



NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

**MACHINE LEARNING, TIME SERIES, AND SURVIVAL
ANALYSIS FOR DESCRIBING SUBSTANCE ABUSE AND
RELAPSE AMONG U.S. NAVY PERSONNEL**

by

Chad J. Minnick

June 2022

Thesis Advisor:
Second Reader:

Ruriko Yoshida
Samuel E. Buttrey

Approved for public release. Distribution is unlimited.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			<i>Form Approved OMB No. 0704-0188</i>	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC, 20503.				
1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE June 2022		3. REPORT TYPE AND DATES COVERED Master's thesis
4. TITLE AND SUBTITLE MACHINE LEARNING, TIME SERIES, AND SURVIVAL ANALYSIS FOR DESCRIBING SUBSTANCE ABUSE AND RELAPSE AMONG U.S. NAVY PERSONNEL			5. FUNDING NUMBERS	
6. AUTHOR(S) Chad J. Minnick				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			10. SPONSORING / MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release. Distribution is unlimited.			12b. DISTRIBUTION CODE A	
13. ABSTRACT (maximum 200 words) The U.S. Navy (USN) has targeted drug and alcohol abuse within its ranks by means of screening, training, and treatment; however, substance abuse remains a threat to the mission and to the personal health and well-being of USN service members. Therefore, we explore new methodology to identify detrimental behaviors associated with relapse and to forecast substance abuse occurrences. Furthermore, survival times for relapses are assessed to inform treatment program enrollment timelines. This research uses the USN Alcohol and Drug Management Information Tracking System (ADMITS) data and classification, time series, and survival analysis machine-learning models to address our research objectives. We apply performance measures to assess a model's accuracy and goodness of fit to distinguish the top-performing models and their consideration for subsequent use. Additionally, we perform supplementary analysis on U.S. duty location, USN rank, and USN occupation subgroups to identify over- or under-representation over time in terms of substance abuse occurrences. Our results identified service members with prior relapses to be of most importance in predicting future relapse, strong seasonality in alcohol abuse occurrences, and a subset of USN service members who relapse faster than observed treatment timelines. Additionally, many U.S. duty location, USN rank, and USN occupation subgroups were discovered to be trending above or below the baseline proportion of alcohol abuse occurrences.				
14. SUBJECT TERMS substance abuse, relapse, ADMITS, machine learning, classification, time series analysis, survival analysis			15. NUMBER OF PAGES 159	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UU	

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release. Distribution is unlimited.

**MACHINE LEARNING, TIME SERIES, AND SURVIVAL ANALYSIS
FOR DESCRIBING SUBSTANCE ABUSE AND RELAPSE AMONG
U.S. NAVY PERSONNEL**

Chad J. Minnick
Captain, United States Air Force
BA, Purdue University – Fort Wayne, 2010

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

**NAVAL POSTGRADUATE SCHOOL
June 2022**

Approved by: Ruriko Yoshida
Advisor

Samuel E. Buttrey
Second Reader

W. Matthew Carlyle
Chair, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The U.S. Navy (USN) has targeted drug and alcohol abuse within its ranks by means of screening, training, and treatment; however, substance abuse remains a threat to the mission and to the personal health and well-being of USN service members. Therefore, we explore new methodology to identify detrimental behaviors associated with relapse and to forecast substance abuse occurrences. Furthermore, survival times for relapses are assessed to inform treatment program enrollment timelines. This research uses the USN Alcohol and Drug Management Information Tracking System (ADMITS) data and classification, time series, and survival analysis machine-learning models to address our research objectives. We apply performance measures to assess a model's accuracy and goodness of fit to distinguish the top-performing models and their consideration for subsequent use. Additionally, we perform supplementary analysis on U.S. duty location, USN rank, and USN occupation subgroups to identify over- or under-representation over time in terms of substance abuse occurrences. Our results identified service members with prior relapses to be of most importance in predicting future relapse, strong seasonality in alcohol abuse occurrences, and a subset of USN service members who relapse faster than observed treatment timelines. Additionally, many U.S. duty location, USN rank, and USN occupation subgroups were discovered to be trending above or below the baseline proportion of alcohol abuse occurrences.

THIS PAGE INTENTIONALLY LEFT BLANK

Table of Contents

1	Introduction	1
1.1	Motivation	1
1.2	Objective	3
1.3	Structure.	5
2	Background and Literature Review	7
2.1	Over- or Under-Representation by State, Rank, and Rating	7
2.2	Machine Learning Methods	10
2.3	Time Series	18
2.4	Survival Analysis	26
2.5	Performance Evaluation.	33
2.6	Literature Review	41
3	Methodology	45
3.1	Methodology Overview.	45
3.2	Data	47
3.3	Model Fitting and Implementation	56
4	Analysis and Results	69
4.1	Model Results and Analysis	69
5	Conclusion	101
5.1	Summary	101
5.2	Future Work	105
	Appendix: State, Rank, and Rating Visuals	107
A.1	AD and Reserve Alcohol Abuse Over- or Under-Representation by State . .	107
A.2	AD and Reserve Alcohol Abuse Over- or Under-Representation by Rank . .	111

A.3 AD Alcohol Abuse Over- or Under-Representation by Rating	117
List of References	127
Initial Distribution List	131

List of Figures

Figure 2.1	Total Alcohol Incidents by Region	8
Figure 2.2	Total Alcohol Incidents by Rank	8
Figure 2.3	Total Alcohol Incidents by Department	9
Figure 2.4	Machine Learning Process Flow	13
Figure 2.5	Bias Variance Trade-off	14
Figure 2.6	k-Nearest Neighbors (kNN) Classification	17
Figure 2.7	Seasonality, Trend, and Cycles	20
Figure 2.8	Train-Test Split for Time Series Data	21
Figure 2.9	Seasonal Decomposition of Time Series Data	22
Figure 2.10	Censoring Methods for Survival Analysis	28
Figure 2.11	Right-censoring Survival Analysis Data	28
Figure 2.12	Kaplan-Meier (KM) Curve for Survival Analysis	30
Figure 2.13	Random Survival Forest (RSF) Classification Tree	32
Figure 2.14	Confusion Matrix for Binary Classification	35
Figure 2.15	Classification Model Threshold Adjustments	36
Figure 2.16	Receiver Operating Characteristic (ROC) Curves and Associated Area Under the (ROC) Curve (AUC) Percentages	37
Figure 2.17	Time Series Rolling Horizons Design	39
Figure 3.1	Total Alcohol Incidents per Month (October 2008–August 2013) .	53
Figure 3.2	Total Alcohol Incidents per Month (January 2015–May 2020) . .	53
Figure 3.3	Total Other Drug Incidents per Month (October 2008–August 2013)	54

Figure 3.4	Total Other Drug Incidents per Month (January 2015–May 2020)	54
Figure 3.5	Alcohol and Drug Management Information Tracking System (ADMITS) Data Set Missing Values	56
Figure 4.1	Active Duty (AD) U.S. Navy (USN) Navy 2020 Alcohol Substance Abuse Over- or Under-Representation	70
Figure 4.2	AD USN E-4 Alcohol Substance Abuse Over- or Under-Representation	71
Figure 4.3	AD USN BU General Rating Alcohol Substance Abuse Over- or Under-Representation	72
Figure 4.4	Alcohol Abuse Pre-System Outage Classification Model ROC Plots	73
Figure 4.5	Alcohol Abuse Pre-System Outage Feature Importance for Random Forest Model	74
Figure 4.6	Alcohol Abuse Post-System Outage Classification Model ROC Plots	75
Figure 4.7	Alcohol Abuse Post-System Outage Feature Importance for Random Forest Model	76
Figure 4.8	Drug and Alcohol Report (DAR) Entries for Abused Substances Other Than Alcohol	77
Figure 4.9	Other-Drug Abuse Pre-System Outage Classification Model ROC Plots	78
Figure 4.10	Other-Drug Abuse Pre-System Outage Feature Importance for Random Forest Model	79
Figure 4.11	Other-Drug Abuse Post-System Outage Classification Model ROC Plots	80
Figure 4.12	Other-Drug Abuse Post-System Outage Feature Importance for Random Forest Model	81
Figure 4.13	Total Non-NA SARP Features for Pre- and Post-System Outage Alcohol Incidents	82
Figure 4.14	Total Non-NA SARP Features for Pre- and Post-System Outage Other-Drug Incidents	82

Figure 4.15	Seasonal Decomposition for Alcohol Time Series Data	84
Figure 4.16	Alcohol Incident Forecasting Model Performance	85
Figure 4.17	Seasonal Autoregressive Integrated Moving Average (SARIMA) Model Alcohol Incident Forecast	86
Figure 4.18	SARIMA Model Residuals from Alcohol Incident Forecast . . .	87
Figure 4.19	Seasonal Decomposition for Other-Drug Time Series Data	88
Figure 4.20	Other-Drug Incident Forecasting Model Performance	89
Figure 4.21	SARIMA Model Other-Drug Incident Forecast	90
Figure 4.22	SARIMA Model Residuals from Other-Drug Incident Forecast . .	90
Figure 4.23	Number of Days Between Repeat Alcohol Incidents Versus Days Between Substance Abuse Rehabilitation Program (SARP) Screening	92
Figure 4.24	Number of Days Between Repeat Other-Drug Incidents Versus Days Between SARP Screening	93
Figure 4.25	KM Survival Curve for Alcohol Incidents	94
Figure 4.26	Cox Proportional-Hazards Regression (CPHR) Survival Curve for Alcohol Incidents	95
Figure 4.27	RSF Survival Curve for Alcohol Incidents	95
Figure 4.28	12-month KM Survival Curve for Alcohol Incidents	96
Figure 4.29	KM Survival Curve for Other-Drug Incidents	97
Figure 4.30	CPHR Survival Curve for Other-Drug Incidents	97
Figure 4.31	RSF Survival Curve for Other-Drug Incidents	98
Figure 4.32	12-month KM Survival Curve for Other-Drug Incidents	99
Figure A.1	AD USN 2008 and 2009 Alcohol Substance Abuse Over- or Under- Representation	107
Figure A.2	AD USN 2010 and 2011 Alcohol Substance Abuse Over- or Under- Representation	107

Figure A.3	AD USN 2012 and 2013 Alcohol Substance Abuse Over- or Under-Representation	108
Figure A.4	AD USN 2015 and 2016 Alcohol Substance Abuse Over- or Under-Representation	108
Figure A.5	AD USN 2017 and 2018 Alcohol Substance Abuse Over- or Under-Representation	108
Figure A.6	AD USN 2019 and 2020 Alcohol Substance Abuse Over- or Under-Representation	109
Figure A.7	Reserve USN 2008 and 2009 Alcohol Substance Abuse Over- or Under-Representation	109
Figure A.8	Reserve USN 2010 and 2011 Alcohol Substance Abuse Over- or Under-Representation	109
Figure A.9	Reserve USN 2012 and 2013 Alcohol Substance Abuse Over- or Under-Representation	110
Figure A.10	Reserve USN 2015 and 2016 Alcohol Substance Abuse Over- or Under-Representation	110
Figure A.11	Reserve USN 2017 and 2018 Alcohol Substance Abuse Over- or Under-Representation	110
Figure A.12	Reserve USN 2019 and 2020 Alcohol Substance Abuse Over- or Under-Representation	111
Figure A.13	AD USN E-1 and E-2 Alcohol Substance Abuse Over- or Under-Representation	111
Figure A.14	AD USN E-3 and E-4 Alcohol Substance Abuse Over- or Under-Representation	112
Figure A.15	AD USN E-5 and E-6 Alcohol Substance Abuse Over- or Under-Representation	112
Figure A.16	AD USN E-7 and E-8 Alcohol Substance Abuse Over- or Under-Representation	112
Figure A.17	AD USN E-9 Alcohol Substance Abuse Over- or Under-Representation	113

Figure A.18	AD USN CW-2 and CW-3 Alcohol Substance Abuse Over- or Under-Representation	113
Figure A.19	AD USN CW-4 Alcohol Substance Abuse Over- or Under-Representation	113
Figure A.20	AD USN O-1 and O-2 Alcohol Substance Abuse Over- or Under-Representation	114
Figure A.21	AD USN O-3 and O-4 Alcohol Substance Abuse Over- or Under-Representation	114
Figure A.22	AD USN O-5 and O-6 Alcohol Substance Abuse Over- or Under-Representation	114
Figure A.23	Reserve USN E-1 and E-2 Alcohol Substance Abuse Over- or Under-Representation	115
Figure A.24	Reserve USN E-3 and E-4 Alcohol Substance Abuse Over- or Under-Representation	115
Figure A.25	Reserve USN E-5 and E-6 Alcohol Substance Abuse Over- or Under-Representation	116
Figure A.26	Reserve USN E-7 and E-8 Alcohol Substance Abuse Over- or Under-Representation	116
Figure A.27	Reserve USN O-3 and O-4 Alcohol Substance Abuse Over- or Under-Representation	116
Figure A.28	AD USN AB and AC General Rating Alcohol Substance Abuse Over- or Under-Representation	117
Figure A.29	AD USN AD and AE General Rating Alcohol Substance Abuse Over- or Under-Representation	117
Figure A.30	AD USN AG and AM General Rating Alcohol Substance Abuse Over- or Under-Representation	118
Figure A.31	AD USN AO and AS General Rating Alcohol Substance Abuse Over- or Under-Representation	118
Figure A.32	AD USN AT and AW General Rating Alcohol Substance Abuse Over- or Under-Representation	118

Figure A.33	AD USN AZ and BM General Rating Alcohol Substance Abuse Over- or Under-Representation	119
Figure A.34	AD USN BU and CE General Rating Alcohol Substance Abuse Over- or Under-Representation	119
Figure A.35	AD USN CM and CS General Rating Alcohol Substance Abuse Over- or Under-Representation	119
Figure A.36	AD USN CT and DC General Rating Alcohol Substance Abuse Over- or Under-Representation	120
Figure A.37	AD USN EA and EM General Rating Alcohol Substance Abuse Over- or Under-Representation	120
Figure A.38	AD USN EN and EO General Rating Alcohol Substance Abuse Over- or Under-Representation	120
Figure A.39	AD USN EOD and ET General Rating Alcohol Substance Abuse Over- or Under-Representation	121
Figure A.40	AD USN FC and FT General Rating Alcohol Substance Abuse Over- or Under-Representation	121
Figure A.41	AD USN GM and GSM General Rating Alcohol Substance Abuse Over- or Under-Representation	121
Figure A.42	AD USN HM and HT General Rating Alcohol Substance Abuse Over- or Under-Representation	122
Figure A.43	AD USN IC and IS General Rating Alcohol Substance Abuse Over- or Under-Representation	122
Figure A.44	AD USN IT and LN General Rating Alcohol Substance Abuse Over- or Under-Representation	122
Figure A.45	AD USN MA and MC General Rating Alcohol Substance Abuse Over- or Under-Representation	123
Figure A.46	AD USN MM and MN General Rating Alcohol Substance Abuse Over- or Under-Representation	123
Figure A.47	AD USN MR and MT General Rating Alcohol Substance Abuse Over- or Under-Representation	123

Figure A.48	AD USN MU and NC General Rating Alcohol Substance Abuse Over- or Under-Representation	124
Figure A.49	AD USN ND and OS General Rating Alcohol Substance Abuse Over- or Under-Representation	124
Figure A.50	AD USN PR and PS General Rating Alcohol Substance Abuse Over- or Under-Representation	124
Figure A.51	AD USN QM and RP General Rating Alcohol Substance Abuse Over- or Under-Representation	125
Figure A.52	AD USN RS and SB General Rating Alcohol Substance Abuse Over- or Under-Representation	125
Figure A.53	AD USN SO and ST General Rating Alcohol Substance Abuse Over- or Under-Representation	126
Figure A.54	AD USN SW and TM General Rating Alcohol Substance Abuse Over- or Under-Representation	126
Figure A.55	AD USN UT and YN General Rating Alcohol Substance Abuse Over- or Under-Representation	126

THIS PAGE INTENTIONALLY LEFT BLANK

List of Tables

Table 3.1	ADMITS Data Set Features and Descriptions	49
Table 3.2	DMDC Rank Data Set Variables and Descriptions	51
Table 3.3	DMDC State Data Set Variables and Descriptions	51
Table 3.4	Navy Enlisted System (NES) Rating Data Set Variables and Descriptions	52

THIS PAGE INTENTIONALLY LEFT BLANK

List of Acronyms and Abbreviations

ACF	Autocorrelation Function
AD	Active Duty
ADMITS	Alcohol and Drug Management Information Tracking System
AIC	Akaike Information Criterion
ARIMA	Autoregressive Integrated Moving Average
AUC	Area Under the (ROC) Curve
AUDIT	Alcohol Use Disorders Identification Test
BIC	Bayesian Information Criterion
BUPERS-34	Bureau of Naval Personnel Metrics and Analytics Branch
CART	Classification and Regression Tree
CNO	Chief of Naval Operations
CPHR	Cox Proportional-Hazards Regression
CHF	Cumulative Hazard Function
DAR	Drug and Alcohol Report
DEERS	Defense Enrollment Eligibility Reporting System
DMDC	Defense Manpower Data Center
DOD	Department of Defense
DON	Department of the Navy
ETS	Error, Trend, and Seasonality

EDA	Exploratory Data Analysis
FNR	False Negative Rate
FPR	False Positive Rate
HW	Holt-Winters
HW-A	Holt-Winters Additive
HW-M	Holt-Winters Multiplicative
KM	Kaplan-Meier
kNN	k-Nearest Neighbors
LGMM	Latent Growth Mixture Model
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MASE	Mean Absolute Scaled Error
NES	Navy Enlisted System
NFL	National Football League
OPNAV	Office of the Chief of Naval Operations
PACF	partial autocorrelation function
PII	Personally Identifiable Information
PTSD	Post-Traumatic Stress Disorder
RAN	Royal Australian Navy
RMSE	Root Mean Squared Error
RSF	Random Survival Forest
ROC	Receiver Operating Characteristic

SARIMA	Seasonal Autoregressive Integrated Moving Average
SARP	Substance Abuse Rehabilitation Program
SECNAV	Office of the Secretary of the Navy
STL	Seasonal and Trend Decomposition using Loess
SUD	Substance Use Disorder
TNR	True Negative Rate
TPR	True Positive Rate
USN	U.S. Navy
WHO	World Health Organization

THIS PAGE INTENTIONALLY LEFT BLANK

Executive Summary

Alcohol abuse is defined by the Office of the Secretary of the Navy (SECNAV) (2019) as the “use of alcohol to an extent that it has an adverse effect on performance, conduct, behavior, mission effectiveness, the user’s health, family, community, Department of the Navy (DON) or leads to one or more acts of alcohol-related misconduct.” With the same adverse effects as alcohol abuse, drug abuse is the “wrongful use of a controlled substance, prescription medication, over-the-counter medication, or intoxicating substance (other than alcohol)” (SECNAV 2019). Despite USN efforts to counter alcohol and drug abuse, substance abuse still remains a threat to the USN mission and the readiness of USN service members.

The Office of the Chief of Naval Operations (OPNAV) N17 directorate has interest in combating substance abuse within the USN. OPNAV N17 bears responsibility for fulfilling objectives in key programs such as personal, family, and physical readiness, and substance abuse prevention (Chief of Naval Operations [CNO] 2013). As such, an OPNAV N17 objective is to proactively counter substance abuse by identifying risk factors for relapses.

This thesis employs machine learning models built using historical data from ADMITS, which stores substance abuse incidents, or DARs, for USN service members. We apply classification models to predict future substance abuse incidents, or relapses, after USN service members have a DAR. We also use time-series analysis to forecast DARs and assess trends and seasonality. Survival analysis models were built to evaluate the survival probability for repeat incidents over time among USN service members with at least one DAR. Lastly, we explored U.S. state, USN rank, and USN general ratings to evaluate whether subgroups were proportionally over- or under- represented in terms of DAR entries.

Our research identified trends in the proportion of alcohol abuse occurrences committed by subgroups within the USN. Seasonality was discovered for alcohol DARs though not as prevalent for other abused substances. Machine learning methods were effective towards classifying relapses and identifying risk factors. We also discovered that while the majority of USN service members with at least one DAR do not relapse, there exists a subset of the population that relapses at faster rates than observed timelines for enrollment into treatment programs (Office of the Chief of Naval Operations 2021a).

References

Chief of Naval Operations (2013) Establishment of Navy's Twenty-First Century Sailor Office (OPNAV N17). Technical Report NAVADMIN 153/13, Washington, DC, <https://navadmin-viewer.fly.dev/NAVADMIN/153/13>.

Office of the Chief of Naval Operations (2021a) Alcohol and Drug Management Information Tracking System (ADMITS). Dataset, years 2008 to 2020. Millington, TN.

Office of the Secretary of the Navy (SECNAV) (2019) Military substance abuse prevention and control. Technical Report SECNAVINST 5300.28F, Washington, DC, <https://www.secnav.navy.mil/doni/Directives/05000%20General%20Management%20Security%20and%20Safety%20Services/05-300%20Manpower%20Personnel%20Support/5300.28F.pdf>

Acknowledgments

This thesis is dedicated to my unbelievably supportive wife, Vanessa Minnick, and the two joys of our life, Calvin and Sofia Minnick. I vow to encourage, support, and be there for you in the pursuit of your dreams as you have exceedingly demonstrated towards me in the pursuit of my own. My praise to God for bringing each of you into my life.

Additionally, I dedicate this thesis to veterans who battle with substances. I pray that you are able to find healing from the afflictions that brought substances into your life. I hope this research is able to contribute in some way towards bringing you the unconditional care and treatment that you, and your family, deserve.

My sincerest gratitude to Professor Ruriko Yoshida and Professor Samuel Buttrey. I consider myself extremely fortunate to have learned from the both of you, and am beyond grateful for your invaluable time, energy, and guidance. Professor Yoshida, the passion that you have for helping others and towards your field of expertise puts wind in the sails of those around you. Professor Buttrey, your brilliance, sense of humor, and genuine desire to share your love of learning is one of a kind, and refreshing.

I would also like to thank the mountain of mentors, leaders, friends, family, and all of those that I have served alongside for believing in me and lifting me up to the opportunities that I have had throughout my career. I truly never consider any successes to be my own, and I hope to be a step up to your success as well.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 1:

Introduction

This chapter sets forth the purpose of our research and calls attention to the narrative of substance abuse within the U.S. Navy (USN). Thereafter, we address our studies' objectives, and present the thesis' framework.

1.1 Motivation

Substance abuse has a long history of being detrimental to populations, and the US military is not immune to its negative effects. The USN's goal is to be rid of the deleterious impact substance abuse inflicts upon the USN, thwarting the readiness of its personnel and hindering the USN's ability to execute its mission at its highest potential (Office of the Chief of Naval Operations [OPNAV] 2009). USN leaders have underscored the personal responsibility of USN service members to act in a lawful manner should they choose to use alcohol (OPNAV 2009; SECNAV 2019). At the same time, USN leadership acknowledges that alcohol abuse and dependence are treatable conditions (OPNAV 2009). Disciplinary measures for drug abuse or dependence can lead to separation from service for any deviation from the USN's strict zero tolerance policy (OPNAV 2009). The USN command and organizational responsibilities are two-fold: upholding a culture that is intolerant to substance abuse, and providing treatment to service members when necessary (OPNAV 2009; SECNAV 2019).

The USN has targeted drug and alcohol abuse within its ranks by means of screening, training, and treatment (OPNAV 2009). Identifying members who abuse substances or have formed dependencies upon them is crucial not just to the mission of the organization that the member belongs to, but even more importantly, to the member. Without proper treatment and care, substance abuse and dependencies can evolve into a Substance Use Disorder (SUD) diagnosis, which can have many negative consequences to the member such as medical or mental health issues, interpersonal complications both within families or workcenters, and increased risk of suicidal thoughts or attempts (Teeters et al. 2017). Research shows that SUD issues are often complex in that they stem from one or many root causes depending upon the individual (Teeters et al. 2017). For military personnel, risk

factors for substance abuse have been connected with deployments, exposure to combat, and reintegration conflicts when returning home (Teeters et al. 2017). Furthermore, Post-Traumatic Stress Disorder (PTSD), depression, physical or sexual abuse as a child, are other links to catalysts for SUDs and there are yet others (Teeters et al. 2017). The impact to the service member and the USN can be extensive by proactively identifying and treating substance abuse issues before they become course-altering.

The Alcohol and Drug Management Information Tracking System (ADMITS) is the primary information management system for the Navy's Substance Abuse Rehabilitation Program (SARP) (OPNAV 2009). ADMITS is utilized by commanders, medical personal and SARP counselors as a central database for all USN Drug and Alcohol Report (DAR), screening, and treatment entries (OPNAV 2021b). Moreover, ADMITS usage is not restricted to reporting substance abuse incidents or DARs, but also for measuring the effectiveness of Navy substance abuse programs and ensuring compliance with the Navy's drug and alcohol policies throughout the USN (OPNAV 2021b). ADMITS is overseen by Office of the Chief of Naval Operations (OPNAV) N17, otherwise known as the 21st Century Sailor Office (CNO 2013). OPNAV N17 is a directorate responsible for ensuring policy objectives are met across the following programs: personal and family readiness, physical readiness, substance abuse prevention, suicide prevention, sexual assault prevention and response, equal opportunity, and hazing prevention (CNO 2013).

Despite USN undertakings to combat substance abuse through education, deterrence, and rehabilitation programs, substance abuse remains an issue within the USN. In 2015, a RAND Corporation survey of Department of Defense (DOD) service members found that of 3,016 USN Active Duty (AD) service members surveyed, about 34% reported binge drinking (Meadows et al. 2018). Furthermore, about 67% of USN AD members reported that the military culture was supportive of drinking (Meadows et al. 2018). The RAND Corporation survey was repeated in 2018 and included responses from 3,675 USN AD service members (Meadows et al. 2021). While about 29% of USN AD members reported that the military culture was supportive of drinking, a 39% decrease from 2015, 43% reported binge drinking within the past month, a 9% increase from 2015 (Meadows et al. 2021). The 2018 RAND Corporation survey also asked DOD service members about their use of drugs such as marijuana, nonprescription steroids, cocaine, methamphetamine, etc. (Meadows et al. 2021). The findings were not broken down by DOD branch of service; however, they

found that 1.3% of AD service members reported use of illicit drugs (Meadows et al. 2021). The most commonly reported drug abused was marijuana or synthetic cannabis with 0.9% of service members reporting use within the past 12 months (Meadows et al. 2021).

To bring the USN closer to the realization of a fleet free of the negative toll substance abuse can inflict, we explore new methodology to potentially augment prior methods. If the USN can understand some of the red flags that can place a service member at higher risk of abusing substances or relapsing, then potentially it can get ahead of the issue before it becomes problematic. Supervised machine learning models have shown success in identifying factors that place individuals at risk for substance abuse and/or relapsing after treatment for a SUD (Islam et al. 2020). Also, machine learning algorithms have been applied to military substance abuse research advancing our understanding of risk-factors associated with unique military experiences and demographics (Betancourt et al. 2021; McKenzie et al. 2006; Fink et al. 2017). As machine learning models enable analysts to predict, forecast, and assess potential risk-factors, there is much to be gained in terms of new perspective and insight that can better inform decision makers to lessen the issue of military substance abuse and better serve the service member.

1.2 Objective

Our first objective, determining over- or under- representation among subgroups, is to attain a better understanding as to the true contribution of alcohol substance abuse occurrences by subsets of the USN population. Specifically, we will examine subgroups within USN ranks, USN general ratings, or U.S. state duty locations by comparing the proportion of occurrences by each subgroup to the subgroup's percentage of the USN population.

For our second objective, we desire to predict future substance abuse for USN service members through classification models and identify important risk factors among the variables used within our best models. Our observations within the ADMITS data consists of USN service members that have a DAR entry for a substance abuse occurrence. Therefore, our aim is to equip decision makers and caregivers to better understand what factors place one at higher risk of relapsing in the future given their current status. Unfortunately, one limitation of the ADMITS data set is that the ADMITS system experienced an outage for about one year. There were some entries during the outage period, which suggests that the

outage was not experienced across all USN servers though these observations were removed to eliminate any potential bias from limited access to the system. Moreover, OPNAV N17 requested separate classification models for the data prior to and after the outage to contrast the risk factors for the differing time periods. Furthermore, OPNAV N17 expressed a strong interest in differentiating alcohol abuse from other substances abused, therefore, we will build separate classification models to classify relapses for alcohol and other drugs.

Thereafter, we seek to forecast alcohol and other-drug substance abuse occurrences through the application of time-series analysis. The ambition of our time-series models is three-fold: (1) visualize the impact of the ADMITS system outage and assess system recovery, (2) identify trends and seasonality in alcohol and other-drug occurrences, and (3) produce our best 12-month forecast for substance abuse occurrences within the USN. By identifying annual seasonality in alcohol or other-drug incidents, leaders can better anticipate spikes in occurrences and proactively prepare for such spikes through various counter-measures.

Lastly, we will generate survival analysis models for both alcohol and other-drug occurrences. This analysis will afford our readers insight into the survival probability of the USN population that is at-risk of relapsing. Survival, in terms of our analysis, equates to a USN service member not experiencing a relapse incident as captured by ADMITS. The survival probability of the population at-risk will be measured as a function of time since the last substance abuse incident. The analysis will be beneficial in assessing the survival of the broader USN population that has experienced at least one substance abuse occurrence. In addition, the discernment of the survival probability curves as a function of survival time can reinforce or focus dialogue concerning appropriate treatment timelines for those at-risk of relapsing.

Explicitly, we examine subgroups within several features, and then we employ classification algorithms, time-series analysis, and survival analysis machine learning models to address the following objectives:

1. Are there subgroups within USN ranks, USN general ratings, or working within U.S. states that are over- or under-represented with respect to their percentage of the USN force and their percentage contribution of total alcohol substance abuse occurrences?
2. Through the application of statistical learning models, by partitioning the ADMITS data set into pre- and post- ADMITS system outage data and separately analyzing

alcohol and other-drug related substance abuse:

- (a) With what degree of accuracy can we correctly classify relapse occurrences among Sailors?
 - (b) What risk-factors are deemed important by classification models for predicting substance abuse relapse incidents?
3. With what degree of accuracy can we forecast alcohol or other-drug substance abuse occurrences?
 4. What is the survival probability over time of USN service members with respect to repeat substance abuse events and the number of days between incidents?

1.3 Structure

The forthcoming chapters in this thesis build up to the results of our analysis in addressing the research objectives. Chapter 2 renders a view of our analytical approach, the foundation of our models, and concludes with a review of pertinent literature to our study. Chapter 3 acquaints the reader with the data sets used in our analysis, as well as the specific functions and parameters used to build our classification, time series, and survival analysis models. Chapter 4 captures the results of our analysis with inferences and visualizations where appropriate. Chapter 5 concludes our research with key findings and recommendations for future research.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 2:

Background and Literature Review

This chapter opens with our motivation for assessing the proportion of occurrences among subgroups of state, rank, and rating populations. In section 2.2, we discuss the machine learning models and our algorithmic approach for classifying future alcohol or other-drug relapses. Section 2.3 defines key terms and the models that were applied for our time-series analysis of alcohol and other-drug occurrences. Within section 2.4 we cover survival analysis terminology and discuss the models that were built to assess survival probabilities in terms of repeat alcohol or other-drug use. The performance measures used to evaluate the machine learning, time series, and survival analysis models are reviewed within section 2.5. We conclude this chapter with a review of literature pertinent to our analysis.

2.1 Over- or Under-Representation by State, Rank, and Rating

As with civilian populations, a snapshot of USN military and personnel demographics at a particular time will reflect a diverse force of various age, racial, and ethnic groups, educational attainment, work experience, places of birth, and so on. Furthermore, USN service members are stationed at bases throughout the United States and abroad, achieve various ranks as they progress throughout their careers, and serve within distinct occupational specialities otherwise known as sailors' ratings. Figures 2.1, 2.2, and 2.3 provide a visual depiction of the total number of alcohol substance abuse occurrences by region, rank, and department. In these figures states were consolidated to regions, and ratings were grouped by overarching command departments.

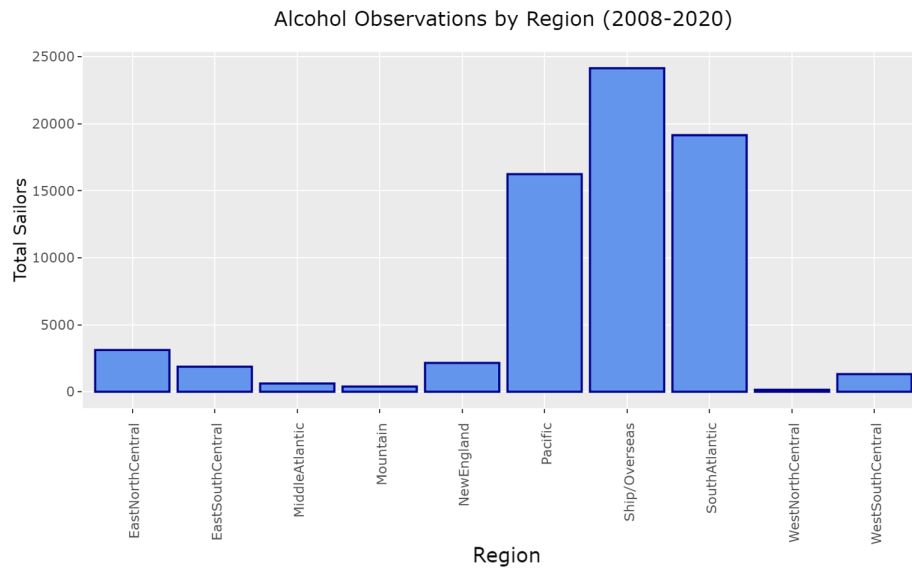


Figure 2.1. Higher incidents of alcohol substance abuse were reported for USN service members at deployed or overseas locations, as well as, Pacific or South Atlantic regions of the United States from 2008–2020. Source: OPNAV (2021a).

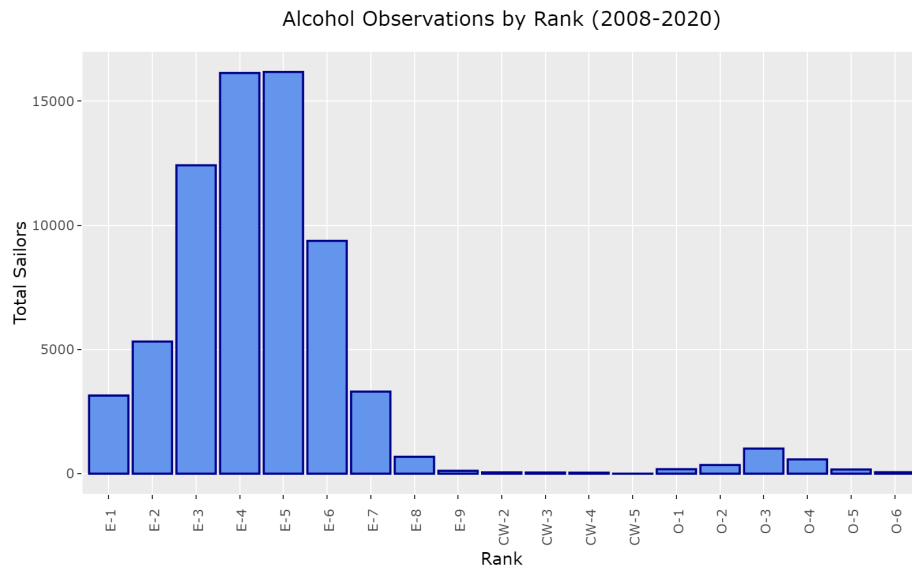


Figure 2.2. Enlisted USN service members, predominantly ranks E-3 to E-5, had the highest number of observed alcohol abuse occurrences from 2008–2020. Source: OPNAV (2021a).

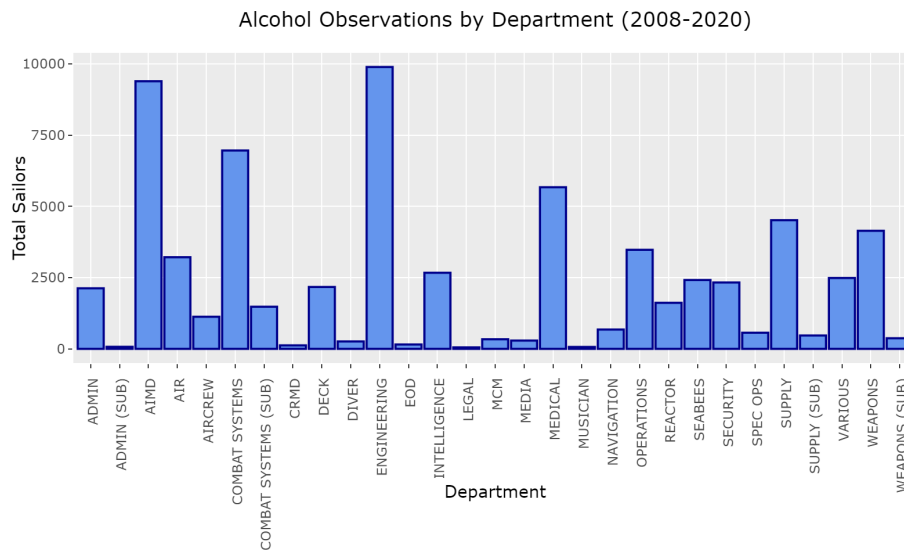


Figure 2.3. The highest number of alcohol substance abuse incidents among USN service members from 2008–2020 were observed in Engineering, Aircraft Intermediate Maintenance, and Combat Systems departments. Source: OP-NAV (2021a).

A surface level view of Figures 2.1, 2.2, and 2.3 does not provide a solid basis for concluding that certain demographics are more susceptible to alcohol abuse for the same reason that one would expect there to be more traffic violations in Texas than North Dakota: different population sizes. State military population data from the Defense Manpower Data Center (DMDC) indicates that approximately 48 percent of the USN AD and Reserve service members who reside within the U.S. were stationed within California or Virginia in 2021 (DMDC 2021). Comparably, DMDC rank population data reflects that 83 percent of USN service members were of enlisted ranks in 2021, and 60 percent of those enlisted held the rank of E-3 to E-5 (DMDC 2021). As for the USN departments of ratings that carried the bulk of the alcohol observations from 2008 to 2020, Engineering, Aircraft Intermediate Maintenance, and Combat Systems contained ratings for approximately 44 percent of the USN enlisted force in 2021 based on data provided by the Bureau of Naval Personnel Metrics and Analytics Branch (BUPERS-34) (BUPERS-34 2021).

The total number of alcohol substance abuse occurrences is presumably larger for certain state, rank, or rating demographics than others as a mere product of their population sizes.

Therefore, the total proportion of occurrences observed by each state, rank, or rating will be compared to the demographic's proportion of the population. For instance, if 20 percent of USN service members are located in Florida, and 30 percent of all USN alcohol incidents occurred in Florida that same year, then this would be categorized as an over-representation. The proportional representations for each state, rank, and rating as a function of time will be provided and discussed further within the results and analysis chapter.

