



**NAVAL
POSTGRADUATE
SCHOOL**

MONTEREY, CALIFORNIA

THESIS

**A HUMAN FACTORS ANALYSIS OF USAF REMOTELY
PILOTED AIRCRAFT MISHAPS**

by

Matthew T. Taranto

June 2013

Thesis Advisor:
Co-Advisor
Second Reader:

Michael E. McCauley
Christian (Kip) Smith
Chad W. Seagren

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			<i>Form Approved OMB No. 0704-0188</i>
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.			
1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE June 2013	3. REPORT TYPE AND DATES COVERED Master's Thesis	
4. TITLE AND SUBTITLE A HUMAN FACTORS ANALYSIS OF USAF REMOTELY PILOTED AIRCRAFT MISHAPS		5. FUNDING NUMBERS	
6. AUTHOR(S) Matthew T. Taranto		8. PERFORMING ORGANIZATION REPORT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000		10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A		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. IRB Protocol number ____N/A____.	
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited		12b. DISTRIBUTION CODE A	
13. ABSTRACT (maximum 200 words) As the effort to demonstrate the viability and effectiveness of Remotely Piloted Aircraft (RPA) systems continues, there is an increasing demand for improved total system performance; specifically, reduced mishap rates. The USAF MQ-1 and MQ-9 have produced lifetime mishap rates of 7.58 and 4.58 mishaps per 100,000 flight hours, respectively. To improve the understanding of RPA mishap epidemiology, an analysis was completed on USAF MQ-1 and MQ-9 RPA mishaps from 2006-2011. The dataset included 88 human error-related mishaps that were coded using the DoD Human Factors Analysis and Classification System. The specific research question was: Do the types of active failures (unsafe acts) and latent failures (preconditions, unsafe supervision, and organizational influences) differ between the MQ-1 and MQ-9 when operated with the same Ground Control Station (GCS)? The single inclusion of Organizational Climate (organizational influence) in the Level II logistic regression model suggests that there is not a statistically significant difference in RPA-type mishaps with regard to human error. These results suggest that human performance requirements should be coupled to the GCS and not aircraft type. The models have the promise to inform RPA certification standards and future system designs.			
14. SUBJECT TERMS Human Systems Integration, Remotely Piloted Aircraft, Safety, Human Factors, Human Error, Human Factors Analysis and Classification System (HFACS), MQ-1, MQ-9, Ground Control Station (GCS)			15. NUMBER OF PAGES 89
			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

**A HUMAN FACTORS ANALYSIS OF USAF REMOTELY PILOTED
AIRCRAFT MISHAPS**

Matthew T. Taranto
Major, United States Air Force
M.S., Embry-Riddle Aeronautical University, 2010
B.S., Northern Arizona University, 2003

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN HUMAN SYSTEMS INTEGRATION

from the

**NAVAL POSTGRADUATE SCHOOL
June 2013**

Author: Matthew T. Taranto

Approved by: Michael E. McCauley
Thesis Advisor

Christian (Kip) Smith
Co-Advisor

Chad W. Seagren
Second Reader

Robert F. Dell
Chair, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

As the effort to demonstrate the viability and effectiveness of Remotely Piloted Aircraft (RPA) systems continues, there is an increasing demand for improved total system performance; specifically, reduced mishap rates. The USAF MQ-1 and MQ-9 have produced lifetime mishap rates of 7.58 and 4.58 mishaps per 100,000 flight hours, respectively. To improve the understanding of RPA mishap epidemiology, an analysis was completed on USAF MQ-1 and MQ-9 RPA mishaps from 2006-2011. The dataset included 88 human error-related mishaps that were coded using the DoD Human Factors Analysis and Classification System. The specific research question was: Do the types of active failures (unsafe acts) and latent failures (preconditions, unsafe supervision, and organizational influences) differ between the MQ-1 and MQ-9 when operated with the same Ground Control Station (GCS)? The single inclusion of Organizational Climate (organizational influence) in the Level II logistic regression model suggests that there is not a statistically significant difference in RPA-type mishaps with regard to human error. These results suggest that human performance requirements should be coupled to the GCS and not aircraft type. The models have the promise to inform RPA certification standards and future system designs.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

I.	INTRODUCTION.....	1
A.	OVERVIEW	1
B.	BACKGROUND	2
1.	The MQ-1Predator	2
2.	The MQ-9 Reaper	3
3.	The Ground Control Station.....	4
C.	OBJECTIVE	5
D.	PROBLEM STATEMENT	5
E.	RESEARCH QUESTION	6
F.	HUMAN SYSTEMS INTEGRATION	6
1.	Personnel.....	7
2.	Human Factors Engineering.....	8
3.	Occupational Health	8
4.	Safety	8
G.	SCOPE AND LIMITATIONS	9
H.	ORGANIZATION	9
II.	LITERATURE REVIEW	11
A.	OVERVIEW	11
B.	ACCIDENT CAUSATION THEORIES	12
1.	Simple Linear Accident Model (Domino Model)	12
2.	Complex Linear Accident Model (Swiss Cheese Model).....	13
3.	Non-Linear or Systemic Models	14
C.	DOD HUMAN FACTORS ANALYSIS AND CLASSIFICATION SYSTEM	14
D.	USAF MQ-1 AND MQ-9 MISHAPS.....	17
E.	USAF SAFETY INVESTIGATIONS	19
1.	Mishap Categories	20
2.	AFSAS.....	20
III.	RESEARCH METHOD	21
A.	APPROACH.....	21
B.	DATABASE AND ACCIDENT CODING	21
1.	HFACS Coding.....	21
C.	DATA ANALYSIS	23
1.	DOD HFACS Category Frequency	23
2.	Inter-Rater Reliability	23
3.	Human Error Pattern Analysis	24
IV.	RESULTS	25
A.	ACCIDENT DATABASE	25
B.	INTER-RATER RELIABILITY.....	29
1.	HFACS Level I	29
2.	HFACS Level II.....	30

C.	HUMAN ERROR PATTERN ANALYSIS	31
1.	HFACS Level I Analysis.....	31
2.	HFACS Level II Analysis	33
D.	LOGISTIC REGRESSION ANALYSIS	35
1.	Covariate Analysis	36
2.	Level I Analysis	37
3.	Level I Cross Validation.....	41
4.	Level II Analysis.....	43
5.	Level II Cross Validation	48
E.	SUMMARY	52
V.	DISCUSSION	55
A.	OVERVIEW.....	55
B.	RESEARCH QUESTION	57
C.	IMPLICATIONS FOR SYSTEM DESIGN	58
D.	CONCLUSION	59
APPENDIX A.	MQ-1 AND MQ-9 SYSTEM CHARACTERISTICS	61
APPENDIX B.	INTER-RATER RELIABILITY SAMPLE SET.....	63
LIST OF REFERENCES	65
INITIAL DISTRIBUTION LIST	67

LIST OF FIGURES

Figure 1.	MQ-1 Predator (From U.S. Air Force, 2012)	3
Figure 2.	MQ-9 Reaper (From U.S. Air Force, 2012).....	4
Figure 3.	Interior of the GCS for the MQ-1 and MQ-9 (From U.S. Air Force, 2012).....	5
Figure 4.	USAF RPA Training (After Taranto, 2012)	7
Figure 5.	Simple Linear Accident Model (From Hollnagel, Woods, & Leveson, 2006)	12
Figure 6.	The Swiss Cheese Model (From Webster, White, & Wurmstein, 2005).....	13
Figure 7.	The Four Tiers of DoD HFACS (From Webster, White, & Wurmstein, 2005)	16
Figure 8.	MQ-1 Mishap History (From USAF, 2012).	18
Figure 9.	MQ-9 Mishap History (From USAF, 2012).	18
Figure 10.	Level I Inter-Rater Gauge Attribute Chart.....	29
Figure 11.	Level I Inter-Rater Reliability Kappa Coefficients.....	30
Figure 12.	Level II Inter-Rater Gauge Attribute Chart	30
Figure 13.	Level II Inter-Rater Reliability Kappa Coefficients	31
Figure 14.	Level I HFACS Coding by RPA.....	32
Figure 15.	Level II HFACS Coding by RPA	33
Figure 16.	Level I Stepwise Fit Results for RPA.....	38
Figure 17.	Level I Nominal Logistic Fit for RPA	39
Figure 18.	Level I Odds Ratio Results	40
Figure 19.	Level I ROC Curve	41
Figure 20.	Level I Cross Validation Training Set Results	42
Figure 21.	Level I Cross Validation Test Set Results	43
Figure 22.	Level II Stepwise Fit Results for RPA.....	45
Figure 23.	Level II Nominal Logistic Fit for RPA.....	46
Figure 24.	Level II Odds Ratio Results.....	47
Figure 25.	Level II ROC Curve.....	48
Figure 26.	Level II Cross Validation Training Set Results	49
Figure 27.	Level II Cross Validation Test Set Results	50
Figure 28.	Level II Cross Validation Training Set Results (Second Iteration)	51
Figure 29.	Level II Cross Validation Test Set Results (Second Iteration).....	52

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

Table 1.	USAF Mishap Categories	20
Table 2.	DoD HFACS Grouping.....	22
Table 3.	RPA-Human Error Mishap Distribution.....	25
Table 4.	Nanocodes Cited by RPA Type	26
Table 5.	Mishaps by Class	26
Table 6.	Mishaps by Class and Nanocode	27
Table 7.	Mishaps by Domain	27
Table 8.	Mishaps by Phase of Flight.....	28
Table 9.	Mishaps by Phase of Flight for Both RPA Types.....	28
Table 10.	Level I Citing Frequency by RPA	32
Table 11.	Level I Citing Frequency (Both).....	33
Table 12.	Level II Citing Frequency by RPA	34
Table 13.	Level II Citing Frequency (Both)	35
Table 14.	Summary of Covariate Chi-Square Tests	37
Table 15.	Summary of Level I Mishap Distribution.....	37
Table 16.	Summary of Level II Mishap Distribution.....	44

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

A	Acts
AE1	Skill-Based Errors
AE2	Judgment and Decision-Making Errors
AE3	Perception Errors
AFB	Air Force Base
AFI	Air Force Instruction
AFMAN	Air Force Manual
AFSAS	Air Force Safety Automated System
AFSEC	Air Force Safety Center
AIC	Akaike Information Criterion
AV	Violations
CAS	Close Air Support
CSAR	Combat Search and Rescue
DoD	Department of Defense
FAA	Federal Aviation Administration
FY	Fiscal Year
GCS	Ground Control Station
HFACS	Human Factors Analysis and Classification System
HFE	Human Factors Engineering
HPW	Human Performance Wing
HSI	Human Systems Integration
IRB	Institution Review Board
ISR	Intelligence, Surveillance, and Reconnaissance
MALE	Medium Altitude Long Endurance
MQ	Multi-Role Unmanned Aircraft
Mx	Maintenance
O	Organizational Influences
OC	Organizational Climate

OP	Organizational Processes
Ops	Operations
OR	Resource/Acquisition Management
OSD	Office of the Secretary of Defense
P	Preconditions
PC1	Cognitive Factors
PC2	Psycho-Behavioral Factors
PC3	Adverse Physiological States
PC4	Physical/Mental Limitations
PC5	Perceptual Factors
PE1	Physical Environment
PE2	Technological Environment
PP1	Coordination/Communication/Planning Factors
PP2	Self-Imposed Stress
ROC	Receiver Operating Characteristic
RPA	Remotely Piloted Aircraft
S	Supervision
SEF	Flight Safety
SF	Failure to Correct Known Problem
SI	Inadequate Supervision
SP	Planned Inappropriate Operations
SUPT	Specialized Undergraduate Pilot Training
SV	Supervisory Violations
TST	Time Sensitive Targets
UAS	Unmanned Aircraft System
USAF	United States Air Force

EXECUTIVE SUMMARY

Human error continues to plague military aviation well into the 21st century and does not appear to discriminate between manned or unmanned aircraft systems. Historical analysis provides evidence that human error is identified as a causal factor in 80 to 90 percent of aviation mishaps, and is therefore the single greatest threat to flight safety. The dramatic increase in Combatant Commanders' requests for Remotely Piloted Aircraft (RPA) systems during the last decade, in addition to the rapidly growing civilian RPA sector, is evidence that these systems are becoming an integral component to our national defense and numerous civil aeronautics sectors.

