC	Individual	Injured in Hostile Situation
Injured in Non-Hostile Situation	DMDC	Individual	Injured in Non-Hostile Situation

Predictor	Source*	Level of Aggregation (based on residence)	Description
Married	DMDC	Individual	Marital Status
Number of Children 1-4 Years Old	DMDC	Individual	Number of Children 1-4 Years Old
Number of Children 15-18 Years Old	DMDC	Individual	Number of Children 15-18 Years Old
Number of Children 19-25 Years Old	DMDC	Individual	Number of Children 19-25 Years Old
Number of Children 5-14 Years Old	DMDC	Individual	Number of Children 5-14 Years Old
Number of Jobs Per Capita	DOC	County	A count of jobs, both full-time and part-time. It includes wage and salary jobs, sole proprietorships, and individual general partners, but not unpaid family workers nor volunteers.
O1-O3	DMDC	Individual	Rank between O1 to O3, inclusive
Per Capita Dividends, Interest, & Rent	DOC	County	Consists of personal dividend income, personal interest income, and rental income of persons with capital consumption adjustment (CCAdj).
Per Capita Income Maintenance Benefits	DOC	County	Income maintenance benefits consists largely of Supplemental Security Income (SSI) benefits, Earned Income Tax Credit (EITC), Additional Child Tax Credit, Supplemental Nutrition Assistance Program (SNAP) benefits, family assistance, and other income maintenance benefits, including general assistance.
Per Capita Net Earnings	DOC	County	Consists of earnings by place of work less contributions for government social insurance plus the adjustment for residence.
Per Capita Personal Current Transfer Receipts	DOC	County	Receipts of persons from government and business for which no current services are performed. Current transfer receipts from government include Social Security benefits, medical benefits, veterans' benefits, and unemployment insurance benefits. Current transfer receipts from business include liability payments for personal injury and corporate gifts to nonprofit institutions.
Per Capita Personal Income	DOC	County	Consists of the income that persons receive in return for their provision of labor, land, and capital used in current production as well as other income, such as personal current transfer receipts. In the state and local personal income accounts the personal income of an area represents the income received by or on behalf of the persons residing

Predictor	Source*	Level of Aggregation (based on residence)	Description
			in that area. It is calculated as the sum of wages and salaries, supplements to wages and salaries, proprietors' income with inventory valuation (IVA) and capital consumption adjustments (CCAdj), rental income of persons with capital consumption adjustment (CCAdj), personal dividend income, personal interest income, and personal current transfer receipts, less contributions for government social insurance plus the adjustment for residence.
Per Capita Retirement & Other	DOC	County	Personal current transfer receipts excluding unemployment insurance compensation and income maintenance benefits. Retirement and other consists of retirement and disability insurance benefits, medical benefits, veterans' benefits, education and training assistance, other transfer receipts of individuals from governments, current transfer receipts of nonprofit institutions, and current transfer receipts of individuals from businesses.
Per Capita Unemployment Compensation	DOC	County	Unemployment insurance compensation is made up of the following: State unemployment compensation are benefits consisting mainly of the payments received by individuals under state-administered unemployment insurance (UI) programs, but they include the special benefits authorized by federal legislation for periods of high unemployment. The provisions that govern the eligibility, timing, and amount of benefit payments vary among the states, but the provisions that govern the coverage and financing are uniform nationally. Unemployment compensation of Federal civilian employees are benefits which are received by former federal civilian employees under a federal program administered by the state employment security agencies acting as agents for the U.S. Government. Unemployment compensation of railroad employees are benefits which are received by railroad workers who are unemployed because of sickness or because work is unavailable in the railroad industry and in related industries, such as carrier affiliates. This UI program is administered by the Railroad Retirement Board (RRB) under a federal formula that is applicable throughout the Nation. Unemployment compensation of veterans are benefits which are received by unemployed veterans who have recently separated from military service and who are not eligible for military retirement benefits. Trade adjustment assistance are benefits received by workers who are unemployed because of the adverse economic effects of international trade arrangements.
Percent Joint Tax Returns	IRS	Zip Code	Percent of Joint Tax Returns
Percent Prepared Returns	IRS	Zip Code	Percent of Prepared Returns
Percent Returns with Child Care Credit	IRS	Zip Code	Percent of Returns with Child Care Credit