2.2 Machine Learning Methods

Statistical machine learning is the use of mathematical algorithms applied to data and trained to identify patterns, structures, and correlations useful in generating models to aid forecasts, predictions, and/or data-driven decisions. There are numerous algorithms that can be employed upon data, sometimes in an ensemble fashion, but generally the choice will be dependent upon the analytical goals and needs of the decision maker. Statistical learning algorithms commonly reside within three paradigms, being supervised, unsupervised, and reinforcement learning (Huddleston and Brown 2019).

Supervised learning algorithms are practical when features or predictors within the data are used to predict either a numerical or categorical response variables. For numerical response variables, the supervised learning algorithm is known as regression, such as linear or non-linear regression or neural networks. For categorical response variables, algorithms are categorized as classification supervised learning algorithms such as classification trees and support vector machines (Huddleston and Brown 2019).

Unsupervised learning algorithms are useful in revealing veiled structure within data, and contrary to supervised learning, response variables are not required. Two prevalent applications of unsupervised learning are clustering and dimension reduction algorithms. Clustering is advantageous for partitioning data into homogeneous subgroups, and dimension reduction algorithms aid in reducing the complexity of data by discarding features with low explanatory power within the data. Data analysts will often employ unsupervised learning algorithms prior to supervised learning algorithms for more interpretive and predictive results (Huddleston and Brown 2019).

Reinforcement learning algorithms employ a trial and error process through the utilization of feedback from prior experience in conjunction with current learning to improve performance. At the onset, the algorithm perhaps has as much predictive power as the flip of a coin, but more iterations and subsequent learning bolsters the model's performance. Reinforcement learning algorithms are associated with Q-learning, deep adversarial networks, and artificial intelligence systems (Huddleston and Brown 2019).

For the purposes of this research, only supervised learning algorithms will be employed to generate classification, forecasting, and survival analysis models. Covered in greater detail within ensuing sections, this thesis utilized the following algorithms:

1. Classification:

- Logistic Regression
- k-Nearest Neighbors (kNN)
- Random Forest

2. Time Series Forecasting:

- Seasonal Decomposition
- Seasonal Autoregressive Integrated Moving Average (SARIMA)
- Holt-Winters (HW)
- Error, Trend, and Seasonality (ETS)

3. Survival Analysis:

- Kaplan-Meier (KM) Curves
- Cox Proportional-Hazards Regression (CPHR)
- Random Survival Forest (RSF)

2.2.1 Algorithmic Modeling Overview

An algorithmic modeling overview developed by Huddleston and Brown (2019) captures the recipe of this studies workflow. Huddleston and Brown (2019) recommend the following steps for deriving predictive solutions from supervised learning algorithms:

1. Data Acquisition and Cleaning:

The data cleaning commences upon conferring with subject matter experts for confidence that the data is appropriate to address the analytical intent of the work to

be conducted. The analyst prudently determines the best means for handling missing values, treating outliers, identifying anomalies, formatting, and standardizing the data. Conventionally, a cleaned n observations by p predictors table with a response variable is the end state for this step and readies the data for the ensuing steps.

2. Feature Engineering and Data Normalization:

Features, otherwise known as predictor or explanatory variables, contain associated values for observations within the data. Supervised learning algorithms utilize these features to predict the response variables. In this step, new features may be constructed via interactions, transformations, or extrapolations of established features to better explain or augment a model's predictive power. Data normalization, also referred to as scaling, standardizes the scale of feature values. This procedure is useful for algorithms which are sensitive to units for numerical variables. Furthermore, when numerical predictors are set to the same scale this also improves the interpretability of the model's coefficients, which are comparable if set to the same scale.

3. Train-Test Split:

Customarily, the use of supervised learning algorithms for generating predictive models is to do just that, predict. While there is definite utility in gleaning insight into current observational data, analysts and decision makers will perhaps be interested in applying developed models to impending data as well. One paramount step for gauging a model's performance and predictive potential is to conduct a train-test split on the data prior to model fitting and selection. Visualized within Figure 2.4, a data set is partitioned into subsets of data known as the training, validation, and test sets.

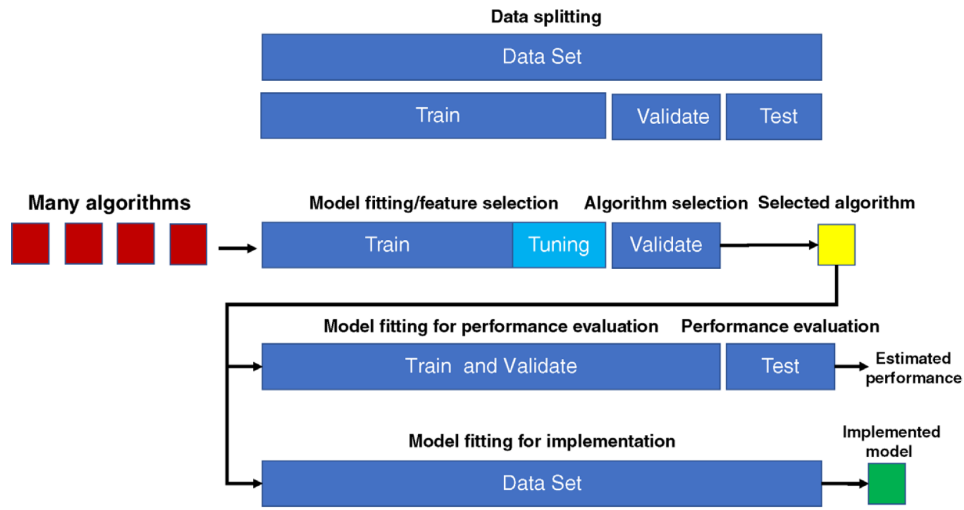


Figure 2.4. Machine learning process flow and the use of partitioned data for model fitting, feature selection, algorithm competition, performance evaluation, and implementation. Source: Huddleston and Brown (2019).

4. Model Fitting (Training) and Feature Selection:

Algorithmic models are then trained using the training set and the performance of a model is measured based upon how well the model fits to the newly introduced validation set observations. The cautionary tale and reasoning for this is that in training a model, the model could acclimate itself so well to the training data, known as “over-fitting” a model, that it does not generalize well to new observations. This process of adding complexity to a model through various means like additional features, interactions, or transformations becomes not a process of adding as much complexity as we can to fit the training data well, but just enough complexity to fit the test data well. This is also known as finding the “sweet spot” as a trade-off between bias and variance, which is observed in Figure 2.5.

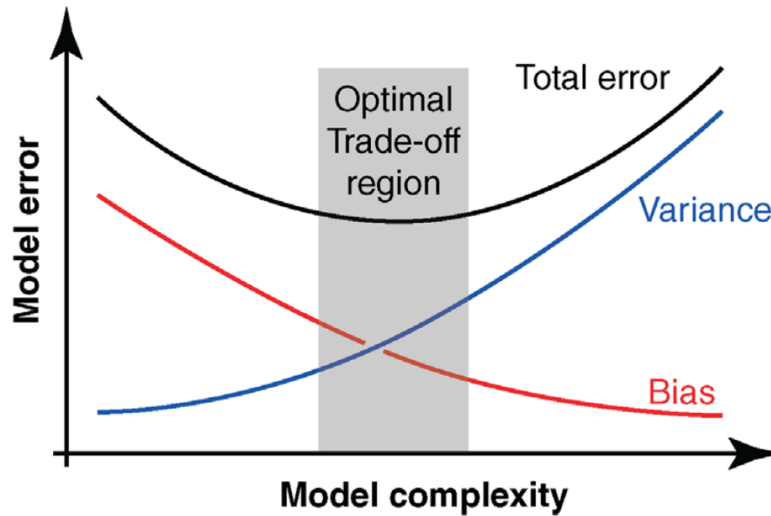


Figure 2.5. The bias-variance trade-off plot reflects the optimal trade-off region for bias and variance when training a model. A model that adds too much complexity to best fit the training data will not generalize well to new observations as indicated by the increase in total model error. Conversely, a model of insufficient complexity is also likely to have high error rates from unexplained bias. The goal is to add just enough complexity to the model, at the equilibrium where the bias and variance intersect, to sufficiently train the model, generalize well to new data, and minimize the total error. Source: Huddleston and Brown (2019).

The training of a model derives the model parameters. A process that is often concurrently employed with model fitting is feature selection. An analyst is equipped with several feature selection tactics to make use of such as step-wise feature selection, principal component analysis, cross validation, or even introducing penalties to induce sparsity in the number of features that are selected. It is also good practice to make use of several alternative measures of a model's goodness-of-fit such as Mallow's C_p , Akaike Information Criterion (AIC), or Bayesian Information Criterion (BIC).

5. Model Selection:

Analysts may use one specific algorithm to build a model, based upon such matters as expertise, computational complexity, or interpretative preference; however, a more common approach is to have multiple algorithms to compare. At this point, algorithms have been trained using the training set and their best-of-class performers advance for further consideration. The trained models are now applied to the validation set to

determine which model has superior predictive performance based upon a suitable performance measure and the interpretative needs of the end-user.

The end-user or decision maker may be, for example, a loan company that employs a model to make customer loan determinations. If your model is a “black box” like that of a neural network, which yields only an output of “Yes” or “No,” then that loan agency will likely have a difficult time explaining the “why” behind a denial. Therefore, the top performing model may not be the ideal selection if interpretability is sacrificed yet imperative to the end-user. Moreover, if a decision maker needs a quick-turn analysis of a large dataset, then the computational complexity of the algorithm could become a factor giving preference to less time-intensive algorithms. Ultimately, the objective of this step is to select the algorithm that demonstrates the best performance given other considerations.

6. Model Performance Assessment:

The best model in terms of interpretability needs and measured performance is now trained using the combined training and validation sets. Subsequently, the trained model is employed upon the test set and the performance measurement is indicative of how well the model will perform on newly introduced observations.

7. Model Implementation:

For the last step, the best model is fit to the full data set (i.e. training, validation, and test partitions) resulting in a model that is ready for deployment. Educating end-users and model sustainment are also critical at this step to equip end-users to effectively employ the model and best ensure its continued success.

2.2.2 Supervised Machine Learning Algorithms

The supervised learning algorithms that were used in this study were logistic regression, kNN, and random forest algorithms.

Logistic Regression

Logistic regression is a classification supervised learning algorithm as the response variable is categorical with observations taking one of two response values (e.g. “Yes/No” or “1/0”)

(James et al. 2017). Logistic regression models the probability of the response variable being “Yes” given the predictors using the logistic function (James et al. 2017) and the conditional probability is modeled as:

$$p(Y|X) = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}}.$$

The logistic function can in turn be manipulated to produce the logit (James et al. 2017):

$$\log \left(\frac{p(Y|X)}{1 - p(Y|X)} \right) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p.$$

The values $\beta_1, \beta_2, \dots, \beta_p$ for each of the p features are the coefficients for the logistic regression model. An interpretation of the coefficients suggest that increasing X_i by one unit change the log odds by β_i (James et al. 2017). More simply, positive coefficients for β_i suggest that increasing X_i will increase $p(Y|X)$ and vice versa.

k-Nearest Neighbors

kNN is a supervised learning algorithm adaptable to both numerical and categorical response variables (James et al. 2017). “Given a positive integer k and a test observation x_0 , the KNN classifier first identifies the k points in the training data that are closest to x_0 , represented by N_0 . It then estimates the conditional probability for class j as the fraction of points in N_0 whose response values are equal to j ” (James et al. 2017, p.39). The parameter k is selected by the analyst either through trial-and-error, tuning, cross-validation, or an exhaustive approach in determining the best value of k as measured by model performance (James et al. 2017). For instance, if the k parameter of the algorithm was set to 10, then the kNN algorithm will identify 10 training set observations that are closest to a test set observation, x_0 , in the real coordinate space of dimension n denoted $\mathbb{R}^n$ (James et al. 2017). The kNN classifier will then classify the response for x_0 based on the response value that has the highest estimated probability among the k neighbors (James et al. 2017). As an example, if 70% of the k nearest training set observations to x_0 have a response value of “1,” and 30% have response “0”, then the kNN algorithm will classify the response for x_0 as “1,” being the class with the highest estimated probability. A visualization of the kNN classification methodology is given in Figure 2.6. As well, the notation for the kNN classifier

is as follows (James et al. 2017):

$$Pr(Y = j|X = x_i) = \frac{1}{k} \sum_{i \in N_i} I(y_i = j),$$

where Y is the response variable for test set observation, X is the vector of explanatory variables, k is the number of nearest neighbors to identify, N_i is the set of k points that are closest to x_i , which is the vector of explanatory variables for the i th observation, and y_i is the response value for the i th observation in the set N_i .

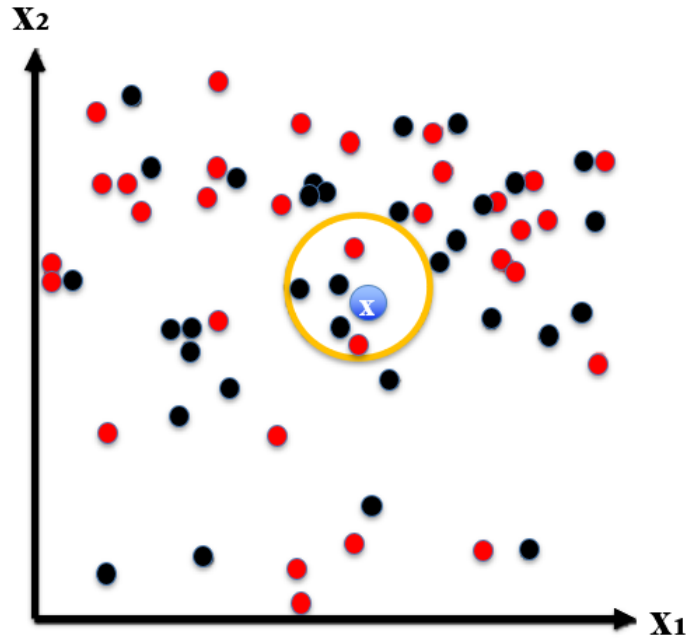


Figure 2.6. kNN classification methodology. The red and black points represent training set observations with response values indicated by their respective color. The blue point labeled x is a newly introduced observation from a test set with unknown response. From the $k = 5$ nearest neighbors to x there is a $3/5 = 60\%$ probability that an observation within the circle has response black, and $2/5 = 40\%$ probability of a training observation having a red response value. The kNN will classify x based on the response with the highest estimated probability for the k nearest neighbors, which in this case is black. Adapted from: Yoshida (2021b).

Random Forest

Random forest, a tree-based model, also falls within the realm of a supervised learning algorithm that can be applied to regression or classification problems. Though not as easily interpreted as a logistic regression model, the random forest algorithm draws its strength from randomness. Multicollinearity, or correlation among features, can make it difficult to know which features have explanatory power (James et al. 2017). The random forest algorithm combats multicollinearity by randomly selecting m of p predictors for each split among d decision trees from bootstrapped samples of training data (James et al. 2017). The algorithm's parameters of m and d are determined by the analyst, but in general $m \approx \sqrt{p}$ and more desirable results stem from increased numbers of decision trees, d , with diminishing returns (James et al. 2017). A further benefit of randomly selecting only a small subset of the available features is that the decision trees themselves are decorrelated and are not dominated by the same strong predictors (James et al. 2017). When assessing feature importance, this method widens the door for middle-of-the-road predictors to be better assessed as their influence in predicting the response variable is not washed out by dominant features. Ultimately, the final randomly-derived ensemble of trees with low correlation, a "random forest," forms a model that predicts the response of new observations in a majority vote fashion amongst the forest's trees.

2.3 Time Series

A time series is a set of chronologically-ordered observations of a variable of interest, often at equally spaced intervals over a prescribed time frame (Hyndman and Athanasopoulos 2018). Time-series analysis entails a statistical examination of the time series data (Hyndman and Athanasopoulos 2018). A common objective of the analysis is to learn from the past to gain insight into the future. In other words, a statistical analysis on the historical evolution of the data points can unearth patterns that may be prescriptive of future forecasts or predictions.

This research makes use of a handful of prevalent time-series analysis algorithms, discussed further in section 2.3.2, to include:

- Seasonal Decomposition;
- Seasonal Autoregressive Integrated Moving Average (SARIMA);

- HW;
- Holt-Winters Additive (HW-A);
- Holt-Winters Multiplicative (HW-M); and
- ETS.

2.3.1 Time Series Definitions

It is important to define some core components and attributes of time series to gain a deeper appreciation for the algorithmic models as described by Hyndman and Athanasopoulos (2018).

1. Time Series Patterns

Trend

A trend is a longstanding increase or decrease in the time series data, which does not necessarily have to be linear.

Seasonal

Seasonality arises when the time series data is impacted by fixed intervals or frequencies throughout a year, or at a more micro level, days of the week.

Cyclic

Rises and falls in the time series that are not of a fixed interval or frequency are demonstrative of a cycle.

Noise

The variance in the time series that cannot be explained by either trends, seasonality, or cyclic behavior is noise.

2. Time Series Properties

Homoscedastic

A homoscedastic characteristic occurs when the variance in a time series is static over time (Yoshida 2021a).

Heteroscedastic

Fluctuations in the variance of a time series is indicative of the heteroscedastic property (Yoshida 2021a).

Stationarity

A time series that is independent of the time period of observation and one without trend, seasonality, or heteroscedastic properties (Yoshida 2021a).

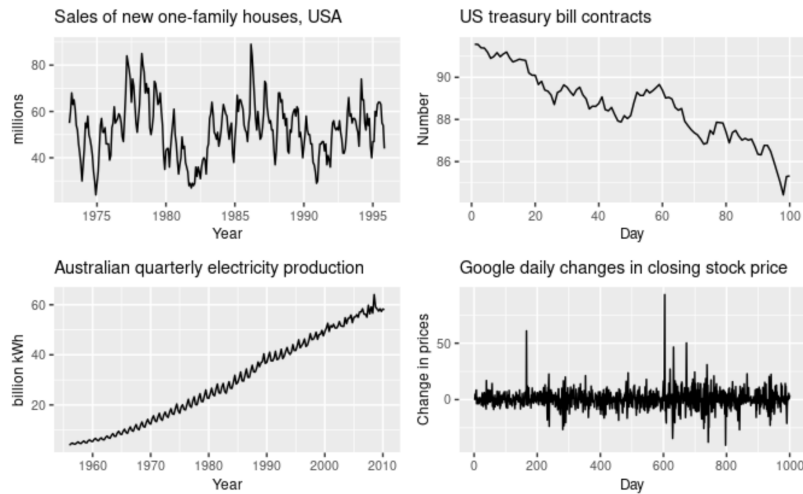


Figure 2.7. Monthly housing sales (top-left) reflects annual seasonality and 6–10 year cycles. Daily treasury contracts (top-right) displays a decreasing trend. Annual electricity production (bottom-left) demonstrates a positive trend and seasonal behavior. Daily stock prices (bottom-right) illustrates random fluctuations that do not fit trend, seasonal, or cycle molds, but are homoscedastic. Source: Hyndman and Athanasopoulos (2018).

2.3.2 Time Series Models and Forecasting Algorithms

The methodology for time series models and forecasting algorithms is quite similar to the algorithmic modeling overview presented within section 2.2.1. Montgomery et al. (2016) propose the following tasks:

1. Problem definition.
2. Data collection.
3. Data analysis.
4. Model selection and fitting.
5. Model validation.

6. Forecasting model deployment.
7. Monitoring forecasting model performance. (Montgomery et al. 2016)

The notable difference between the supervised learning and time series processes resides within the Exploratory Data Analysis (EDA) and performance, or validation, steps. EDA for a time-series analysis, as with supervised learning models, involves checking for anomalies and/or outliers within the data set. However, time series EDA also involves the preliminary assessment of trends, seasonality, and/or cyclic patterns (Montgomery et al. 2016). Concurrently, visually examining the data for heteroscedasticity, homoscedasticity, and stationarity further enhances ones insight into the data's structure.

A distinction between the supervised learning and time series algorithmic steps pertains to the partitioning of the training and testing sets. While this is a recommended step for both algorithmic endeavors to essentially measure model performance, there is less randomness for time series partitioning of the data set as the partitions are not randomly selected samples of observations, but rather samples of observations that are indexed by time (Hyndman and Athanasopoulos 2018). Hyndman and Athanasopoulos (2018) visualize the training and test data in Figure 2.8, not as randomly selected samples of observations, but specifically as training data preceding test data as indexed by time. Furthermore, the train-test split is generally an 80–20 split of the data set, yet the duration of time for the test data should at least span the desired forecast length (Hyndman and Athanasopoulos 2018).



Figure 2.8. Train-Test split of time series data for model performance evaluation. Source: Hyndman and Athanasopoulos (2018).

Seasonal Decomposition

For time-series analysis, the process of breaking a time series down into its trend, seasonal, and cyclic components is often referred to as seasonal decomposition. There is a dual-purpose nature of seasonal decomposition befitting both EDA and forecast improvement, though we primarily used seasonal decomposition as an EDA tool within this research (Hyndman and Athanasopoulos 2018). Considering that seasonal decomposition is a prominent

tactic there are several methods available to analysts. This study opted for the Seasonal and Trend Decomposition using Loess (STL) method for its advantages in both robustness and adaptability to various types of seasonality. Figure 2.9 visualizes time series data and its respective trend-cycle, seasonal behavior, and remainder (i.e. noise) components from the STL seasonal decomposition.

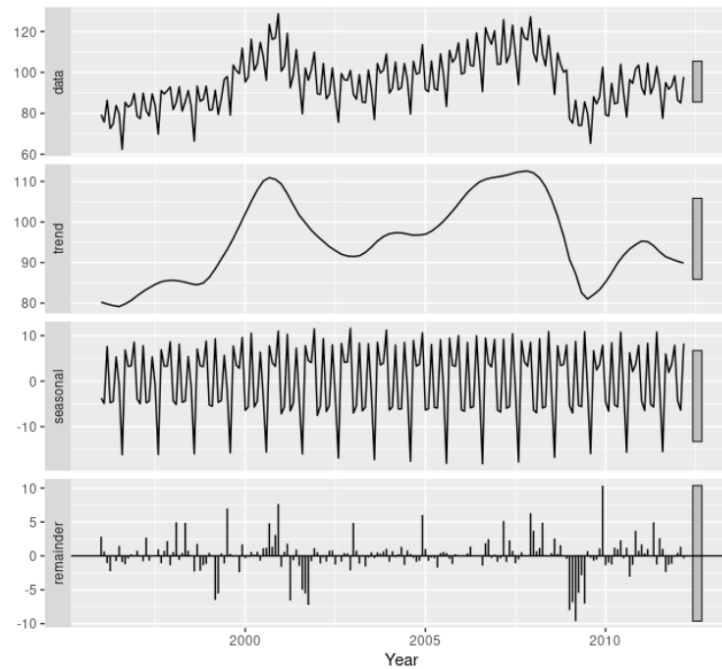


Figure 2.9. Seasonal decomposition of the time series data (top sector) broken down into trend-cycle, seasonal, and remainder components. The components can be merged to restore the initial data (top sector). Source: Hyndman and Athanasopoulos (2018).

Naïve and Seasonal Naïve

Naïve time series models establish the baseline for gauging the performance of other models. As the name implies, the model is quite simply derived in that the next sequential observation in a time series is predicted to be equivalent to the preceding observation (Yoshida 2021a). Seasonal Naïve models take a similar approach in that forecast values are equivalent to prior observations, yet the prior observation is dependent upon the seasonal period (Yoshida

2021a). For instance, if the seasonal period was 12 months and a forecast value for June 2022 was being determined, the model would impute the June 2022 forecast with the June 2021 observed value.

SARIMA

The SARIMA model is comprised of two components: non-seasonal and seasonal. Not surprisingly, if the model were restricted to just its non-seasonal component, then it would be known as a non-seasonal Autoregressive Integrated Moving Average (ARIMA) model taking the following form (Hyndman and Athanasopoulos 2018):

$$y'_t = c + \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} + \epsilon_t,$$

where

y'_t : Value at time t .

c : Numerical coefficient selected by algorithm representing the average differences between sequential observations. Constant term may or may not be 0.

ϵ : Error terms or white noise.

θ : Weight coefficient for moving average component of model. Determines how much weight to place on errors for distant versus recent observations.

ϕ : Autocorrelation coefficient, which is a measure of the relationship between y_t and y_{t-k} after removing the effects of lags 1, 2, 3, ..., $k-1$.

p : order of the autoregressive component.

q : order of the moving average component.

ARIMA models use a concept known as differencing, which is a method for making a non-stationary time series stationary (Hyndman and Athanasopoulos 2018). Fundamentally, the non-seasonal ARIMA algorithm will produce an ARIMA(p,d,q) model represented by the following parameters (Hyndman and Athanasopoulos 2018):

- p = order for the autoregressive aspect,

- d = first differencing degree, and
- q = order for the moving average aspect.

The derivation of p , d , and q by ARIMA algorithms can be assessed through Autocorrelation Function (ACF) and partial autocorrelation function (PACF) plots or calculations (Hyndman and Athanasopoulos 2018). The ACF measures autocorrelations between the observed value y at time t , y_t , and the observed value of the previous period, y_{t-k} (Hyndman and Athanasopoulos 2018). The value k pertains to the time gap or lag between the observations. Whereas, the PACF assesses the similarity between y_t and y_{t-k} once the impacts of lags $1, 2, \dots, k-1$ have been removed (Hyndman and Athanasopoulos 2018).

SARIMA models are an adaptation of the the non-seasonal ARIMA model where a seasonal aspect is built into the model. Consequently, the SARIMA model takes the form $SARIMA(p, d, q)(P, D, Q)_m$ where the newly incorporated values of P , D , and Q are seasonal representations of p , d , and q (Hyndman and Athanasopoulos 2018). Furthermore, a period m is determined by the number of annual seasonal periods for a time series (Hyndman and Athanasopoulos 2018).

Holt-Winters

The HW seasonal method is an exponential smoothing method that has an interesting derivation with popular adaptations. The Holt method, primarily concerned with the trend of a time series, was originated by Holt (1957) with the development of a forecast equation containing two smoothing equations: l_t indicative of the level or height of the time series at time t , and b_t representing the time series' trend at time t (Hyndman and Athanasopoulos 2018). Correspondingly, l_{t-1} and b_{t-1} would represent the level and trend equations for the time period immediately preceding time period t . The smoothing equations incorporate two smoothing parameters, α and β , for the level equation, l_t , and trend equation, b_t , respectively (Hyndman and Athanasopoulos 2018). As determined by the model, α and β assume values within the $[0, 1]$ interval (Hyndman and Athanasopoulos 2018). At a given time t , the forecast value at time $t + h$ -time steps, denoted $\hat{y}_{t+h|t}$, is determined as follows (Hyndman and Athanasopoulos 2018):

$$\hat{y}_{t+h|t} = l_t + hb_t,$$

where

$$l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + b_{t-1}),$$

with the smoothing constant α and

$$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1}.$$

The Holt method was extended with a seasonal component, s_t , by P.R. Winters in 1960 resulting in two seasonal adaptations: HW-A and HW-M (Hyndman and Athanasopoulos 2018). HW-A and HW-M introduce a new variable m , to account for the seasonal period of the time series, as well as an additional smoothing parameter $\gamma \in [0, 1]$ for the seasonal equation, s_t (Hyndman and Athanasopoulos 2018). HW-A finds its use for relatively consistent seasonal patterns and take the following form (Hyndman and Athanasopoulos 2018):

$$\hat{y}_{t+h|t} = l_t + hb_t + s_{t+h-m(k+1)},$$

where

$$l_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(l_{t-1} + b_{t-1}),$$

$$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1},$$

and

$$s_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m}.$$

HW-M, on the other hand, is more appropriate for seasonal patterns that fluctuate based on the height of the time series values (Hyndman and Athanasopoulos 2018). The HW-M equations include:

$$\hat{y}_{t+h|t} = (l_t + hb_t)s_{t+h-m(k+1)},$$

where

$$l_t = \alpha \left(\frac{y_t}{s_{t-m}} \right) + (1 - \alpha)(l_{t-1} + b_{t-1}),$$

$$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1},$$

and

$$s_t = \gamma \left(\frac{y_t}{(l_{t-1} + b_{t-1})} \right) + (1 - \gamma)s_{t-m}.$$

ETS

The ETS algorithm is an Exponential Smoothing method based upon its Error, Trend, and Seasonal components. The ETS (Error, Trend, Seasonal) model is built to adapt to various trend or seasonal patterns that are present in time series data by setting its Error, Trend, and Seasonal properties to best fit the time series. The component property options include (Hyndman and Athanasopoulos 2018):

- **Error:** Additive or Multiplicative;
- **Trend:** None, Additive, Multiplicative, Additive-damped, or Multiplicative-damped; and
- **Seasonal:** None, Additive, or Multiplicative.

As observed, there are many potential combinations or ETS models that the algorithm can select to best fit a time series. For instance, an ETS “Additive-None-None” model is a simple exponential smoothing with additive errors, “Multiplicative-Additive-Multiplicative” follows the HW-M method with multiplicative errors, and so on (Hyndman and Athanasopoulos 2018). There are as many mathematical models as feasible ETS (Error, Trend, Seasonal) combinations.

2.4 Survival Analysis

Survival analysis uses statistical methods to analyze the length of survival time from a specified starting point in time until an event, or “death”, takes place (Kleinbaum and Klein 2005). Commonly connected with biomedical studies, the application of survival analysis is appropriate for use within many fields as the “death” does not necessarily need to be an actual physical death, but it certainly can be. Reasonably, the time until a part fails, a person relapses after treatment, an individual separates from the military, a National Football League (NFL) player returns to the field after an injury, or a patient dies after receiving a certain medical diagnosis are all candidates for survival analysis.

2.4.1 Survival Analysis Definitions

Fundamental to survival analysis are the following terms as presented by Kleinbaum and Klein (2005):

Exposure: The incident that starts the clock for the variable of interest.

Time: The number of days, weeks, months, or years from exposure to event, often referred to as survival time.

Event: The incident that stops the clock for the variable of interest.

Truncation: Truncation occurs when observations are excluded by virtue of the time-to-event. The two types of truncation include:

1. **Left Truncation:** Observations with short time-to-event are excluded.
2. **Right Truncation:** Observations with long time-to-event are excluded.

Censoring: A method for handling survival times of observations for which the exact time to event is unknown. Depicted within Figure 2.10, there are three types of censoring:

1. **Left-censoring:** The time-to-event, x , is less than some period of time t such that $x < t$.
2. **Interval-censoring:** The time-to-event, x , is between some period of time such that $t_1 < x < t_2$.
3. **Right-censoring:** The time-to-event, x , is greater than some period of time t such that $x > t$.

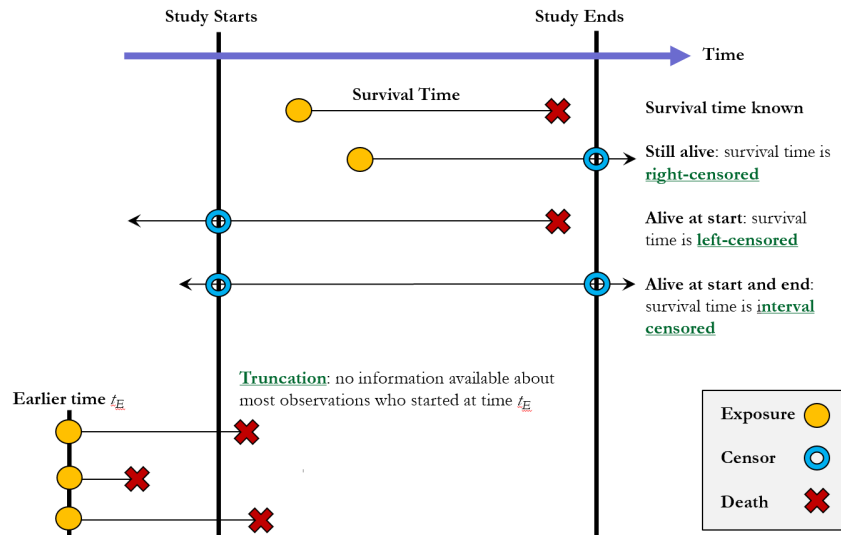


Figure 2.10. Plausible survival times and censoring methods that can be applied when the initial exposure and/or time-to-event is unknown. Adapted from: Buttrey (2022).

Right-censoring is frequently selected in survival analysis studies in which the time-to-event is often unknown (Kleinbaum and Klein 2005). Visualized within Figure 2.11, right-censoring was the method used for this study.

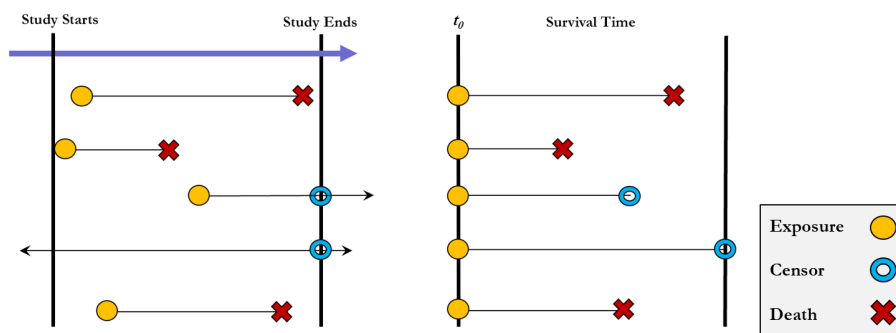


Figure 2.11. The left side represents the survival times for an observational period for known and/or unknown exposures and/or events. The right side sets all observations to start at time t_0 and extend the length of their known survival time or right-censored point. Adapted from: Buttrey (2022).

Survival Function: The survival function, $S(t)$, is the probability that an observations survival time, T , is greater than a specified time t denoted (Kleinbaum and Klein 2005):

$$S(t) = Pr(T > t).$$

Hazard Function: The hazard function is the momentary probability per unit of time for an observation to experience an event at time t given a survival up to time t (Kleinbaum and Klein 2005).

2.4.2 Survival Analysis Models Overview

KM Method

KM curves are based upon life-tables, which illustrate the pattern of survival of a population by considering the probability of death at each point in time $t_0, t_1, \dots, t_j$ (Kleinbaum and Klein 2005). A life-table can account for numerous variables depending upon the interests of the analysis, but will at least contain the following (Kleinbaum and Klein 2005):

t_j : the survival time since time t_0 ;

n_j : the number of observations still alive at time t_j ;

m_j : the number of deaths between time t_{j-1} and t_j ;

q_j : the number of observations that were censored between time t_{j-1} and t_j ; and

$\hat{S}(t_j)$: KM survival probability for the alive observations surviving beyond time t_j calculated as follows (Kleinbaum and Klein 2005):

$$\hat{S}(t_j) = \hat{S}(t_{j-1}) \left(\frac{n_j - m_j}{n_j} \right).$$

The survival probability, $\hat{S}(t_j)$, plotted from time t_0 results in a KM curve as demonstrated within Figure 2.12. For smaller data sets, the KM curve will tend take on the appearance of a step function, and for larger data sets the steps gravitate towards a continuous KM curve.

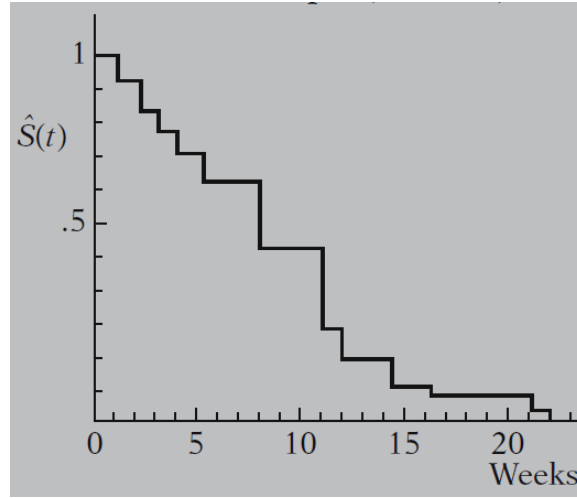


Figure 2.12. Plot representation of a KM curve for the survival probability, $\hat{S}(t_j)$, of a population over time. Source: Kleinbaum and Klein (2005).

CPHR

The CPHR model is a prominent mathematical model for conducting survival analysis. The CPHR model is a regression based model that predicts the hazard rate for an observation as a function of both time and the observation's respective predictor values (Kleinbaum and Klein 2005). The CPHR model hazard function typically takes the following form (Kleinbaum and Klein 2005):

$$h(t, \vec{X}) = h_0(t) e^{\sum_{i=1}^p \beta_i X_i},$$

where

$h_0(t)$: baseline hazard function,

X_i : predictor, or explanatory variables of a model, for each of p predictors ranging from $X_1, X_2, \dots, X_p$, or collectively written as the vector $\vec{X}$, and

β_i : coefficient ascribed to explanatory variable X_i as determined by the algorithm for the model.

From the CPHR model hazard function it is observed that while the baseline hazard function is time-dependent, the exponential summation for each predictor is time-independent (Kleinbaum and Klein 2005). The exponential component of the equation ensures that the predicted hazard rates are within the feasible bounds of $[0, \infty)$, thus non-negative (Kleinbaum and Klein 2005).

The CPHR hazard function can be conveniently adjusted into the CPHR model's estimated survival function as follows (Kleinbaum and Klein 2005):

$$\hat{S}(t, \vec{X}) = [\hat{S}_0(t)]^{e^{\sum_{i=0}^P \hat{\beta}_i X_i}}.$$

As with a linear regression model, a CPHR model's output indicates the coefficients of the predictor variables as well as p-values from significance tests where the null hypothesis, $H_0 : \beta_i = 0$ is tested against the alternative hypothesis, $H_1 : \beta_i \neq 0$ (Kleinbaum and Klein 2005). Unlike a regression model, the CPHR model also produces a hazard ratio for each predictor variable. The hazard ratio exhibits the variables' effect on the hazard rate after being adjusted to account for other variables within the model (Kleinbaum and Klein 2005). Ultimately, the CPHR model survival curves $\hat{S}(t, \vec{X})$ are often close in appearance to KM curves, but differ from KM models in that they are adjusted for covariates (Kleinbaum and Klein 2005).

RSF

The RSF algorithm is quite similar in principal to the random forest supervised learning algorithm discussed in Section 2.2.2. The RSF algorithm randomly selects m of p features from bootstrapped training data samples to form n decision trees, such as the tree in Figure 2.13, and the individual survival curves are computed based upon statistical averages of the terminal nodes (Wang and Li 2017). The primary distinctions pertain to the RSF algorithm's ability to consider both time and censored data (Wang and Li 2017).