Along with the rapid increase in RPA use, a high mishap rate has followed. The cost associated with human error-related RPA mishaps is significant. RPAs provide a unique challenge to developers of certification standards (e.g., FAA) because the cockpit, also referred to as the ground control station (GCS), and the aircraft are separate and it is theoretically possible to mix and match GCS' and aircraft. So what matters in terms of human performance: the GCS or the aircraft? This question is a significant point of debate in policy and worthy of analysis. As such, adequate incorporation of Human Systems Integration early in the system acquisition phases is dependent on quantitative and relevant data to serve as forcing functions in designing and building smart human-centered systems that optimize total system performance.

The analysis and understanding of where human error contributes to RPA mishaps is lacking in the current literature. In an effort to improve the understanding of RPA mishap epidemiology, an analysis was completed on USAF MQ-1 and MQ-9 RPA mishaps from 2006–2011. The dataset provided the opportunity to gain insight into this question as a natural experiment in which the GCS is controlled and the aircraft is varied. The pattern of human performance failures provide evidence supporting the development of aircraft certification standards or the standards on the GCS used in the RPA system. The dataset included 88 human error-related mishaps that were coded using DoD Human Factors Analysis Classification System (HFACS), an evolution of Reason's (1990) complex linear accident model, known as the Swiss Cheese Model. The MQ-1 and MQ-9

are the premier operational RPA systems for the USAF and are highly valued operational assets. The aircraft have different flight characteristics but are controlled using the same GCS. Do the types of active failures (unsafe acts) and latent failures (preconditions, unsafe supervision, and organizational influences) differ between the MQ-1 and MQ-9 when operated with the same GCS? The present analysis of the human error data sheds light on that issue.

Human error coding was assigned by the original mishap investigators and was validated by conducting inter-rater reliability analyses of the mishaps. The moderate to good agreement identified between Rater 1 (original mishap investigator) and Raters 2 (aerospace medicine specialist) and 3 (aerospace physiologist) provided sufficient evidence to support validation of the study dataset.

The initial exploration of the data involved the organization of the data into two levels of the DoD HFACS hierarchy, Level I (Acts, Preconditions, Supervision, Organization) and Level II (20 subcategories of Level I), referred to as categories of nanocodes. Covariates evaluated in the dataset included Phase of Flight (Ground Operations, Takeoff, Climb, Enroute, Landing, and Other), Mishap Domain (Operations, Logistics/Maintenance, and Miscellaneous), and Mishap Class (A, B, and C) by RPA type. The application of chi-square tests to evaluate the observed and expected frequencies at both Levels for the MQ-1 and MQ-9 provided statistical rationale for selecting nanocodes and covariates for inclusion in the logistic regression analysis. The analysis at Level I did not identify any latent or active failures, as defined in DoD HFACS, for inclusion in the model. The analysis at Level I suggests that the binary response variable (RPA type) was not associated with human error (DoD HFACS). The Level II results of the logistic regression are consistent with the results from Level I and included only one DoD HFACS category, Organizational Climate. The analyses rejected the hypothesis that there is an effect of human performance concerns on RPA type while operating RPA systems with the same GCS. These results provide additional evidence that human performance requirements need to be closely coupled to the GCS and not necessarily to the aircraft type. Current and future RPA systems should consider and

prioritize the impact of GCS design, policy, and procedures with regard to RPA total system performance.

The unique patterns, or lack thereof, of human performance failures provide evidence supporting the development of GCS standards used in RPA systems. Further exploration and analysis must be accomplished to transition to a more comprehensive understanding of human error-related RPA mishap patterns. By using the analysis in the present research, the USAF may be able to develop effective system design strategies with the objective to reduce the growing cost of these mishaps. The efforts presented in this study have contributed to the understanding of this relatively new realm in aviation history, the RPA.

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

First, I would like to thank my wife, whose love; sacrifice, support, and encouragement are what really made this journey possible. You are a true blessing and the love of my life. I would not be the man I am today without you. The goal of spending time with you is what motivated me to keep putting “one boot in front of the other.”

I would also like to express sincere appreciation for several individuals who guided, mentored, and coached me through the last two years. I would first like to thank my thesis committee, advisor Dr. Michael McCauley, for his exceptional mentorship and support, co-advisor Dr. “Kip” Smith for providing unwavering support, passion, and creativity, as well as Major Chad Seagren, USMC for his analytical instruction and guidance. You endured not one proposal, but two; beverages are on me next time we meet. I would also like to extend my appreciation for the faculty at NPS, your dedication and willingness to guide students through this process matters. I also extend my gratitude to Col. Anthony Tvaryanas, USAF for his willingness to sponsor the dataset and assist with the validation efforts, which was critical for the timely completion of this thesis. Additionally, I would like to thank Col. Charles “Tuna” Carlton, USAF for his support in pursuing this incredible opportunity. Last, and most importantly, I thank my heavenly Father for the gift of His Son and numerous blessings I have been given in this life, it’s more than I deserve.

Isaiah 6:8

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. OVERVIEW

A primary technique for evaluating fielded systems is to analyze historical mishaps. This study is a quantitative analysis of the distribution of human error in six years (2006–2011) of mishap data involving Remotely Piloted Aircraft (RPA) within the United States Air Force (USAF). The archival data on mishaps involving USAF RPA was provided by the Air Force Safety Center (AFSEC) to the student author for this thesis. The data set consists of codes generated by mishap investigation teams using the DoD Human Factors Analysis and Classification System (HFACS). HFACS provides a hierarchical approach to identifying the root cause of mishaps. The archival nature of the data set afforded a quasi-experimental study of two USAF RPA airframes, the MQ-1 Predator and the MQ-9 Reaper. This thesis analyzed these data to identify patterns of human error by airframe type and developed guidance for designing safer systems.

The MQ-1 and MQ-9 are the premier operational RPAs for the USAF and are highly valued operational assets. The aircraft have different flight characteristics but are controlled using the same Ground Control Station (GCS). Is the same GCS appropriate for such different aircraft? The analysis of the human error data sheds light on that issue.

Current and future RPA systems must consider the impact of brittle engineering on the ability of an individual and/or an aircraft to conduct sense making, and ultimately understand the path to returning to dynamic stability. This thesis reviews human performance in such environments, and recommends a solution aimed at proactive mishap prevention. This study explored the potential human error patterns in the USAF MQ-1 and MQ-9 communities, and recommends a solution aimed at proactive mishap prevention. New technologies have been introduced with the intent that they will eliminate known issues, only to find that the potential for new error types has been overlooked, and that new error may be worse than those being eliminated (Hollnagel, Woods, & Leveson, 2006).

The many complex factors that exist within the context of RPA operations are dynamic and interdependent. The fragile tension within the envelope of human performance provides clear boundaries that define constraints that must be met to ensure the safety of flight. As aircraft technology has advanced, RPA have provided increasingly impressive capabilities. With the addition of the MQ-9, different ingredients for human performance threats may have been introduced into the system.

The cost associated with human error-related RPA mishaps is significant. Future system designs need to incorporate the identified patterns of human error as seen in historical mishaps to improve the total system performance of future RPA systems.

B. BACKGROUND

1. The MQ-1 Predator

The Predator RPA system, shown in Figure 1, was designed in response to a DoD requirement to provide the warfighter with persistent intelligence, surveillance, and reconnaissance (ISR) information combined with a kill capability. In April 1996, the Secretary of Defense selected the USAF as the operating service for the RQ-1 Predator system. The “R” is the DoD designation for reconnaissance, and “Q” means Unmanned Aircraft System (UAS). The “1” refers to the aircraft being the first of the series of RPA systems. A change in designation from “RQ-1” to “MQ-1” occurred in 2002. The “M” is the DoD designation for multi-role, reflecting the addition of the capabilities to carry Hellfire missiles and to fire them autonomously. The MQ-1 provides armed ISR capabilities to overseas contingency operations. In August 2011, the MQ-1 passed a major milestone- one million total operating hours, a significant accomplishment for the USAF. The system characteristics are shown in Appendix A (U.S. Air Force, 2012a).



Figure 1. MQ-1 Predator (From U.S. Air Force, 2012)

The MQ-1 Predator is an armed, multi-mission, medium-altitude, long endurance (MALE) RPA that is employed primarily in a killer/scout role as an intelligence collection asset and secondarily against dynamic execution targets. Given its significant loiter time, wide-range sensors, multi-mode communications suite, and precision weapons, it provides the capability to execute the kill chain (find, fix, track, target, engage, and assess) against high-value, fleeting, and time-sensitive targets (TSTs) autonomously. The MQ-1 also can perform the following missions and tasks: ISR, close air support (CAS), combat search and rescue (CSAR), precision strike, buddy-lase, convoy/raid overwatch, route clearance, target development, and terminal air guidance. The MQ-1's capabilities qualify it to conduct irregular warfare operations.

2. The MQ-9 Reaper

The USAF proposed the MQ-9 Reaper system, shown in Figure 2, in response to the DoD direction to support overseas contingency operations. It is larger and more powerful than the MQ-1. It is capable of flying higher, faster, and farther than the MQ-1. Like the MQ-1, it is designed to prosecute time-sensitive targets with persistence and precision, and to destroy or disable those targets. The “9” indicates it is the ninth in the series of remotely piloted aircraft systems. The system characteristics are shown in Appendix A (U.S. Air Force, 2012b).



Figure 2. MQ-9 Reaper (From U.S. Air Force, 2012)

The MQ-9 is an armed, multi-mission MALE RPA that is employed primarily in a hunter/killer role against dynamic execution targets and secondarily as an intelligence collection asset. Given its significant loiter time, wide-range sensors, multi-mode communications suite, and precision weapons, it provides a capability to execute the kill chain (find, fix, track, target, execute, and assess) against high value, fleeting TSTs autonomously.

The MQ-9 also can perform the following missions and tasks: ISR, CAS, CSAR, precision strike, buddy-laser, convoy/raid overwatch, route clearance, target development, and terminal air guidance. The MQ-9's capabilities qualify it to conduct irregular warfare operations.

3. The Ground Control Station

The GCS for both the MQ-1 and the MQ-9 is shown in Figure 3. The GCS is a self-contained operations center that includes seats, computers, keyboards, screens, flight controls, and audio equipment. The two operators in the GCS are the pilot and the sensor operator.



Figure 3. Interior of the GCS for the MQ-1 and MQ-9 (From U.S. Air Force, 2012)

C. OBJECTIVE

RPAs provide a unique challenge to developers of certification standards (e.g., FAA) because the GCS and the aircraft are separate and it is theoretically possible to mix and match GCS' and aircraft. So what matters in terms of human performance: the GCS or the aircraft? This question is a significant point of debate in policy and worthy of analysis.