Predictor	Source*	Level of Aggregation (based on residence)	Description
Percent Returns with Earned Income Credit	IRS	Zip Code	Percent of Returns with Earned Income Credit
Percent Returns with Pension or Annuity Income	IRS	Zip Code	Percent of Returns with Pension or Annuity Income
Percent Returns with Retirement Distributions	IRS	Zip Code	Percent of Returns with Retirement Distributions
Percent Returns with Self Employment Income	IRS	Zip Code	Percent of Returns with Self Employment Income
Percent Returns with Unemployment Income	IRS	Zip Code	Percent of Returns with Unemployment Income
Property Crime Rate	DOJ	County	Property crime rate within county of residence
Received Special Pay	DMDC	Individual	Bonuses, housing allowances, hazardous duty, and other special duty payments.
Serving in Air Force	DMDC	Individual	Serving in Air Force
Serving in Army	DMDC	Individual	Serving in Army
Serving in Navy	DMDC	Individual	Serving in Navy
Violent Crime Rate	DOJ	County	Violent crime rate within county of residence

*DMDC=Defense Manpower Data Center; DOC=Department of Commerce; DOJ=Department of Justice; IRS=Internal Revenue Service

APPENDIX B:

CRITERION MEASURE RULES

Table B-1
Criterion Measure Rules

Criterion Measure	Source	Rule
Unsuitability Discharge	Active Duty Transaction File	• Inter-Service Separation Codes = • 1060: Character or Behavior Disorder • 1061: Motivational Problems (Apathy) • 1064: Alcoholism • 1065: Discreditable Incidents - Civilian or Military • 1066: Shirking • 1067: Drugs • 1068: Financial Irresponsibility • 1069: Lack of Dependent Support • 1071: Civil Court Conviction • 1072: Security • 1073: Court Martial • 1074: Fraudulent Entry • 1075: AWOL, Desertion • 1077: Sexual Perversion • 1078: Good of the Service (In lieu of Court-Martial) • 1079: Juvenile Offender • 1080: Misconduct (Reason Unknown) • 1081: Unfitness (Reason Unknown) • 1082: Unsuitability (Reason Unknown) • 1083: Pattern of Minor Disciplinary Infractions • 1084: Commission of a Serious Offense • 1086: Expeditious Discharge/Unsatisfactory Performance • 1087: Trainee Discharge/Entry Level Performance and Conduct • 1098: Breach of Contract • 1101: Dropped from Strength for Desertion • 1102: Dropped from Strength for Imprisonment
DCII Investigation	DCII	Context Code = S: Subject of investigation
Security Incident	JPAS Incident Table	Existence of record in the Incident table. Example 1: Subj was convicted by a Special Courts-Martial of 3 counts of wrongful use of marijuana and sentenced to reduction to E-1, confinement for 45 days, forfeiture of $759. Example 2: Subj counseled for having a bottle of Jim Beam is his room, a violation of a written order. Example 3: Subj on Military Police blotter for Domestic assault on spouse and damage to Government property.

Criterion Measure	Source	Rule
Access Suspended	JPAS Access Table	Access Code = 3: Pending Reply to SOR X: Action Pending Y: Access Suspended

APPENDIX C:

CATEGORICAL PREDICTOR COMPARISON GROUPS

Table C-1
Predictors and Reference Groups

Predictor	Reference Group
College Graduate	UNKNOWN*
Deployed	Not Deployed
Divorced	UNKNOWN
E1-E4	W1 - W5 (i.e., rank of Warrant Officer)
E5-E9	W1 - W5 (i.e., rank of Warrant Officer)
Never Married	UNKNOWN
High School Graduate	UNKNOWN
Injured in Hostile Situation	Not Injured in Hostile Situation
Injured in Non-Hostile Situation	Not Injured in Non-Hostile Situation
Married	UNKNOWN
O1-O3	W1 - W5 (i.e., rank of Warrant Officer)
Received Special Pay	Did Not Receive Special Pay
Serving in Air Force	Serving in Marines
Serving in Army	Serving in Marines
Serving in Navy	Serving in Marines

* UNKNOWN is a valid value in some DMDC mainframe files. Its use in this report reflects the coded value in place in the DMDC files.