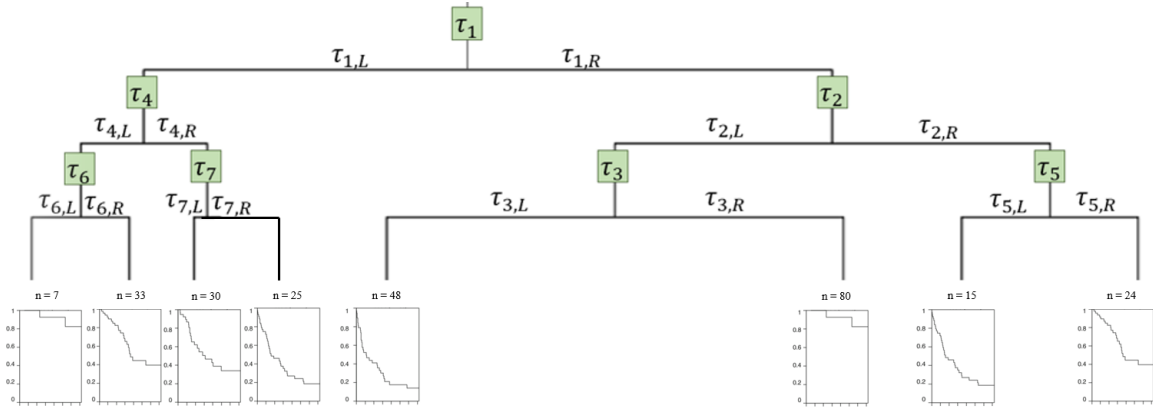


Figure 2.13. Classification tree for RSF reflecting splits at nodes $\tau_1, \tau_2, \dots, \tau_k$ based upon a specified split criterion, and the estimated survival curves for each leaf. Adapted from: Zhang and Singer (2010).

For each of the n decision trees, the randomly selected m of p predictors are split into a tree-based structure at nodes $\tau_1, \tau_2, \dots, \tau_k$ as demonstrated within Figure 2.13 (Zhang and Singer 2010). There are several splitting rules (e.g. log-rank, concordance index, etc.) that can be utilized to test the difference in KM survival distributions amongst the explanatory variables in a data set. In essence, of the m predictors selected, τ_1 will be the predictor that is able to be split into two groups, group L and R , based on a split criterion such that the survival curves for the groups are the most different out of all the possible predictor splits. Then, node τ_2 will be the predictor that when split contains group L and R survival curves that are the second most different, and so on. For this study, our splitting rule was based on the log-rank test (Zhang and Singer 2010):

$$\text{Log Rank} = \frac{\sum_{t=1}^k (a_t - E_t)}{\sqrt{\sum_{t=1}^k V_t}},$$

where

$$E_t = \frac{d_t r_t}{K_t},$$

$$V_t = \left(\frac{d_t (K_t - r_t) r_t}{K_t (K_t - 1)} \right) \left(1 - \frac{d_t}{K_t} \right),$$

t : the survival time since time 0,

k : the number of distinct event times,

K_t : the number of observations still alive, or at risk, at time t ,

d_t : the number of events, or deaths, at time t ,

a_t : the number of deaths for group L at time t ,

r_t : the number of observations still alive for group L at time t ,

E_t : expected number of events at time t , and

V_t : variance of expected number of events at time t .

Once all of the n decision trees have been grown, a Cumulative Hazard Function (CHF) is then computed for each decision tree as follows (Wang and Li 2017):

$$\hat{H}(t) = \sum_{t \leq T} \frac{d_t}{K_t}.$$

The ensemble of decision tree CHFs are brought together giving rise to the RSF model (Wang and Li 2017):

$$\hat{H}(t|x) = \frac{1}{n} \sum_{i=1}^n \hat{H}_i(t|x).$$

The RSF model predicts the hazard for a newly introduced observation, x , as an average across each of the n decision trees' CHFs (Wang and Li 2017).

2.5 Performance Evaluation

This study makes use of the following measures of performance for the models specified:

Classification:

- Sensitivity;
- Specificity;

- Receiver Operating Characteristic (ROC) Curve; and
- Area Under the (ROC) Curve (AUC).

Time-Series Analysis:

- Mean Absolute Percentage Error (MAPE);
- Mean Absolute Scaled Error (MASE); and
- Residual Auto-correlation.

Survival Analysis:

- Concordance.

2.5.1 Classification Model Performance Evaluation

Classification algorithms, or classifiers, are gauged on their ability to predict the binary or multi-level response value of new observations (e.g. test data) after learning from training data. Initial insight into a model's performance is often gleaned from a 2×2 matrix known as the confusion matrix. A binary classification model's confusion matrix, represented in Figure 2.14, offers a comparison of the model's predictions against the true response values by providing the number of true positives, false positives (i.e. type I errors), true negatives, and false negatives (i.e. type II errors) (Huddleston and Brown 2019). For a binary response, the output from a model for an observation is typically a computed probability that an observation has a "1" or "0" response. The default cut-off threshold for most algorithms is 0.50, but may vary (Huddleston and Brown 2019). Therefore, if a computed probability of being "1" is greater than 0.50, then the model predicts a "1" response for that observation or "0" otherwise.

		Actual Class	
		Positive (1)	Negative (0)
Predicted Class	Positive (1)	True Positives	False Positives
	Negative (0)	False Negatives	True Negatives

Figure 2.14. Confusion matrix for a classification model predicting a binary response. The values of the 2×2 matrix will indicate counts for the four listed outcomes from which sensitivity and specificity calculations can be derived. Adapted from: Huddleston and Brown (2019).

From the confusion matrix, the classification model's sensitivity, or True Positive Rate (TPR), and specificity, or True Negative Rate (TNR) are calculated (Huddleston and Brown 2019):

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}},$$

and

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}}.$$

The classification model's specificity and sensitivity are a measure of the model's ability to correctly classify true positives or true negatives (Huddleston and Brown 2019). Understandably, higher rates of specificity and sensitivity are good indicators of model performance. Contrarily, higher False Positive Rate (FPR) and/or False Negative Rate (FNR) reduce a model's credibility. The FPR and FNR are calculated in the following manner (Huddleston and Brown 2019):

$$\text{FPR} = \frac{\text{False Positives}}{\text{False Positives} + \text{True Negatives}},$$

and

$$\text{FNR} = \frac{\text{False Negatives}}{\text{False Negatives} + \text{True Positives}}.$$

An overall assessment of a model's performance based upon the confusion matrix can be captured by calculating both the accuracy and misclassification rates (Huddleston and Brown 2019):

$$\text{Accuracy Rate} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total}},$$

and

$$\text{Misclassification Rate} = \frac{\text{False Positives} + \text{False Negatives}}{\text{Total}}.$$

ROC curves are an attractive measure of a model's performance known for their interpretability, visualization, and ability to assimilate multiple ROC curves within a single plot for a comparative analysis when competing models. As depicted within Figure 2.15, a classifier will predict that the response of an observation is positive if to the right of a threshold or negative if to the threshold's left (Yoshida 2021c). For the two hypothetical distributions within Figure 2.15, intuition tells us that there will be varying counts of true positives or negatives, as well as false positives or negatives, depending upon the similarity of the distributions and the classifiers' threshold adjustment.

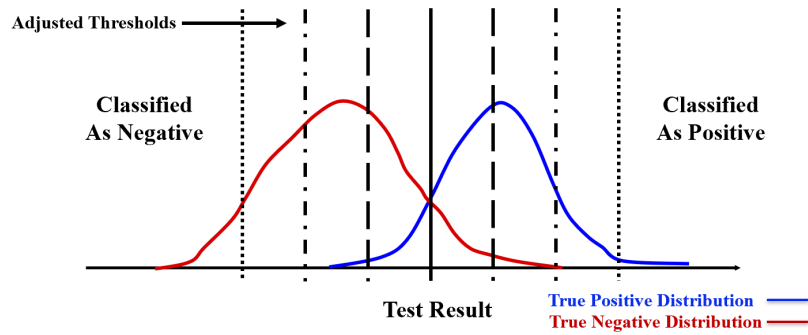


Figure 2.15. Classifier threshold adjustments and the effect upon two hypothetical distributions of observations having true positive or negative response values. As the thresholds shift from left to right, the classification model predicts an observation's response as negative for observations to the left of the threshold, and positive for observations to the threshold's right. Adapted from: Yoshida (2021c).

ROC curves plot the classification model's false positive rate against its true positive rate as the threshold adjusts (Huddleston and Brown 2019). As observed within Figure 2.16, the ideal ROC curve will gravitate towards the upper-left corner of the plot indicating the model has good sensitivity and specificity (Yoshida 2021c). A ROC curve for a poor performing model will gravitate towards a 45° line extending from the bottom-left to the upper-right corner, which basically represents the flipping of a coin to predict an observation's response as being positive or negative (Yoshida 2021c). The AUC, as its name suggests, is the total area under the ROC curve. If the total area of the plot is 1, or 100%, then the AUC of an ROC curve extending along the 45° line would be 0.5, or 50% of the total plot area. Therefore, models producing higher AUCs tend to be more favored when AUCs are used as a performance measure.

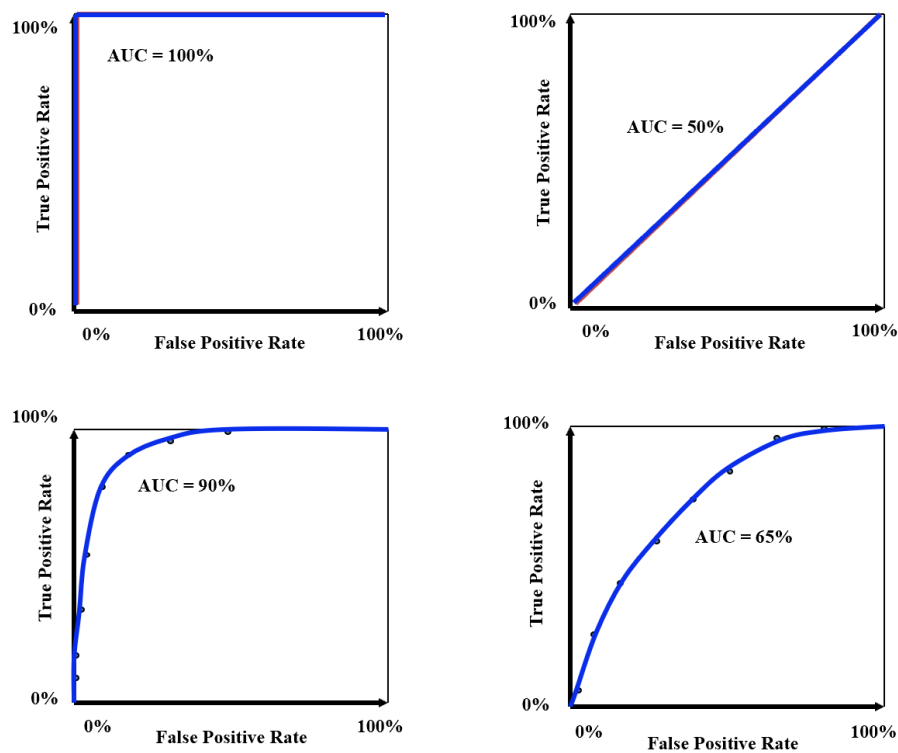


Figure 2.16. ROC and their respective AUC. As the classifier's threshold adjusts from left to right, ROCs that tend toward the upper-left corner are indicative of higher specificity and sensitivity than ROCs that gravitate to a 45° line extending from the bottom-left to the upper-right corner. Adapted from: Yoshida (2021c).

2.5.2 Time Series Model Performance Evaluation

To ascertain the performance of a time series model we measure the difference between our model's forecast and the observed value at time periods $t_1, t_2, \dots, t_n$. This disparity between a model's forecast, $\hat{y}_t$, and the known value, y_t , is referred to as a forecast error, ϵ_t , calculated as follows (Yoshida 2021a):

$$\epsilon_t = y_t - \hat{y}_t$$

Prior to discussing the quantitative measures of evaluating the forecast errors across all time periods of the analysis, it is important to understand the rolling-horizons design. As discussed in section 2.3.2, a prominent approach to time series that is employed within this study makes use of partitioning the time series data into training and test data sets. The training data set precedes the test data as indexed by time. Moreover, this practice is often bolstered through use of a rolling-horizons design which generates a series of training and test data sets (Hyndman and Athanasopoulos 2018). The concept for the rolling-horizons methodology is represented by Figure 2.17, which depicts the initial training and test data sets at the top of the figure to the final partition at the bottom. Notice that the test data covers only one time period. Furthermore, successive partitions of the data entail an absorption of prior test data into the training data set and the new test data becomes the next period's observation as indexed by time. This ensuing "rolling" forward of the test data is where the "rolling-horizons" design derives its name.

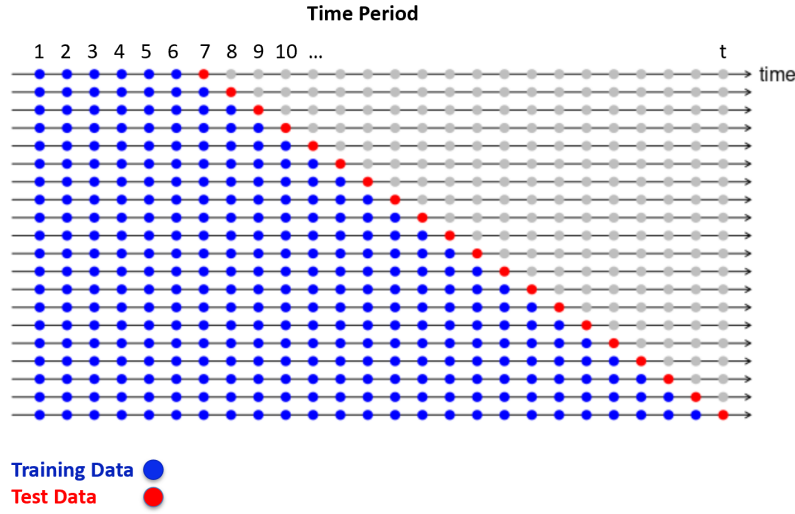


Figure 2.17. Rolling-horizon design methodology. From top to bottom, the partitioning of the training and test data sets for a 1-step ahead rolling-horizons design for successive iterations of the algorithm. Adapted from: Hyndman and Athanasopoulos (2018).

Regarding the time period for the test data, Figure 2.17 shows a 1-step forecast. While a 1-step forecast is useful for measuring a model's ability to predict the value of interest for the next time period, there may be instances in which forecasts extending further out in time are of more importance such as a 4-step or 12-step forecast. This is also possible with a rolling-horizons design such that the test periods would simply be shifted to the right n steps from the last training period.

There are many statistical approaches available for measuring the forecasting capability of a time series model. Two well established measures include MAPE and MASE as they are scale-free, that is, they do not depend upon the scale for observations y_t (Yoshida 2021c). In other words, MAPE and MASE may be used to consider time series that have differing units. Whereas, such measures as Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE) may also be employed, but are scale-dependent (Hyndman and Athanasopoulos 2018).

This study makes use of MAPE and MASE measures of performance for both a 1-step and 12-step forecast. MAPE values are contained within the interval $[0, 1]$ where lower values

are preferred when competing models and more indicative of predictive power (Yoshida 2021c). MASE values assess a forecasting model's performance against a naïve model with respect to their errors. Contrary to MAPE, MASE values may exceed 1 signifying that the model performs worse than the naïve model. Hence, as with MAPE, lower MASE values are preferred. As described by Yoshida (2021c), MAPE and MASE calculations take the following form:

$$\text{MAPE} = \frac{\sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right|}{n},$$

and

$$\text{MASE} = \frac{\sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t - y_{t-1}} \right|}{n},$$

where y_t is a value at time period t , $\hat{y}_t$ is a forecast value at time period t , and n is the number of time periods t .

2.5.3 Survival Model Performance Evaluation

The prevailing statistical measure for assessing the performance of a RSF model's predictive power is a concordance statistic, also known as a C-statistic or Harrell's C (Schmid et al. 2016). The concordance value measures a model's fit when such a model is built for generating risks, or hazards for observations (Schmid et al. 2016). Thereby, making concordance values a good measure for survival analysis. A concordance value is equivalent in interpretation to that of an AUC in that a concordance of 0.5 represents a model that predicts based on random chance, and a concordance of 1 is perfect prediction accuracy (Schmid et al. 2016). For biomedical studies, concordance values tend to hover between 0.6 and 0.75 (Schmid et al. 2016).

The formula behind a concordance value concerns itself with the ratio of concordant pairs to comparable pairs (Schmid et al. 2016). For the concordance equation, η represents the model's level of risk assigned to an observation and T is the observation's survival time (Schmid et al. 2016). The comparable pairs are essentially the number of concordant and discordant pairs in which a pair (i, j) is concordant if $T_i > T_j$ and $\eta_j > \eta_i$ and discordant if $T_i > T_j$ and $\eta_i > \eta_j$ (Schmid et al. 2016). In words, this tells us that if the model's assessed risk, for instance, of patient j is greater than the assessed risk for patient i AND patient j 's survival time is less than patient i 's survival time, then the pair of observations is

concordant. The concordance formula is the ratio of concordant to comparable pairs written as follows (Schmid et al. 2016):

$$\text{Concordance} = \frac{\sum_{i,j} I(T_i > T_j) \cdot I(\eta_j > \eta_i) \cdot \delta_j}{\sum_{i,j} I(T_i > T_j) \cdot \delta_j},$$

where T_i and T_j are independent pairs of survival times, or time-to-event, η_i and η_j are hazard, or risk, predicted values, and

$$\delta_j = \begin{cases} 0 & \text{if one of compared observations has a shortened survival time due to censoring} \\ 1 & \text{otherwise.} \end{cases}$$

2.6 Literature Review

The focus of this review centers upon the use of machine learning algorithms in studies that examine substance abuse within military populations. The aim is to assess the predictive accuracy of machine learning models when applied to military substance abuse studies. We hone in on military members specifically as that is the population of interest for our research objectives. Furthermore, military members often go through unique experiences that are not common to civilian populations. For instance, deployments, combat, long periods of separation from immediate or extended family, and frequent moves are experiences that are more descriptive of military than civilian populations. Gaining a deeper insight into the military experiences that can potentially contribute to substance abuse issues can aid efforts to combat substance abuse within military ranks.

A recent study by Betancourt et al. (2021) used logistic regression and Classification and Regression Tree (CART) models to assess possible predictors contributing to a veteran relapsing after being previously enrolled into a substance abuse treatment program. The data contained a sample size of $n = 40,909$ veterans and included 10 socio-demographic and treatment-specific predictor variables (Betancourt et al. 2021). The results reflected an AUC of 0.713 for their best logistic regression model and a 0.65 AUC for the CART model (Betancourt et al. 2021). For the socio-demographic variables, veterans that fell into any of the following categories were associated with higher risks of relapse: 35 to 64 years of

age, males, unemployed, college educations, “other” race/ethnicity, and at least one arrest (Betancourt et al. 2021). As well, veterans who dropped out of treatment, were treated by detox services, and/or had lengths of stay in treatment greater than 365 days were found to be at higher risk of relapsing (Betancourt et al. 2021). The CART model indicated that, in order of importance, the living arrangement at discharge from treatment, reason for discharge, and length of stay reflected the highest importance in explaining relapse (Betancourt et al. 2021). Of note, age, ethnicity, and race ranked the lowest in terms of variable importance (Betancourt et al. 2021).

A study by McKenzie et al. (2006) also employed logistic regression to identify predictors of substance abuse among $n = 1201$ male Royal Australian Navy (RAN) Gulf War veterans. For their data collection efforts, an Alcohol Use Disorders Identification Test (AUDIT) questionnaire was provided to RAN service members. The AUDIT questionnaire was developed by the World Health Organization (WHO) for identifying individuals with hazardous or harmful drinking habits (McKenzie et al. 2006). Their study reports an AUC of 0.88 for their logistic model in predicting the response, AUDIT “caseness,” referring to prior hazardous or harmful alcohol consumption that has classified a member as being an AUDIT case.

The significant predictors in classifying veterans exhibiting AUDIT caseness were “separated/divorced/widowed, single/never married, former smokers, current smokers, those with current DSM-IV diagnoses of major depression, and pre-deployment alcohol use or dependence” (McKenzie et al. 2006). Furthermore, service members in non-supervisory roles, similar to USN E-1 through E-3’s, were also associated with AUDIT caseness. Dissimilar to the results found in (Betancourt et al. 2021), college educated service members had lower risk of AUDIT caseness. Of interest, the number of deployments were considered yet determined to be insignificant in classifying AUDIT caseness and PTSD would have been determined to be significant at an $\alpha = 0.10$ test level.

On the topic of deployments, Fink et al. (2017) contrasted alcohol use between a sample of service members who deployed, “deployers,” and a control group of service members that did not deploy, “non-deployers,” to analyze the effect of deployments on alcohol consumption. The data for the study was obtained by surveying Ohio US National Guard service members on three separate occasions throughout a 2009 to 2014 time span (Fink

et al. 2017). The sample included $d = 203$ respondents who deployed to a combat area, namely Iraq or Afghanistan, and $n = 2,141$ non-deployers (Fink et al. 2017). To identify a control group of four non-deployers for each deployer, logistic regression was utilized for generating propensity scores for each service member to gauge the probability of a member deploying (Fink et al. 2017). In conjunction with the propensity score, an optimal-matching method was employed to best identify homogeneous subgroups (Fink et al. 2017). Once matched, a linear regression model and Latent Growth Mixture Model (LGMM) were used to analyze the differences in alcohol consumption between deployers and their control groups, as well as to examine the effect of deployment (Fink et al. 2017). The results reflected a non-significant increase in alcohol consumption for deployers prior to deployment and up to one year thereafter, with consumption stabilizing to pre-deployment levels at the two year mark (Fink et al. 2017).

Advancing our focus on deployments, Shen et al. (2012) explored SUDs and major depression diagnoses after a deployment with a concentration on deployed location and member service affiliation factors. Their data was derived through a compiled effort, merging Defense Enrollment Eligibility Reporting System (DEERS) demographic data with TRICARE medical encounter data. The study examined a random sample of $n = 678,382$ AD service members who proportionally represented the DOD with respect to the sizes of the U.S. military service branches (Shen et al. 2012). As with the observational period for Fink et al. (2017), a period of five years from 2001 to 2006 was used for the study (Shen et al. 2012).

Concentrating their focus on the analysis of SUDs, Shen et al. (2012) used a logistic regression model consisting of various personal and military demographics as predictors for a binary response variable indicative of being diagnosed with a SUD after deployment. The results for all branches of service reflected significant increases in the odds of a member receiving a SUD diagnosis after being deployed to Afghanistan or Iraq, to a non-classified locations such as Kuwait or Qatar, or to a classified or unknown location (Shen et al. 2012). As our research objectives are narrowed in scope to USN personnel, it is noted that USN members had odds of SUD diagnosis that were 1.77, 1.94, and 1.82 times higher than non-deployed USN members following a deployment to Afghanistan or Iraq, a non-classified location (e.g. Kuwait or Qatar), or a classified or unknown location respectively (Shen et al. 2012). For deployments categorized as short (< 120 days), medium (120 to 180 days), and long (> 180 days), when compared to short deployments all branches with exception to the

Marines reflected increased odds of SUD diagnosis after the deployment (Shen et al. 2012). For USN personnel, service members were 1.12 and 1.06 times more likely to have a SUD diagnosis than short-term deployers for medium or long duration deployments respectively (Shen et al. 2012).

Across these studies, there were many similarities and differences. A shared nature was the use of machine learning algorithms, specifically, logistic regression to evaluate substance abuse in veteran populations. The studies also examined many similar personal and military demographic predictor variables. Furthermore, aside from the study by Fink et al. (2017) significant results were found giving credence to the effective use of machine learning algorithms when examining substance abuse in military populations.

The main contrast across all of the studies pertained to the response variables, which is as expected given the differing research objectives. Even the substance abuse studies that focused on deployments by Fink et al. (2017) and Shen et al. (2012) differed in that while the former examined more of a temporal effect of substance abuse after deployment, the later was more concerned with deployment locations, duration, and a member's branch of service. Furthermore, it is observed that the sample sizes of all the studies were quite widespread, ranging from 1,201 to 678,382 observations. The data collection efforts varied between pulls from large databases or by means of questionnaire.

Our analytical approach and research objectives seek to expand upon the aforementioned research by incorporating some common threads, but also contributing some new. Our research objectives are complementary in a broad sense, being centered upon substance abuse within a military population. Our contribution to the topic of military substance abuse will be unique in its primary concentration, being future substance abuse relapse. Furthermore, while our study does employ the logistic regression machine learning algorithm, we also compare the logistic regression model to kNN and random forest models. Moreover, we also incorporate time-series analysis, survival analysis, and an in-depth look at key military demographics to further bolster our insight into the issue of substance abuse by military members.

CHAPTER 3: Methodology

This chapter commences with the methodology behind our models. Section 3.2 introduces the data sets that were used for this research and describes its structure, features, and limitations. The chapter concludes with discussion of the specific algorithms and associated parameters that were used to build our classification, time series, and survival analysis models.

3.1 Methodology Overview

Machine learning models are renowned for their usefulness in understanding the story and latent structures of a data set, further augmented by their predictive power and applicability to numerous fields of study. Moreover, as discussed by Huddleston and Brown (2019) the aim is to develop an algorithm that can learn from and adapt to newly introduced observations to address real-world interests. As outlined in Section 2.2.1, the methodology for developing our machine learning algorithms adheres to the following steps prescribed by Huddleston and Brown (2019):

1. **Data Acquisition and Cleaning:**

In this step, we first acquired data from OPNAV N17. The data, to be discussed in the next section, will be referred to as ADMITS data since it was retrieved from the ADMITS system. The ADMITS data was then cleaned by carefully handling missing values, anomalies, and outliers within the data. Moreover, the ADMITS data was prepared and formatted to ensure its structure was appropriate for our proposed algorithms.

2. **Feature Engineering and Scaling:**

For this component of the algorithmic recipe, we created new features from the existing data. As will be seen within the following section, the raw ADMITS data contained 33 variables, though for many of them it would be implausible for our algorithms to find them useful in addressing our research objectives (OPNAV 2021a). For this reason, we built new predictors from the raw data to bolster our algorithm's

performance and improve our analysis. By the same token, the response variable was also required to be extrapolated from the raw data. With respect to data normalization, all of our variables were categorical; therefore, there was no need to standardize any of our predictors.

3. **Train-Test Split:**

Once all of our features were established, the data was partitioned into training and test data sets in a manner befitting the algorithm employed. Specifically, for the classification algorithms, an 80–20 split was enforced such that we drew a random sample equivalent to 80% of the full data set's size with the remaining 20% set aside as test data. Furthermore, we ensured that the two partitions contained an equal proportion of 0's and 1's with respect to the response values. The time series algorithms require training data to precede test data as indexed by time. As proposed by Hyndman and Athanasopoulos (2018), an 80–20 split is commonly used, though the test data must at least cover the forecasting period. For our analysis, we adhered to the 80–20 split as that enabled us to have two years for test data to support our one-year forecast.

4. **Model Fitting (Training) and Feature Selection:**

The training data sets were then used as input to the classification, time series, and survival analysis algorithms. Within this step, for classification we also utilized forward step-wise feature selection, in which the model starts with an intercept-only model and adds additional features one at a time based on a specific performance measure (e.g. AIC) (R Core Team 2021). This process results in a combination of features which generally explains the response for a particular model well. For our analysis, we compared models that were identified to be the best performers using both AIC and BIC performance measures. Cross-validation was also implemented to tune the k parameter for the kNN model.

5. **Model Selection:** As depicted by Figure 2.5 within Chapter 2, the goal of a model is to find the optimal bias-variance trade-off region by the model learning from the training data and being evaluated based on its prediction accuracy of the test data response or value (Huddleston and Brown 2019). In this stage, we compared our best performing logistic regression, kNN, and random forest models based on their performance as

measured by AUC. For time series, we compared SARIMA, HW, HW-A, HW-M, and ETS models evaluated by MAPE and MASE performance measures. Our CPHR and RSF survival analysis models were also compared based on their concordance measure.

6. Performance Assessment:

The performance measures of our classification, time series, and survival analysis models were then assessed to determine our best models for each type of analysis.

7. Model Implementation:

Our top-performing models for classification, time-series analysis, and survival analysis were then analyzed to determine variable importance, forecast quality, and other items of interest to address our research questions and the sponsor's needs. For our study, it was important to visualize the results to make known the story behind the data in a manner that would be readily understood and explained. While the other steps are indeed of great importance, if the results of the analysis are not expressed in a manner that can be understood, then the efforts are meaningless.

Of note, the train-test split process was discussed by Huddleston and Brown (2019) though not exclusively as a distinct step. For our approach, the step was amended as a clear-cut action to emphasize the importance of this procedure.

3.2 Data

The structure of a data set can contain different data types. The most common of which are numeric (e.g. 33 or 3.8), logical (i.e. "TRUE" or "FALSE"), and character for text values (Downs 2019). A sequence of characters is referred to as a string (Buttrey and Whitaker 2018). For machine learning, features (i.e columns of data) that contain characters or strings are typically converted to categorical data, also known as factors, with distinct levels representing each unique string (Buttrey and Whitaker 2018). For instance, if a feature of a data set is U.S. states and territories, then there may be 50 or more levels for this feature with each observation containing 1 of the n levels in the absence of missing values. One potential exception to this pertains to features with date strings, which customarily are converted to date objects (Buttrey and Whitaker 2018).

3.2.1 Data Overview

There were three data sets that were used for this study, specifically ADMITS substance abuse data from OPNAV N17, state and rank military population data retrieved from DMDC, and USN Navy Enlisted System (NES) rating data received from BUPERS-34 (OPNAV 2021a; DMDC 2021; BUPERS-34 2021). The raw ADMITS data as received from OPNAV N17 contains 33 categorical variables that are associated with personal or military demographics, abused substance, or SARP information (OPNAV 2021a). There are 101,963 observations representing substance abuse occurrences among USN service members that were entered into the ADMITS system between October 2008 and May 2020 (OPNAV 2021a). While each observation, or DAR, represents one Sailor, not all observations are unique in that some Sailors had repeat observations—a concentration of this research. In the interest of protecting the Sailor’s identities, the ADMITS data was de-identified by OPNAV N17 in that any Personally Identifiable Information (PII) was either removed entirely or given a unique identification number (OPNAV 2021a).

Not all of the 33 categorical variables were practical for our research objectives and analytical methods. However, there was a total of 16 additional variables that were constructed from the raw ADMITS data for predictive purposes. Table 3.1 provides an overview of the features of the enhanced ADMITS data set (OPNAV 2021a). The new variables that were established are in **bold** font. The response variables that were extrapolated are also in **bold** font and designated by an asterisk (*) symbol.

Table 3.1. **ADMITS Data Set Features and Descriptions.** Adapted from: (OPNAV 2021a).

Start of Table	
Feature	Description
ID	Member's unique identifier
SerBranchID	Member's branch of service
mSerBranchID	Member's AD or Reserve Status
CurrentUIC	Command's 5-digit identifier
City	Command's location by city
State	Command's location by state
Zip	Command's location by ZIP code
CommandName	Command's name
cSerBranchID	Command's branch of service
UICRUC	Command's 5-digit identifier
Sex	Member's gender
BirthDate	Member's date of birth
Race	Member's race
EthnicGroup	Member's ethnic group
CurrentMember	Member currently serving in military
PayGrade	Member's pay grade
Rate	Member's USN occupation with rank
GenRate	Member's USN occupation without rank
Dept	Member's command department
IncidentDate	Date of substance abuse incident
ReportDate	Date DAR input into ADMITS system
IncidentInfoName	Incident information for DAR
SubstanceName	Substance used
DUIDWI	Driving under the influence or while intoxicated
ScreeningDate	Date member was screened by SARP facility
sFacilityName	SARP Facility Name
ReferralReason	Reason member referred to SARP
ReferralSource	Quality of referral
RecommendedAction	SARP recommendation for treatment
tFacilityName	SARP treatment facility

Continuation of Table	
Feature	Description
tFacilityType	SARP treatment facility type
ProgramType	SARP program type
ProgramAction	Action or agency that led to SARP referral
Prognosis	SARP duty status suggestion for member's recovery
Counselor	SARP facility counselor
CompletionDate	Date member completed SARP treatment
RepeatBeforeTreat	Member had repeat DAR before SARP screening
nDaysBwIncTreat	Number of days between Incident and Screening
Region	Geographic region of DAR incident
IncidentYear	Year of DAR incident
Season	Season DAR incident was committed
HolidayBirthday	Incident during holiday/birthday (1-week window)
PriorDUIDWI	Member had prior DUIDWI reported
Age	Members age in years
RankTier	Tier of member's rank
nTreatmentDays	Number of days member in SARP treatment
*FutureIncident	Future DAR for member past current incident date
*IsRepeat	Member had previous DAR
*nDaysLastIncident	Number of days since last DAR
End of Table	

There were two types of data retrieved from DMDC: rank and state military population data (DMDC 2021). The annual data sets came partitioned into AD and Reserve data from 2008 to 2020 for all U.S. service branches (DMDC 2021). As our study concerns USN members, the DMDC data was merged to form two 2008 to 2020 USN-only data sets, one for rank and another for state data. Table 3.2 and Table 3.3 indicate the variables for the final DMDC data sets. The rank data consists of 24 levels for the ranks Seaman Recruit to Admiral, to include warrant officers, for each calendar year (DMDC 2021). The state data has 51 levels per calendar year including the District of Columbia (DMDC 2021).

Table 3.2. **DMDC Rank Data Set Variables and Descriptions.** Adapted from DMDC (2021).

DMDC Rank Data	
Variable	Description
Year	Year of DMDC assessment
Rank	Military Rank
ADNavy	Total AD USN members by Year and Rank
ReserveNavy	Total Reserve USN members by Year and Rank

Table 3.3. **DMDC State Data Set Variables and Descriptions.** Adapted from DMDC (2021).

DMDC Rank Data	
Variable	Description
Year	Year of DMDC assessment
State	State within the U.S.
ADNavy	Total AD USN members by Year and State
ReserveNavy	Total Reserve USN members by Year and State

The NES navy rate data sets were obtained from BUPERS-34 and covered the period from 2015 to 2020 (BUPERS-34 2021). As with the DMDC data sets, the NES data was also compartmentalized into annual data sets that were merged together. However, this data was restricted to AD Navy enlisted occupations. The DMDC and NES data sets were only used for our over- or under-representation analysis; therefore, we could not indicate whether certain reserve or officer occupations were over- or under-represented in terms of the number of substance abuse occurrences per month. Table 3.4 provides the variables and respective descriptions for the final NES data set which contained counts for 57 general ratings per calendar year (BUPERS-34 2021).

Table 3.4. **NES General Rating Data Set Variables and Descriptions.**
Adapted from BUPERS-34 (2021).

NES Rating Data	
Variable	Description
Year	Year of assessment
GeneralRating	General rating for USN occupations
ADNavy	Total AD USN members by Year and Rating

3.2.2 Data Limitations

There are two primary limitations with respect to the ADMITS data set: an ADMITS system outage and observations with missing values (OPNAV 2021a). ADMITS experienced a system issue which required the system to be off-line at most locations for nearly a year, from late 2013 through 2014 (Office of the Chief of Naval Operations 2021a). The outage resulted in a significant decrease in DARs input into the ADMITS database during the outage period and the DARs have still not returned to pre-outage levels (OPNAV 2021a). As a result of the ADMITS outage, OPNAV N17 requested a partitioning of the data-set into pre-system and post-system outage data sets for the classification analysis. For the time series and survival analyses, we combined the pre- and post-system outage data sets to remove the affect of the outage. Furthermore, OPNAV N17 called for an independent analysis of alcohol DARs and other drug DARs as clinicians expressed concern with the credibility of other drug reporting in reference to timeliness and frequency. The omission of the DARs during the outage period is observed within Figures 3.1, 3.2, 3.3, and 3.4 which reflect the alcohol and other drug incident rates leading up to and after the system outage. For this analysis, other drug DARs will refer to DARs that were input for substances abused other than alcohol.

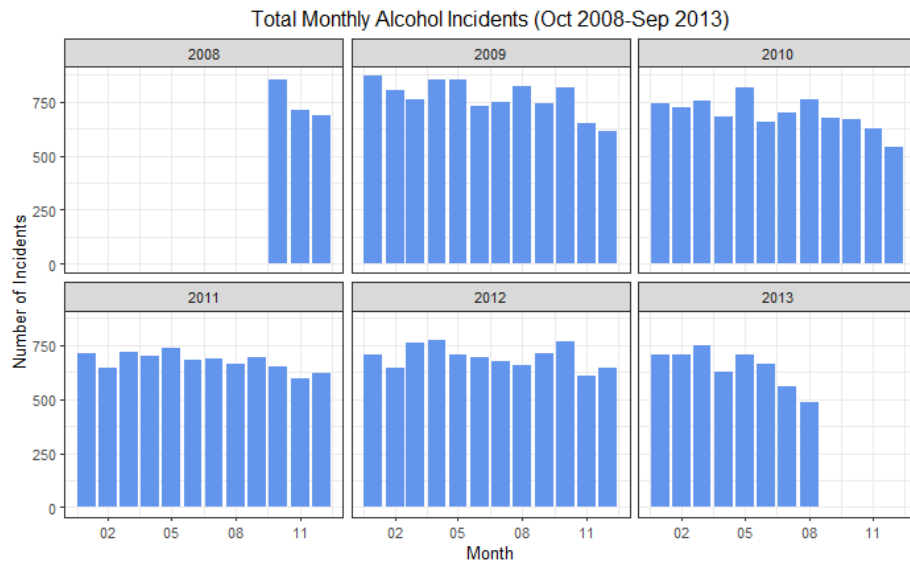


Figure 3.1. The total number of USN personnel with alcohol DARs per month from October 2008 to August 2013 (pre-system outage). The number of observations from June 2013 to August 2013 are observed to decrease near the onset of the system outage period. Adapted from: OPNAV (2021a).

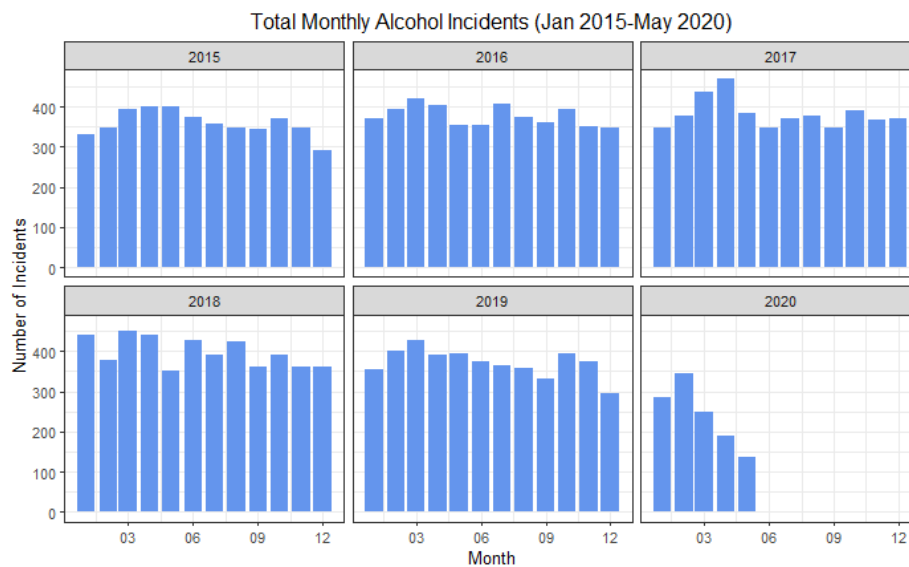


Figure 3.2. The total number of USN personnel with alcohol DARs per month from January 2015 to May 2020 (post-system outage). Adapted from: OPNAV (2021a).