The dataset provided the opportunity to gain insight into this question as a natural experiment in which the GCS is controlled and the aircraft is varied. The pattern of human performance failures provide evidence supporting the development of aircraft certification standards or GCS standards used in the RPA system.

D. PROBLEM STATEMENT

The use of the same GCS to control both the MQ-1 and MQ-9 aircraft creates an opportunity to explore and identify the human factors issues underlying RPA safety. The study identifies mishap issues that are unique to each aircraft and those that are shared by both. The analysis focuses on characteristics of the aircraft and their missions and on how these factors may define patterns of human performance failures.

E. RESEARCH QUESTION

This research is driven by the need to improve the understanding of human performance patterns in the realm of RPA operations. The specific research question is: Do the types of active failures (unsafe acts) and latent failures (preconditions, unsafe supervision, and organizational influences) differ between the MQ-1 and MQ-9 when operated with the same GCS? The research analyzed the archive of HFACS data to identify human factors issues that are unique to each aircraft and those that are shared by both. It developed logistic regression models to predict aircraft type given the HFACS coding scheme (discussed in Chapter II). The models have the promise to inform RPA certification standards and future system designs.

F. HUMAN SYSTEMS INTEGRATION

Air Force Instruction 63-1201, Life Cycle Systems Engineering, defines Human Systems Integration as a disciplined, unified, and interactive systems engineering approach to integrate human considerations into system development, design, and life cycle management to improve total system performance and reduce costs of ownership.

The major categories or domains of Air Force HSI are:

- Manpower
- Personnel
- Training
- Environment
- Safety
- Occupational Health
- Human Factors Engineering
- Survivability
- Habitability

This section discusses how the research in this thesis impacts four domains of HSI. Several of the HSI domains are involved in any human error mishap. This study impacts Personnel, Human Factors Engineering (HFE), Occupational Health, and Safety within the USAF RPA community.

1. Personnel

Personnel considers the type of human knowledge, skills, abilities, experience levels, and human aptitudes required to operate, maintain, and support a system; and the means to provide such people. Personnel recruitment, testing, qualification, and selection are driven by system requirements (USAF HSI Office, 2009).

The USAF began filling RPA manpower billets with rated fighter pilots during the initial phases of the RPA mission. Most of the initial cadre migrated from the F-16 and F-15 communities. Following an increase in pilot demand, the USAF began selecting RPA pilots from Specialized Undergraduate Pilot Training (SUPT) upon completion of manned flight school. As the demand continued to increase, the USAF developed an independent career field (18X) and a formal training pipeline. The RPA pilot pipeline is shown in Figure 4 (Taranto, 2012). RPA pilots must complete about 140 hours of academics for RPA instrument qualification at Randolph Air Force Base (AFB). Additionally, they must pass seven tests and accomplish 36 missions on T-6 simulators during 48 hours of training. Once they complete instrument qualification, the students move on to the four-week RPA fundamentals course, also at Randolph AFB. They then move to the basic qualifications course at Creech AFB, NV or Holloman AFB, NM. In all, the RPA pilot pipeline takes approximately one year to complete.

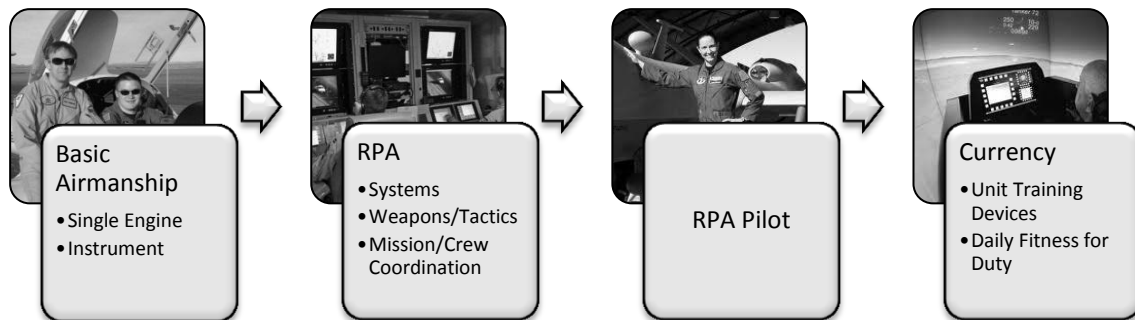


Figure 4. USAF RPA Training (After Taranto, 2012)

2. *Human Factors Engineering*

HFE involves understanding and comprehensive integration of human capabilities (cognitive, physical, sensory, and team dynamic) into system design. A major concern for HFE is creating integration of human-system interfaces to achieve optimal total system performance (USAF HSI Office, 2009).

The evolution of RPA technology and integration into USAF operations has increased military capabilities. As with most new systems, known and unknown trade-offs for both the human and the system occur spanning the entire lifecycle. In the case of RPA operations, human performance boundaries and limitations may have been unintentionally exceeded. Further, the potential for RPA specific mismatches between system design and operator training and capabilities may exist. While these advanced systems are very attractive, inevitable gaps in the system design are likely to exist between work as imagined and work as practiced. Anything that obscures this gap will make it impossible for the organization (or system) to calibrate its understanding or model itself and thereby undermine processes of learning and improvement (Hollnagel, Woods, & Leveson, 2006).

3. *Occupational Health*

Occupational Health promotes system design features that serve to minimize the risk of injury, acute or chronic illness, disability, and enhance the job performance of personnel who operate, maintain, or support the system (USAF HSI Office, 2009).

RPAs provide a unique challenge to developers of certification standards (e.g., FAA) because the cockpit and the aircraft are separate and it is theoretically possible to mix and match GCS' and aircraft. The pattern of human performance failures provide evidence supporting the development of aircraft certification standards or the standards on the GCS used in the RPA system.

4. *Safety*

Safety promotes system design characteristics and procedures to minimize the potential for accidents or mishaps that: cause death or injury to operators, maintainers,

and support personnel; threaten the operation of the system; or cause cascading failures in other systems. Using safety analyses and lessons learned from predecessor systems, the Safety community prompts design features to prevent safety hazards where possible and to manage safety hazards that cannot be avoided. The focus is on designs that have back-up systems, and, where an interface with humans exists, to alert them when problems arise and also to help to avoid and recover from errors (USAF HSI Office, 2009).

G. SCOPE AND LIMITATIONS

Although mishaps related to human error are a systemic problem throughout the DoD, this research focuses on the USAF MQ-1 and MQ-9 human error-related mishaps from 2006-2011.

H. ORGANIZATION

The remainder of this thesis is organized in the following manner: Chapter II describes a review of the applicable literature, while Chapter III outlines the methodological approach of research. Chapter IV describes the results of the researcher's analysis and findings, and Chapter V describes the conclusions and recommendations.

THIS PAGE INTENTIONALLY LEFT BLANK

II. LITERATURE REVIEW

A. OVERVIEW

This chapter provides an overview of human error, accident causation, DoD HFACS, and USAF RPA human error-related mishaps. The literature review consisted of published papers, research reports, and publications written by human factors professionals.

Human error continues to plague military aviation. Analysis provides evidence that human error is identified as a causal factor in 80 to 90 percent of mishaps, and is therefore the single greatest mishap hazard. Further, it is well established that mishaps are rarely attributed to a single cause, or in most instances, even a single individual. The goal of a mishap or event investigation is to identify these failures and conditions in order to understand why the mishap occurred and how it might be prevented from happening again (Webster, White, & Wurmstein, 2005).

The DoD HFACS categorizations of human error have been completed following mishaps, where the outcome is identified and the human operator is assigned the blame (Salmon, Regan, & Johnston, 2005). Rasmussen's view was that if the system performs less satisfactorily because of a human act, then it is likely human error (Rasmussen, 1986). In contrast, Woods (2006) describes the labeling of "human error" as prejudicial. Using "human error" hides much more than it reveals about how a system functions or malfunctions (Woods, Dekker, Cook, Johannesen, & Starter, 2010). This study accepts Reason's definition of an error: a symptom that reveals the presence of latent conditions in the system at large (Reason, 1997).

The word "error" is often vaguely used to describe action or inaction on part of the human. A clear understanding of the definition for the purposes of this study is consistent with that of Reason. Error is split into two main categories: errors and violations. Violations differ in that they are considered intentional acts (Reason, 1990).

B. ACCIDENT CAUSATION THEORIES

The various perceptions of the accident phenomenon are what present day terminology call “accident models.” The genesis of these models was single-factor models, e.g., accident proneness (Greenwood & Woods, 1919). These models developed from simple and complex linear causation models to present-day systemic and functional resonance models.

1. *Simple Linear Accident Model (Domino Model)*

The archetype and most commonly known simple linear model is Heinrich’s (1931) Domino model, which uses linear propagation of a chain of causes and effects to explain accidents (Figure 5). The focus of the Domino model is that accidents are the result of a sequence of events. He viewed the dominos as unsafe conditions or unsafe acts, where their respective removal would prevent a chain reaction from propagating, thus preventing the accident. This model is associated with one of the first attempts at formulating a comprehensive safety theory. This view suggests that accidents are basically disturbances inflicted on an otherwise stable system. While this model has been highly useful by providing a concrete approach to understanding accidents, it has also reinforced the misunderstanding that accidents have a root cause and that this root cause can be identified by simply working backwards from the event through the chain of events that precede it (Hollnagel, Woods, & Leveson, 2006).

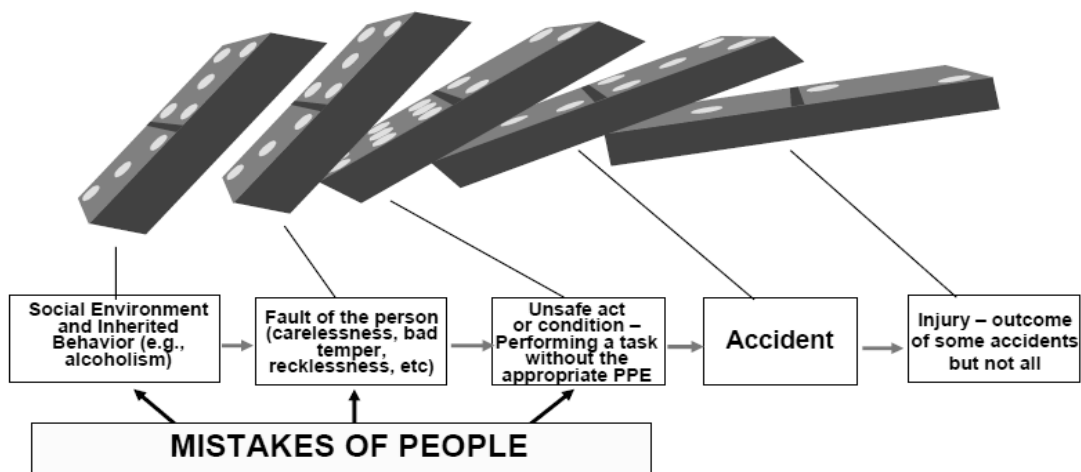


Figure 5. Simple Linear Accident Model (From Hollnagel, Woods, & Leveson, 2006)

2. *Complex Linear Accident Model (Swiss Cheese Model)*

The well-known Swiss Cheese Model (Reason, 1990) is an archetype complex linear accident model. Reason's model focuses on the structure or hierarchy of the organization to illustrate how a mishap or accident can occur. According to this model, accidents can be seen as the result of interrelations between real time "unsafe acts" by the operator and "latent conditions" upstream in the hierarchy. The hierarchical layers of defense are the "cheese". The unsafe acts and latent conditions are the holes in the "cheese" (Figure 6).