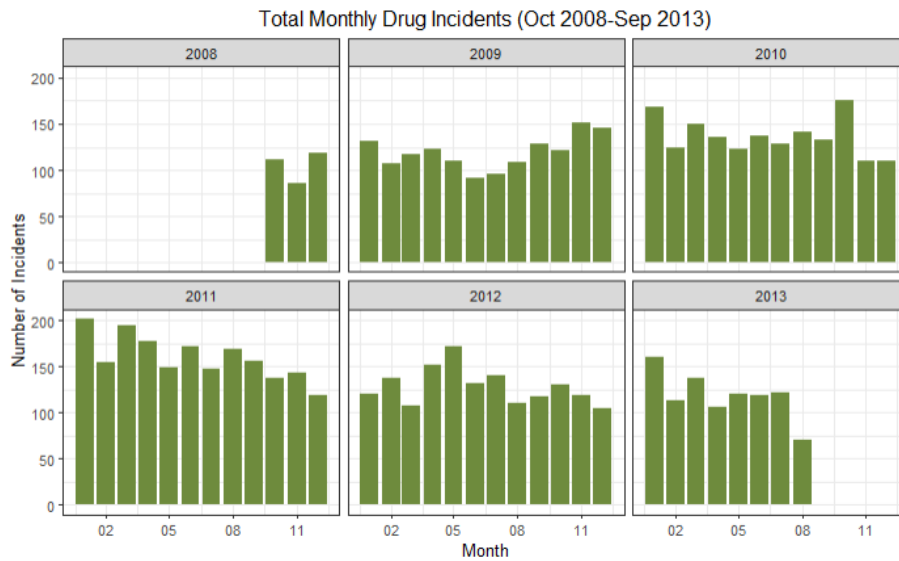


Figure 3.3. The total number of USN personnel with other drug DARs per month from October 2008 to August 2013 (pre-system outage). Adapted from: OPNAV (2021a).

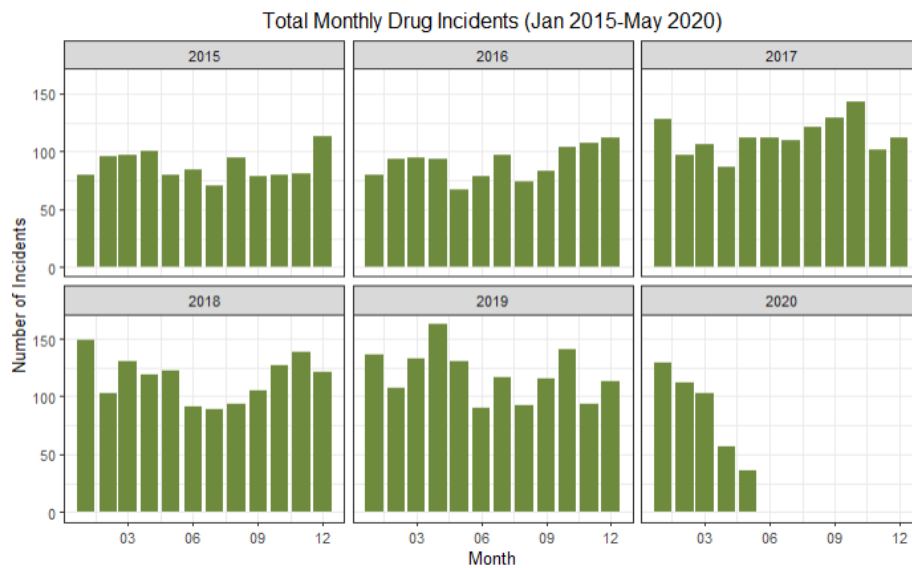


Figure 3.4. The total number of USN personnel with other drug DARs per month from January 2015 to May 2020 (post-system outage). Adapted from: OPNAV (2021a).

Missing values are frequently observed in data sets for many reasons, often a by-product of human error or unknown information (Downs 2019). Regardless of the cause, it is imperative to assess the impact missing values can have in potentially distorting the representation of the data in relation to the population (Downs 2019). Downs (2019) suggests that if the proportion of missing values to sample size is small, then tossing out those observations will have a negligible impact on the analytical results. Whereas, if there are a substantial amount of missing values, then imputation may be required (Downs 2019).

An initial assessment of the missing values within the ADMITS data set was conducted using the `vis_mis()` function from the `naniar` package (Tierney et al. 2021). As reflected within Figure 3.5, 35.4% of the values within the ADMITS data set were missing entries. From Figure 3.5 this is starkly apparent for the SARP features near the right-side of the data set and for IncidentInfoName and Current.UIC. Fortunately, the 13 SARP variables including IncidentInfoName are categorical, and the missing values are justified. Entries into these fields are only required for USN service members who are enrolled into SARP treatment. Therefore, as not all USN personnel are referred to SARP based on clinical criteria and/or command determinations, missing entries are valid and were imputed with a value of NoSARP. The Current.UIC field contains 3,507 levels, and ultimately, was of no use for our analysis. Other fields with missing values were also of no use to our analysis such as UICRUC, CommandName, City, and Zip; therefore, these features were removed. The remaining observations with missing values, 5% of the alcohol observations and 4% of the drug observations, were negligible with respect to the sample size, so the observations were removed. Thereafter, there were approximately 21K observations with a DAR submitted from either a deployed or overseas location which extended beyond the scope of our analysis and were also removed. The resulting data sets for our analysis include (OPNAV 2021a):

- Alcohol Pre- and Post-System Outage: 65,623 observations
 - Alcohol Pre-System Outage: 41,752 observations
 - Alcohol Post-System Outage: 23,871 observations
- Other Drug Pre- and Post-System Outage: 14,610 observations
 - Other Drug Pre-System Outage: 7,853 observations
 - Other Drug Post-System Outage: 6,757 observations

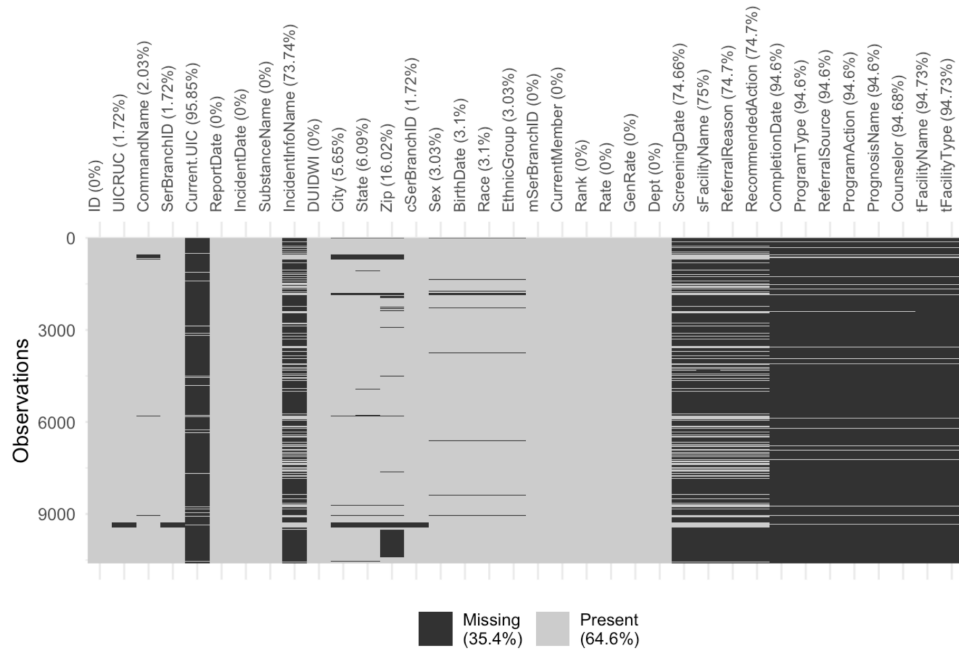


Figure 3.5. The raw ADMITS data had missing values in 35.4% of the entries. Adapted from OPNAV (2021a).

3.3 Model Fitting and Implementation

In this section, we address the algorithmic tools, processes, and parameters used for each of the four analyses we undertook in this study to address our research objectives. Recall from Section 2.2.1 that training and test sets are used to identify the optimal trade-off region between variance and bias (Huddleston and Brown 2019). The model fitting stage is where we make use of the train-test sets to train our models, employ feature selection techniques, and then measure the model’s performance as a function of how well it performs on the test set observations. The better performing models should be closer to or within the optimal trade-off region as depicted in Figure 2.5 (Huddleston and Brown 2019). While the over- or under-representation analysis was EDA, the classification, time series, and survival analysis were statistical machine learning algorithms that made use of the train-test data to build models with classification, forecasting, or predictive power.

3.3.1 Over- or Under-Representation

This analysis was used to gauge whether certain subgroups were over- or under-represented with respect to their proportional size in relation to their proportional contribution of alcohol DARs. There were four subgroups that were assessed: USN duty locations within the U.S. by state, USN ranks, and USN general ratings or occupations. This analysis made use of all three data sets, these being ADMITS, DMDC, and NES data. OPNAV N17 was invested in this analysis for the insight that was able to be drawn when comparing proportions rather than counts.

States

For the AD state-level analysis, we utilized the ADMITS and DMDC data sets (OPNAV 2021a; DMDC 2021). The following parameters were used to determine a state's over- or under-representation ratio for each year from 2008 to 2020 (OPNAV 2021a; DMDC 2021):

- **TotalADNavyPers** = total number of AD Navy personnel per year from 2008 to 2020
- **TotalADNavyState** = total number of AD Navy personnel for each state per year from 2008 to 2020
- **StatePercent** = $\text{TotalADNavyState} / \text{TotalADNavyPers}$
- **TotalOccurrences** = total AD alcohol substance abuse incidents per year from 2008 to 2020
- **TotalStateOccurrences** = total AD alcohol substance abuse incidents for each state per year from 2008 to 2020
- **StateOccurrencePercent** = $\text{TotalStateOccurrences} / \text{TotalOccurrences}$
- **StateRepresentationRatio** = $\text{StateOccurrencePercent} / \text{StatePercent}$

The calculations for the Reserves mirrored the AD calculations.

For the percent over- or under-representation of each state per year, a state with $\text{StateRepresentationRatio} < 1$, indicated an under-representation, whereas $\text{StateRepresentationRatio} > 1$ revealed an over-representation. In this context, states that are under-represented based on their annual population sizes and substance abuse occurrences are performing better than over-represented states.

To visualize the results, we made use of the `plot_geo()` function from the `plotly` pack-

age to generate AD and reserve interactive choropleth maps from 2008 to 2020 (Sievert 2020). The choropleth visualizations enable a swift comparative assessment of a state's performance in relation to other states by its color designation. For our choropleth's color scheme, we employed a sliding scale ranging from light brown to black. The `plot_geo()` function shaded states with the lowest StateRepresentationRatio values light brown, and alternatively, black for the states with the highest StateRepresentationRatio values (Sievert 2020). Therefore, light brown and black shades are indicative of the best and worst performing states respectively. Correspondingly, moderate performers will have either darker shades of brown or red applied. The choropleth maps are a useful visual aid for a quick assessment of a states relative performance as well as for the identification of under- or over-performing regions.

For our results, there are 24 choropleth maps for AD and reserve populations, 28 plots consisting of AD and reserve ranks, and 57 AD rating plots. The visualizations may be found within Appendix 1:

- State choropleth maps: Section A.1
- Rank plots: Section A.2
- Rating plots: Section A.3

Ranks

The same data sets were used for the rank subgroup analysis as in the state analysis, being the ADMITS and DMDC data sets (OPNAV 2021a; DMDC 2021). To determine the percent over- or under-representation amongst USN AD rank subgroups we computed the following (OPNAV 2021a; DMDC 2021):

- **TotalADNavyPers** = total number of AD Navy personnel per year from 2008 to 2020
- **TotalADNavyRank** = total number of AD Navy personnel for each rank per year from 2008 to 2020
- **RankPercent** = TotalADNavyRank / TotalADNavyPers
- **TotalOccurrences** = total AD alcohol substance abuse incidents per month from 2008 to 2020
- **TotalRankOccurrences** = total number of alcohol substance abuse incidents per month for each AD rank

- **RankOccurrencePercent** = TotalRankOccurrences / TotalOccurrences
- **RankRepresentationRatio** = RankOccurrencePercent / RankPercent

As with the state analysis, both AD and reserve populations were analyzed. The reserve rank calculations mirror the AD rank calculations.

By computing the ratio of RankOccurrencePercent to RankPercent, it follows that if a specific rank subgroup is measured to be a proportion of the population that is equivalent to the proportion of their occurrences, then their assessed value would be 1. Therefore, under-representation is indicated by RankRepresentationRatio values less than 1 and over-representation is reflected by RankRepresentationRatio values greater than 1.

The results for this analysis was visualized by using the `ggplot()` function from the `ggplot2` package (Wickham 2016). Furthermore, the `ggplot()` function was augmented with a loess smoothing method, which is a non-parametric method for fitting a trend line to the observations (Wickham 2016). Collectively, 18 plots for the paygrades of a Seaman Recruit to Captain were created for AD USN personnel to include several warrant officer ranks. Likewise, 11 rank plots were created for reserve USN personnel ranging from Seaman Recruit to Senior Chief Petty Officer, and for Lieutenant and Lieutenant Commander officer ranks. Fewer plots were produced for the reserves than AD as many of the ranks had an inadequate number of observations to develop meaningful results that were also not misleading.

Ratings

For the AD rating subgroup analysis, we utilized the ADMITS and NES data sets (OPNAV 2021a; BUPERS-34 2021). As with the state and rank analyses, the percent over- or under-representation was based on a ratio with the baseline set at 1. As well, similar to the rank plots, the `ggplot()` function was used to produce the visualizations, of which 57 AD rating plots were generated (Wickham 2016). However, we did not have reserve rating data; therefore, this analysis was limited to AD USN personnel. To determine the percent over- or under-representation amongst USN AD general rating subgroups we computed the following (OPNAV 2021a; BUPERS-34 2021):

- **TotalADNavyPers** = total number of AD Navy personnel per year from 2015 to 2020

- **TotalADNavyRating** = total number of AD Navy personnel for each general rating per year from 2015 to 2020
- **RatingPercent** = $\text{TotalADNavyRating} / \text{TotalADNavyPers}$
- **TotalOccurrences** = total AD alcohol substance abuse incidents per month from 2015 to 2020
- **TotalRatingOccurrences** = total number of alcohol substance abuse incidents per month for each AD general rating
- **RatingOccurrencePercent** = $\text{TotalRatingOccurrences} / \text{TotalOccurrences}$
- **RatingRepresentationRatio** = $\text{RatingOccurrencePercent} / \text{RatingPercent}$

3.3.2 Classification

This study compared three classification algorithms, logistic regression, kNN, and random forest, which were trained and built into models used to classify our relapse response variable of interest. As mentioned, OPNAV N17 requested a separate analysis for alcohol and for other drugs (OPNAV 2021a). Hence, the response was either future alcohol repeat incident or future drug repeat incident, depending upon whether alcohol or other drug data sets were used. For the purposes of our study, the response values will be referred to as FutureIncident.

Furthermore, OPNAV N17 expressed its concern with the data during the system outage period and desired assessments of the data prior to and after the system outage (OPNAV 2021a). Throughout the remainder of this thesis, the distinct ADMITS data sets that were enhanced with the new features discussed in section 3.2.1 will be referenced as follows:

- Alcohol Pre- and Post-System Outage: alcohol enhanced full
 - Alcohol Pre-System Outage: alcohol enhanced 1
 - Alcohol Post-System Outage: alcohol enhanced 2
- Other Drug Pre- and Post-System Outage: drug enhanced full
 - Other Drug Pre-System Outage: drug enhanced 1
 - Other Drug Post-System Outage: drug enhanced 2

To address our research interests, logistic regression, kNN, and random forest models were built separately using the alcohol enhanced 1, alcohol enhanced 2, drug enhanced 1, and drug enhanced 2 data sets. The models were produced independently such that the

features selected, and parameter settings used were dependent upon the specific data set being analyzed. Each data set was also partitioned into a training and test set to train our models with the training data and assess their performance by how well the model predicted the response of the newly introduced test set observations. The `createDataPartition()` function from the `caret` package was used to create the 80–20 split in a manner such that the proportion of response values being 1 or 0 was equally represented in both partitions (Kuhn 2008). For each of our classification algorithms, we used the associated functions, feature selection techniques, tuning methods and parameters described:

- **Logistic Regression:**

The `glm()` function from the `stats` package was employed with `FutureIncident` as the response value of interest and the `family` parameter was set to `logit` for logistic. The model produced from the `glm()` function was then used as input into the `step()` function, which also comes from the `stats` package, and the `direction` parameter was set to `both` (R Core Team 2021). By using `both` as the directional parameter, the `step()` function employs forward and backward feature selection based on the AIC criterion as default. The `step()` function was also utilized to test for interactions among the features that could possibly improve performance. Any interactions that were found resulted in negligible improvement to the AIC performance measure; therefore, they were deemed not to be worth the additional model complexity.

- **kNN:**

The ADMITS data set contained many categorical features with more than 2 levels for each feature. Therefore, while there are other approaches, the method we chose to employ to make the data set work with the kNN algorithm was to create a sparse data set consisting of 1's and 0's with each level instantiated as a feature. The `dummyVars()` from the `caret` package was used to perform this operation (Kuhn 2008). For instance, if the initial feature was `Substance` containing the values `marijuana`, `cocaine`, and `morphine`, then the new dataset would have `Substance_marijuana`, `Substance_cocaine`, and `Substance_morphine` as new features. Moreover, if a prior observation had a value of `marijuana` listed under the `Substance` feature, then a value of 1 would be contributed to the new feature `Substance_marijuana`, and a 0 would be annotated for `Substance_cocaine` and `Substance_morphine`. As can be imagined, this method can create a sparse data set depending upon the number of distinct levels in a data set,

but is able to be fielded appropriately by the kNN algorithm to build a classification model. With sparsity in mind, we chose to use features such as Region instead of State, and RankTier rather than Rank to reduce the amount of levels.

Additionally, the `tune.knn()` function from the `e1071` package was used to determine the optimal value for the parameter k , or the number of nearest neighbors to be used by the kNN algorithm (Meyer et al. 2021). The `tune.knn()` employs a 10-fold cross-validation method and searches across a specified range of potential k values (Meyer et al. 2021). The ranges that we used to search for an optimal k were dependent upon the size of the data set. In general, we searched across approximately 50 values of k to determine which k value produced the lowest error rate. The best identified k value was then taken as input by the `knn()` function from the `class` package to build the model and evaluate its performance against the test set (R Core Team 2021).

- **Random Forest:**

The `randomForest()` function from the `randomForest` package was used to build the random forest model (Breiman and Cutler 2022). Like the other classification functions, the `randomForest()` function takes the training set as input. The parameters that were used within the `randomForest()` function were `ntrees`, `mtry`, and `nodesize` (Breiman and Cutler 2022). The parameter `ntrees` refers to the number of bootstrap sampled trees to grow, `mtry` is the number of randomly selected features to consider as split candidates, and `nodesize` set the minimum number of observations for terminal nodes (Breiman and Cutler 2022). The parameter settings were dependent upon the data set being analyzed, but in general, `ntrees` was set to be greater than or equal to 500, `mtry` altered between 3 and 4, and `nodesize` ranged from 5 to 70.

For all algorithmic models, the `predict()` function from the `stats` package was used to predict the classes of the observations from the test set using the trained models (R Core Team 2021). Confusion matrices were then produced using the `confusionMatrix()` function from the `caret` package (Kuhn 2008). The performances of models were visually compared by making use of the `performance()` function from the `ROCR` package and feeding the output into the `ggplot()` function (Sing et al. 2005). The model's performances were assessed based on the AUC criterion that was also retrieved from the output of the `performance()` function.

3.3.3 Seasonal Decomposition and Time Series Forecasting

We implemented time series algorithms to determine the degree of accuracy with which we can forecast alcohol or other drug substance abuse incidents given the removal of observations during the system outage period. As with the classification analysis, alcohol and drug DARs were independently analyzed. However, the ADMITS data sets used for the time series algorithms were the combined pre-system outage and post-system outage data: alcohol enhanced full and drug enhanced full. The structure for the alcohol and drug data was also altered to be compatible with the time series algorithms such that the data was compressed to reflect only the total number of monthly alcohol or other drug observations for the time period of study.

The monthly alcohol and other drug data sets were partitioned into an 80–20 train-test split. The time periods of the observations spanned from October 2008 to August 2013 and September 2015 to March 2020. The removal of observations from September 2013 to August 2014 not only removed observations during the system outage, but also allowed for any annual seasonality or cyclic behavior to be maintained as September data followed August data, granted in different years. Therefore, the train-test split was as follows for both the alcohol and other drug time series data:

- **80–20 Train-Test Split:**

Training data: October 2008–August 2013, and September 2015–January 2018

Test data: February 2018–March 2020

For our analysis of the time series data, we first generated a STL model using the `stl()` function from the `stats` package to identify any seasonal patterns or trends in the time series (R Core Team 2021). We then proceeded to build seven forecasting models: naïve, seasonal naïve, SARIMA, HW, HW-A, HW-M, and ETS. To train our models and measure performance, we used a rolling-horizons design in which each period was one month and a month from the test data was added to the training data sequentially with each iteration. The models were taken as input into the `forecast()` function from the `forecast` package (Hyndman et al. 2022). The models were then compared in performance as determined by their ability to accurately forecast test data observations, which was numerically assessed by 1- and 12-step MAPE and MASE performance measures. To produce the models, the following functions and associated parameters were used:

- **Naïve and Seasonal Naïve:** To build the Naïve and Seasonal Naïve models, we used the `naive()` and `snaive()` functions from the `forecast` package (Hyndman et al. 2022). The parameters `y` and `h` were used for the training data and length of the forecasting periods respectively (Hyndman et al. 2022).
- **SARIMA:** The SARIMA model was established using the `auto.arima()` function from the `forecast` package (Hyndman et al. 2022). The function searches across all non-seasonal (p,d,q) and seasonal (P,D,Q) parameters to best fit the data for the $ARIMA(p,d,q)(P,D,Q)$ model based on AIC or BIC criteria (Hyndman et al. 2022).
- **HW:** Unlike the previous two models, the HW model was built using the `stats` package (R Core Team 2021). The `HoltWinters()` function was used taking the training data as input and best determines the two smoothing parameters, α and β , pertaining to the level and slope as discussed in section 2.3.2. (Hyndman and Athanasopoulos 2018).
- **HW-A and HW-M:** The HW-A and HW-M models were both derived from the `hw()` function, which comes from the `forecast` package (Hyndman et al. 2022). As in the other models, the training data is taken as input. The `hw()` function also contains a `seasonal` parameter that is set to additive for HW-A models and multiplicative for the HW-M model (Hyndman et al. 2022). For the HW-A and HW-M models, in addition to finding the best exponential smoothing parameters α and β , the `hw()` function also attempts to find the best parameter γ for their seasonal component calculations (Hyndman and Athanasopoulos 2018).
- **ETS:** Like the SARIMA model, the ETS model can be set to search across various parameter settings to identify a best fitting model. The ETS model is built using the `ets()` function from the `forecast` package (Hyndman et al. 2022). For this function, in addition to taking the training data as input, we also made use of the `model` parameter, which is able to be set as `ZZZ` to automatically search across the various Error, Trend, and Seasonal component properties (Hyndman et al. 2022).

3.3.4 Survival Analysis

We undertook survival analysis to better understand the duration of time between repeat substance abuse incidents among USN personnel. More specifically, as our observations come from ADMITS data, we concern our efforts on the survival of USN personnel that have at least one DAR entry from October 2008 to May 2020. In this regard, the pattern of survival of USN personnel is illustrated by considering the probability of a repeat substance abuse occurrence at each consecutive period of time.

In the context of survival analysis terminology, an exposure, or the incident that starts the clock for an observation, is the DAR IncidentDate. A repeat substance abuse occurrence, which is the submission of another DAR entry for a unique member that is beyond the IncidentDate of a prior DAR entry, is considered an event or “death” that stops the clock for an observation. If such an event occurred, then we set the IsRepeat feature value for the observation to be 1, or 0 otherwise. The time period of interest for the survival analysis is the time between an initial DAR entry’s IncidentDate and a subsequent DAR entry’s IncidentDate. The time between incidents was extrapolated from the raw ADMITS data and built into the ADMITS enhanced data sets with a given feature name of nDaysLastIncident. Therefore, if there was a new DAR entry for a USN service member who already had a previous DAR entry, then the new observation would have a value in the nDaysLastIncident column of the data set. Likewise, if a new DAR entry is associated with a USN service member who did not have a previous substance abuse incident, the observation’s nDaysLastIncident value would be “NA.”

We analyzed alcohol and other drug related DARs separately for the survival analysis. The response variable for the analysis, nMonthsLastIncident, was extrapolated from the nDaysLastIncident feature by dividing the nDaysLastIncident values by 30.4167 days, which is an approximate average number of days per month. This was done to reduce the computational complexity on the algorithm by reducing the number of time periods from 3,712 to 122, which directly correspond to the number of observational days and months respectively. As discussed in Section 2.4.1, survival and hazard functions compute the probability of surviving or experiencing an event given an observations survival up to time t . Therefore, we reduced the complexity of the computations required by reducing the number of time periods. For our analysis, the impact was not necessarily observed for the KM Method algorithm, but for the CPHR and RSF algorithms as their hazard and survival functions also

take an observation's features into consideration as discussed in Section 2.4.2.

Observations were right-censored for our analysis. Specifically, if an observation was for a service member who did not relapse during the observational period, then that observation was right-censored. For service members who did relapse during the observational period, the survival time was computed to be the number of days between incidents. One limitation of the analysis is that some USN service members may have had DAR observations in ADMITS and then either separated or retired prior to May 2020. However, the ADMITS data set does not contain separation or retirement dates that could be used to right-censor these observations and reduce the number of personnel at risk. As well, it is plausible that some members who separated or retired may have relapsed during the observational period though this was not captured by ADMITS given that the system is used only for AD or Reserve service members.

KM Method

The KM model was fit using the `survfit()` and `Surv()` functions from the `survival` package (Therneau 2021). The `survfit()` function takes either a formula or previously fitted model as input. For our analysis, we used the `Surv()` function as the formula for input into the `survfit()` function (Therneau 2021). Within the `Surv()` function we used the parameters `time` for survival time, `event`, `type` for censoring method, `conf.type` or confidence interval type, and `data` (Therneau 2021). The `time` parameter was set to be the vector `nMonthsLastIncident`, `event` was the vector `IsRepeat`, `type` was set to "right", `conf.type` was established as "log-log", and the `data` was either the alcohol or drug enhanced full ADMITS data set.

CPHR

Our CPHR model was built using the `coxph()` and `Surv()` functions from the `survival` package (Therneau 2021). The `coxph()` function is quite similar to the `survfit()` with respect to parameters. The `coxph()` function has a parameter `formula`, which we set to be the `Surv()` function (Therneau 2021). For the `Surv()` function we used the same parameter inputs as the KM model with exception to the `conf.type` parameter as this is done by the `coxph()` function (Therneau 2021). The `time` parameter was set to be the vector `nMonthsLastIncident`, `event` was the vector `IsRepeat`, and `type` was set to "right".

Although, dissimilar to the KM model, in the `Surv()` function we now use the features of the data set to predict the hazard probabilities of the observations. The `data` parameter took either the alcohol or drug enhanced full ADMITS data set as input (Therneau 2021).

RSF

The RSF model was developed using the `ranger()` and `Surv()` functions. The `ranger()` function comes from the `ranger` package and can be used for classification, regression, and survival forests (Wright and Ziegler 2017). The `ranger` function has a `formula` parameter for which we used the `Surv()` functions taking the `nMonthsLastIncident` and `IsRepeat` vectors as the response inputs and the features of the data set for predictors as in the `coxph()` model (Wright and Ziegler 2017; Therneau 2021). The `ranger` function also has parameters built in that are useful for random forest and RSF models such as `mtry`, `num.trees`, and `min.node.size` (Wright and Ziegler 2017). The `mtry` parameter was set to be 3 specifying the number of features to randomly select as split candidates for each decision tree. The `num.trees` parameter was set to be 500 regarding the number of survival trees to grow. We set the `min.node.size` parameter to be equal to 5, that being the minimum number of observations in a node that the algorithm would consider for further splits.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 4:

Analysis and Results

This chapter presents the results of our analysis with respect to the proportional measurements for state, rank, and rating subgroups, as well as for our classification, time series, and survival analysis models. The results include important visualizations, key observations and findings, as well as performance measures for model comparison.

4.1 Model Results and Analysis

4.1.1 Alcohol Abuse Over- or Under-Representation

The distinct subgroups that were assessed for over- or under- representation with respect to alcohol substance abuse occurrences were drawn from three demographics: the USN service member's state of residence, his or her rank based on pay-grade, and occupational category (i.e. general rating). Of note, the rank, state of residence, and general rating of a USN service member was determined based upon the observation's DAR submission date into ADMITS. Furthermore, any USN service member substance abuse incidents that occurred during the ADMITS system outage period were excluded. The results bring into perspective the proportion of alcohol occurrences by a subset of the population measured against its proportional make-up of the population.

For our results, there are 24 choropleth maps for AD and reserve populations, 28 plots consisting of AD and reserve ranks, and 57 AD rating plots. Therefore, we only display one visual for each demographic within this section, selected in an unsystematic way for explanatory purposes only. The remaining plots may be found within Appendix 1:

- State choropleth maps: Section A.1
- Rank plots: Section A.2
- Rating plots: Section A.3

State

The choropleth map within Figure 4.1 displays the U.S. states that were over- or under-represented with respect to the proportion of USN AD service members residing in a particular state, and the proportion of DARs for AD service members that were submitted in the same state in 2020. A over- or under-ratio value of 1 is the baseline for the state choropleth maps indicating that the population proportion is equivalent to the proportion of alcohol substance abuse occurrences for a particular state. Therefore, ratios less than 1 illustrate an under-representation, while ratios greater than 1 represent an over-representation. As choropleth maps shade regions relative to the highest and lowest observed values, it is evident in Figure 4.1 that in 2020 North Carolina and Louisiana had the lowest (i.e. 0.22) and highest (i.e. 3.79) ratios respectively. With the scale set to the observed values of North Carolina and Louisiana, the other states will be shaded based on their ratios position within the numerical scale provided to the right of the choropleth map. We also observed that some states did not have USN AD populations and/or alcohol DAR submissions – these states will have no shade rendered, being white.

2020 AD Alcohol Occurrences Percent Over- or Under-Represented by State

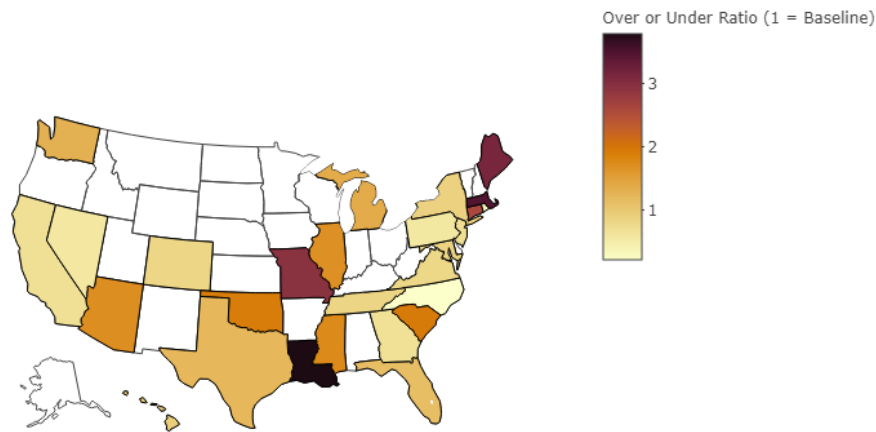


Figure 4.1. AD USN 2020 Alcohol Substance Abuse Over- or Under-Representation. North Carolina had the lowest ratio at 0.22 and Louisiana had the highest ratio at 3.79. Other states are shaded based on their position between the highest and lowest ratios as indicated by the sliding scale to the right of the choropleth map. States above the baseline ratio of 1.00 are over-represented, and states below the baseline ratio are under-represented. Source: OPNAV (2021a); DMDC (2021).

Rank

The visual in Figure 4.2 reflects the over- or under- representation ratio for an AD USN service member of pay-grade E-4, which is a Petty Officer Third Class rank. The observations span the time periods October 2008 to August 2013 and January 2015 to May 2020. The red baseline on the plot is static at 1. As discussed in section 3.3.1, each monthly observation, RankRepresentationRatio, is a ratio of the proportion of a subgroup's alcohol incidents to the subgroup's proportion of the population. Therefore, in Figure 4.2, observed points below the red baseline represent an under-representation in alcohol substance abuse incidents by AD USN E-4's. In contrast, monthly observations above the baseline designate an over-representation for this subgroup.

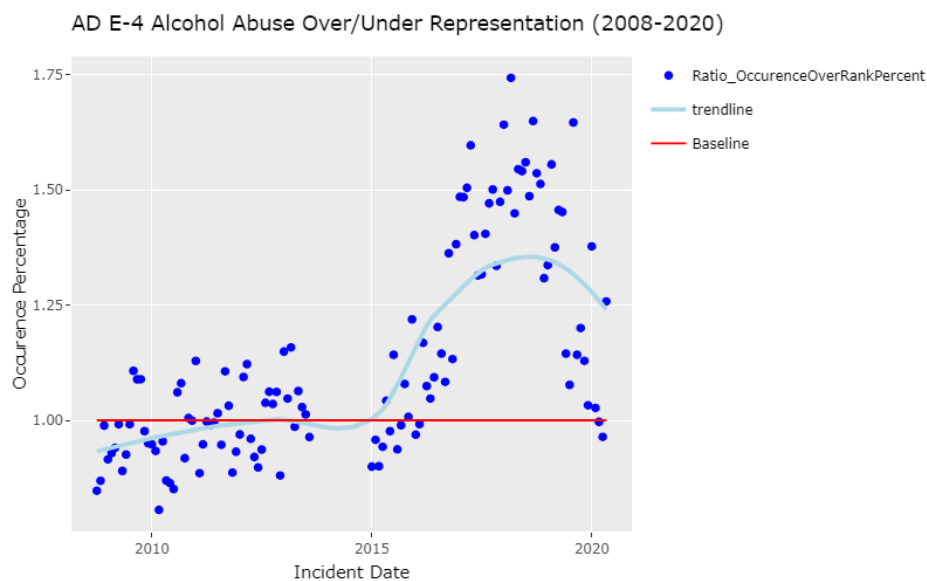


Figure 4.2. AD USN E-4 Alcohol Substance Abuse Over- or Under-Representation. The monthly observations reflect observed representation ratios that hover near the baseline from October 2008 to August 2013, trend upward from 2015 to 2019, and trend downward into May 2020. Source: OP-NAV (2021a); DMDC (2021).

Rating

The AD USN general rating plots are interpreted in the same manner as the rank plots. The general rating plots have monthly points that were computed based on the general rating's proportion of alcohol substance abuse occurrences to the rating's proportion of the AD USN enlisted occupations. We labeled this ratio the RatingRepresentationRatio, which are the monthly observed points within the general rating plots. Figure 4.3 reflects the over- or under-representation ratio for USN builders, or BU rated personnel from 2015 to 2020.

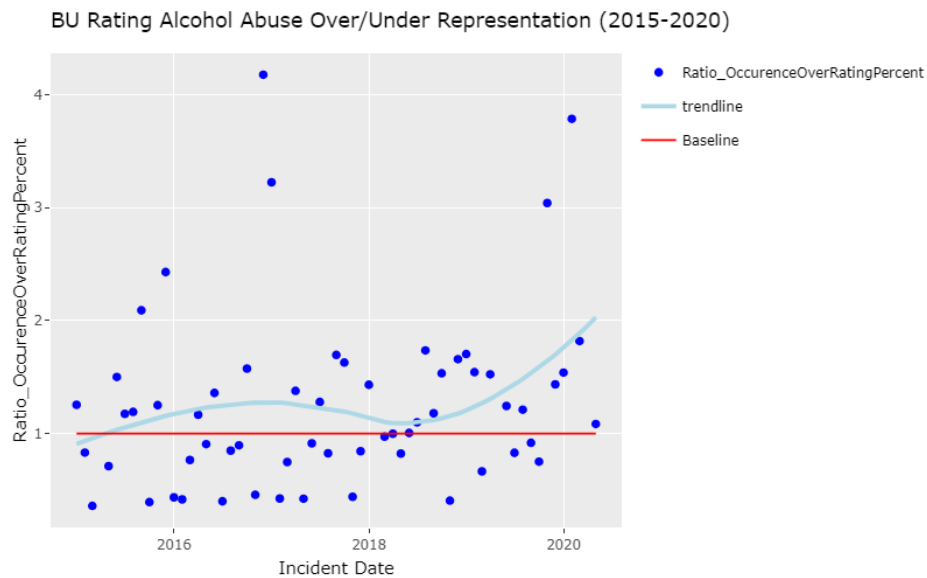


Figure 4.3. AD USN BU General Rating Alcohol Substance Abuse Over- or Under-Representation. The observations display a trend that holds near to the baseline from 2015 to late 2019. An upward trend towards an over-representation value of ≈ 1.75 or 75% above the baseline starts in 2019. Source: OPNAV (2021a); BUPERS-34 (2021).

4.1.2 Classification

Future Repeat Alcohol Abuse Pre-System Outage

For the analysis of alcohol incidents prior to the ADMITS system outage date, the classification models performed fairly well in predicting that a USN service member would have

a future relapse. As observed at the bottom of Figure 4.4, all of the models had an AUC value above 0.7, with the random forest model having the highest AUC value at 0.774. Correspondingly, the higher measured performance for the random forest model is also observed in Figure 4.4 with a ROC curve that rests above the logistic regression and kNN ROC curves.

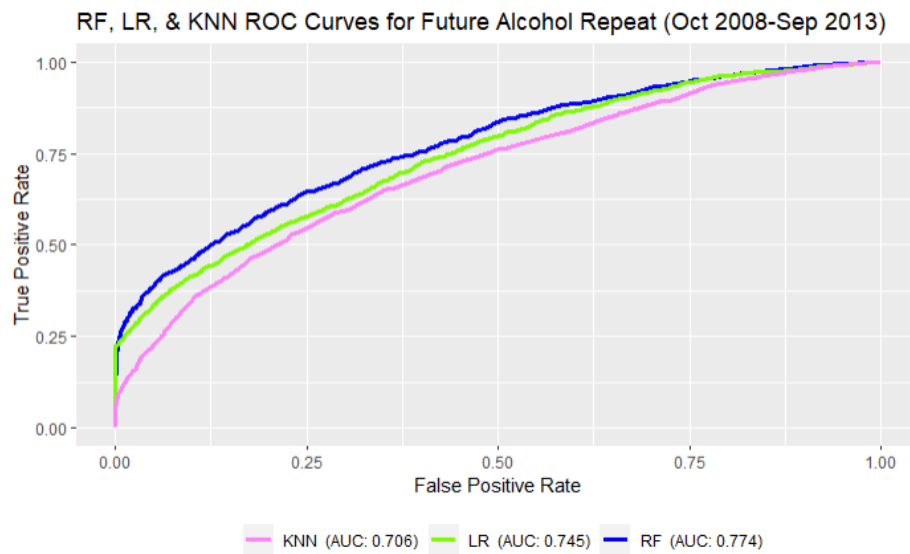


Figure 4.4. ROC curves for all classification models that were built to predict a future repeat alcohol incident using post-system outage data. As ideal ROC curves draw near to the top-left corner, it is observed that the random forest model outperforms both the logistic regression and kNN models. Source: OPNAV (2021a).

As the random forest model was determined to be the best model, we present a chart in Figure 4.5 displaying each feature's measured importance for classifying whether or not a service member will relapse. From the top of the chart to the bottom, the features are ordered from most to least influential as measured by the model's mean decrease in accuracy when that feature is not considered. The IsRepeat feature was identified to be the most important feature. Therefore, when IsRepeat was not selected as one of the m of p predictors this led to the highest decrease in model accuracy. The IsRepeat predictor identifies if a service member at the time of an incident had a prior substance abuse incident. Since our model is attempting to classify whether someone will relapse in the future, the importance of the

IsRepeat predictor informs us that knowing whether an individual has already relapsed is predictive of relapsing again in the future.

The next important feature was State, which refers to the member’s duty location at the time of the substance abuse incident. One word of caution when considering variable importance is that random forest algorithms can favor features with a high number of distinct levels. While many of the features with a substantial number of levels were not considered, the GenRate, State, Dept and EthnicGroup features have the highest number of levels for those considered with 70, 51, 31, and 23 respectively. In figure 4.5 we notice that State, GenRate, and Dept have relatively high importance. While the features may have been beneficial to the random forest model in correctly classifying the relapse response for an observation, the importance may be somewhat overstated.

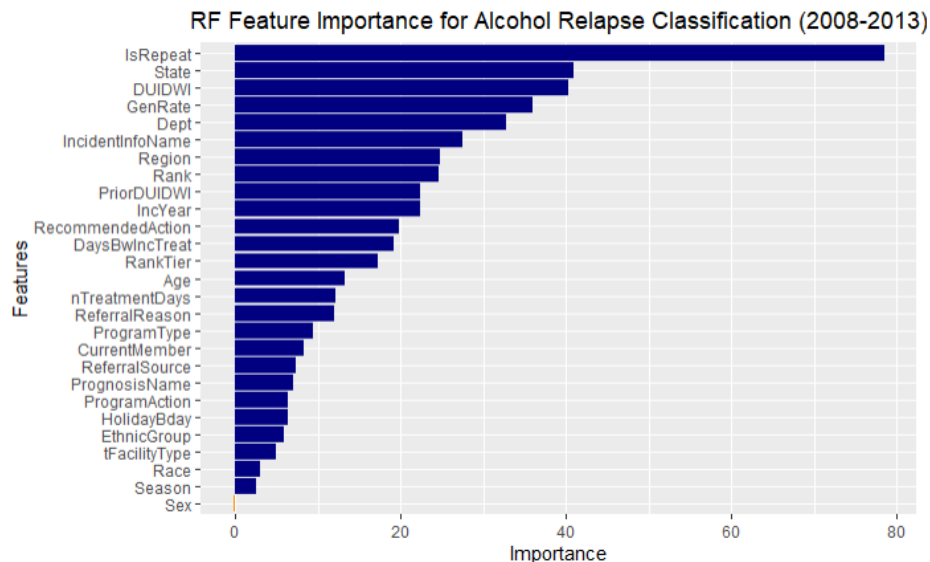


Figure 4.5. Random forest model features of importance using pre-system outage alcohol data. The features are ordered in a top-down fashion of measured importance with respect to their usefulness to the random forest model in correctly classifying the response—future repeat alcohol abuse. Source: OP-NAV (2021a).

Future Repeat Alcohol Abuse Post-System Outage

Now we contrast the classification models using the pre- versus post-system outage data sets. As observed in Figure 4.6, the results were similar with respect to the ordering of the algorithms by measured performance. The random forest model performed the best, followed by the logistic regression model, and the kNN on the tail-end. Although, while the random forest and logistic regression models still performed fair with AUCs of 0.71 and 0.70 respectively, their classification accuracy dropped by $\approx 5.9\%$ and 4.3% respectively. The kNN algorithm performed poorly yielding a 0.645 AUC.

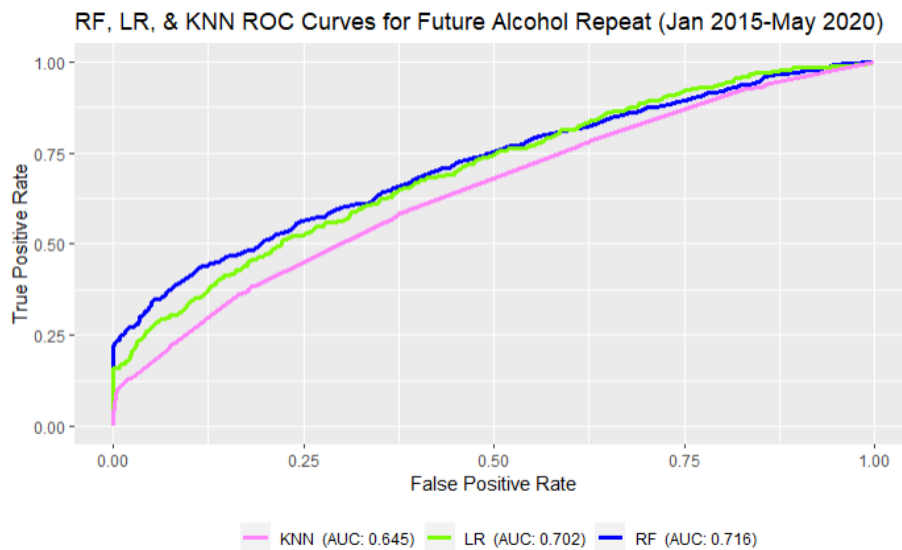


Figure 4.6. ROC curves for all classification models that were built to predict a future repeat alcohol incident using post-system outage data. As ideal ROC curves draw near to the top-left corner, it is observed that the random forest model outperforms both the logistic regression and kNN models. Source: OPNAV (2021a).

As the random forest model outperformed the logistic regression and kNN models, again we display feature importance for the model in Figure 4.7. It is noteworthy that for both pre- and post-system outage models, IsRepeat has the highest importance and GenRate, Dept, IncYear, Rank, and State are relatively high. This further bolsters our prior analysis into the over- or under-representation among state, rank, and general rating subgroups.

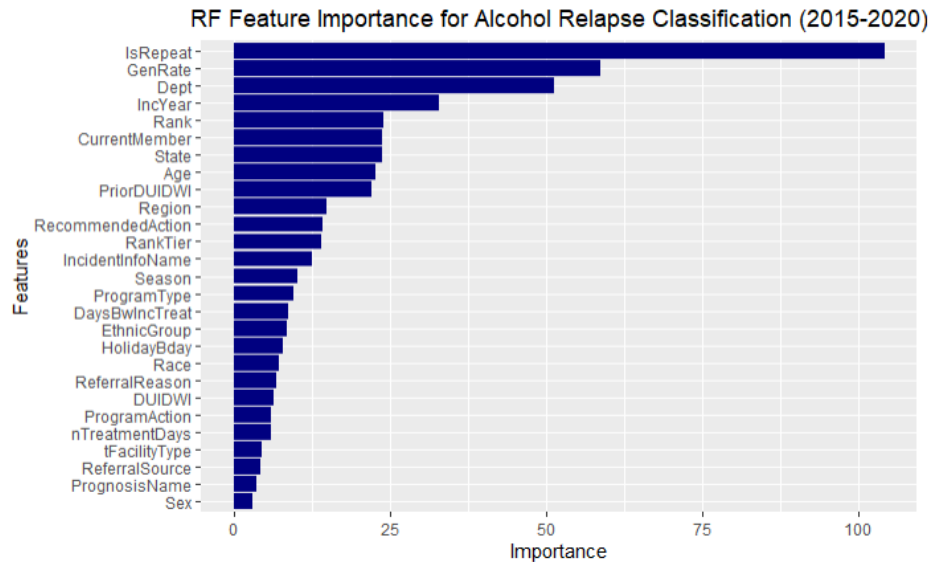


Figure 4.7. Random forest model features of importance using post-system outage alcohol data. The features are ordered in a top-down fashion of measured importance with respect to their usefulness to the random forest model in correctly classifying the response—future repeat alcohol abuse. Source: OP-NAV (2021a).

Future Repeat Other-Drug Abuse Pre-System Outage

We now shift our focus to the classification models that were built to predict a future other-drug relapse incident. For the purpose of our research, we classified substance abuse incidents that were not associated with alcohol as “other-drug.” This category consists of such substances as marijuana, cocaine, amphetamines, etc., and are reflected within Figure 4.8. Analogously to the alcohol classification models, the ADMITS data was partitioned into pre- and post-system outage data sets. Just as the alcohol data sets were filtered to only alcohol substance DAR entries, the other-drug data sets were restricted to substances other than alcohol.

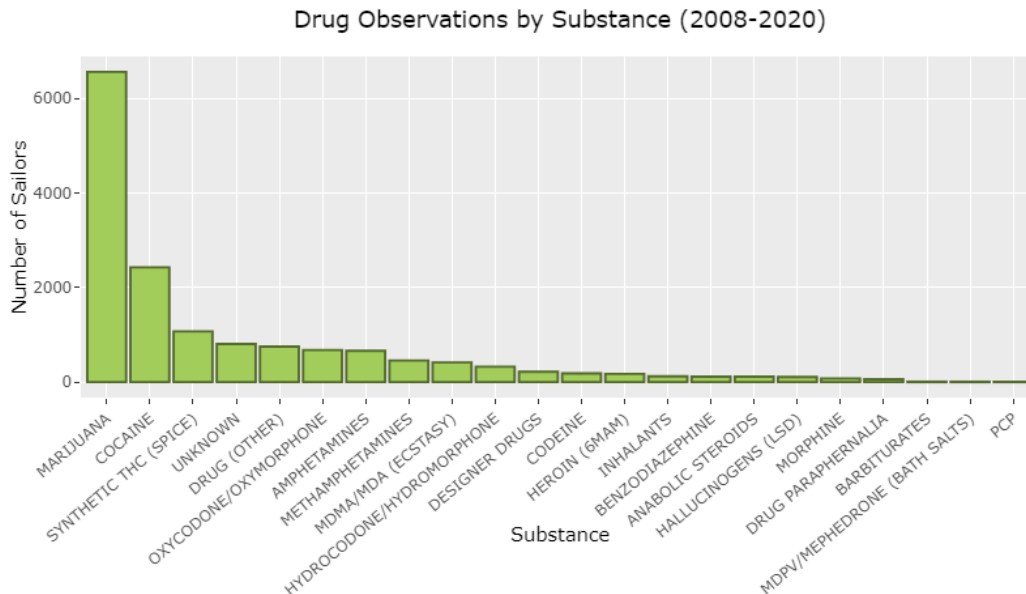


Figure 4.8. Abused substances other than alcohol by USN service members from 2008 to 2020. The substances depicted were termed “other-drugs” by this study. It is evident that marijuana was the most used substance within the other-drug category, followed by cocaine and spice. Source: OPNAV (2021a).

We commence with the classification models that were constructed using the pre-system outage data. The ROCs and AUCs for the classification models are depicted within Figure 4.9. The performance was fair for both the random forest and logistic regression models in their ability to accurately predict a future repeat other-drug incident. As observed, the random forest and logistic regression models had AUC performance measures of 0.77 and 0.71. The kNN model performed poorly yielding a 0.67 AUC.

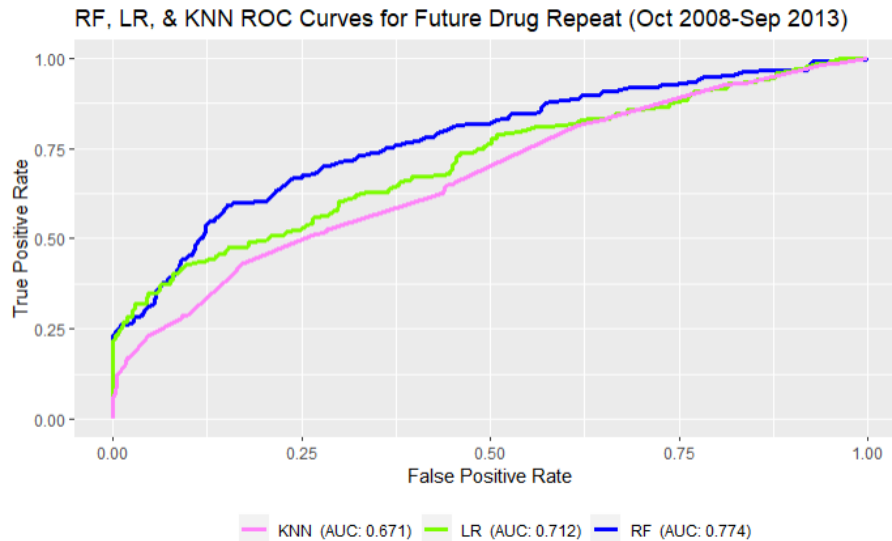


Figure 4.9. ROC curves for all classification models that were built to predict a future repeat other-drug incident using pre-system outage data. As ideal ROC curves draw near to the top-left corner, it is observed that the random forest model outperforms both the logistic regression and kNN models. Source: OPNAV (2021a).

Once again, the random forest model was the best model for predicting future substance abuse based on AUC performance measures. As before, we visualize the random forest model's features in order of importance in Figure 4.10. Once again, IsRepeat crowns the chart. We also observe that akin with the alcohol random forest classification models, GenRate, Dept, and Region are deemed to be of high importance. A feature that wasn't used in the alcohol studies, SubstanceName, also is of high importance. This feature contains the name of the substance that was abused as indicated in Table 3.1, hence the exclusion for alcohol abuse models.

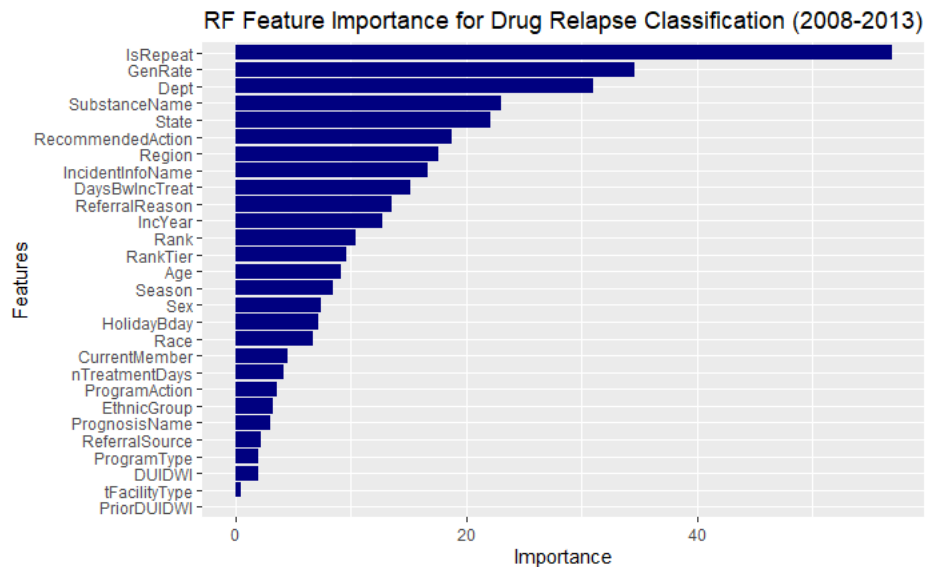


Figure 4.10. Random forest model features of importance using pre-system outage other-drug data. The features are ordered in a top-down fashion of measured importance with respect to their usefulness to the random forest model in correctly classifying the response—future repeat other-drug abuse. Source: OPNAV (2021a).

Future Repeat Other-Drug Abuse Post-System Outage

As in the pre- and post-alcohol comparison, there are some noticeable similarities and distinctions between the other-drug models. In contrast to the alcohol models, all of the post-system outage classification models performed better than the pre-system outage models. As observed in Figure 4.11 the random forest, logistic regression, and kNN models returned AUCs of ≈ 0.78 , 0.72 , and 0.69 respectively. In terms of best performing models, we found the random forest algorithm to be the best for all of the alcohol or other-drug partitioned data sets, followed by logistic regression and then kNN.

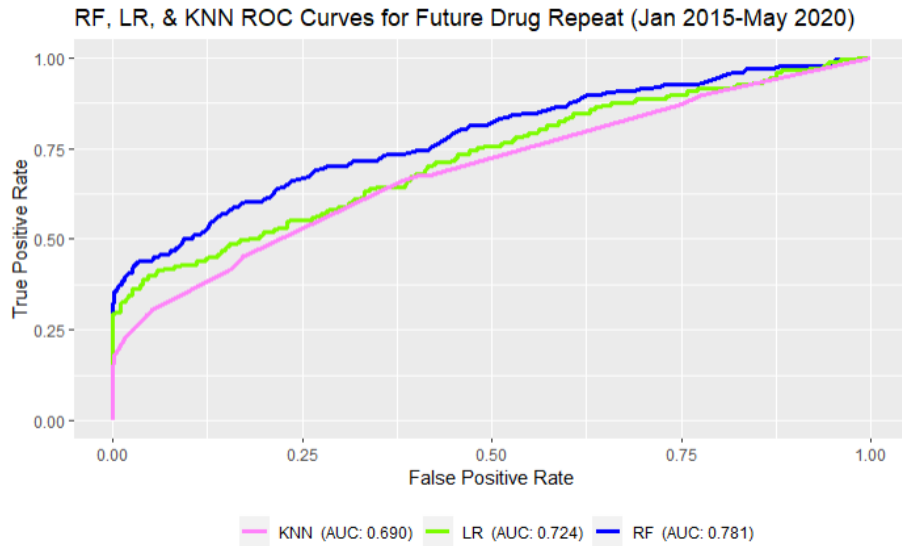


Figure 4.11. ROC curves for all classification models that were built to predict a future repeat other-drug incident using post-system outage data. As ideal ROC curves draw near to the top-left corner, it is observed that the random forest model outperforms both the logistic regression and kNN models. Source: OPNAV (2021a).

The random forest model's features of importance for the post-system outage other-drug classification model are depicted in Figure 4.12. As a good indication of features that are important in predicting future repeat other-drug abuse, the top 3 features for the random algorithm remained unchanged: IsRepeat, GenRate, and Dept. The State, Region, and SubstanceName features have shifted slightly in importance, yet remain relatively high. We also observe that Rank, Age, and HolidayBday jumped up in precedence as compared with the pre-system outage model. Alternatively, we notice that the features IncidentInfoName, RecommendedAction, and DaysBwIncTreat noticeably declined in importance.

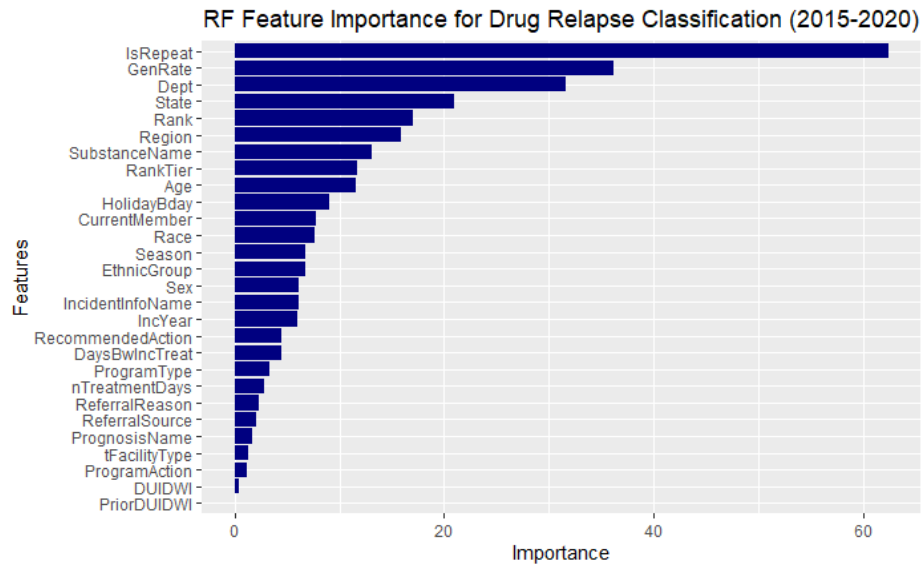


Figure 4.12. Random forest model features of importance using post-system outage other-drug data. The features are ordered in a top-down fashion of measured importance with respect to their usefulness to the random forest model in correctly classifying the response—future repeat alcohol abuse. Source: OPNAV (2021a).

As we were assessing the ordering of the features with respect to their measured importance across all of the random forest models, it was ascertained that many of the SARP features diminished in their relative importance after the system-outage period. One possible explanation for this was discovered after generating plots of the total number of non-NA values prior to and after the system outage. As observed in Figures 4.13 and 4.14, there is a substantial reduction in the number of observations that have observed values within SARP related features. Furthermore, this cannot be attributed to differing lengths in the observed periods, as the post-system outage period actually has 181 more days than the pre-system outage period.

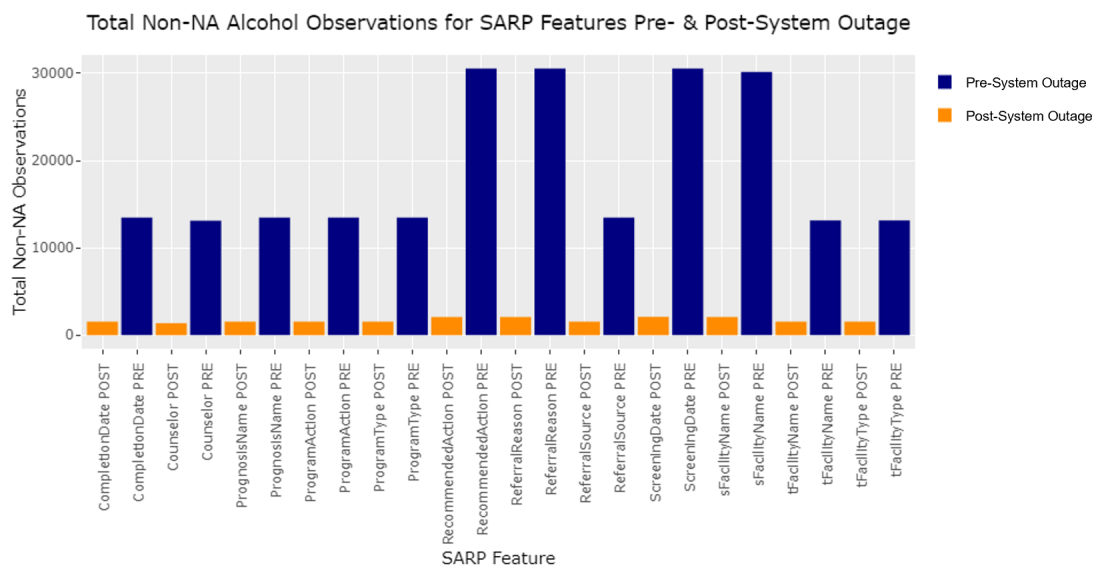


Figure 4.13. Total Non-NA SARP features for pre- and post-system outage alcohol observations. Source: OPNAV (2021a).

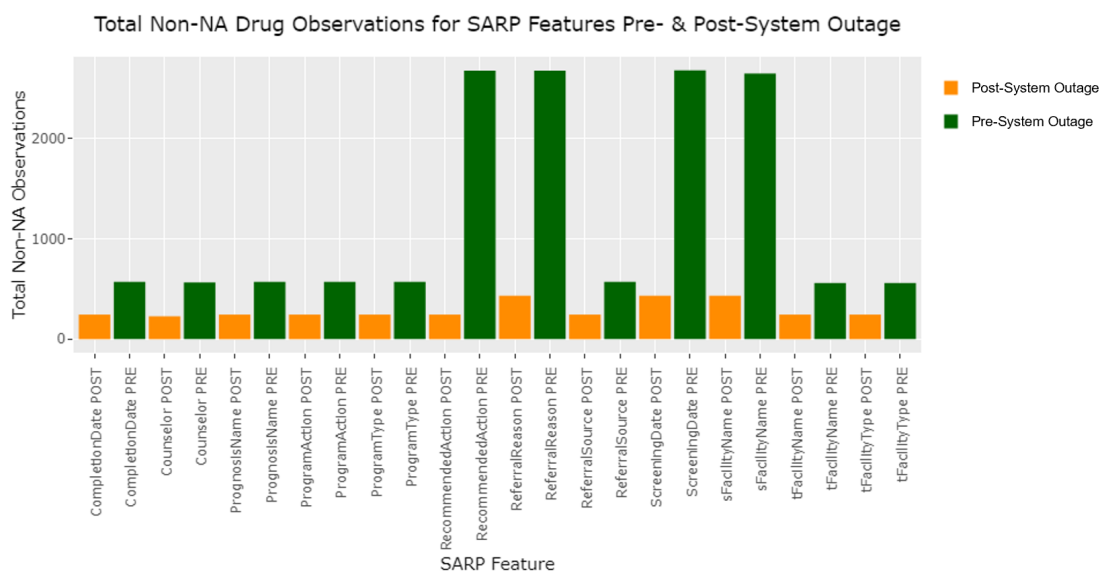


Figure 4.14. Total Non-NA SARP features for pre- and post-system outage other-drug observations. Source: OPNAV (2021a).

4.1.3 Seasonal Decomposition and Time Series Forecasting

Alcohol Substance Abuse Seasonal Decomposition and Forecasting

Alcohol DARs Seasonal Decomposition

As illustrated in Figure 4.15, the STL chart shows the combined pre- and post-system outage data as well as the trend, seasonal, and remainder components. The data component is merely the total number of alcohol substance abuse incidents by USN service members for the merged time periods October 2008 to August 2013 and September 2014 to February 2020. As reflected by the trend, we notice a gradual downward trend from 2008 to 2011 that shifts to a lesser upward trend that plateaus late-2012. The stark downward trend starting in 2013 is a result of the system outage period beginning to commence. As observed, the trend subsides and does not return to normal pre-system outage levels. We observe a slight upward trend in the data from early 2014 into 2017 that plateaus and trends lightly downward for the remainder of the observational period.

We also observe the seasonality of the data through the fluctuating patterns within the seasonal component of Figure 4.15. The seasonality for the annual periods suggests that the alcohol time series data decreases slightly for the start of the year, and rises sharply after about 1 month. The peak in the data slowly subsides over the next 3–4 months returning to roughly pre-peak levels half-way through the year. There is a noticeable decline near month 7 that quickly peaks around month 9. Beyond the last peak, the seasonal component picks up on an acute reduction in observations for the later part of the year.

As depicted in the data component of Figure 4.15, substance abuse occurrences experienced a sharp decline near the end of 2013 as a result of the system outage and have yet to bounce back to pre-system outage levels. We suspect that either the ADMITS system and/or previous processes for DAR entries have yet to be fully restored across all USN sites. This notion seems to be much more credible than assuming that a drastic reduction in alcohol substance abuse occurrences has occurred throughout the USN. The number of alcohol substance abuse incidents observed in the 2016 to 2020 time-frame would be more credible if there was at least a recovery period following the ADMITS system outage. However, after the system outage period the number of monthly incidents only rose to about 50% of previously observed levels and stabilized.

Decomposition of Alcohol Incidents (Oct 2008-Aug 2013 and Sept 2014-Feb 2020)

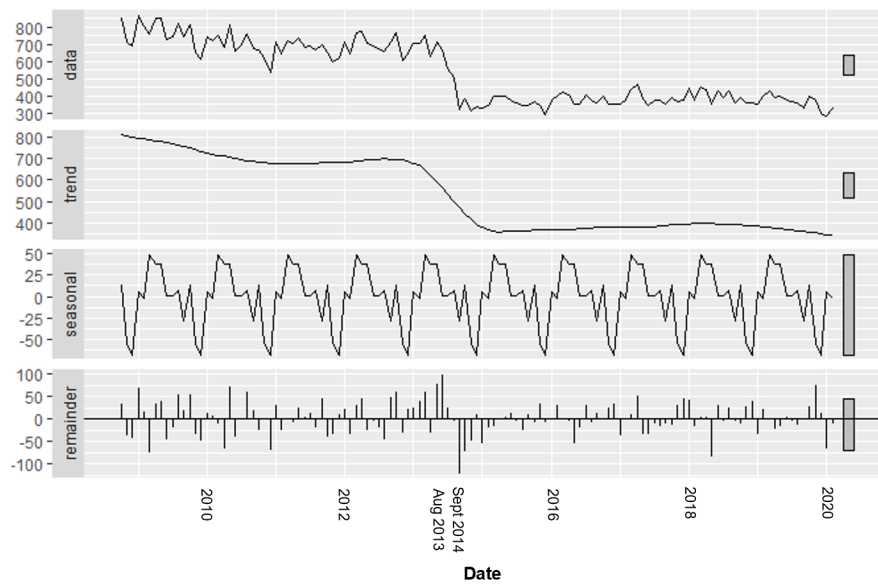


Figure 4.15. Seasonal decomposition for alcohol time series data reflecting the data, trend, seasonal, and remainder components. The time series excludes observations during the ADMITS system-outage period. The sharp decline in the trend is a by-product of the system outage and indicates that observations have not returned to pre-system outage levels. The seasonal component reflects the seasonal peaks and valleys identified by the STL over a 12-month period. The four gray boxes at the right of each component are of the same weight and can be used for scale. Source: OPNAV (2021a).

Alcohol DARs Forecasting Models

The 1-step and 12-step MAPE and MASE values, as well as the residual Ljung-Box test p-value for each forecasting model are provided in Figure 4.16. As discussed in Section 2.5.2, MAPE is a measurement for forecasting error of a model, such that lower values are better. MASE values are a reflection of how well a model performed in comparison to the naïve model. MASE values less than one are preferred, as this indicates that the forecasting model outperformed the naïve model. Furthermore, since the residual Ljung-Box test determines whether residuals are representative of white-noise (i.e. null hypothesis) or correlated (i.e. alternative hypothesis), we desire p-values to be greater than 0.05 (Hyndman and Athanasopoulos 2018).

Based on the observed measures of performance, it is clear that in our study the SARIMA model is the best model for forecasting alcohol abuse incidents among USN service members. As depicted in Figure 4.16, while all of the models outperformed the naïve model with exception to the HW-M, the SARIMA model had the best performance as measured by the 12-step MASE. The MASE value for the SARIMA model was 0.65, indicating that the SARIMA model was ≈ 1.5 times better in performance than the naïve model. Furthermore, even though the SARIMA model did not have the lowest 1-step MASE value, we place greater importance on the 12-step MASE as this is a gauge for a 12-month as opposed to 1-month forecast. Moreover, the SARIMA model was also the only model to pass the residual Ljung-Box test, such that its residuals were indistinguishable from white-noise.

MAPE & MASE Comparison for Alcohol Forecasting Models (Rolling Horizon)

Model	1-Step MAPE	12-Step MAPE	1-Step MASE	12-Step MASE	Residual Ljung-Box Test P-Value
Naive	0.099	0.100	1.00	1.00	0.0021
Seasonal Naive	0.097	0.093	0.94	0.92	0.0000
SARIMA	0.084	0.066	0.82	0.65	0.5300
Holt-Winters	0.077	0.092	0.76	0.94	0.0170
Holt-Winters (A)	0.092	0.096	0.90	0.96	0.0038
Holt-Winters (M)	0.091	0.110	0.88	1.10	0.0000
ETS	0.077	0.076	0.74	0.73	0.0400

Figure 4.16. Performance measures for seven alcohol substance abuse forecasting models. The SARIMA model was determined to be the optimal forecasting model based on its low MAPE, best-of-class 12-step MASE, and it was the only model to pass the residual Ljung-Box test. Source: OPNAV (2021a).

The best performing SARIMA model identified by the `auto.arima()` function was the $ARIMA(1, 1, 0)(1, 0, 1)_{12}$ model. The predicted 12-month forecast for the ARIMA model is provided within Figure 4.17. The chart displays the total number of observations per month of the time series. Furthermore, the chart reflects the fit of the model to the test data, as well as the March 2020 to February 2021 forecast with an 80% confidence interval.

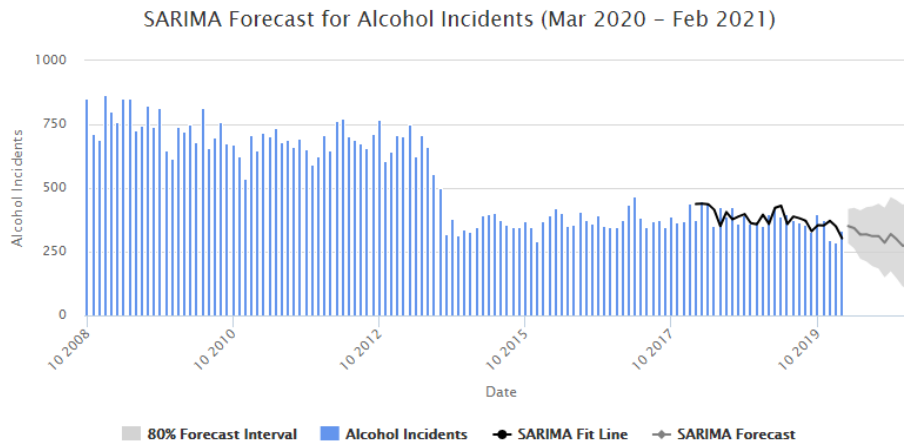


Figure 4.17. The total monthly alcohol DAR entries are reflected by the blue bars in the chart. The time series spans from October 2008 to February 2020 excluding data during the ADMITS system-outage period from September 2013 to August 2014. The chart reflects the fit of the model to the test data, as well as a 12-month forecast from March 2020 to February 2021 with an 80% confidence interval. Source: OPNAV (2021a).

Aside from the Ljung-Box test, we can also take a closer look at the residuals for the model using the `checkresiduals()` function from the `forecast` package (Hyndman et al. 2022). The residuals from the $ARIMA(1, 1, 0)(1, 0, 1)_{12}$ model are displayed in Figure 4.18. As discussed by (Hyndman and Athanasopoulos 2018), it is beneficial to observe the residuals to assess the model's ability to capture the appropriate structure from the time series. Furthermore, a good forecasting model will have residuals that demonstrate the following properties: uncorrelated, mean zero, constant variance, and normally distributed (Hyndman and Athanasopoulos 2018). From Figure 4.18 we observe that the residuals are close to mean zero though slightly shifted rightward. The ACF plot demonstrates that the residuals do not appear to be auto-correlated and the spikes are within the significance limits denoted by the blue-dashed lines.

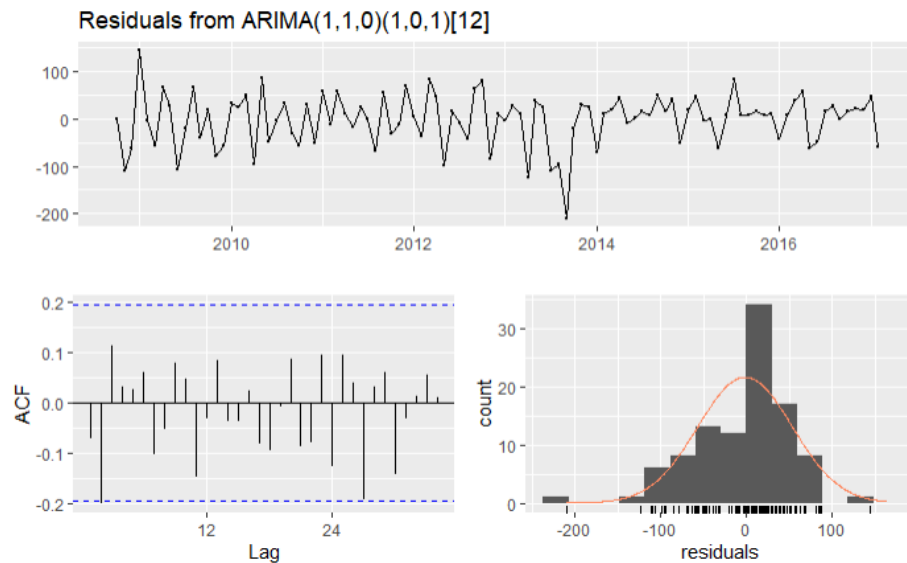


Figure 4.18. The SARIMA model residuals for the alcohol DAR forecast have a slight rightward shift, but are near mean zero. The ACF plot demonstrates that the residuals are uncorrelated and any spikes are within significance limits. Source: OPNAV (2021a).

Other-Drug Substance Abuse Seasonal Decomposition and Forecasting

Other-Drug DARs Seasonal Decomposition

The STL provided in Figure 4.19 displays a trend line with many ebbs and flows throughout the observational period. The trend line reflects a steady upward trend from 2008 that peaks near mid-2011. The trend line has a downward trend that nearly mirrors the initial upward trend; however, the ADMITS system outage period is observed as the trend line plummets to the baseline from 2013 through 2015. The recovery in observations for the other-drug DARs is noticeably better than the previously observed alcohol trend line. Beyond 2015, the other-drug trend reflects a steady rise that appears to stabilize near 120 observations per month.

The seasonal component of the STL in Figure 4.19 captured several seasonal peaks and valleys for roughly the first half of the year. As observed, the year starts at a peak that quickly subsides during the first month. Thereafter, the seasonal component suggests that there is a steady increase in observations for a few months that is met by a swift decline near

mid-year. Subsequently, a lower in magnitude rise in the observations for about a month is followed by a steep downward shift for the remainder of the year.