The Swiss Cheese Model suggests that a layered defense would not have any holes, forming a blockade that prevents any hazards that may lead to an accident. The breakdown of the conspicuous defenses comprises the components of risk and failures. With this model, causality is not considered a single linear propagation of effects; it is still the result of precipitating events and the failure of a barriers still the failure of an individual component (Hollnagel, Woods, & Leveson, 2006). Complex linear models, such as the Swiss Cheese Model, are designed to describe how coincidences occur, but are bound to a rigid, hierarchic structure that fails to account for dynamic relations between agents, host, barriers and environments. Many accidents defy the explanatory ability of these complex linear models. More sophisticated explanations are required.

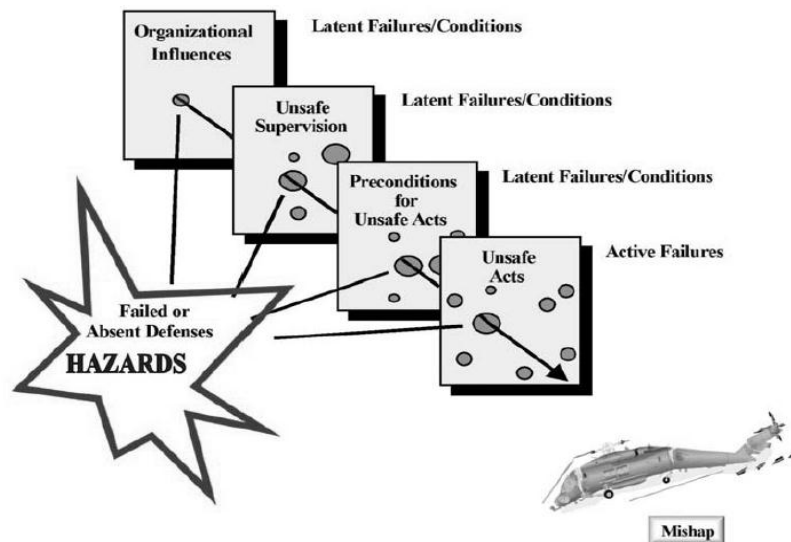


Figure 6. The Swiss Cheese Model (From Webster, White, & Wurmstein, 2005)

3. *Non-Linear or Systemic Models*

Authors, researchers, and investigators have concluded that accidents can be due to an unexpected combination or aggregation of conditions or events known as concurrence. The acknowledgement that two or more events happening at the same time can affect each other has led to the development of non-linear “systemic models.” These models focus on the non-linear phenomena that emerge in a complex system. This perspective admits that variability in system performance is influenced by both constituent subsystems and the operating environment, that is, by both endogenous and exogenous variability, respectively. The systemic model selects a functional point of view where resilience is an organization’s or system’s ability to adequately adjust to destabilizing influences. The strength of resilience comes from the ability to adapt and adjust rather than the power to resist or blockade. A dangerous state may evolve due to system adjustments being inadequate or wrong, rather than due to “human error” or failure. This perspective views failure as the flip side of success, and therefore a normal phenomenon (Hollnagel, Woods, & Leveson, 2006).

C. **DOD HUMAN FACTORS ANALYSIS AND CLASSIFICATION SYSTEM**

A taxonomy called DoD HFACS has been developed and is used to characterize the root causes of mishaps. HFACS draws upon Reason's (1990) Swiss Cheese Model of system failure and Wiegmann and Shappell's (2003) concept of active failures and latent failures/conditions. It describes the four tiers of failures/conditions shown in Figure 7. Wiegmann and Shappell created a taxonomy of codes that define various aspects of human error that may lead to mishaps. These classification codes are termed “nanocodes” (Wiegmann and Shappell, 2003).

As described by Reason (1990), *active failures* are the actions or inactions of operators that are believed to cause the mishap. Traditionally referred to as “error,” they are the last “acts” committed by individuals, often with immediate and devastating consequences. For example, an aviator forgetting to lower the landing gear before touch down will have relatively immediate, and potentially grave, consequences. In contrast, *latent failures or conditions* are errors that exist within the organization or elsewhere in

the supervisory chain of command that affect the sequence of events of a mishap. For example, it is not difficult to understand how tasking crews or teams at the expense of quality crew rest can lead to fatigue and ultimately to errors (active failures) in the cockpit. Viewed from this perspective, the actions of individuals are the end result of a chain of factors originating in other parts (often the upper echelons) of the organization. Unfortunately, these latent failures or conditions may lie dormant or undetected for some period of time prior to their manifestation as a mishap (Webster, White, & Wurmstein, 2005).

DoD HFACS describes four levels at which active failures and latent failures/conditions may occur within complex operations (Figure 7). DoD HFACS is particularly useful in mishap investigation because it forces investigators to address latent failures and conditions within the causal sequence of events. DoD HFACS does not stop at supervision; it also considers *Organizational Influences* that can impact performance at all levels.

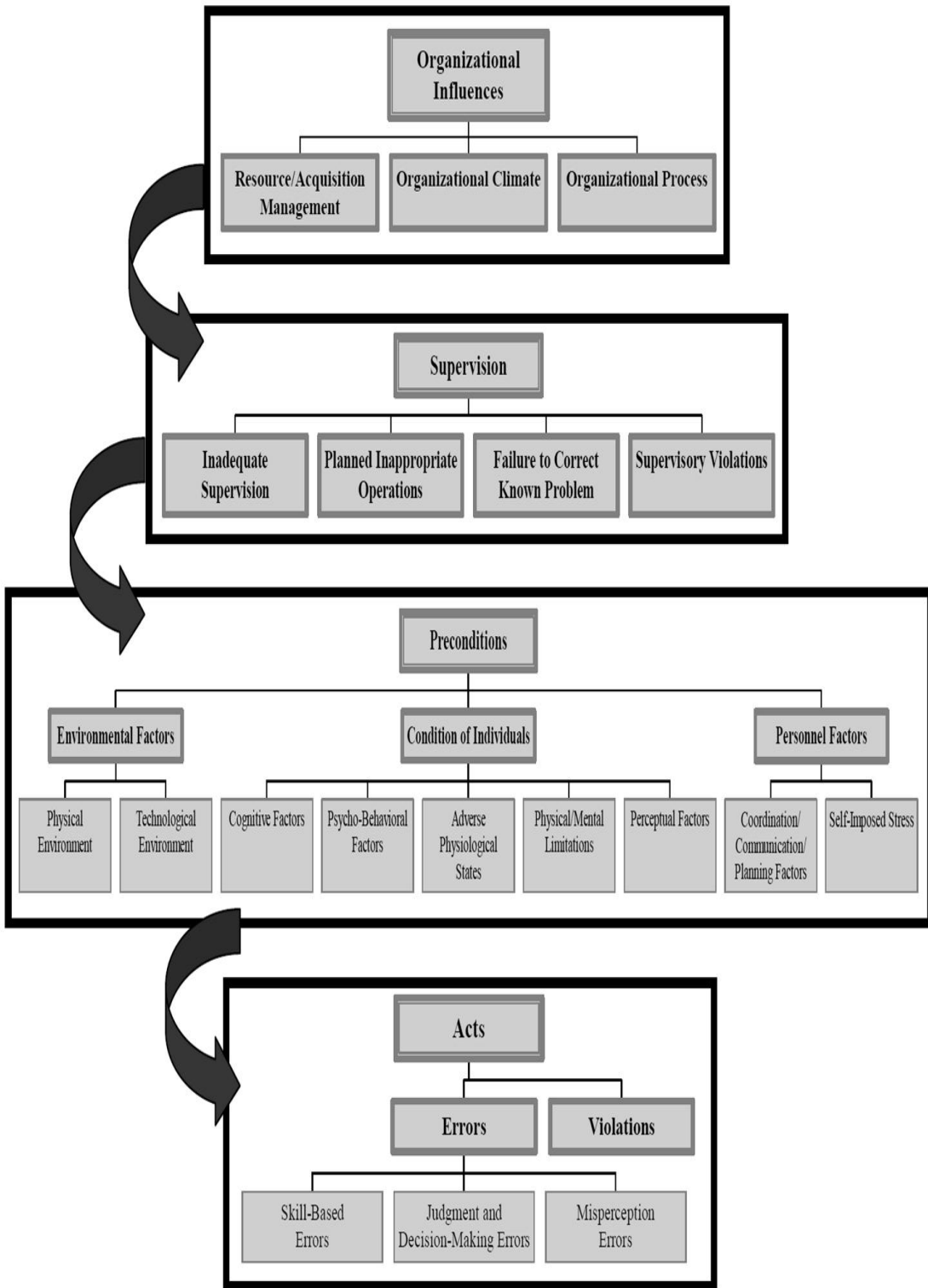


Figure 7. The Four Tiers of DoD HFACS (From Webster, White, & Wurmstein, 2005)

According to AFI 91-204 paragraph A5.1., the USAF requires the use of DoD HFACS as described in this excerpt:

The DoD Instruction directs DoD components to “Establish procedures to provide for the cross-feed of human error data using a common human error categorization system that involves human factors taxonomy accepted among the DoD Components and U.S. Coast Guard.” *All investigators who report and analyze DoD mishaps will use DoD HFACS.* Human Factors is not just about humans. It is about how features of people’s tools, tasks and working environment systemically influence human performance. This model is designed to present a systematic, multidimensional approach to error analysis. (USAF/SEF, 2008)

D. USAF MQ-1 AND MQ-9 MISHAPS

As the effort to demonstrate the viability and effectiveness of RPA systems continues, there is an increasing demand for improved total system performance; specifically reduced mishap rates. The dramatic increase in Combatant Commander’s requests for these mission critical systems during the last decade, in addition to the rapidly growing civilian RPA sector, it is evident these systems are becoming an integral component to our national defense and numerous civil aeronautics sectors. Along with the rapid increase in RPA use, a high mishap rate has followed. The USAF MQ-1 has produced a lifetime mishap rate of 7.58 mishaps per 100,000 flight hours (Figure 8) and the USAF MQ-9 is currently at 4.58 per 100,000 flight hours (Figure 9). While these rates have been reduced significantly in the last several years, there is still room for improved performance. The USAF fighter aircraft rate is typically between one and two mishaps per 100,000 flight hours and general aviation boasts a rate of only 1 mishap per 100,000 flight hours.

MQ-1 RPA MISHAP HISTORY								
YEAR	CLASS A		CLASS B		DESTROY		HOURS	Cum HOURS
	#	RATE	#	RATE	A/C	RATE		
FY97	3	112.91	0	0.00	3	112.91	2657	2657
FY98	0	0.00	0	0.00	0	0.00	3258	5915
FY99	2	38.95	0	0.00	2	38.95	5135	11050
FY00	1	15.56	1	15.56	1	15.56	6426	17476
FY01	4	52.83	1	13.21	4	52.83	7571	25047
FY02	7	36.25	0	0.00	6	31.07	19313	44360
FY03	2	9.75	0	0.00	2	9.75	20507	64867
FY04	6	19.12	0	0.00	5	15.93	31383	96250
FY05	10	24.38	1	2.44	9	21.94	41024	137274
FY06	5	8.65	0	0.00	3	5.19	57798	195072
FY07	7	8.84	0	0.00	5	6.31	79193	274265
FY08	10	6.76	3	2.03	9	6.08	147980	422245
FY09	13	6.94	4	2.13	10	5.34	187393	609638
FY10	7	3.46	3	1.48	6	2.97	202330	811968
FY11	11	4.60	5	2.09	10	4.18	239304	1051272
FY12	8	3.71	3	1.39	8	3.71	215560	1266832
5 YR AVG	9.8	4.94	3.6	1.81	8.6	4.33	198513	
10 YR AVG	7.9	6.46	1.9	1.55	6.7	5.48	122247	
LIFETIME	96	7.58	21	1.66	83	6.55	1266832	

Figure 8. MQ-1 Mishap History (From USAF, 2012).