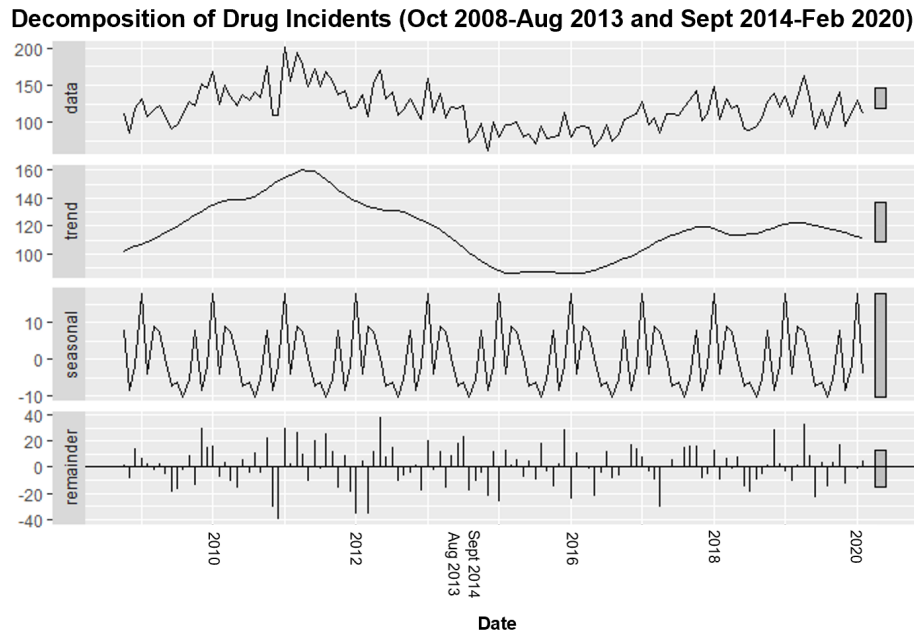


Figure 4.19. Seasonal decomposition for other-drug time series data reflecting the data, trend, seasonal, and remainder components. The time series excludes observations during the ADMITS system-outage period. The trend flat lines from 2014 to 2015 as a result of the system outage. Thereafter, observations are seen to be nearly restored to pre-outage levels. The seasonal component reflects the seasonal peaks and valleys identified by the STL over a 12-month period. The four gray boxes at the right of each component are of the same weight and can be used for scale. Source: OPNAV (2021a).

Other-Drug DARs Forecasting Models

As observed in Figure 4.20, all of the forecasting models outperformed the naïve model as assessed by their 1- and 12-step MASE values. The SARIMA, HW, and ETS models stand-out as their residuals passed the Ljung-Box test. Of the three models with uncorrelated residuals, once again the SARIMA model is preferred. The SARIMA models yielded the best 12-step MASE with an even better 1-step MASE value coupled with low error rates as determined by the 1- and 12-step MAPE values.

MAPE & MASE Comparison for Drug Forecasting Models (Rolling Horizon)

Model	1-Step MAPE	12-Step MAPE	1-Step MASE	12-Step MASE	Residual Ljung-Box Test P-Value
Naive	0.20	0.19	1.00	1.00	0.0034
Seasonal Naive	0.15	0.14	0.75	0.74	0.0000
SARIMA	0.13	0.14	0.67	0.73	0.7100
Holt-Winters	0.13	0.16	0.65	0.82	0.8400
Holt-Winters (A)	0.12	0.14	0.57	0.74	0.0057
Holt-Winters (M)	0.12	0.14	0.61	0.74	0.0006
ETS	0.15	0.16	0.74	0.81	0.8400

Figure 4.20. Performance measures for seven other-drug substance abuse forecasting models. The SARIMA model was determined to be the optimal forecasting model as it had low forecasting error and the best 12-step MASE among models passing the Ljung-Box test at $p > 0.05$. Source: OPNAV (2021a).

The monthly other-drug DAR observations are reflected within Figure 4.21. As with the alcohol time series, the observations from September 2013 to August 2014 were removed to mitigate the impact of the ADMITS system outage. Figure 4.21 also displays the SARIMA models fit to the test data and resulting 12-month forecast. The SARIMA model that was selected by the `auto.arima()` function was an $ARIMA(1, 0, 1)(2, 1, 0)_{12}$ model. The residuals for the models are provided in Figure 4.22 which reflect good normality in the residuals near mean 0. From the ACF plot we observe that there were two spikes in the residuals suggestive of auto-correlation that was not picked up by the SARIMA model; however, one or two spikes over 36 time periods is somewhat expected. Additionally, the majority of the residuals do appear to be white-noise and the model still passed the Ljung-Box test.

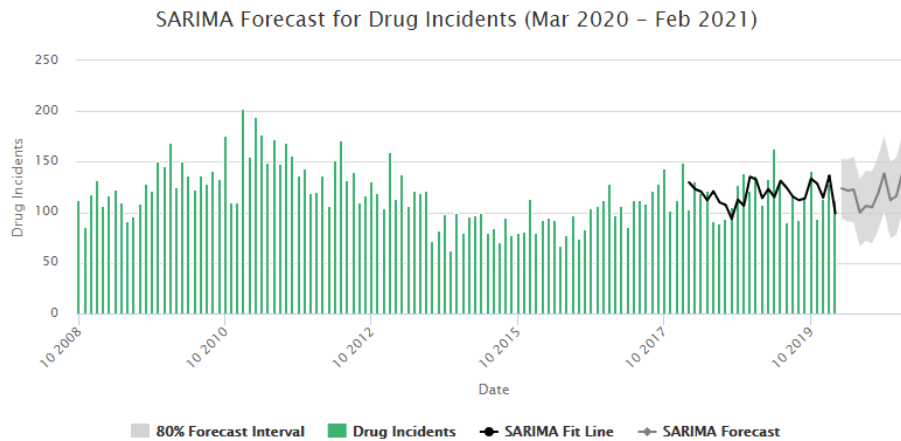


Figure 4.21. The total monthly other-drug DAR entries are reflected by the blue bars in the chart. The time series spans from October 2008 to February 2020 excluding data during the ADMITS system-outage period from September 2013 to August 2014. The chart reflects the fit of the model to the test data, as well as a 12-month forecast from March 2020 to February 2021 with an 80% confidence interval. Source: OPNAV (2021a).

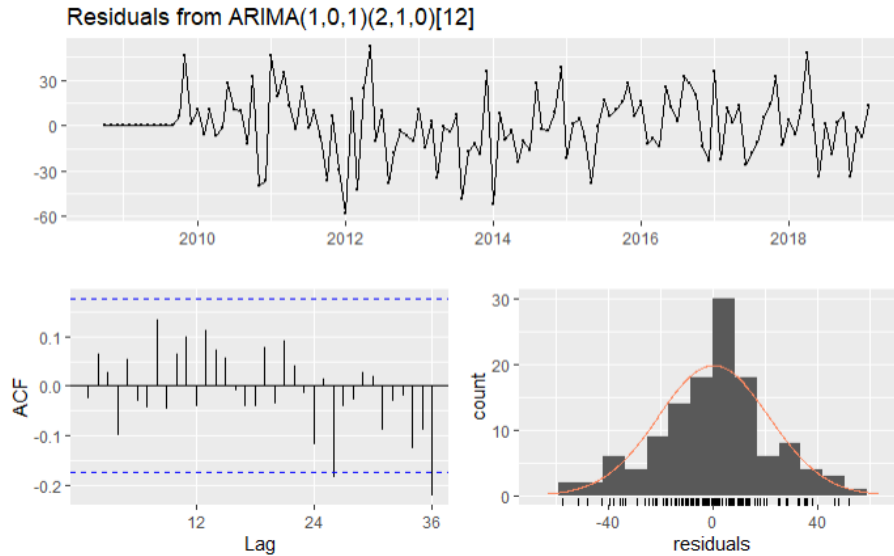


Figure 4.22. The SARIMA model residuals for the other-drug DAR forecast seem approximately normally distributed near mean zero. The ACF plot demonstrates that the residuals are primarily uncorrelated, with 2 of 35 five spikes exceeding significance limits. Source: OPNAV (2021a).

4.1.4 Survival Analysis

During our exploratory data analysis phase, we wanted to assess the total number of days between substance abuse incidents compared to the total number of days between an incident and SARP screening date for enrollment into a treatment program. Essentially, we wanted to compare the two distributions to see if there could potentially be a large number of sailors that are relapsing before they are entered into a treatment program. For instance, if a USN service member had an alcohol substance abuse incident on February 1st, 2020, was screened for enrollment into a treatment program on March 30th, 2020, but had an incident on February 21st, 2020, then that member relapsed before they were seen for treatment of a prior incident.

For alcohol incidents, as depicted in Figure 4.23, we built a bar chart reflecting the total number of days between alcohol incidents along the x-axis with the total number of sailors indicated by the y-axis. We then built another bar chart indicating the total number of days between an alcohol incident and the SARP screening date for that specific incident, which is represented by the lower bar chart within Figure 4.23. To compare the two distributions for the number of days between repeat incidents and the number of days between an incident and enrollment into a treatment program, the two bar charts were stacked on top of each other with the same number of days along the x-axis. The same charts were also produced for substances other than alcohol as observed in Figure 4.24.

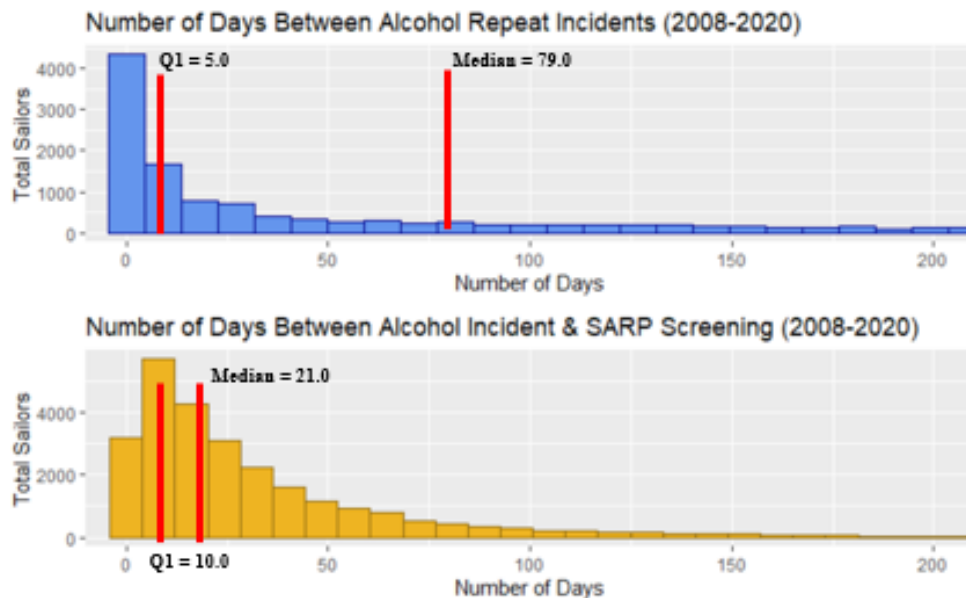


Figure 4.23. The upper-chart reflects the number of days between repeat alcohol incidents and the frequency of USN personnel. The lower-chart indicates the number of days between an alcohol incident and SARP screening when applicable. Source: OPNAV (2021a).

For the alcohol incidents observed in Figure 4.23, we computed the median number of days between repeat incidents to be 79 days and the 1st quantile to be 5 days. Similarly, we computed the median number of days between an incident and screening for treatment to be 21 days and the 1st quantile to be 10 days. When we contrast the median number of days between repeat incidents and the median number of days between an incident and screening for treatment, it does appear that many service members are repeating well beyond observed treatment timelines. However, since 25% of USN service members with repeat incidents relapse with 5 days compared to 25% of screenings for treatment occurring within 10 days, there does appear to be a large number of Sailors that potentially relapse before being seen for treatment. In fact, when we only considered Sailors ($n = 8,408$) that were enrolled into SARP treatment programs for an incident with a subsequent relapse, we discovered that 45% of them relapsed prior to being screened for the prior incident. Therefore, while the majority of USN service members who repeat, do so well beyond normal SARP screening timelines, there is a subset of the population that repeats quickly, prior to normal SARP screening timelines.

The other-drug incidents reflected within Figure 4.24 depict a similar story. We observe that the median number of days between repeat incidents and the median number of days between an incident and treatment for that specific incident are quite close, at 25 and 24 days respectively. Furthermore, the 1st quantiles are also close in proximity for the number of days between repeat incidents and treatment timelines for an incident at 7 and 10 days respectively. Nevertheless, it appears that USN personnel who abuse other-drug substances have repeat timelines that are shorter in duration than those involved with alcohol abuse. In fact, when we consider only those USN personnel who were enrolled into a SARP program and had a subsequent relapse, we found that 73% of those service members repeated prior to being screened for the prior incident.

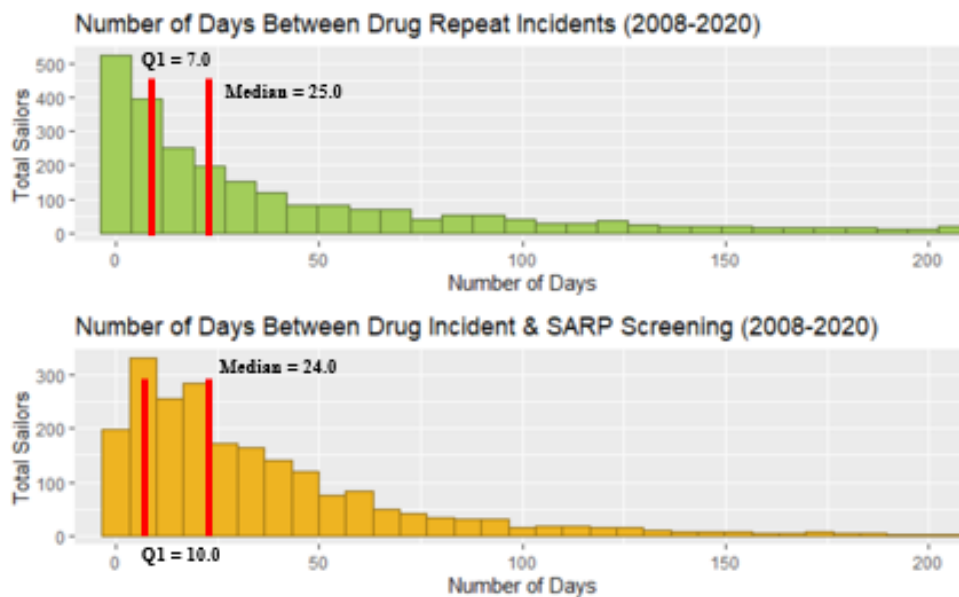


Figure 4.24. The upper-chart reflects the number of days between repeat other-drug incidents and the frequency of USN personnel. The lower-chart indicates the number of days between an other-drug incident and SARP screening when applicable. Source: OPNAV (2021a).

Repeat Alcohol Abuse Survival Analysis

The observations within Figures 4.23 and 4.24 led us into our survival analysis, in which we examined the survival of USN service members – that is, the time between repeat

incidents. The results of our KM, CPHR, and RSF survival analysis models examining the time between repeat alcohol substance abuse incidents are displayed within Figures 4.25, 4.26, and 4.27. The concordance values, defined within Section 2.5.3, for the CPHR and RSF models were 0.70 and 0.68 respectively.

An initial look at the KM, CPHR, and RSF survival curves in Figures 4.25, 4.26, and 4.27 suggests that they more-or-less echo one another, with the KM and average line for the RSF being nearly equivalent and the CPHR survival curve being slightly shifted upward. As displayed, each survival curve spans 120 months, with each time period being of length 1 month. The survival probability of the population ranges from 0.00 to 1.00 along the y-axis and the months are indexed sequentially by time from 0 to 120 months. Since the months along the x-axis relate to survival time, the survival curve will drop when events, or relapses, occur with the same duration of survival time. The KM and RSF survival curves estimate that survival at the end of the observational period is about 75% for USN personnel with at least one alcohol substance abuse incident. The survival probability is estimated to be about 86% based on the CPHR model. Thus, the majority of the USN population that experiences an alcohol abuse incident that is submitted into ADMITS as a DAR entry does not have a repeat incident.

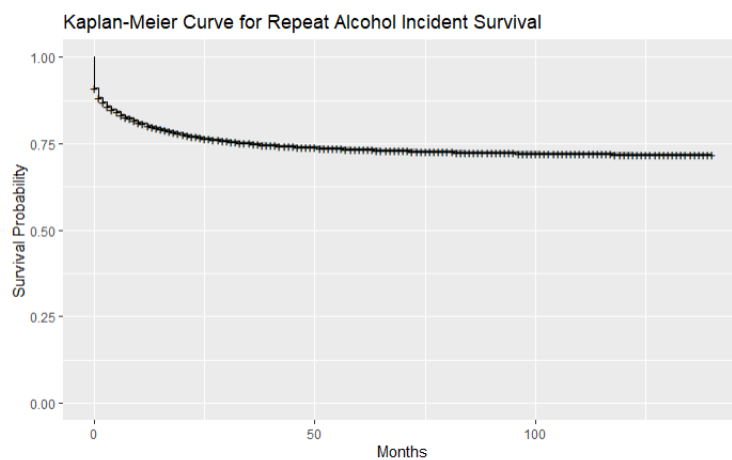


Figure 4.25. KM survival curve for USN personnel with at least one alcohol DAR. The vertical bars along the survival curve indicate that some observations were right-censored at the period of time and removed from the population considered to be at risk. Source: OPNAV (2021a).

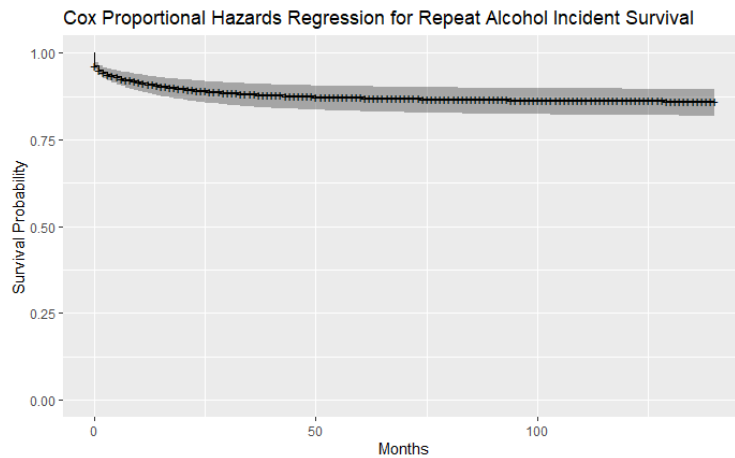


Figure 4.26. CPHR survival curve for USN personnel with at least one alcohol DAR. Source: OPNAV (2021a).

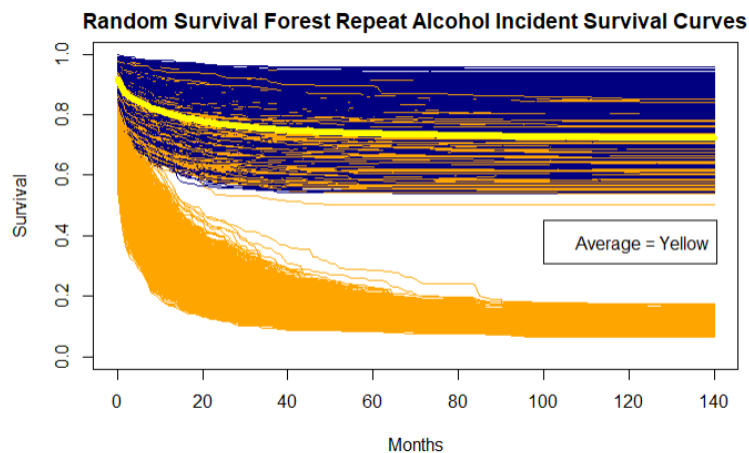


Figure 4.27. The `ranger()` function was used to estimate a survival curve for each USN service member with at least one alcohol abuse incident. A sample of 20,000 survival curves are depicted within the plot. To see if the estimated survival curves were lower for those who relapsed compared to those who did not, we chose to color code the estimated survival curves. For USN service member who did relapse, their estimated survival curves are colored orange, whereas the estimated survival curves are colored navy for those who did not relapse. We observe that the estimated survival curves for those who relapsed are generally lower than those who did not relapse. The mean of the survival curves is displayed by the yellow line, which is useful for comparison to the KM and CPHR models. Source: Wright and Ziegler (2017).

While we anticipated the survival to be high with respect to USN service members not relapsing, our primary interest was in the survival time for those that do. Figure 4.28 displays the same KM survival curve as in Figure 4.25, but over a shorter time period for clarity. The KM survival curve indicates that the survival probability experienced an $\approx 10\%$ drop in month 0, which includes relapses between 0 and 30.4167 days, that subsides to nearly an 87.5% survival at month 1. The survival analysis for alcohol events demonstrates that within the population of sailors with one alcohol DAR observation, there exists a subset of the population that will repeat fast and prior to observed SARP screening timelines.

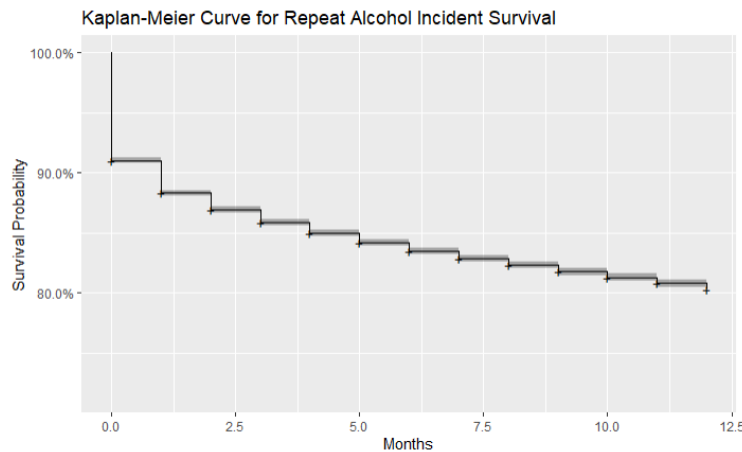


Figure 4.28. 12-month KM survival curve for USN personnel with at least one alcohol DAR. Source: OPNAV (2021a).

Repeat Other-Drug Abuse Survival Analysis

The KM, CPHR, and RSF survival curves for repeat other-drug substance abuse incidents are displayed within Figures 4.25, 4.26, and 4.27. The concordance values for the CPHR and RSF models were 0.65, and 0.64 respectively. As with the survival analysis results for repeat alcohol incidents, the other-drug RSF models produced survival curves that are quite similar as displayed in Figures 4.29, 4.30, and 4.31. The KM and mean survival curve from the RSF model are nearly identical reflecting a survival probability that is estimated to be about 80% after a short period of time. The CPHR model is slightly shifted downward with an estimated survival probability terminating at about 75%. As observed, relapses for other-drug substance abuse incidents does not apply for the majority of the USN population.

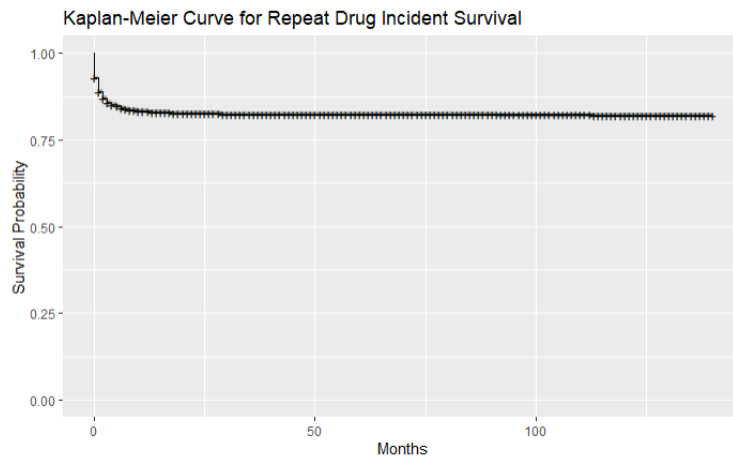


Figure 4.29. KM survival curve for USN personnel with at least one other-drug DAR. Source: OPNAV (2021a).

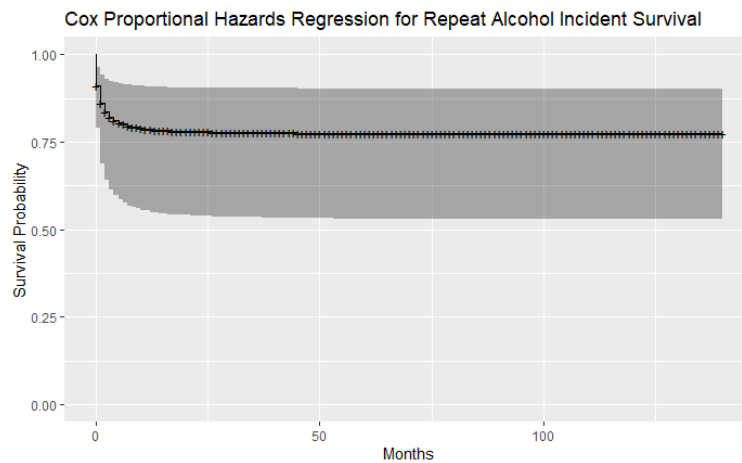


Figure 4.30. CPHR survival curve for USN personnel with at least one other-drug DAR. Source: OPNAV (2021a).

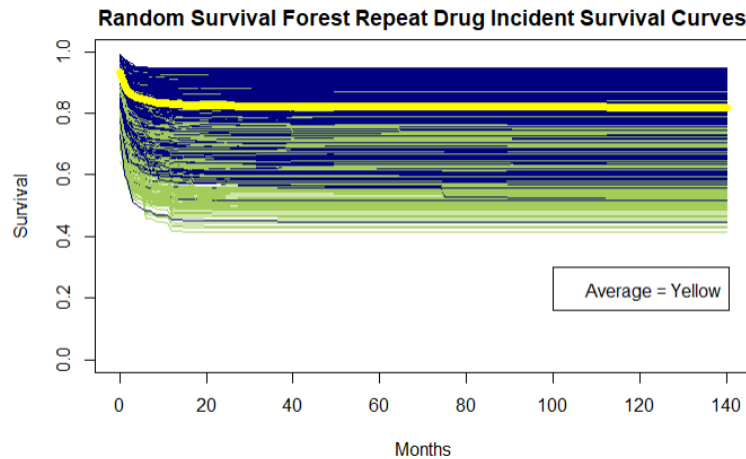


Figure 4.31. The `ranger()` function was used to estimate a survival curve for each USN service member with at least one non-alcohol substance abuse incident. A sample of 15,000 survival curves are depicted within the plot. To see if the estimated survival curves were lower for those who relapsed compared to those who did not, we chose to color code the estimated survival curves. For USN service member who did relapse, their estimated survival curves are colored green, whereas the estimated survival curves are colored navy for those who did not relapse. We observe that the estimated survival curves for those who relapsed are generally lower than those who did not relapse. The mean of the survival curves is displayed by the yellow line, which is useful for comparison to the KM and CPHR models. Source: Wright and Ziegler (2017).

As with the alcohol survival analysis, we narrow our focus by tuning in to the first 12-months of survival time as seen in Figure 4.32. Figure 4.32 is the same survival curve displayed in Figure 4.29, but over a shorter time period for clarity. The KM survival curve indicates that the survival probability experienced an $\approx 7.5\%$ drop in month 0 to $\approx 92.5\%$. Again, month 0 includes all USN service members observed to relapse within 0 to 30.4167 days of survival time. Subsequent reductions in estimated survival probability are observed for months 1 and 2, falling to $\approx 88.5\%$ and $\approx 87.5\%$ respectively. By contrasting the KM survival curves for alcohol and other-drugs in Figures 4.28 and 4.32 we notice that by month 2, the survival probability for both subgroups is $\approx 87.5\%$. As with alcohol, the survival analysis for non-alcohol relapse events demonstrates that within the population of sailors with one non-alcohol substance abuse event, there exists a subset of the population that will relapse

quickly, potentially before observed SARP screening timelines. If treatment timelines are unable to be shifted to the left due to administrative processes or medical protocols, then it may be beneficial for the USN to provide other means of support to service members before they are seen for treatment.

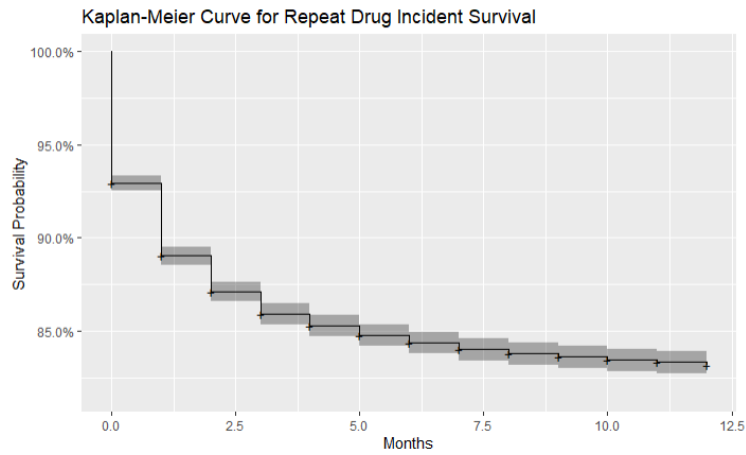


Figure 4.32. 12-month KM survival curve for USN personnel with at least one other-drug DAR. Source: OPNAV (2021a).

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 5:

Conclusion

This final chapter starts with a recap of the intent of our research and summary of the models employed to address our research objectives. We also highlight the study's key findings for each analysis. Section 5.2 concludes our report with some specific opportunities for future work and advancement.

5.1 Summary

The point of this thesis is to understand substance abuse by USN service members, in order to empower those who can effect change or take this research further. Results of our analysis may unveil new perspectives and opportunities to support USN service members battling with substance abuse issues before they become problematic.

To advance our understanding of substance abuse by service members within the USN, we availed ourselves of classification, time series, and survival analysis models. Furthermore, we bolstered our analysis by delving into USN duty locations, USN rank, and USN general rating subgroups by measuring their proportion of substance abuse occurrences over time. The models used were selected to address our proposed research questions:

1. Are there subgroups within USN ranks, USN general ratings, or working within U.S. states that are over- or under-represented with respect to their percentage of the USN force and their percentage contribution of total alcohol substance abuse occurrences?
2. Through the application of statistical learning models, by partitioning the ADMITS data set into pre- and post- ADMITS system outage data and separately analyzing alcohol and other-drug related substance abuse:
 - (a) With what degree of accuracy can we correctly classify relapse occurrences among Sailors?
 - (b) What risk-factors are deemed important by classification models for predicting substance abuse relapse incidents?
3. With what degree of accuracy can we forecast alcohol or other-drug substance abuse occurrences?

4. What is the survival probability over time of USN service members with respect to repeat substance abuse events and the number of days between incidents?

Over- or Under-Representation

The over- or under-representation analysis of U.S. states, USN ranks, and USN general ratings enabled us to examine segments of these subgroups using a proportional measure. On the surface, assessing total occurrences by segments of a population may misrepresent large or small groups as more or less occurrences can be anticipated. Therefore, we chose to assess the proportional size of a group (e.g. AD USN members residing in California) with respect to the larger population (e.g. AD USN members residing in the U.S.) against their proportional contribution of alcohol substance abuse occurrences. In principal, this removed the effect of the group size and permitted us to display whether that segment was proportionally over- or under-represented with respect to alcohol substance abuse occurrences. The results are found in Appendix sections A.1, A.2, and A.3.

As the over- or under-representation analysis range from 2008 to 2020 for U.S. state and USN rank subgroups, and from 2015 to 2020 for USN general rating subgroups, our summary concentrates on more current results. With respect to subgroups that are currently over- or under-represented we identified the following subgroups:

- **Over-Represented Subgroups:** Proportionally more alcohol substance abuse occurrences than their proportional make-up of the population concerned.
 - AD USN military members stationed in Louisiana
 - Reserve USN military members stationed in North Carolina
 - AD USN ranks E-1, E-2, E-3, and O-2
 - AD USN warrant officer ranks.
 - * AD USN warrant officer make up a small percentage of the officer force. As such, alcohol substance abuse occurrences within warrant officer ranks results in high proportions, especially in months with low alcohol abuse incidents among officers.
 - Reserve USN ranks E-1, E-2, E-3, E-4, E-8, O-3, and O-4
 - AD USN general ratings CE, EA, EO, HM, MN, MR, RS, SW, and UT
- **Under-Represented Subgroups:** Proportionally less alcohol substance abuse occurrences than their proportional make-up of the population concerned.

- AD USN military members stationed in North Carolina
- Reserve USN military members stationed in Texas
- AD USN ranks E-5, E-6, E-7, E-8, and E-9
- AD USN general ratings AW, CT, ET, IS, IT, MA, NC, PS, SO, and YN

Classification Models

The classification models performed fairly well in their ability to correctly classify future alcohol or other-drug substance abuse relapse occurrences both prior to and after the ADMITS system outage. The random forest models were determined to be the best for all of the components of our classification analysis compared to the logistic regression and kNN models as assessed by the AUC performance measure. We found that the random forest classification model for future alcohol incidents had an AUC measurement of $\approx 6\%$ less when using the post-ADMITS system outage data than when using pre-system outage data. However, the random forest model for future other-drug incident classification performed $\approx 1\%$ better when using the post-system outage data.

With respect to feature importance for the random forest models, we discovered the feature IsRepeat to be the leading risk-factor for future repeat alcohol and other-drug incidents. This suggests that if a member currently has repeat alcohol or other drug related incidents, then he or she is at a higher risk of relapsing. Other features of importance for predicting future alcohol relapse incidents included GenRate, Dept, IncYear, Rank, State, and Age. With respect to classifying future other drug relapse occurrences, GenRate, Dept, State, Region, SubstanceName, Rank, and Age were determined to be of importance.

Time-Series Analysis

Forecasting the total number of USN monthly alcohol or other-drug incidents was found to be best performed by the SARIMA model. The SARIMA model was one of seven models compared with respect to their 1- and 12-step MAPE and MASE performance measures accompanied with a residual assessment using the Ljung-Box test. With our interest in generating a 12-month forecast, the SARIMA model performed ≈ 1.5 times better than the naïve model when forecasting alcohol incidents and 1.3 times better than the naïve model when forecasting other-drug incidents. The STL charts for alcohol and other-drug incidents displayed annual seasonality with peaks throughout the first half of the year that drop below

normal observed levels in the later half of the year. The trend-lines are also suggestive of the ADMITS system-outage having a more substantial impact on alcohol DAR entries than other-drug DAR entries. While the system-outage period is clearly observed by the steep reduction in alcohol or other-drug substance abuse occurrences, the alcohol incidents stabilized near $\approx 50\%$ of pre-system outage levels compared to an $\approx 80\%$ recovery for other-drug observations. While it is true that fewer observations are desired, we believe the immediacy of the stark decline in observations is more likely to be a result of the ADMITS system outage than a dramatic reduction in alcohol substance abuse occurrences.

Survival Analysis

Our survival analysis looked at time-to-event, being the time between alcohol or other-drug substance abuse occurrences with the event being a repeat incident. The populations for our survival analysis consisted of USN service members with at least one alcohol or other-drug substance abuse occurrence. The alcohol and other-drug KM, CPHR, and RSF survival curves dropped and stabilized near observably high survival probability rates at $\approx 80\%$ for alcohol relapses and $\approx 77\%$ for other-drug relapses.

While the survival curves indicate that the majority of USN service members with at least one substance abuse incident do not relapse, there is more to the story near the onset of the survival curves. When we took a closer look at the KM survival curves within Figures 4.28 and 4.32, we observed abrupt reductions in the survival probabilities between 0 and 2 months of survival time and a steady decline to 12 months of survival time. As discussed within Section 4.1.4, month 0 contains members who relapsed within 0 to 30.4167 days following an initial incident, month 1 captures members who relapsed within 30.4167 and 60.8334 days, and so on. For alcohol relapses, within Figure 4.28 we notice that the survival probabilities for months 0, 1, and 12 are $\approx 91\%$, 87.5% , and 80% respectively. Therefore, even though we observe a relatively high survival probability long-term, there is a subset of the USN population under study that relapses quickly after an alcohol substance incident. For other drug relapses as observed within Figure 4.32, the survival probabilities were quite similar for months 0, 1, and 12 at $\approx 92.5\%$, 88.5% , and 83% respectively.

One concern with the data that requires validation, is that there were many identified relapses that occurred within the first 0 to 7 days. When assessing these observations it is difficult to discern whether there were multiple incidents that occurred within close proximity in time,

multiple incidents occurred on the same day, or if there were data entry errors (e.g. duplicate entries). If there were duplicate entries, then removing these erroneous occurrences would shift the survival curves upward resulting in a higher survival probability over time.

Overall, the survival analysis indicates that while the majority of the USN population considered does not relapse after one substance abuse occurrence, there is a subset of the population that relapses quickly. Furthermore, when we considered only USN service members who were screened by a SARP for a prior incident and had a subsequent repeat incident, we found that 45% of alcohol and 73% of other-drug repeat incidents occurred before the prior incident's SARP screening. It is possible that the repeat incident drove the SARP screening date to be entered for a prior occurrence; however, the data still suggests that members who repeat do so with little time between incidents. Therefore, it may be beneficial to assess standard SARP screening timelines and/or provide members with additional means of support immediately following an alcohol or other drug substance abuse incident.

5.2 Future Work

There are several notable opportunities to expand upon our analysis and advance the issue of substance abuse within the USN. The sponsor of this research, OPNAV N17, has made reference to additional features that may be able to be included within the ADMITS data set. Furthermore, it is possible that ADMITS and NES data from BUPERS-34 could be merged prior to de-identifying the data sets to capture additional features that may be of use to the supervised learning models. As discussed in Section 3.3.4, a limitation of the survival analysis models is that plausible separation or retirement dates were unknown within the ADMITS data. As such, the population considered to be at risk was very likely to be overstated. Therefore, if separation and retirements dates were known for the ADMITS observations, then the observations could be right-censored at the appropriate survival times.

Another feasible option to explore, perhaps with a multinomial classification algorithm, would be a predictive model for the outcome of treatment. The Prognosis feature contains four levels: completed program return to duty, completed program separate, program not completed return to duty, and program not completed separate. In short, a classification model could be built to best predict the prognosis using features of the data set. In like

manner, as there are many types of substances abused, a model could be built to classify the substance abused in conjunction with an assessment of risk factors.

Lastly, the current ADMITS data set spans from October 2008 to May 2020; therefore, it would be advisable to obtain more current data. Furthermore, a reasonable amount of time has now passed since the ADMITS system outage, so it may behoove those interested to restrict analysis to post-2014 data. Moreover, as substance abuse occurrences are also observed in the other branches of service, merging similar substance abuse data sets across the DOD to identify if risk-factors contributing to substance abuse are systemic or isolated within the DOD could be a worthy endeavor attracting a broader audience.

APPENDIX: State, Rank, and Rating Visuals

A.1 AD and Reserve Alcohol Abuse Over- or Under-Representation by State

A.1.1 AD USN Over- or Under-Representation by State

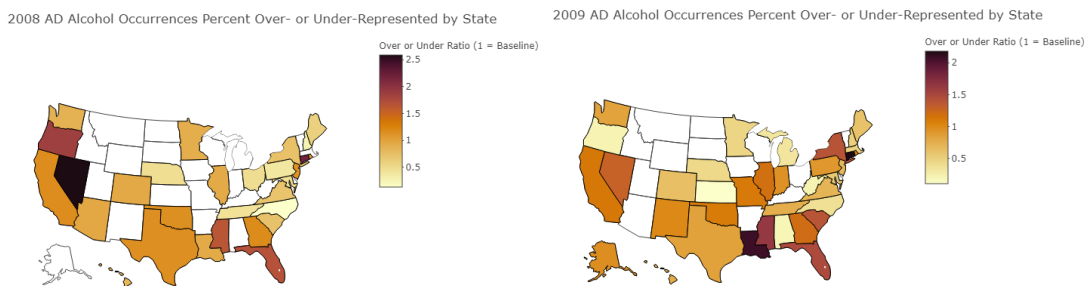


Figure A.1. AD USN 2008 and 2009 Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

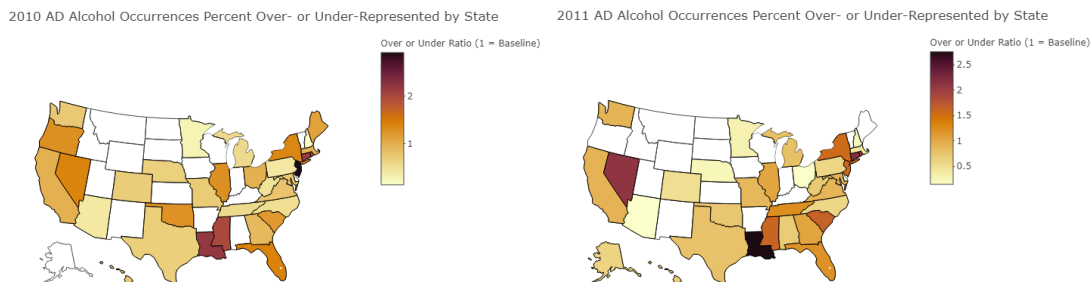
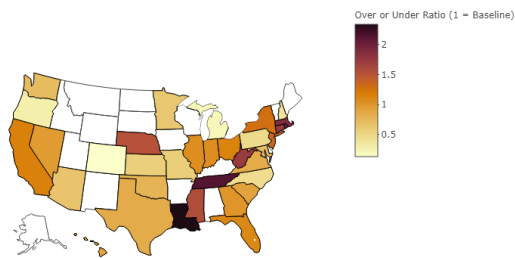


Figure A.2. AD USN 2010 and 2011 Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

2012 AD Alcohol Occurrences Percent Over- or Under-Represented by State



2013 AD Alcohol Occurrences Percent Over- or Under-Represented by State

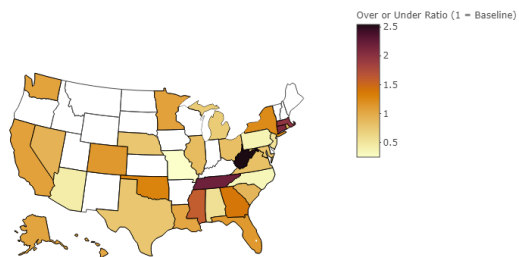
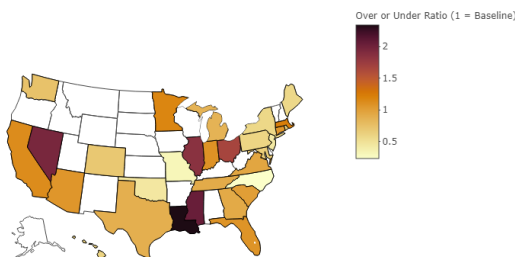


Figure A.3. AD USN 2012 and 2013 Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

2015 AD Alcohol Occurrences Percent Over- or Under-Represented by State



2016 AD Alcohol Occurrences Percent Over- or Under-Represented by State

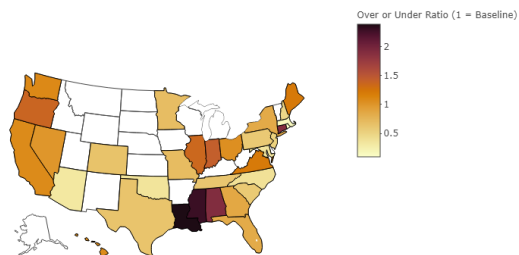
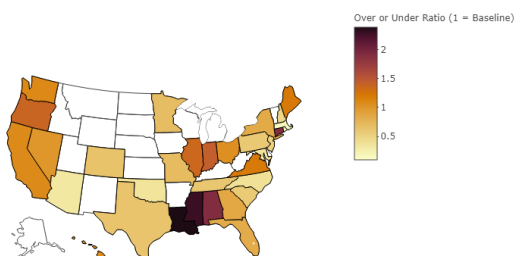


Figure A.4. AD USN 2015 and 2016 Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

2016 AD Alcohol Occurrences Percent Over- or Under-Represented by State



2018 AD Alcohol Occurrences Percent Over- or Under-Represented by State

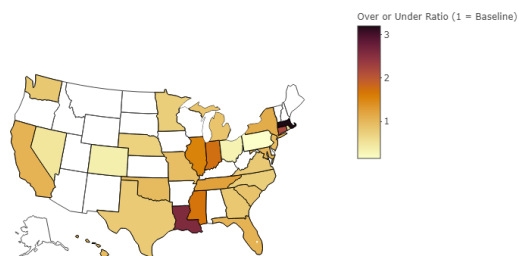


Figure A.5. AD USN 2017 and 2018 Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

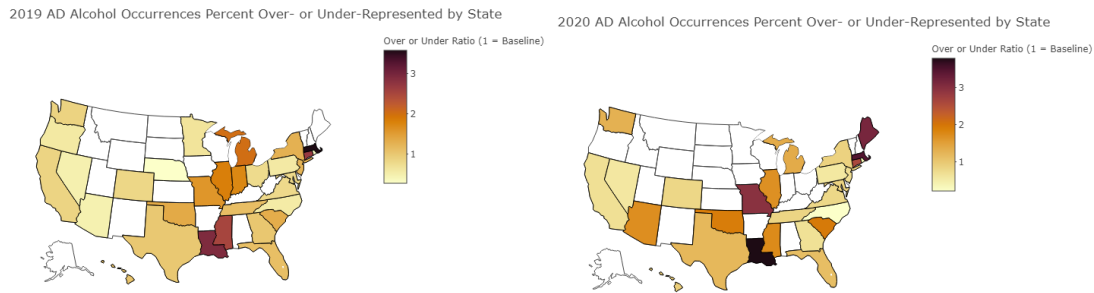


Figure A.6. AD USN 2019 and 2020 Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

A.1.2 Reserve USN Over- or Under-Representation by State

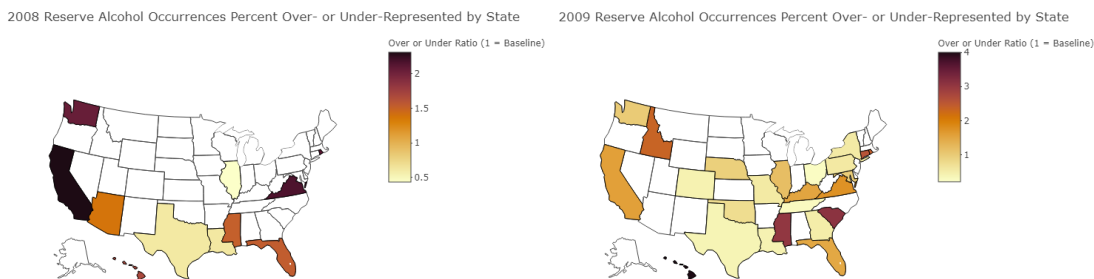


Figure A.7. Reserve USN 2008 and 2009 Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

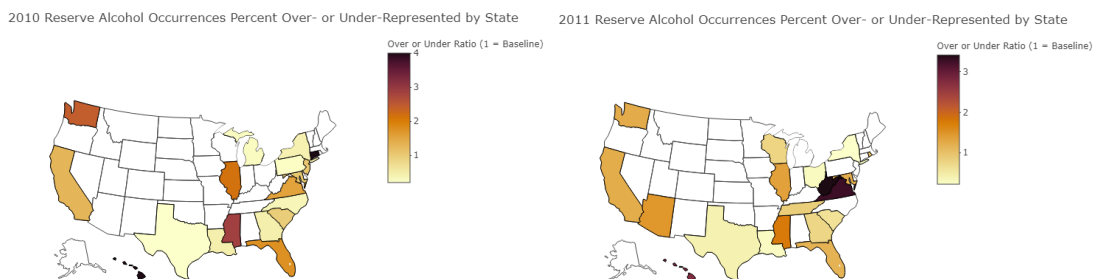
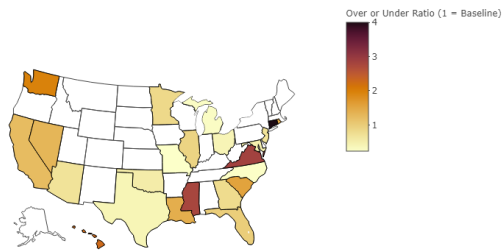


Figure A.8. Reserve USN 2010 and 2011 Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

2012 Reserve Alcohol Occurrences Percent Over- or Under-Represented by State



2013 Reserve Alcohol Occurrences Percent Over- or Under-Represented by State

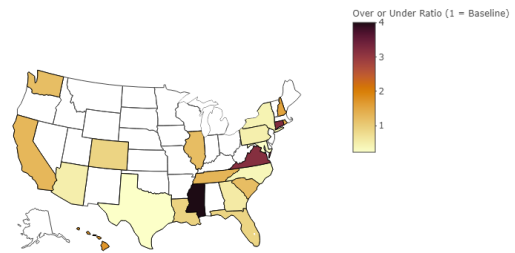
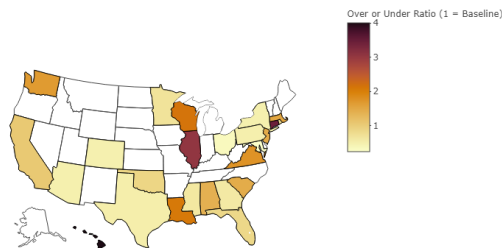


Figure A.9. Reserve USN 2012 and 2013 Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

2015 Reserve Alcohol Occurrences Percent Over- or Under-Represented by State



2016 Reserve Alcohol Occurrences Percent Over- or Under-Represented by State

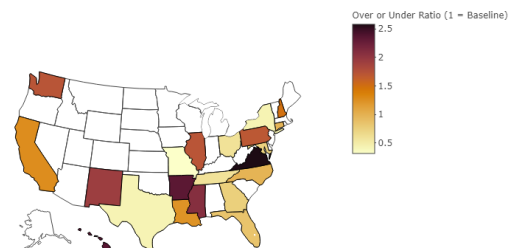
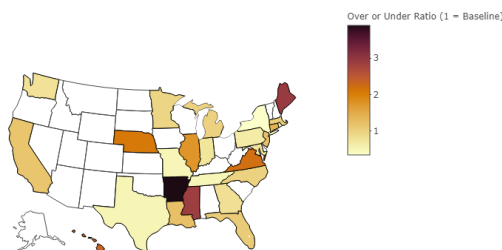


Figure A.10. Reserve USN 2015 and 2016 Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

2017 Reserve Alcohol Occurrences Percent Over- or Under-Represented by State



2018 Reserve Alcohol Occurrences Percent Over- or Under-Represented by State

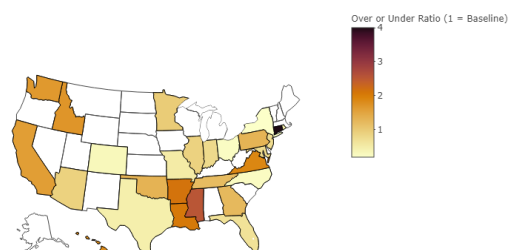


Figure A.11. Reserve USN 2017 and 2018 Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

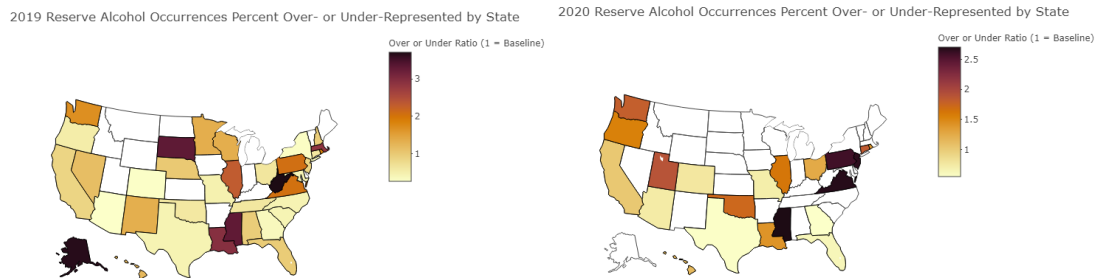


Figure A.12. Reserve USN 2019 and 2020 Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

A.2 AD and Reserve Alcohol Abuse Over- or Under-Representation by Rank

A.2.1 AD USN Over- or Under-Representation by Rank

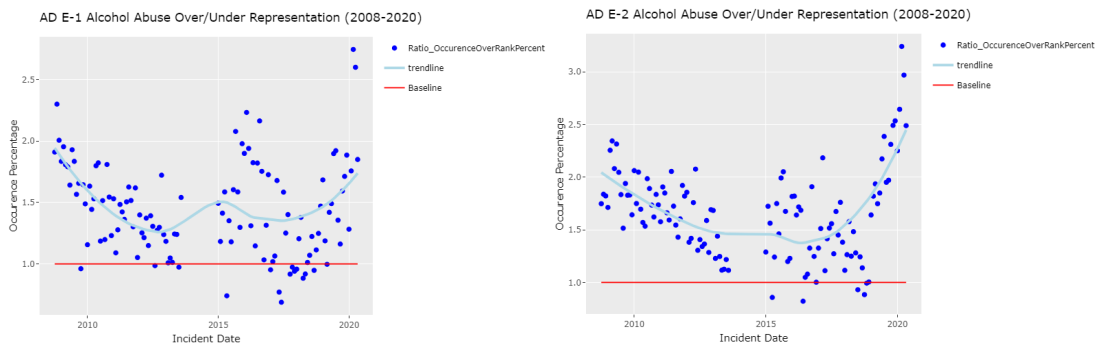


Figure A.13. AD USN E-1 and E-2 Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

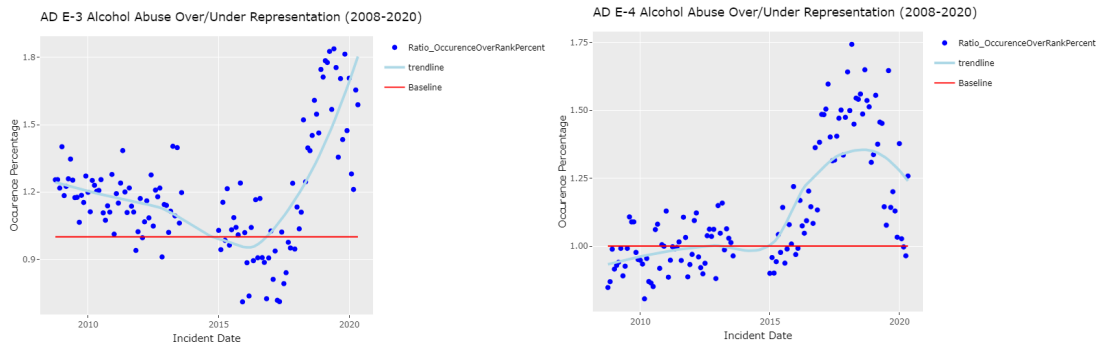


Figure A.14. AD USN E-3 and E-4 Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

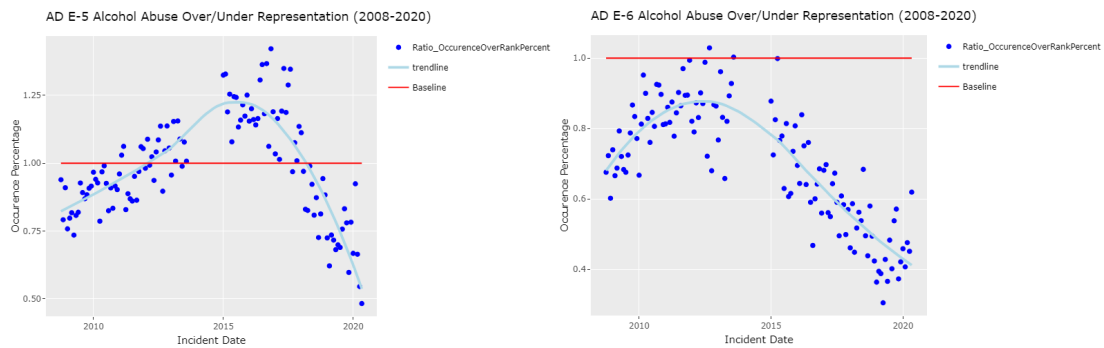


Figure A.15. AD USN E-5 and E-6 Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

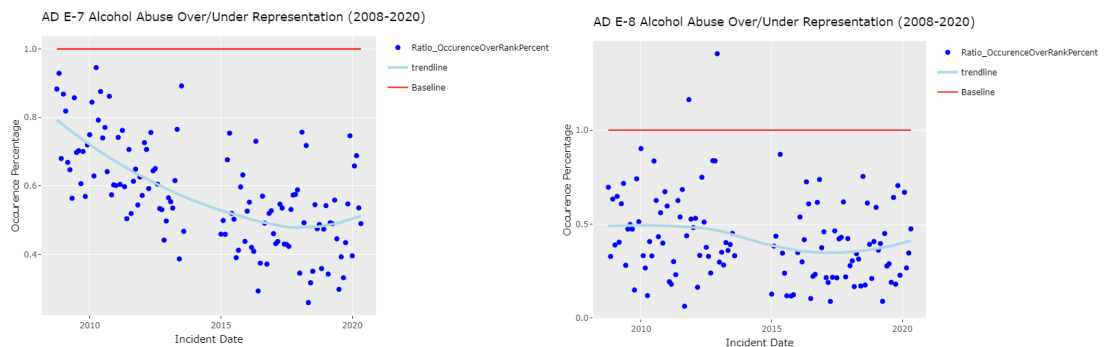


Figure A.16. AD USN E-7 and E-8 Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

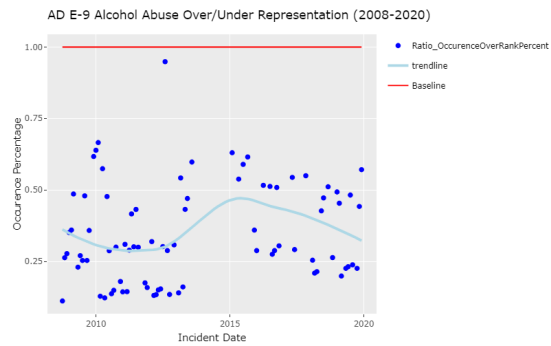


Figure A.17. AD USN E-9 Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

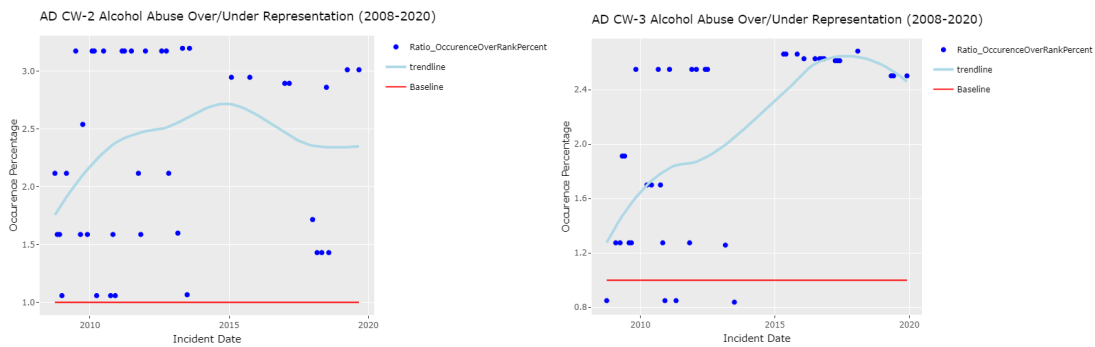


Figure A.18. AD USN CW-2 and CW-3 Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

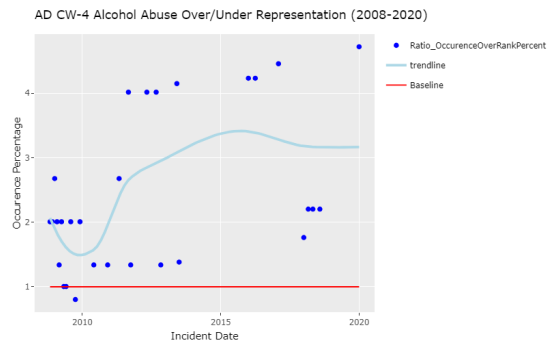


Figure A.19. AD USN CW-4 Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

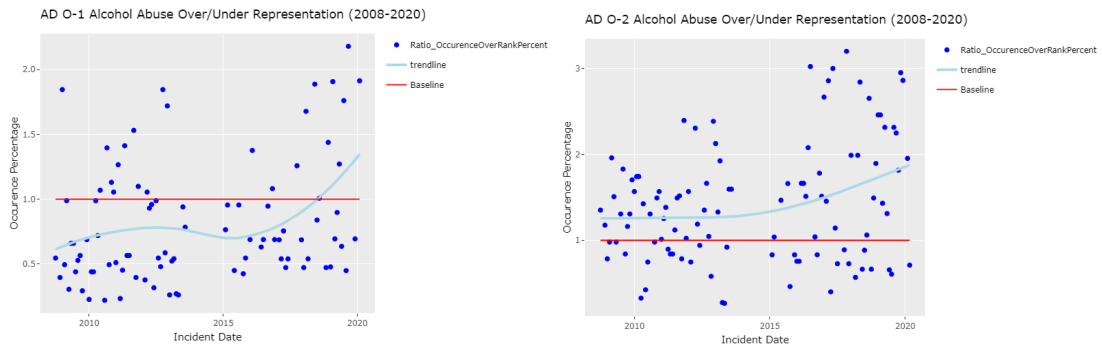


Figure A.20. AD USN O-1 and O-2 Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

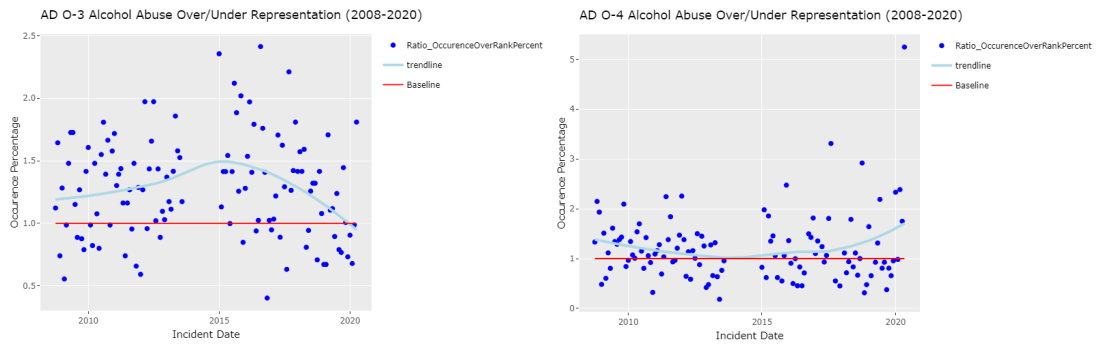


Figure A.21. AD USN O-3 and O-4 Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

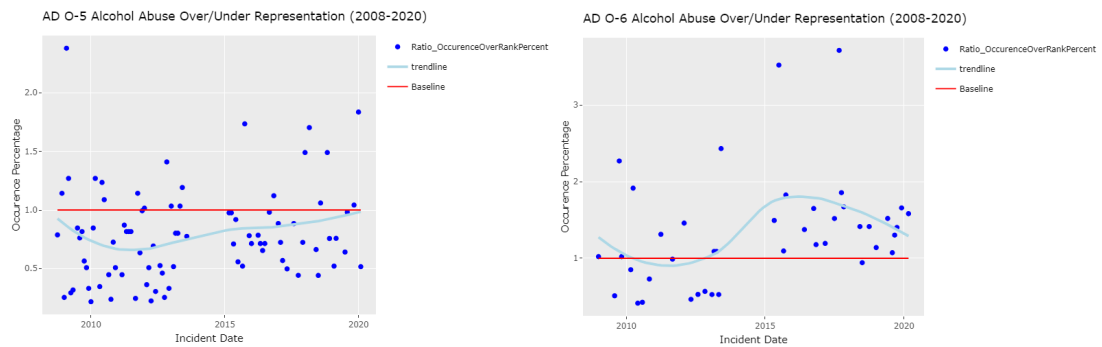


Figure A.22. AD USN O-5 and O-6 Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

A.2.2 Reserve USN Over- or Under-Representation by Rank

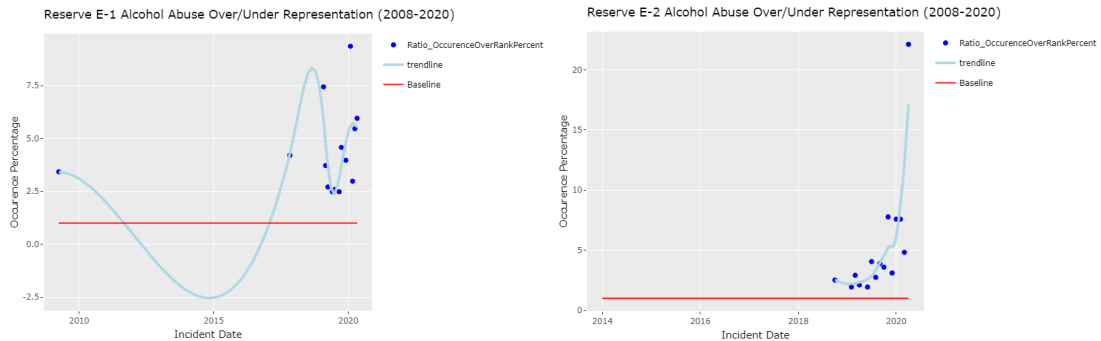


Figure A.23. Reserve USN E-1 and E-2 Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

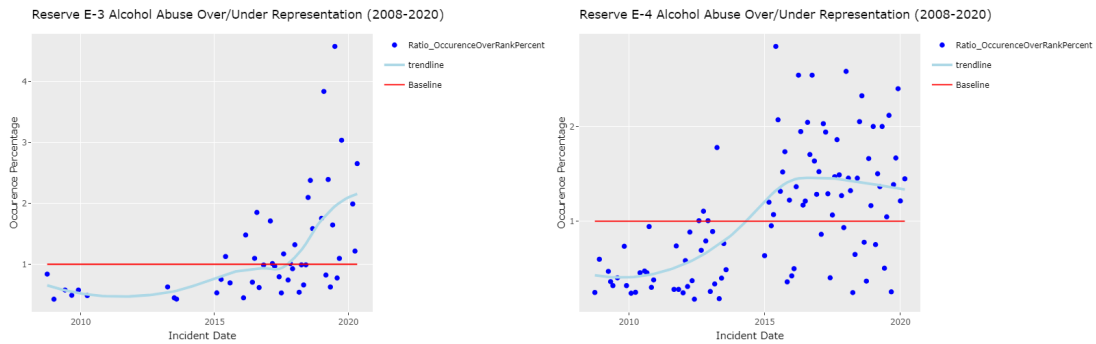


Figure A.24. Reserve USN E-3 and E-4 Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

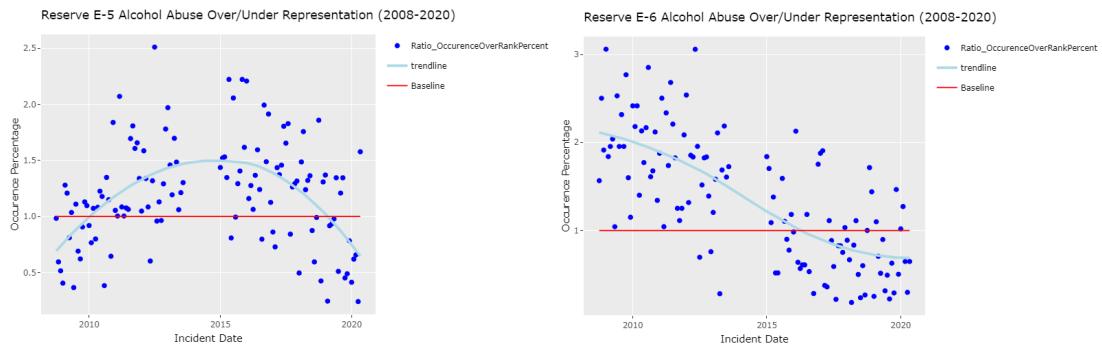


Figure A.25. Reserve USN E-5 and E-6 Alcohol Substance Abuse Over-or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

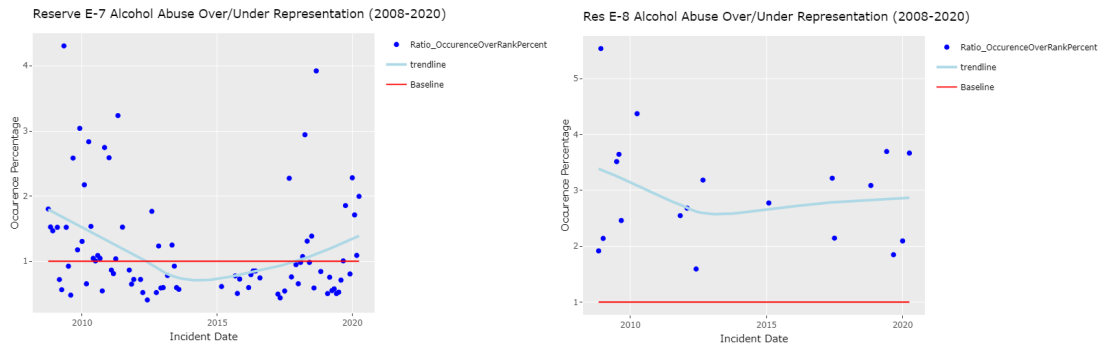


Figure A.26. Reserve USN E-7 and E-8 Alcohol Substance Abuse Over-or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

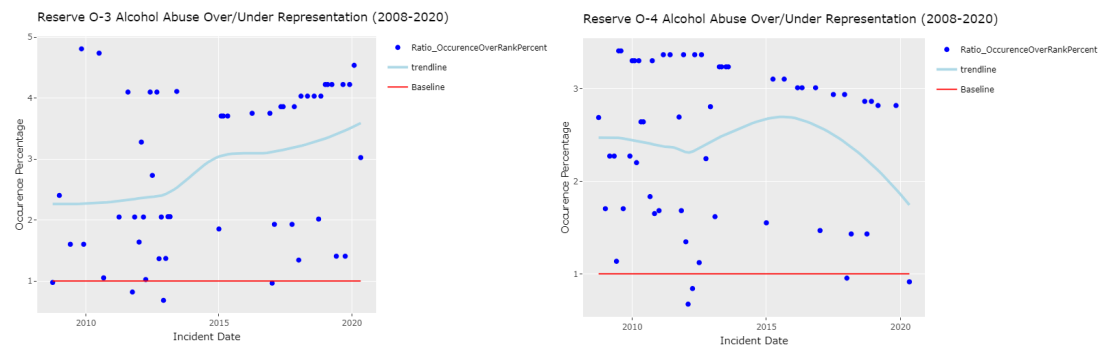


Figure A.27. Reserve USN O-3 and O-4 Alcohol Substance Abuse Over-or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

A.3 AD Alcohol Abuse Over- or Under-Representation by Rating

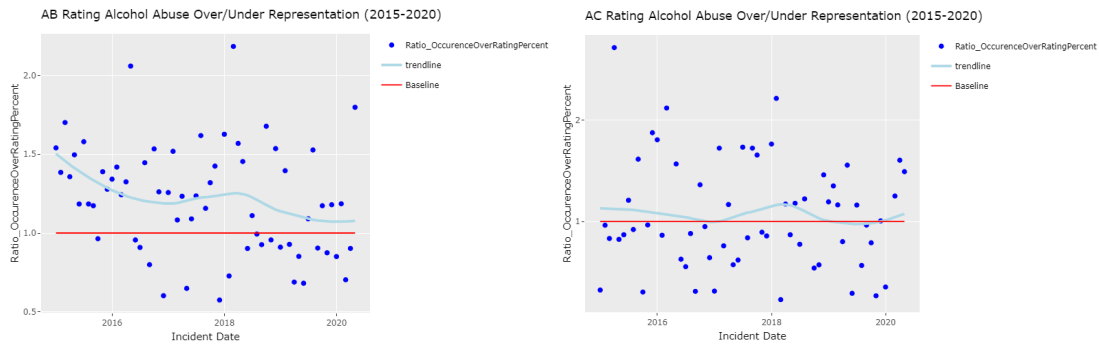


Figure A.28. AD USN AB and AC General Rating Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

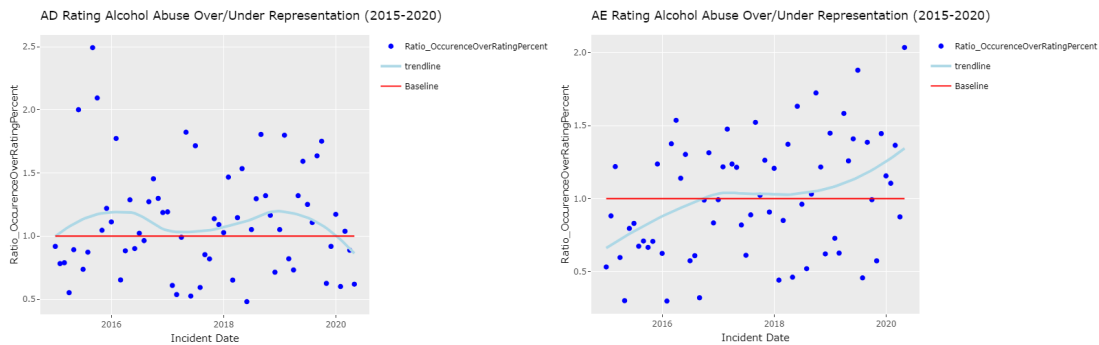


Figure A.29. AD USN AD and AE General Rating Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

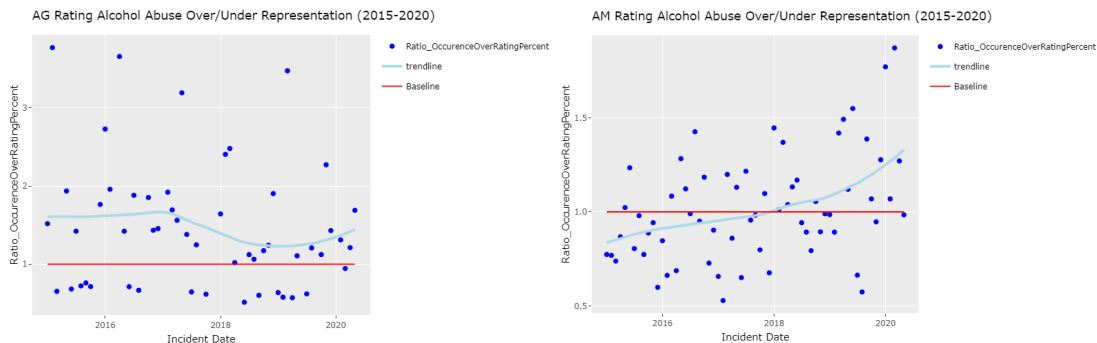


Figure A.30. AD USN AG and AM General Rating Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

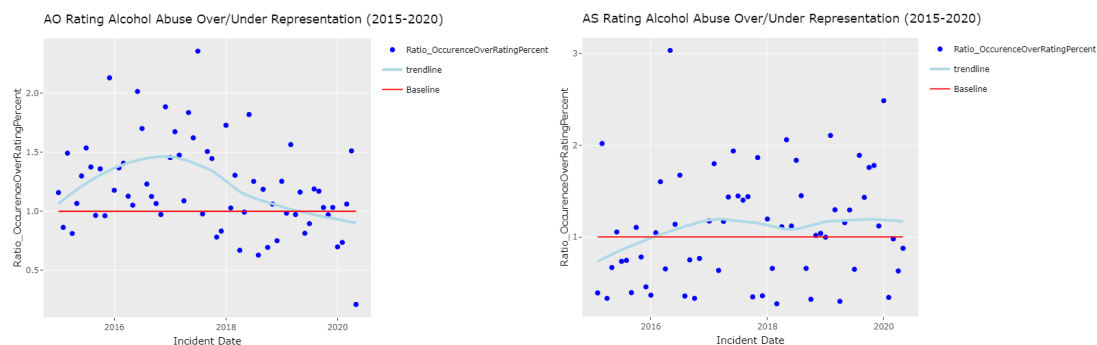


Figure A.31. AD USN AO and AS General Rating Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

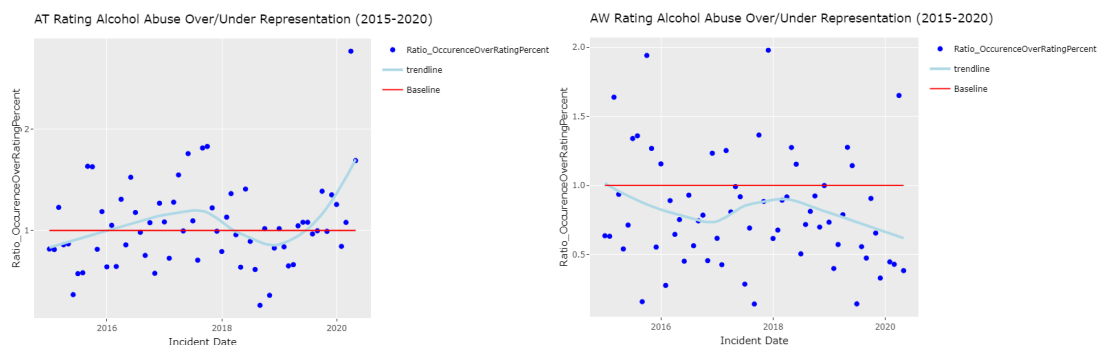


Figure A.32. AD USN AT and AW General Rating Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