MQ-9 RPA MISHAP HISTORY								
YEAR	CLASS A		CLASS B		DESTROY		HOURS	Cum HOURS
	#	RATE	#	RATE	A/C	RATE		
FY01	0	0.00	0	0.00	0	0.00	30	30
FY02	0	0.00	0	0.00	0	0.00	191	221
FY03	0	0.00	0	0.00	0	0.00	100	321
FY04	0	0.00	0	0.00	0	0.00	767	1088
FY05	0	0.00	0	0.00	0	0.00	2373	3461
FY06	2	62.89	0	0.00	0	0.00	3180	6641
FY07	1	14.55	0	0.00	0	0.00	6872	13513
FY08	3	22.24	0	0.00	0	0.00	13490	27003
FY09	4	15.75	0	0.00	1	3.94	25391	52394
FY10	1	1.78	0	0.00	1	1.78	56109	108503
FY11	1	1.16	3	3.47	0	0.00	86526	195029
FY12	3	2.54	1	0.85	2	1.69	118039	313068
5 YR AVG	2.4	4.01	0.8	1.34	0.8	1.34	59911	
10 YR AVG	1.5	4.79	0.4	1.28	0.4	1.28	31285	
LIFETIME	15	4.79	4	1.28	4	1.28	313068	

Figure 9. MQ-9 Mishap History (From USAF, 2012).

Results from a recent study including 221 DoD RPA mishaps spanning a 10-year period found that 79 percent of USAF RPA mishaps were human error-related (Tvaryanas, Thompson, & Constable, 2006). The DoD demonstrated a human error rate of 60 percent in the same study. Air Force Col. Anthony Tvaryanas stated that “If you

really wanted to make a dent in preventing RPA accidents, the DoD needs to look at how they do RPA systems acquisition” (Defense Daily, 2005). He suggests that the human error problem in the RPA community originates before the systems take off on the first mission. He also suggests that the decisions made early on in RPA development likely played a crucial role in the mishap rates. Fielding systems without fully developed requirements; incomplete testing; and buying cheaper components all contribute to the higher mishap rates. In a rush to field RPAs, the services failed to adequately weigh the Human Systems Integration issues that affect RPA total system performance (Defense Daily, 2005).

The unprecedented success with regard to the absence of physical human injury associated with RPA operations is a positive outcome of the system. The current and foreseeable DoD fiscal climate suggests that there are still significant reasons for concern. According to two reports by the Office of the Secretary of Defense (OSD), “the reliability and sustainability of RPAs is vitally important because it underlines their affordability (and acquisition concern), their mission availability (an operations and logistics concern), and their acceptance into civil airspace (an FAA regulatory concern)”. Additionally, a Defense Scientific Advisory Board effort on RPAs issued in February 2004 identified “high mishap rates” as one of the largest threats to RPA potential (as cited in Tyvaryanas, 2006).

E. USAF SAFETY INVESTIGATIONS

A mishap is an unplanned occurrence or series of occurrences that results in damage or injury and meets Class A, B, C, or D mishap reporting criteria as defined by Air Force Instruction (AFI) 91-204 and Air Force Manual (AFMAN) 91-223. All mishaps require a safety investigation and report. The USAF conducts safety investigations for all reportable aircraft events to prevent future mishaps. These reports take priority over any corresponding legal investigations.

The Air Force categorizes mishaps based upon the material involved (e.g., space systems, weapons, aircraft, motor vehicles, person, etc.) and the state of the involved material (e.g., launch, orbit, existence of intent for flight, on- or off-duty, etc.) when the mishap occurs (USAF/SEF, 2008).

1. Mishap Categories

Aircraft Flight: Any mishap in which there is intent for flight and reportable damage to a DoD aircraft. As shown in Table 1, USAF uses thresholds measured in dollars to define three categories of mishaps. The dollar amounts increased in FY2010 (USAF/SEF, 2008). The data set provided by the Air Force Safety Automated System (AFSAS) records the category for each mishap.

Table 1. USAF Mishap Categories

Mishap Type		
Class A	Class B	Class C
Direct mishap cost totaling \$1,000,000 or more	Direct mishap cost totaling \$200,000 or more but less than \$1,000,000	Direct mishap cost totaling \$20,000 or more but less than \$200,000
A fatality or permanent total disability	A permanent partial disability	Any injury or occupational illness or disease that causes loss of one or more days away from work beyond the day or shift it occurred
	Inpatient hospitalization of three or more personnel	

*NOTE: The dollar amounts changed in FY2010 to \$2,000,000 (Class A), \$500,000–\$1,000,000 (Class B), \$50,000–\$500,000 (Class C).

2. AFSAS

AFSAS is a web-based program that provides a mishap reporting capability for all safety disciplines throughout the U.S. Air Force. This system provides a reporting, analysis and trending capability and maintains a comprehensive Air Force safety database. This database enables the AFSEC to respond rapidly to both internal and external customer requests for mishap and safety data (Air Force Safety Center, 2012).

Mishap reporting requires a written narrative be included in the final report and uploaded into AFSAS. The narrative provides important qualitative and quantitative information from which a majority of the DoD HFACS coding can be mapped. The author validated that the mapping accuracy of the reported HFACS codes to their mishap narratives by selecting a random subset of the reports and applying individual expert evaluation by coding each mishap and comparing the results using a Cohen’s Kappa.

III. RESEARCH METHOD

A. APPROACH

This study protocol was reviewed and approved by the 711 HPW/IR (AFRL IRB) in accordance with 32 CFR 219, DoDD 3216.2, and AFI-40-402 and by the Naval Postgraduate School Operations Research Department thesis approval process. The IRB determined that this study was exempt and considered not to be Human Subjects Research. The study design was a quantitative analysis of DoD HFACS nanocodes for six years of RPA mishap data. The inclusion criteria for this study were USAF MQ-1 and MQ-9 mishaps occurring during fiscal years 2006-2011 that resulted in more than \$20,000 in damage. The data were retrieved from AFSAS under a formal request from the 711th Human Performance Wing (HPW) for the purpose of this research. The author was granted an AFSAS account for the purposes of validating all HFACS nanocodes assigned by the investigators. This effort was assisted by Col. Anthony Tvaryanas of the 711th HPW to ensure a balanced and non-biased validation pursuant to DoD HFACS instructions. Additional information in the dataset include relevant parameters such as Phase of Flight, Mishap Domain (Logistics/Maintenance, Miscellaneous, and Operations), and Mishap Class (A, B, and C) by airframe and year. Some of the mishaps were determined not to be human error-related. In total, 88 mishaps were extracted for analysis.

B. DATABASE AND ACCIDENT CODING

1. HFACS Coding

The raw data were produced and validated by three separate raters; all USAF officers (the assigned investigator, an aerospace medicine specialist, and an aerospace physiologist) who analyzed each mishap independently and classified each human causal factor using the DoD HFACS associated nanocodes. The investigator was likely different for each event. Following the coding, inter-rater reliability was calculated using Cohen's Kappa. During the validation effort, databases were constructed using Excel and statistical software package JMP Pro10. Cohen's Kappa, Chi-square, and binary logistic

regression tests were conducted to identify significant human error patterns. The nanocodes were the predictor variables in the logistic regression analyses and aircraft type, MQ-1 or MQ-9, were the binary response variable. A stepwise comparison was executed on the nanocodes and covariates to identify statistically significant variables for each RPA type and thus constructed models for predicting mishap RPA type.

The DoD HFACS were applied at the nanocode level during the investigation and validation phases. Due to historically poor inter-rater reliability at the nanocode level (Level III), the DoD HFACS nanocodes were considered at the top two levels (Level I and II). Table 2 illustrates the organization of HFACS at these levels. Level I is divided into Acts, Preconditions, Supervision, and Organizational Influences. Level II groups the Level III nanocodes into 20 different Level II subcategories.

Table 2. DoD HFACS Grouping

Level I		Level II	
Categories	Code	Subcategories	Code
Acts	A	Skill-Based Errors	AE1
		Judgment and Decision-Making Errors	AE2
		Perception Errors	AE3
		Violations	AV
Preconditions	P	Physical Environment	PE1
		Technological Environment	PE2
		Cognitive Factors	PC1
		Psycho-Behavioral Factors	PC2
		Adverse Physiological State	PC3
		Physical/Mental Limitations	PC4
		Perceptual Factors	PC5
		Coordination/Communication/Planning Factors	PP1
Supervision	S	Inadequate Supervision	SI
		Failure to Correct Known Problem	SF
		Planned Inappropriate Operation	SP
		Supervisory Violations	SV
Organization	O	Resource/Acquisition Management	OR
		Organizational Climate	OC
		Organizational Process	OP
Not Applicable	N/A	Not Applicable	N/A

C. DATA ANALYSIS

1. *DOD HFACS Category Frequency*

The frequency of occurrence for the DoD HFACS categories was evaluated for each of the mishaps within the dataset. The presence of a DoD HFACS nanocode was annotated with a one (1) and the absence of a nanocode was annotated with a zero (0). No code was used more than once in any mishap. The codes were used to determine how often the categories were used in the mishap dataset. The resulting database was analyzed using a Cohen's Kappa to determine inter-rater reliability and was the foundation for the logistic regression to construct the models.

2. *Inter-Rater Reliability*

A Cohen's Kappa analysis and evaluation was conducted to quantify inter-rater reliability among the three raters. The Kappa coefficient is noted as the preferred statistical measurement for determining agreement or disagreement between raters (Ubersax, 1987). It enables identification of statistically significant disagreements between any of the raters within the dataset. Cohen's Kappa was utilized to measure the proportion of agreement versus alignment by chance between each of the three different pairs of raters.

A value of +1.0 indicates 100 percent agreement between the two raters. A kappa value of 0 means there is not a relationship between the two raters, while a kappa of -1.0 is considered to be a 100 percent disagreement. Additional interpretations of the values were defined as follows (Curdy, 2009):

- between 0.8 and 1 is considered Very Good
- between 0.6 and 0.8 is considered Good
- between 0.4 and 0.6 is considered Moderate Agreement
- between 0.2 and 0.4 is considered Fair Agreement
- between 0 and 0.2 is considered Slight Agreement

3. *Human Error Pattern Analysis*

Following the validation efforts, a pattern analysis was conducted to identify the most prevalent causal factors. Logistic regression and the chi-square test were applied to the data to examine the hypothesis among the MQ-1 and MQ-9 mishaps at Level I and II in the DoD HFACS hierarchy. The logistic regression was applied at both levels of dichotomous coded variables (HFACS nanocodes). From this prospective, the response variable can be considered to have a probability between zero and one. The data consist of individual records (mishap nanocodes) that were classified as a success or failure (1 or 0). All nominal covariates with k levels were coded using $k-1$ dummy variables.