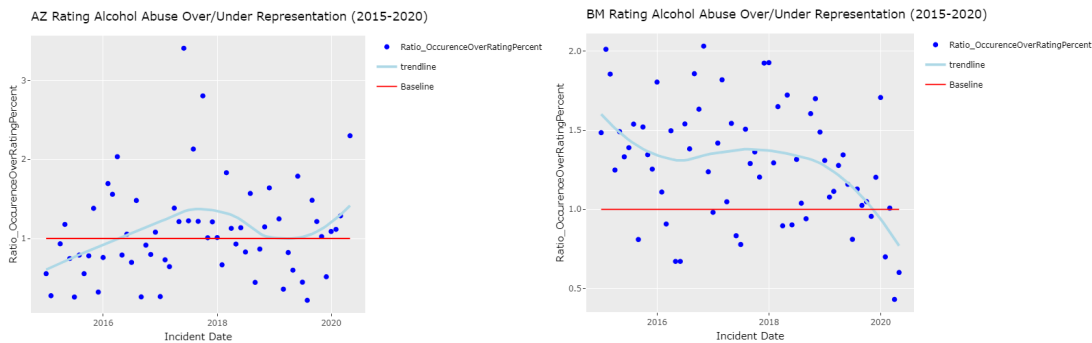


Figure A.33. AD USN AZ and BM General Rating Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

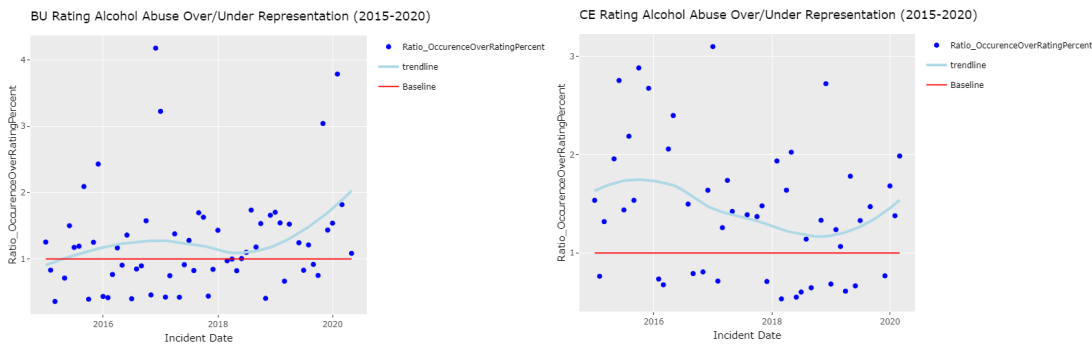


Figure A.34. AD USN BU and CE General Rating Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

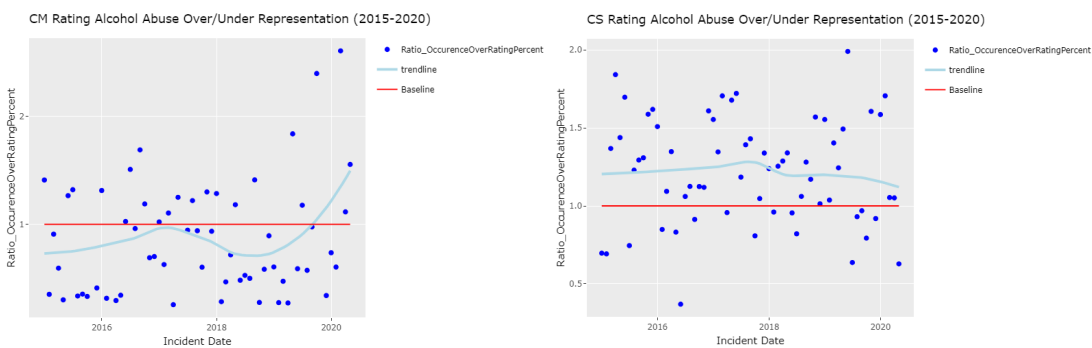


Figure A.35. AD USN CM and CS General Rating Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

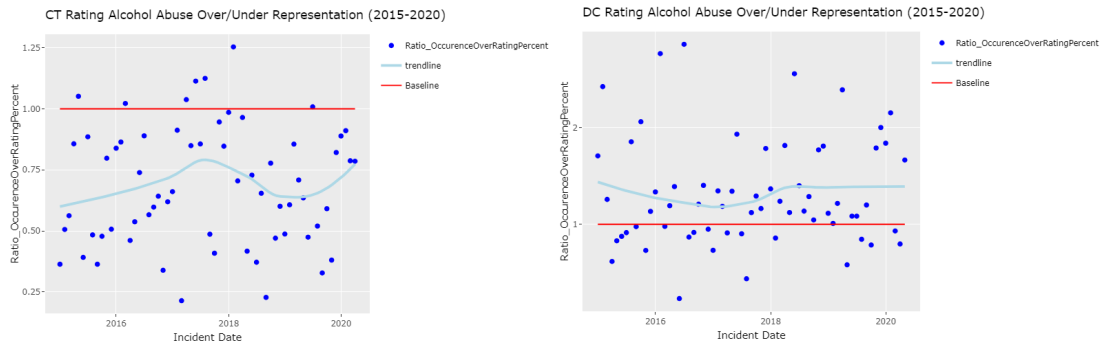


Figure A.36. AD USN CT and DC General Rating Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

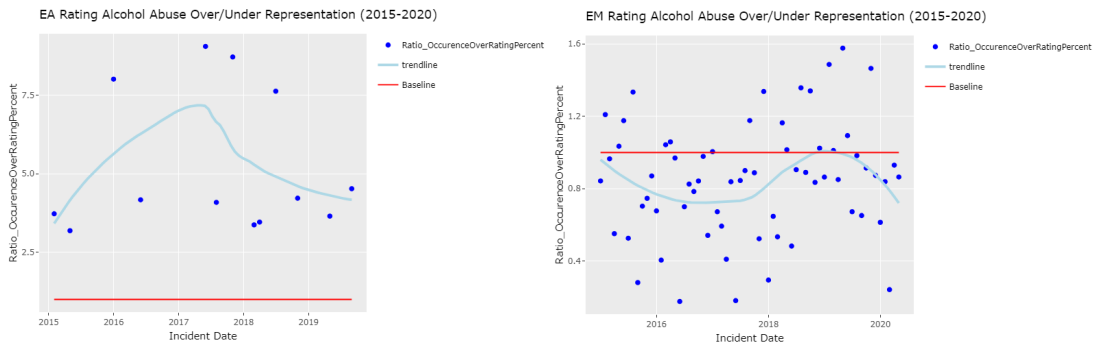


Figure A.37. AD USN EA and EM General Rating Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

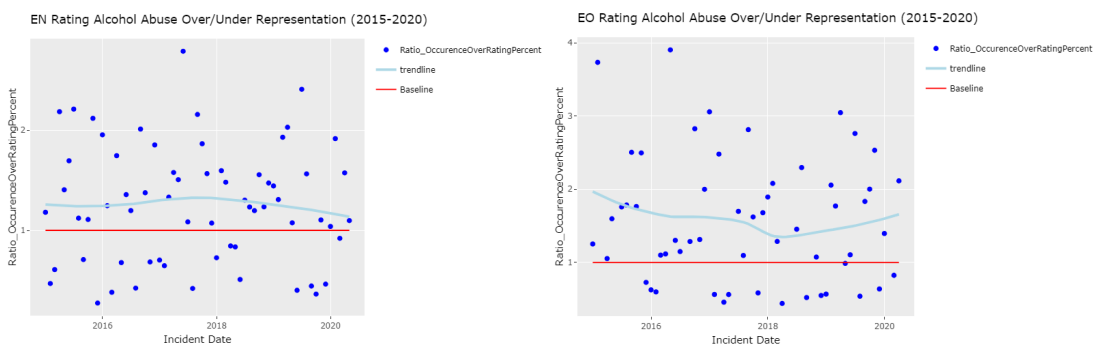


Figure A.38. AD USN EN and EO General Rating Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

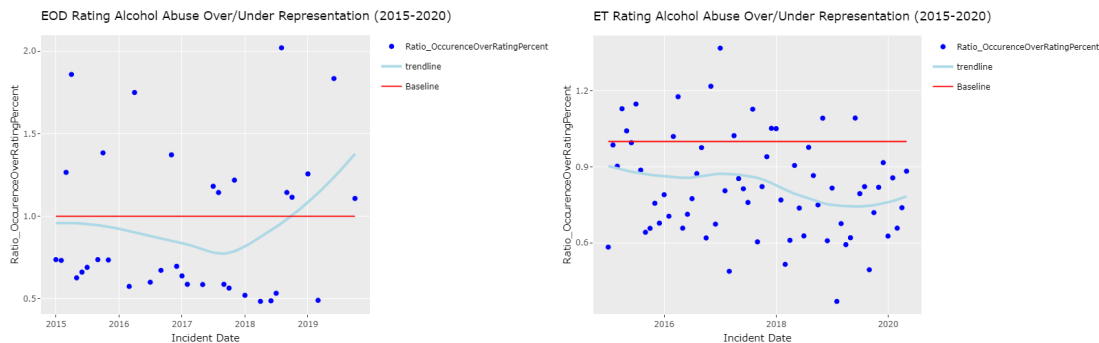


Figure A.39. AD USN EOD and ET General Rating Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

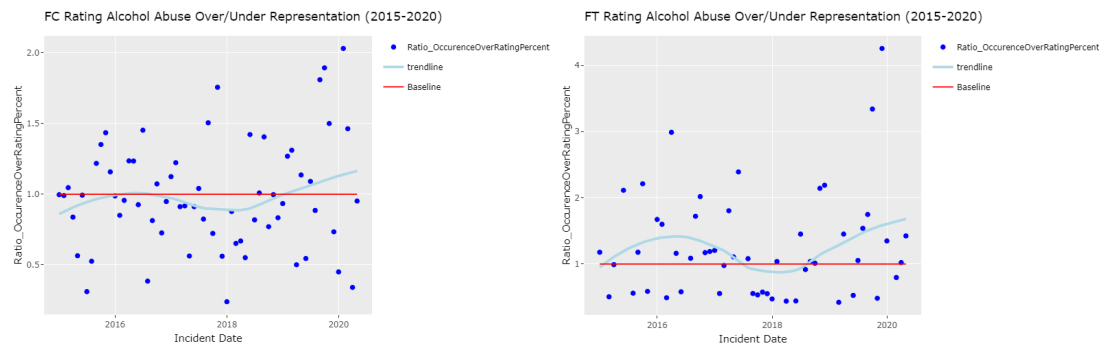


Figure A.40. AD USN FC and FT General Rating Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

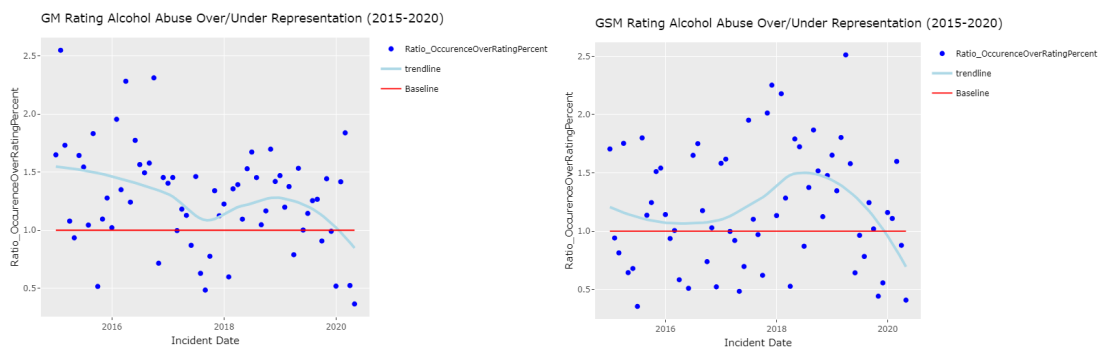


Figure A.41. AD USN GM and GSM General Rating Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

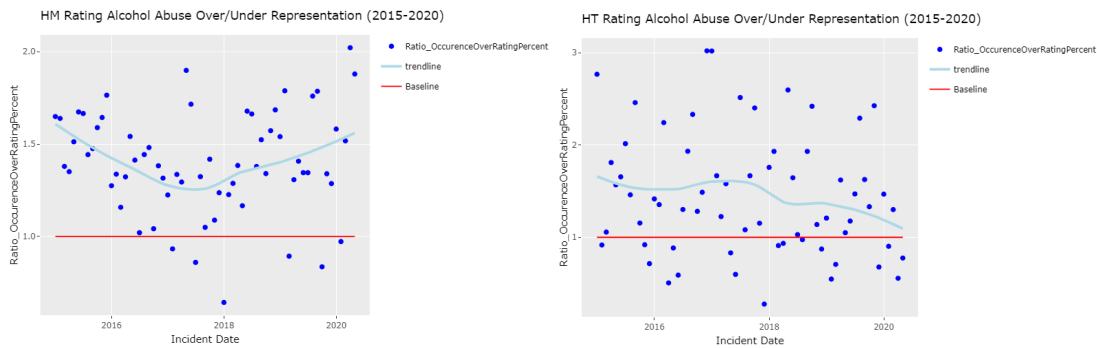


Figure A.42. AD USN HM and HT General Rating Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

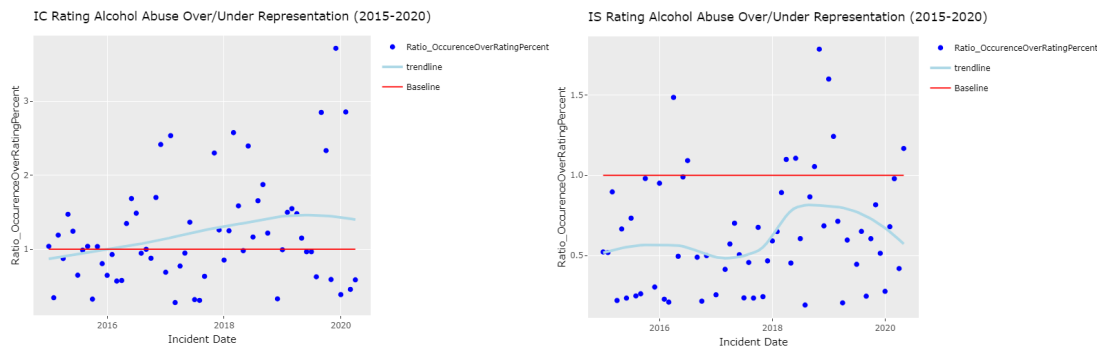


Figure A.43. AD USN IC and IS General Rating Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

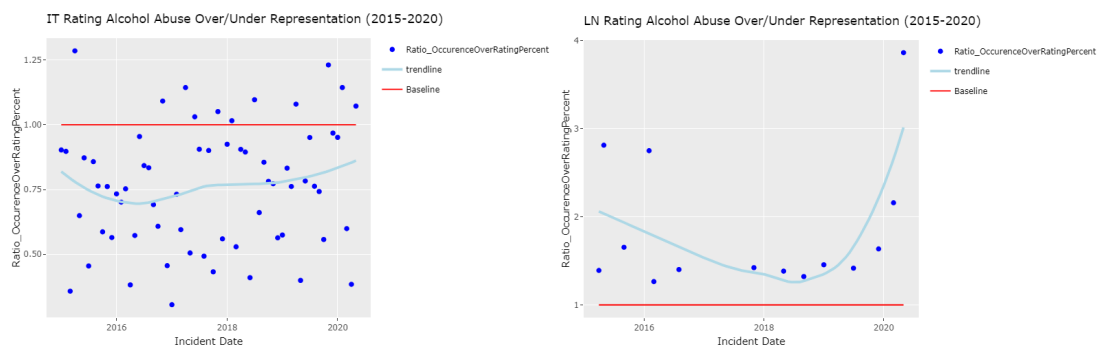


Figure A.44. AD USN FC and FT General Rating Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

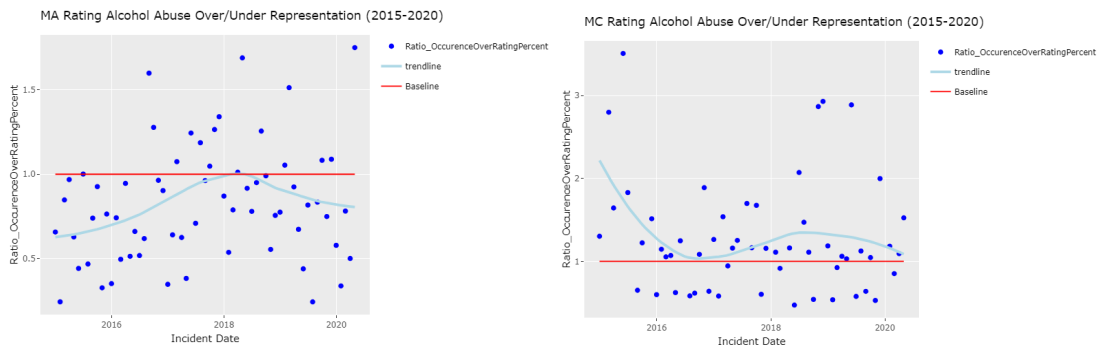


Figure A.45. AD USN MA and MC General Rating Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

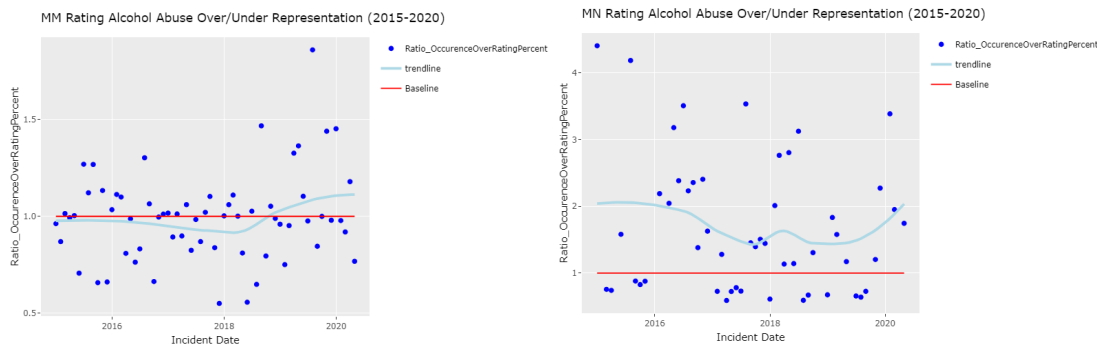


Figure A.46. AD USN MM and MN General Rating Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

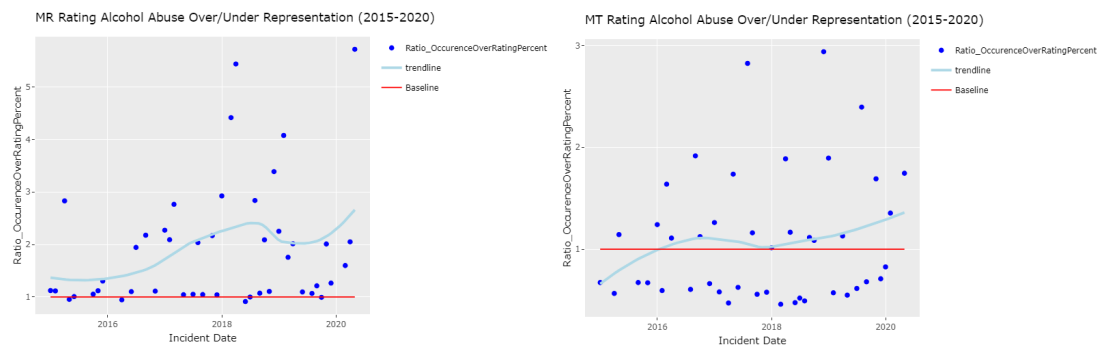


Figure A.47. AD USN MR and MT General Rating Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

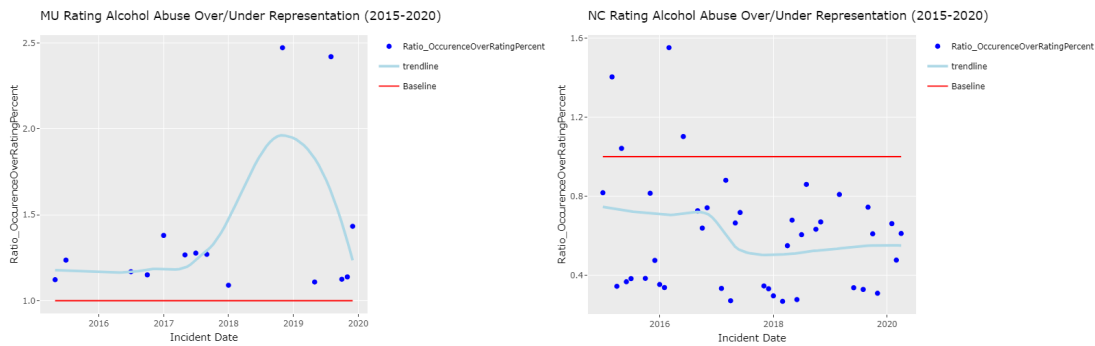


Figure A.48. AD USN MU and NC General Rating Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

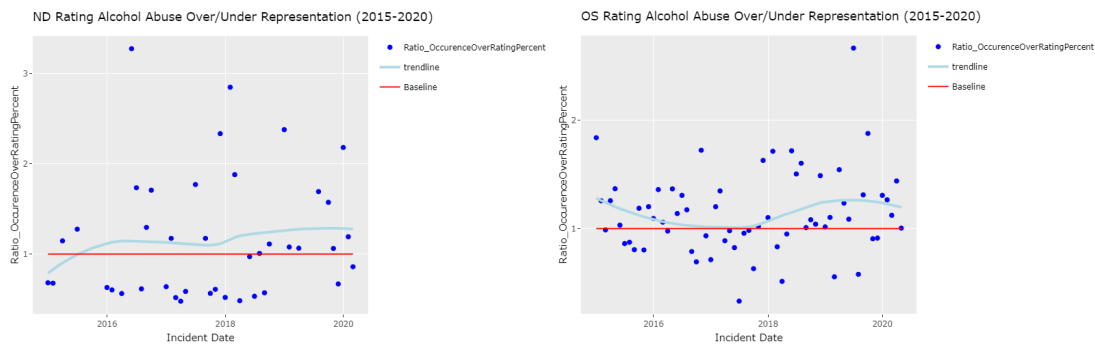


Figure A.49. AD USN ND and OS General Rating Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

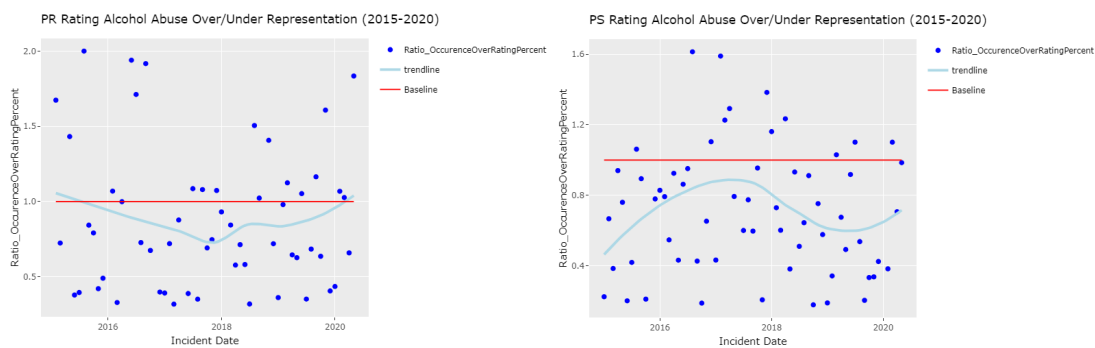


Figure A.50. AD USN PR and PS General Rating Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

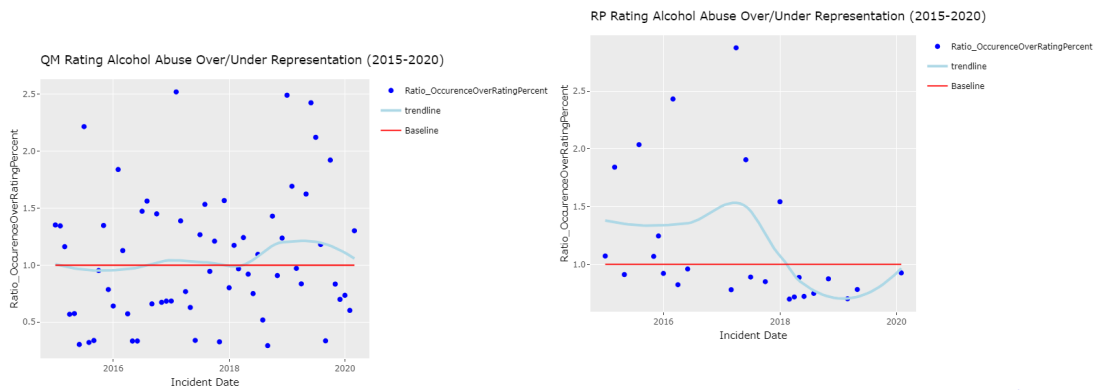


Figure A.51. AD USN QM and RP General Rating Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

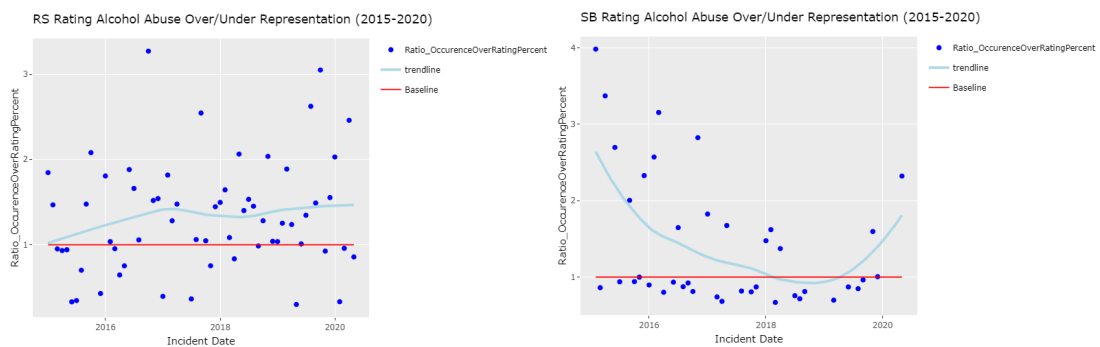


Figure A.52. AD USN RS and SB General Rating Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

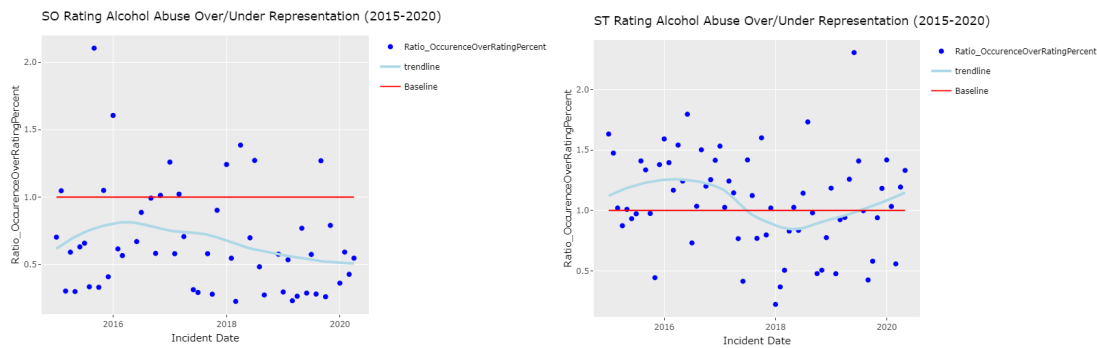


Figure A.53. AD USN SO and ST General Rating Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

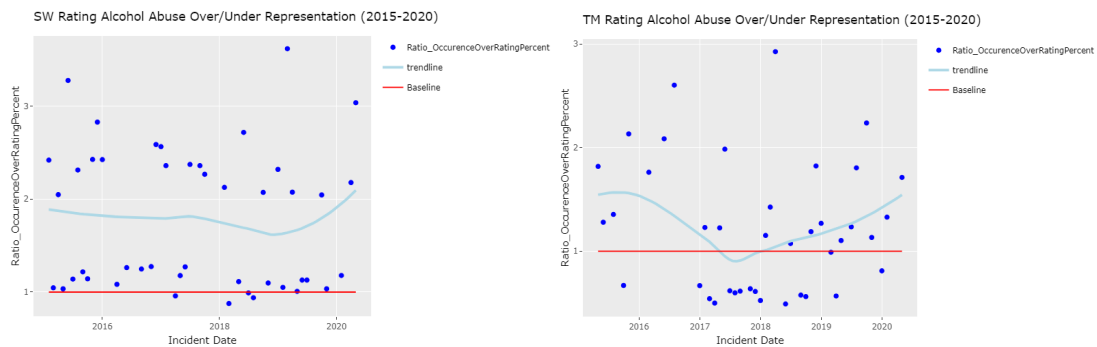


Figure A.54. AD USN SW and TM General Rating Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

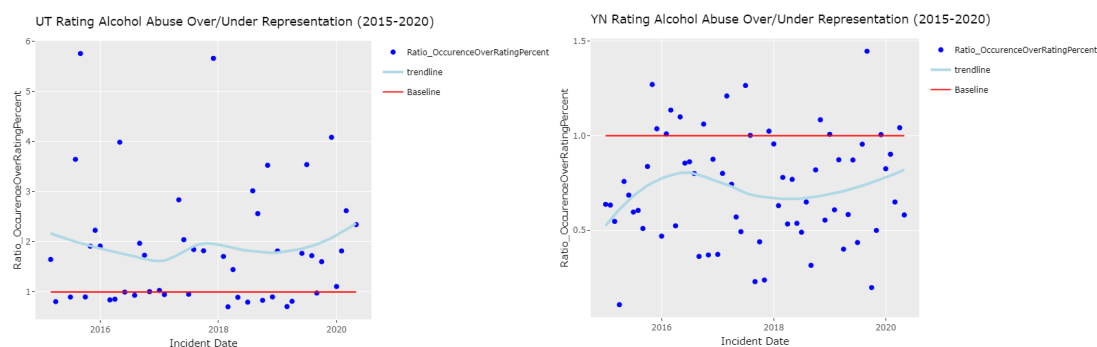


Figure A.55. AD USN UT and YN General Rating Alcohol Substance Abuse Over- or Under-Representation. Source: Office of the Chief of Naval Operations (2021a).

List of References

- DMDC (2021) Active duty & reserve military personnel by service by rank/grade. Accessed November 02, 2021, <https://dwp.dmdc.osd.mil/dwp/app/dod-data-reports/workforce-reports>.
- Office of the Secretary of the Navy (SECNAV) (2019) Military substance abuse prevention and control. Technical Report SECNAVINST 5300.28F, Washington, DC, <https://www.secnave.navy.mil/doni/Directives/05000%20General%20Management%20Security%20and%20Safety%20Services/05-300%20Manpower%20Personnel%20Support/5300.28F.pdf>.
- Betancourt C, Kitsantas P, Goldberg D, Hawks B (2021) Substance use relapse among veterans at termination of treatment for substance use disorders. *Military Medicine* 186(9-10), <https://doi.org/10.1093/milmed/usab280>.
- Breiman L, Cutler A (2022) randomForest: Breiman and cutler’s random forests for classification and regression. R package version 4.6-14, accessed December 02, 2021, <https://cran.r-project.org/web/packages/randomForest/randomForest.pdf>.
- BUPERS-34 (2021) Navy Enlisted System (NES). Dataset, years 2015 to 2020. Millington, TN.
- Buttrey S (2022) Introduction. Class notes, Statistical and Machine Learning, Naval Postgraduate School, January 3, Monterey, CA.
- Buttrey SE, Whitaker LR (2018) *A Data Scientist’s Guide to Acquiring, Cleaning, and Managing Data in R* (John Wiley and Sons Ltd, Hoboken, New Jersey).
- Chief of Naval Operations (2013) Establishment of Navy’s Twenty-First Century Sailor Office (OPNAV N17). Technical Report NAVADMIN 153/13, Washington, DC, <https://navadmin-viewer.fly.dev/NAVADMIN/153/13>.
- Downs B (2019) The data. Cochran J, ed., *INFORMS Analytics Body of Knowledge*, 77–98 (John Wiley and Sons, Inc., Hoboken, NJ).
- Fink D, Keyes K, Calabrese J, Liberzon I, Tamburrino M, Cohen G, Sampson L, Galea S (2017) Deployment and alcohol use in a military cohort: Use of combined methods to account for exposure-related covariates and heterogeneous response to exposure. *American Journal of Epidemiology* 186(4), <https://doi.org/10.1093/aje/kww230>.
- Holt C (1957) Forecasting seasonals and trends by exponentially weighted moving averages. ONR Memorandum, Volume 52, Carnegie Institute of Technology, Pittsburgh.

- Huddleston S, Brown G (2019) Machine learning. Cochran J, ed., *INFORMS Analytics Body of Knowledge*, 231–274 (John Wiley and Sons, Inc., Hoboken, NJ).
- Hyndman R, Athanasopoulos G (2018) *Forecasting: Principles and Practice*. <https://otexts.com/fpp2/>.
- Hyndman R, Athanasopoulos G, Bergmeir C, Caceres G, Chhay L, O’Hara-Wild M, Petropoulos F, Razbash S, Wang E, Yasmeeen F (2022) forecast: Forecasting functions for time series and linear models. R package version 8.16, accessed November 12, 2021, <https://pkg.robjhyndman.com/forecast/>.
- Islam U, Sarker I, Haque E, Hoque M (2020) Predicting individual substance abuse vulnerability using machine learning techniques. Abraham A, Hanne T, Castillo O, Gandhi N, Nogueira R, Hong T, eds., *Hybrid Intelligent Systems. HIS 2020*, 412–421 (Springer, Cham, Switzerland, AG).
- James G, Witten D, Hastie T, Tibshirani R (2017) *An Introduction to Statistical Learning with Applications in R* (Springer Texts in Statistics, New York).
- Kleinbaum DG, Klein M (2005) *Survival Analysis A Self-Learning Text* (Springer Science+Business Media, Inc., New York).
- Kuhn M (2008) Building predictive models in R using the caret package. *Journal of Statistical Software* 28(5), <https://doi.org/10.18637/jss.v028.i05>.
- McKenzie D, McFarlane A, Creamer M, Ikin J, Forbes A, Kelsall H, Clarke D, Glass D, Ittak P, Sim M (2006) Hazardous or harmful alcohol use in Royal Australian Navy veterans of the 1991 Gulf War: Identification of high risk subgroups. *Elsevier Ltd* 31(1683-1694), <https://doi.org/10.1016/j.addbeh.2005.12.027>.
- Meadows S, Engel C, Collins R, Beckman R, Breslau J, Bloom E, Dunbar M, Gilbert M, Grant D, Hawes-Dawson J, Holliday S, MacCarthy S, Pedersen E, Robbins M, Rose A, Ryan J, Schell T, Simmons M (2021) 2018 Health related behaviors survey: Substance use among the active component. rb-10116/3-osd. Rand Corporation, Santa Monica, CA, https://www.rand.org/pubs/research_briefs/RB10116Z3.html.
- Meadows S, Engel C, Collins R, Beckman R, Cefalu M, Hawes-Dawson J, Doyle M, Kress A, Sontag-Padilla L, Ramchand R (2018) 2015 Health related behaviors survey: Substance use among u.s. active-duty service members. rb-9955/7-osd. Rand Corporation, Santa Monica, CA, https://www.rand.org/pubs/research_briefs/RB9955Z7.html.
- Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F (2021) e1071: Misc functions of the department of statistics, probability theory group. R package version 1.9, accessed November 20, 2021, <https://cran.r-project.org/web/packages/e1071/index.html>.

- Montgomery DC, Jennings CL, Kulahci M (2016) *Introduction to Time Series Analysis and Forecasting* (John Wiley and Sons, Inc., Hoboken, New Jersey).
- Office of the Chief of Naval Operations (2009) Navy alcohol and drug abuse prevention and control. Technical Report OPNAVINST 5340.4D, Washington, DC, <https://navytribe.com/wp-content/uploads/2015/11/53504c.pdf>.
- Office of the Chief of Naval Operations (2021a) Alcohol and Drug Management Information Tracking System (ADMITS). Dataset, years 2008 to 2020. Millington, TN.
- Office of the Chief of Naval Operations (2021b) Alcohol and Drug Management Information Tracking System (ADMITS). Department of the Navy, accessed February 11, 2022, <https://www.mynavyhr.navy.mil/support-services/21st-Century-Sailor/NAAP/ADMITS/>.
- R Core Team (2021) R: A language and environment for statistical computing. Vienna, Austria, <https://www.R-project.org>.
- Schmid M, Wright M, Ziegler A (2016) On the use of Harrell's C for clinical risk prediction via random survival forests. *Expert Systems with Applications* 63(30), <https://doi.org/10.1016/j.eswa.2016.07.018>.
- Shen Y, Arkes J, Williams T (2012) Effects of Iraq/Afghanistan deployments on major depression and substance use disorder: Analysis of active duty personnel in the U.S. military. *Research and Practice* 102(S1), <https://doi.org/10.2105/AJPH.2011.300425>.
- Sievert C (2020) Interactive Web-Based data visualization with R, plotly, and shiny. R package version 4.9.4.1, accessed April 15, 2022, <https://plotly-r.com>.
- Sing T, Sander O, Beerenwinkel N, Lengauer T (2005) ROCr: Visualizing classifier performance in R. *Bioinformatics* 21(20), <https://doi.org/10.1093/bioinformatics/bti623>.
- Teeters J, Lancaster C, Brown D, Back S (2017) Substance use disorders in military veterans: Prevalence and treatment challenges. *Substance Abuse and Rehabilitation* 8(69-77), <https://doi.org/10.2147/SAR.S116720>.
- Therneau T (2021) survival: A package for survival analysis in R. R package version 3.2-13, accessed December 07, 2021, <https://cran.r-project.org/package=survival>.
- Tierney N, Cook D, McBain M, Fay C (2021) naniar: Data structures, summaries, and visualisations for missing data. R package version 0.6.1, accessed November 03, 2021, <https://cran.r-project.org/web/packages/naniar/index.html>.
- Wang H, Li G (2017) A selective review on random survival forests for high dimensional data. *Quant Biosci* 36(2), <https://doi.org/10.22283/qbs.2017.36.2.85>.

- Wickham H (2016) ggplot2: Elegant graphics for data analysis. R package version 3.3.3, accessed April 15, 2022, <https://ggplot2.tidyverse.org>.
- Wright M, Ziegler A (2017) ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software* 77(1), <https://doi.org/10.18637/jss.v077.i01>.
- Yoshida R (2021a) (Univariate) Time series overview. Class notes, Advanced Data Analysis, Naval Postgraduate School, July 6, Monterey, CA.
- Yoshida R (2021b) kNN and CART. Class notes, Advanced Data Analysis, Naval Postgraduate School, July 27, Monterey, CA.
- Yoshida R (2021c) Receiver operating characteristic. Class notes, Advanced Data Analysis, Naval Postgraduate School, August 17, Monterey, CA.
- Zhang H, Singer BH (2010) *Recursive Partitioning and Applications* (Springer Science+Business Media, LLC, New York).

Initial Distribution List

1. Defense Technical Information Center
Ft. Belvoir, Virginia
2. Dudley Knox Library
Naval Postgraduate School
Monterey, California