In searching for potentially important covariates, a univariate regression model for each nanocode and covariate was created. Those with p-values less than 0.25 were deemed close enough to be included in subsequent iterations. Those with p-values greater than 0.25 are unlikely to be important and may be safely discarded. Chi-square analysis was conducted at each level followed by a full logistic regression analysis. A stepwise regression was conducted in an effort to fit and select a feasible model. The Odds Ratios were calculated to measure the effect size and to describe the strength of association between the data. Model validation was completed using the Receiver Operating Characteristic (ROC) curve to show the tradeoff between successfully identifying True Positive values and mistakenly identifying False Positives. Cross-Validation was performed to assess how well the model classifies records outside of the data. This process provides a sense of the fit of the model and was executed by assigning training and test sets from the data.

IV. RESULTS

A. ACCIDENT DATABASE

The initial dataset contained a total of 149 USAF Class A, B, and C MQ-1 and MQ-9 mishap reports from fiscal years 2006-2011. Of the 149 reports, DoD HFACS was applied to 88 (59.1 percent) and were events considered to be related to human error suitable for inclusion in the study. The remaining 61 mishap reports were verified to be events that were not related to human error. Table 3 presents the distribution of mishaps by RPA type and human factors applicability with associated rates. The percentage of human error mishaps was not statistically different across RPA type ($\chi^2_{(1)} = 0.021$, $p = 0.886$).

Table 3. RPA-Human Error Mishap Distribution

Total Mishaps		Human Error Mishaps	Rate
MQ1	118	69	58.5%
MQ9	31	19	61.3%
Total	149	88	59.1%

A total of 573 DOD HFACS nanocodes were cited by the mishap investigators in the 88 mishaps. The number of mishap reports by RPA type and respective HFACS codes are listed in Table 4. The MQ-1 and MQ-9 averaged 6.4 and 6.9 nanocodes per mishap respectively. The number of nanocodes per mishap is not statistically different across RPA type ($\chi^2_{(1)} = 0.764$, $p = 0.090$).

Table 4. Nanocodes Cited by RPA Type

Total Mishaps		Nanocodes Cited	Nanocodes per Mishap
MQ1	69	441	6.4
MQ9	19	132	6.9
Total	88	573	6.5

The dataset was categorized by USAF mishap classification (Class A, B, and C) and is presented in Table 5. The MQ-1 breakdown showed there were 46 (66.7 percent) Class A mishaps, 11 (15.9 percent) Class B mishaps, and 12 (17.4 percent) Class C mishaps. The MQ-9 breakdown showed there were nine (47.4 percent) Class A mishaps, five (26.3 percent) Class B mishaps, and five (26.3 percent) Class C mishaps. The distribution of mishaps across class is not statistically different ($\chi^2_{(2)} = 2.384$, $p = 0.304$). In the logistic regression analysis, this polychotomous variable was coded with two dummy variables that assigned Class C as the baseline. All FY 2010 and 2011 mishaps were evaluated for actual cost and were categorized as defined by pre-FY 2010 dollar amounts as listed in Table 1 to standardize the data. In total, five Class C mishaps were re-categorized as Class B mishaps and five Class B mishaps were re-categorized as Class A mishaps for the purpose of data standardization.

Table 5. Mishaps by Class

RPA	Mishap Class	Number of Mishaps	Rate
MQ-1	A	46	66.7%
	B	11	15.9%
	C	12	17.4%
MQ-9	A	9	47.4%
	B	5	26.3%
	C	5	26.3%

Additionally, the distribution of mishaps by nanocodes was analyzed and found to be statistically different across RPA type ($\chi^2_{(2)} = 11.144$, $p = 0.0038$), as shown in Table 6. The greatest departures from the expected distribution were the number of observed MQ-9 nanocodes used in Class B and Class C mishaps.

Table 6. Mishaps by Class and Nanocode

RPA	Mishap Class	Observed Nanocodes	Expected Nanocodes
MQ-1	A	298	302
	B	105	94
	C	38	45
MQ-9	A	94	90
	B	17	28
	C	21	14

The dataset also was organized by Mishap Domain (Operations, Logistics/Maintenance, and Miscellaneous) in Table 7. The MQ-1 mishaps were identified as 31 (44.9 percent) Operations, 33 (47.8 percent) Logistics/Maintenance, and 5 (7.2 percent) Miscellaneous. The MQ-9 mishaps were identified as 17 (89.5 percent) Operations, 2 (10.5 percent) Logistics/Maintenance, and 0 (0 percent) Miscellaneous. For both RPA types, the highest use of nanocodes was in the Operations domain with an average of 8.5 and 7.6 codes cited per mishap for the MQ-1 and MQ-9 respectively. The distribution of mishaps was analyzed and found to be statistically different across the two RPA types ($\chi^2_{(2)} = 14.708$, $p = 0.001$). In the logistic regression section, this polychotomous variable was coded with two dummy variables that used Operations as the baseline.

Table 7. Mishaps by Domain

RPA	Mishap Domain	Observed Mishaps	Expected Mishaps
MQ-1	Operations	31	38
	Logistics/Maintenance	5	5
	Misc	33	26
MQ-9	Operations	17	10
	Logistics/Maintenance	2	2
	Misc	0	7

The dataset was further organized by mishap phase of flight (ground operations, take off, climb, enroute, landing, and other) as shown in Table 8. The MQ-1 mishaps were concentrated as 37 (53.6 percent) enroute and 24 (34.8 percent), landing. The MQ-9

mishap phases of flight were concentrated as 13 (68.4 percent) landing and 2 (10.5 percent) enroute. The distribution of mishaps was analyzed and found to be statistically different across RPA type ($\chi^2_{(5)} = 18.607$, $p = 0.002$). The greatest departures from the expected were during the enroute and landing phases.

Table 8. Mishaps by Phase of Flight

RPA	Mishap Phase of Flight	Observed Mishaps	Expected Mishaps
MQ-1	Ground Ops	1	3.1
	Takeoff	2	2.4
	Climb	4	3.1
	Enroute	37	30.6
	Landing	24	29.0
	Other	1	0.8
MQ-9	Ground Ops	3	0.9
	Takeoff	1	0.6
	Climb	0	0.9
	Enroute	2	8.4
	Landing	13	8.0
	Other	0	0.2

The mishap Phase of Flight was examined for statistical differences between RPA types (Table 9). The distribution of mishaps was found to be statistically different across Phase of Flight ($\chi^2_{(5)} = 89.298$, $p = 0.000$). Significant differences from a uniform distribution exist for every phase. In the logistic regression section, this polychotomous variable was coded with five dummy variables that used Landing as the baseline.

Table 9. Mishaps by Phase of Flight for Both RPA Types

RPA	Mishap Phase of Flight	Observed Mishaps	Expected Mishaps
Both	Ground Ops	4.0	22.0
	Takeoff	3.0	22.0
	Climb	4.0	22.0
	Enroute	39.0	22.0
	Landing	37.0	22.0
	Other	1.0	22.0

B. INTER-RATER RELIABILITY

A sample of 12 mishaps from the 88 in the database was randomly selected for validation and assessed for inter-rater reliability. The aerospace medicine specialist (Rater 2) and aerospace physiologist (Rater 3) conducted independent validations of the sample by reading each mishap report and coding each event adhering to the procedures specified by Webster, White, & Wurmstein (2005). The independent coding data from the random sample are located in Appendix B.

1. HFACS Level I

At Level I there was strong agreement across the categories Acts, Preconditions and Organization. There was partial agreement in the Supervision category. The primary locus of divergence was between Rater 1 (the original accident investigator) and Raters 2 and 3 (Figure 10).

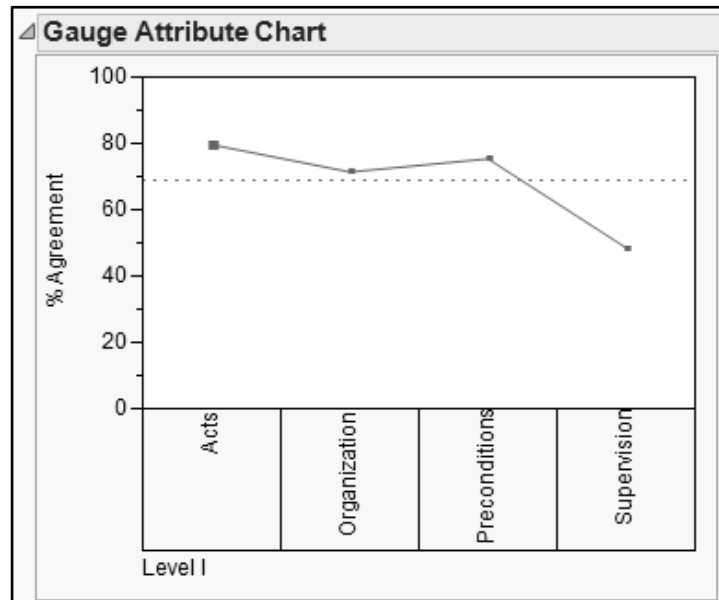


Figure 10. Level I Inter-Rater Gauge Attribute Chart

Cohen's Kappa was calculated for each pair of raters at Level I and is presented in Figure 11. The Kappa for Raters 2 and 3 indicates very good agreement (Cohen's Kappa = 1.00) at Level I. The level of agreement between Rater 1 and Raters 2 and 3 is only

moderate (Cohen’s Kappa = 0.53). This lower agreement is likely attributable to Rater 1 being a different individual during each mishap investigation.

Agreement Comparisons							
Compared		Kappa					Standard Error
Rater	with Rater		.2	.4	.6	.8	
Rater 1	Rater 2	0.5345					0.1238
Rater 1	Rater 3	0.5345					0.1238
Rater 2	Rater 3	1.0000					0.0000

Figure 11. Level I Inter-Rater Reliability Kappa Coefficients

2. HFACS Level II

At Level II there was improved percent agreement across the categories. There was little divergence across raters in any category (Figure 12).

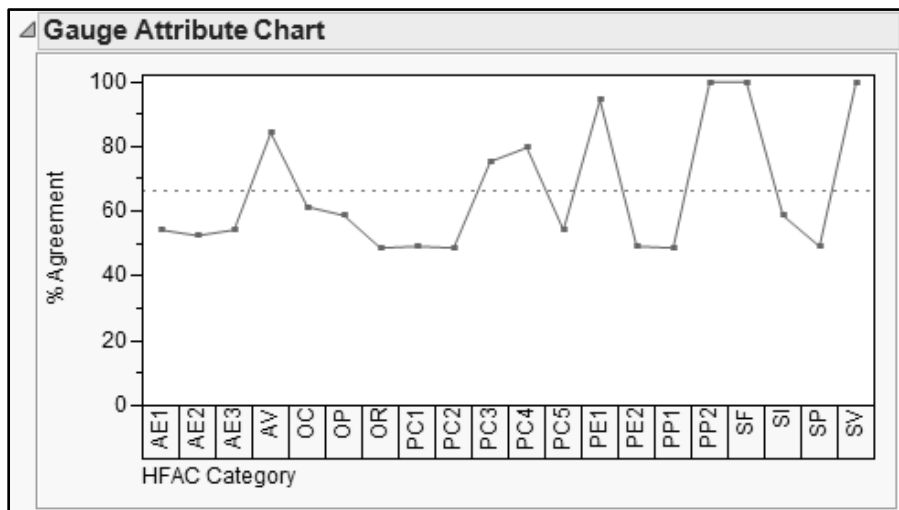


Figure 12. Level II Inter-Rater Gauge Attribute Chart

Cohen’s Kappa was calculated for each pair of raters at Level II and is presented in Figure 13. On average, Level II agreement was stronger than at Level I. Agreement likely improved due to the larger data table used to calculate the Kappa coefficient and some divergence between Raters 2 and 3 at Level II. Agreement between Raters 1 and 2 is considered Good (Cohen’s Kappa = 0.67). Agreement between Raters 1 and 3 is

considered Moderate (Cohen’s Kappa = 0.59). Agreement between Raters 2 and 3 is considered Very Good (Cohen’s Kappa = 0.84).

Agreement Comparisons							
Compared		Kappa					Standard Error
Rater	with Rater		.2	.4	.6	.8	
Rater 1	Rater 2	0.6739					0.0510
Rater 1	Rater 3	0.5859					0.0560
Rater 2	Rater 3	0.8430					0.0367

Figure 13. Level II Inter-Rater Reliability Kappa Coefficients

In sum, there was Moderate to Very Good agreement between the raters for the sample dataset of 12 mishaps. The Moderate to Good agreement between Rater 1 and Raters 2 and 3 provided sufficient evidence to support validation of the dataset. The remaining mishaps were therefore assumed to have been coded correctly by the accident investigators (Rater 1). As a result, the analysis of the data includes all 88 mishaps.

C. HUMAN ERROR PATTERN ANALYSIS

1. HFACS Level I Analysis

Organization (76.8 percent) was cited more often than any other category and Supervision (37.7 percent) was cited the least with regard to the MQ-1 (Figure 14). Acts (84.2 percent) were cited more often than any other category and Supervision (47.4 percent) was cited the least with regard to the MQ-9 (Figure 14). The null hypothesis states that the RPA type is equally likely to be cited as Acts (A), Preconditions (P), Supervision (S), or Organization (O).

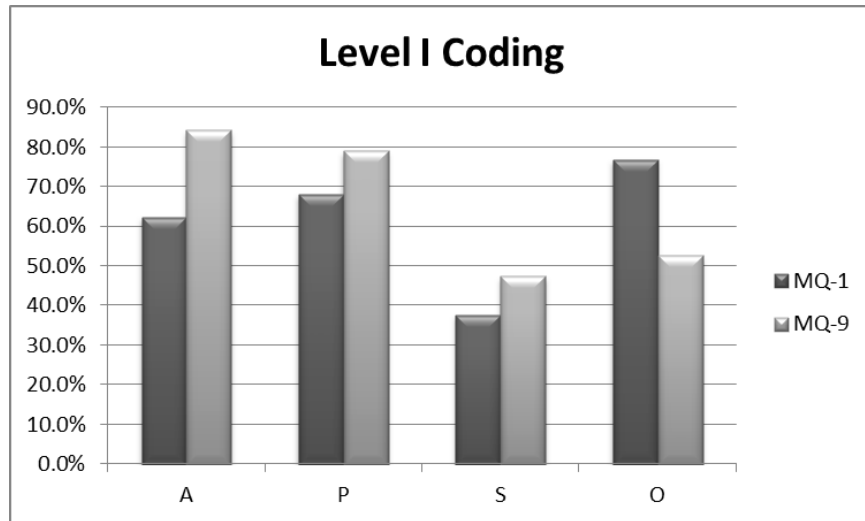


Figure 14. Level I HFACS Coding by RPA

The mishap events are organized by Level I categories and are presented in Table 10. The distribution of mishaps across category was not found to be statistically different ($\chi^2_{(3)} = 2.581, p = 0.461$).

Table 10. Level I Citing Frequency by RPA

RPA	Level I	Observed Citing Frequency	Expected Citing Frequency
MQ-1	Acts	43.0	45.5
	Preconditions	47.0	47.8
	Supervision	26.0	27.0
	Organization	53.0	48.6
MQ-9	Acts	16.0	13.5
	Preconditions	15.0	14.1
	Supervision	9.0	8.0
	Organization	10.0	14.4

The mishap Level I categories were examined for statistical differences between both RPA types (Table 11). The distribution of mishaps was analyzed and found to be statistically different across DoD HFACS Level I categories ($\chi^2_{(3)} = 9.633, p = 0.022$). The number of observed mishaps that were cited as Supervision appears to differ from the expected values for that category.

Table 11. Level I Citing Frequency (Both)

RPA	Level I	Observed Citing Frequency	Expected Citing Frequency
Both	Acts	59.0	54.7
	Preconditions	62.0	54.7
	Supervision	35.0	54.7
	Organization	63.0	54.7

2. *HFACS Level II Analysis*

With regard to the MQ-1, Organizational Processes (60.9 percent) was cited more often than any other category and Violations (1.4 percent) and Physical Environment (1.4 percent) were cited the least (Figure 15). With regard to the MQ-9, Skill Based Errors (63.2 percent) and Cognitive Factors (63.2 percent) were cited more often than any other category and Violations (0.0 percent), Physical Environment (0.0 percent), and Self-Imposed Stress (0.0 percent) were cited the least with regard to the MQ-9 (Figure 15). The null hypothesis states that the RPA type is equally likely to be cited across the 20 categories associated with Level II.

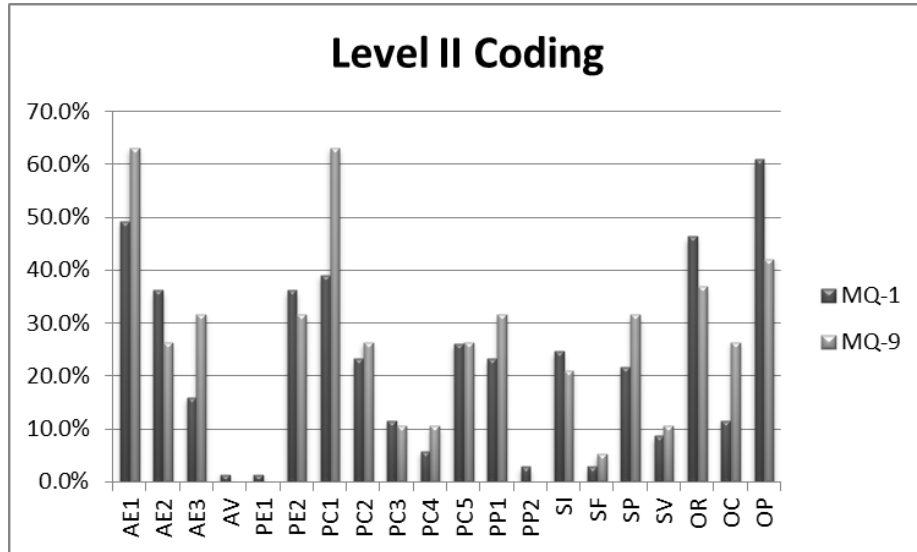


Figure 15. Level II HFACS Coding by RPA

The mishap events are organized by Level II categories and are presented in Table 12. The distribution of mishaps across category was not found to be statistically different ($\chi^2_{(19)} = 10.156, p = 0.949$).

Table 12. Level II Citing Frequency by RPA

RPA	Level II	Observed Citing Frequency	Expected Citing Frequency	RPA	Level II	Observed Citing Frequency	Expected Citing Frequency
MQ-1	AE1	34	35	MQ-9	AE1	12	11
	AE2	25	23		AE2	5	7
	AE3	11	13		AE3	6	4
	AV	1	1		AV	0	0
	PE1	1	1		PE1	0	0
	PE2	25	24		PE2	6	7
	PC1	27	30		PC1	12	9
	PC2	16	16		PC2	5	5
	PC3	8	8		PC3	2	2
	PC4	4	5		PC4	2	1
	PC5	18	18		PC5	5	5
	PP1	16	17		PP1	6	5
	PP2	2	2		PP2	0	0
	SI	17	16		SI	4	5
	SF	2	2		SF	1	1
	SP	15	16		SP	6	5
	SV	6	6		SV	2	2
OR	32	30	OR	7	9		
OC	8	10	OC	5	3		
OP	42	38	OP	8	12		

The mishap Level II categories were examined for statistical differences between RPA types (Table 13). The distribution of mishaps was analyzed and found to be statistically different across DoD HFACS Level II categories ($\chi^2_{(19)} = 216.198, p = 0.000$). Significant differences appear to exist between many of the counts of observed mishaps that were cited and the expected uniform frequency.

Table 13. Level II Citing Frequency (Both)

RPA	Level II	Observed Citing Frequency	Expected Citing Frequency
Both	AE1	46	20.2
	AE2	30	20.2
	AE3	17	20.2
	AV	1	20.2
	PE1	1	20.2
	PE2	31	20.2
	PC1	39	20.2
	PC2	21	20.2
	PC3	10	20.2
	PC4	6	20.2
	PC5	23	20.2
	PP1	22	20.2
	PP2	2	20.2
	SI	21	20.2
	SF	3	20.2
	SP	21	20.2
	SV	8	20.2
	OR	39	20.2
OC	13	20.2	
OP	50	20.2	

D. LOGISTIC REGRESSION ANALYSIS

In searching for potentially important covariates within the mishap reports, a univariate regression model for each category (Level I and II), Class, Domain, and Phase was completed. Factors with p-values less than 0.25 were deemed sufficiently significant to be included in subsequent iterations. Those with p-values greater than 0.25 are unlikely to be statistically significant in the subsequent logistic analysis and were safely discarded. The one exception was AE1 at Level II ($p = 0.28$), which was included due to its proximity to the 0.25 threshold.

The logistic regression was applied at Levels I and II using dichotomously coded predictor variables (0 if absent, 1 if present) for the applicable category/nanocode at each level. Additional variables included in the analysis were Mishap Class, Mishap Domain, and Mishap Phase of Flight. Predictors with k levels were coded using $k-1$ dummy variables. The predicted response varies between zero and one from this perspective.

A stepwise regression was conducted in an effort to fit and select a feasible model. The minimum Akaike Information Criterion (AIC) was set as the stopping rule at both levels (Seagren C, Naval Postgraduate School. Personal communication, 2013). The Odds Ratios were calculated to measure the effect size and to describe the strength of association between the data. Model validation was completed using the ROC curve to show the tradeoff between successfully identifying True Positive values and mistakenly identifying False Positives. Cross-Validation was performed to assess how well the model classifies records outside of the data. Cross validation was executed by assigning training and test sets from the data. Due to the small size of the dataset, the model was cross validated twice at Level II to ensure that valid results were obtained for the dataset. The model was fit once with the test set excluded and once with the test set included. This two stage validation process provides a sense of the fit of the model. Contingency table analysis was conducted at each level to assess the misclassification rate.

The categories AV, PE1, and PP2 were removed from the Level II logistic regression analysis due to the unstable nature of the small sample sizes.

All logistic analysis tables and figures were built in JMP Pro10 statistical software. Cohen's Kappa, Chi-square, and binary logistic regression tests were used to identify human error patterns at Level I and II. The Nanocodes, Domain, and Phase were the predictor variables in the logistic regression analyses. Aircraft type, MQ-1 or MQ-9, was the binary response variable (MQ-1 = 1 and MQ-9 = 0).

1. Covariate Analysis

Three covariates - Class, Domain, and Phase - were analyzed for statistical differences using the chi-square test. A covariate analysis between the three covariates and both RPA types is summarized in Table 14. In evaluating the Mishap Class across both RPA types, the chi-square test resulted in a failure to reject H_0 ($p = 0.313$) and was therefore safely discarded. In consideration of Mishap Domain (Logistics/Maintenance = 1, Miscellaneous = 2, Operations = 3) across both RPA types, the logistic regression analysis coding of the dataset resulted in sufficient evidence to reject H_0 ($p = 0.001$). Operations was selected as the baseline variable in the analysis. The Mishap Phase of

Flight (Ground Operations = 1, Takeoff = 2, Climb = 3, Enroute = 4, Landing = 6, Other = 7) across both RPA types was assessed, the logistic regression analysis coding of the dataset resulted in sufficient evidence to reject H_0 ($p = 0.001$). Landing was selected as the baseline variable. (Note: There were no mishaps coded 5 in the study dataset).

Table 14. Summary of Covariate Chi-Square Tests

Covariate	χ^2	df	<i>p</i>-value
Domain	14.085	2	0.001
Phase	19.748	5	0.001
Class	2.322	2	0.313

2. *Level I Analysis*

All four categories of Level I data were tested for homogeneity with regard to RPA type (Table 15). Acts and Organization met the defined threshold (p -values less than 0.25), $p = 0.059$ and $p = 0.045$, respectively, while Preconditions and Supervision were discarded from the logistic analysis because their p -values exceed the threshold.

Table 15. Summary of Level I Mishap Distribution

Level I Category	χ^2	df	<i>p</i>-value
Acts	3.562	1	0.059
Preconditions	0.882	1	0.348
Supervision	0.577	1	0.448
Organization	4.013	1	0.045

Stepwise logistic regression was run with the four Level I factors found to be statistically significant: Acts, Organization, Domain, and Phase. The baselines for Domain and Phase were Operations and Landing respectively. The stopping rule for this fit was defined by minimum AIC. The model identifies (Figure 16) Domain (Logistics/Maintenance), Phase (Ground Operations), and Phase (Enroute) as parameters to include in the logistic model.

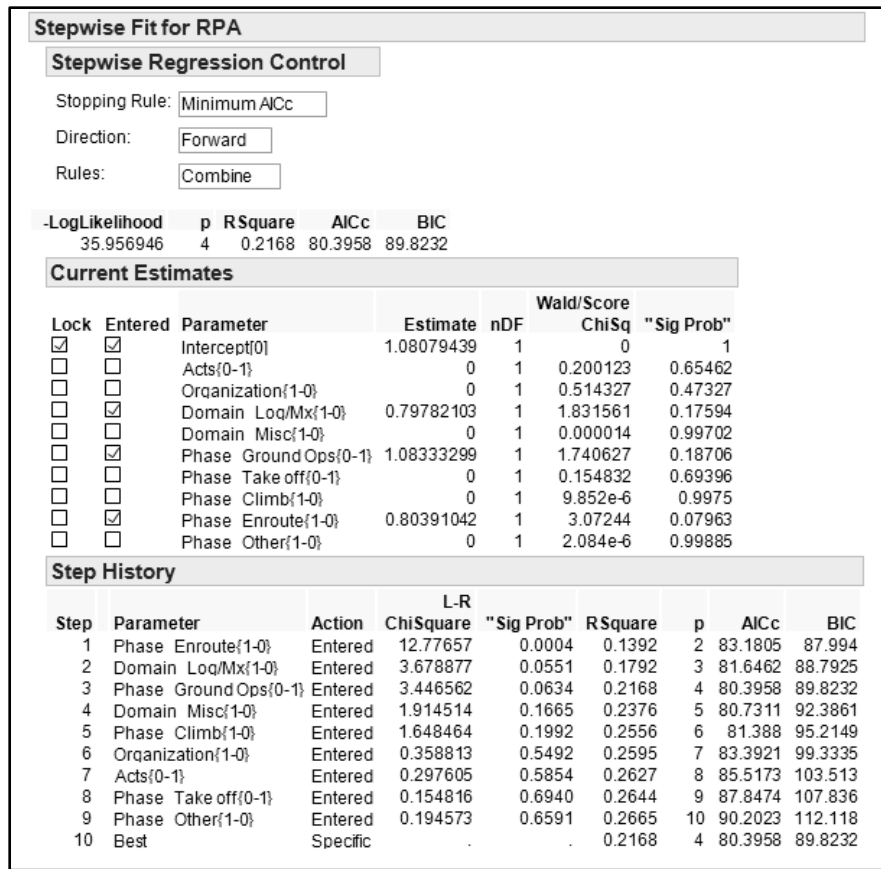


Figure 16. Level I Stepwise Fit Results for RPA

A nominal logistic model was fit to the data identified by the stepwise regression (Figure 17). The Whole Model Test reveals that there was statistically significant evidence to suggest that the model is useful in differentiating between RPA type ($\chi^2_{(3)} = 19.9, p = 0.000$). The Lack of Fit test suggests there was little evidence to support a lack of fit with the selected model ($p = .112$). Phase (Enroute) was identified as the most statistically significant parameter ($p = .052$) in the model. The resulting model for Level I is:

$$\text{logit}(\hat{p}) = 1.08 + .80(\text{Log} / \text{Mx}) - 1.08(\text{GroundOps}) + .81(\text{Enroute})$$

This model implies that Logistics/Maintenance, and Enroute related RPA mishaps are associated with MQ-1 mishaps, and Ground Operations related mishaps are associated with MQ-9 mishaps.

Nominal Logistic Fit for RPA				
Converged in Gradient, 6 iterations				
Whole Model Test				
Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	9.951004	3	19.90201	0.0002*
Full	35.956946			
Reduced	45.907950			
RSquare (U)	0.2168			
AICc	80.3958			
BIC	89.8232			
Observations (or Sum Wqts)	88			
Measure	Training	Definition		
Entropy RSquare	0.2168	1-Loqlike(model)/Loqlike(0)		
Generalized RSquare	0.3125	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))		
Mean -Log p	0.4086	$\sum -\text{Log}(p_{i j})/n$		
RMSE	0.3628	$\sqrt{\sum (y_{i j}-p_{i j})^2/n}$		
Mean Abs Dev	0.2642	$\sum y_{i j}-p_{i j} /n$		
Misclassification Rate	0.1932	$\sum (p_{i j} \neq p_{\text{Max}})/n$		
N	88	n		
Lack Of Fit				
Source	DF	-LogLikelihood	ChiSquare	Prob>ChiSq
Lack Of Fit	2	2.189394	4.378788	
Saturated	5	33.767552		Prob>ChiSq
Fitted	3	35.956946		0.1120
Parameter Estimates				
Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	1.08079442	0.7282659	2.20	0.1378
Domain Loq/Mx[1]	0.79782106	0.438086	3.32	0.0686
Phase Ground Ops[1]	-1.083333	0.6456432	2.82	0.0934
Phase Enroute[1]	0.80391043	0.4141754	3.77	0.0523
For log odds of 1/0				
Effect Likelihood Ratio Tests				
Source	Nparm	DF	L-R ChiSquare	Prob>ChiSq
Domain LoqMx	1	1	4.16089469	0.0414*
Phase Ground Ops	1	1	3.4465622	0.0634
Phase Enroute	1	1	4.7310571	0.0296*

Figure 17. Level I Nominal Logistic Fit for RPA

The Odds Ratios (Figure 18) summarize the effect size and to describe the strength of association between the data. The Odds Ratio for Domain (Logistics/Maintenance) is 4.93. Logistics/Maintenance related RPA mishaps are associated with greater likelihood of an MQ-1 mishap relative to an MQ-9 mishap. The

Odds Ratio for Phase (Ground Operations) is 8.73. Ground Operations related mishaps are associated with greater likelihood of an MQ-9 mishap relative to an MQ-1 mishap. The Odds Ratio for Phase (Enroute) is 4.99. Enroute related mishaps are associated with greater likelihood of an MQ-1 mishap relative to an MQ-9 mishap.

Nominal Logistic Fit for RPA					
Converged in Gradient, 6 iterations					
Lack Of Fit					
Source	DF	-LogLikelihood	ChiSquare		
Lack Of Fit	2	2.189394	4.378788		
Saturated	5	33.767552	Prob>ChiSq		
Fitted	3	35.956946	0.1120		
Odds Ratios					
For RPA odds of 1 versus 0					
Tests and confidence intervals on odds ratios are likelihood ratio based.					
Odds Ratios for Domain_Log/Mx					
Level1	/Level2	Odds Ratio	Prob>Chisq	Lower 95%	Upper 95%
0	1	0.2027783	0.0414*	0.0257136	0.9439918
1	0	4.9314946	0.0414*	1.0593312	38.889878
Odds Ratios for Phase_Ground Ops					
Level1	/Level2	Odds Ratio	Prob>Chisq	Lower 95%	Upper 95%
0	1	8.7291326	0.0634	0.890645	216.22087
1	0	0.1145589	0.0634	0.0046249	1.1227818
Odds Ratios for Phase_Enroute					
Level1	/Level2	Odds Ratio	Prob>Chisq	Lower 95%	Upper 95%
0	1	0.2003237	0.0296*	0.0286872	0.8631961
1	0	4.9919213	0.0296*	1.1584853	34.858778

Figure 18. Level I Odds Ratio Results

The ROC curve identified the tradeoffs between successfully identifying True Positive values and mistakenly identifying False Positives. The resulting ROC curve value for the dataset at Level I was .797 (Figure 19) which suggests that the model may have some trouble with misclassification.

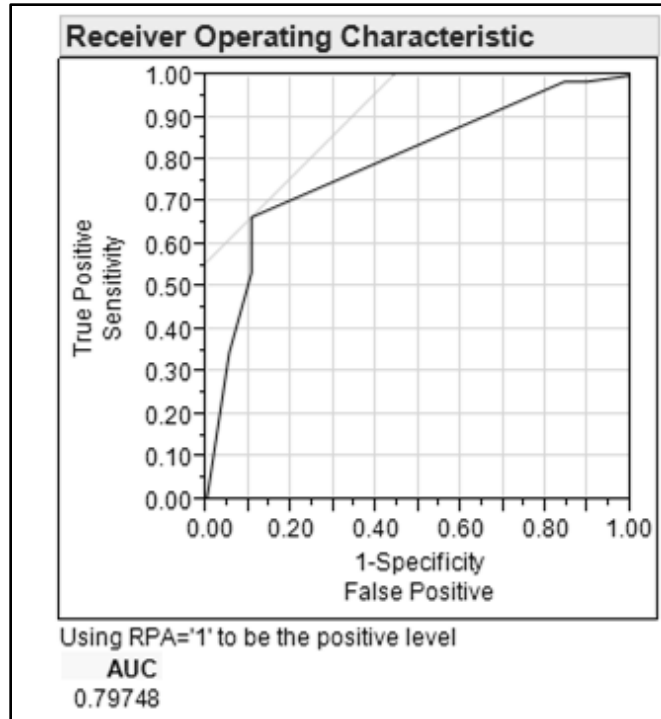


Figure 19. Level I ROC Curve

3. *Level I Cross Validation*

Cross-Validation was performed to assess how well the model classifies records outside of the data. Cross validation was executed by assigning a training set ($n = 72$) and a test set ($n = 16$) from the data. The analysis from the training set (Figure 20) shows that the model misclassified 14 of the 72 mishaps (19.4 percent). The results; however, further indicate the fit of the model is strong for predicting MQ-1 mishaps (98.3 percent) and relatively weak for predicting MQ-9 mishaps (13.3 percent).

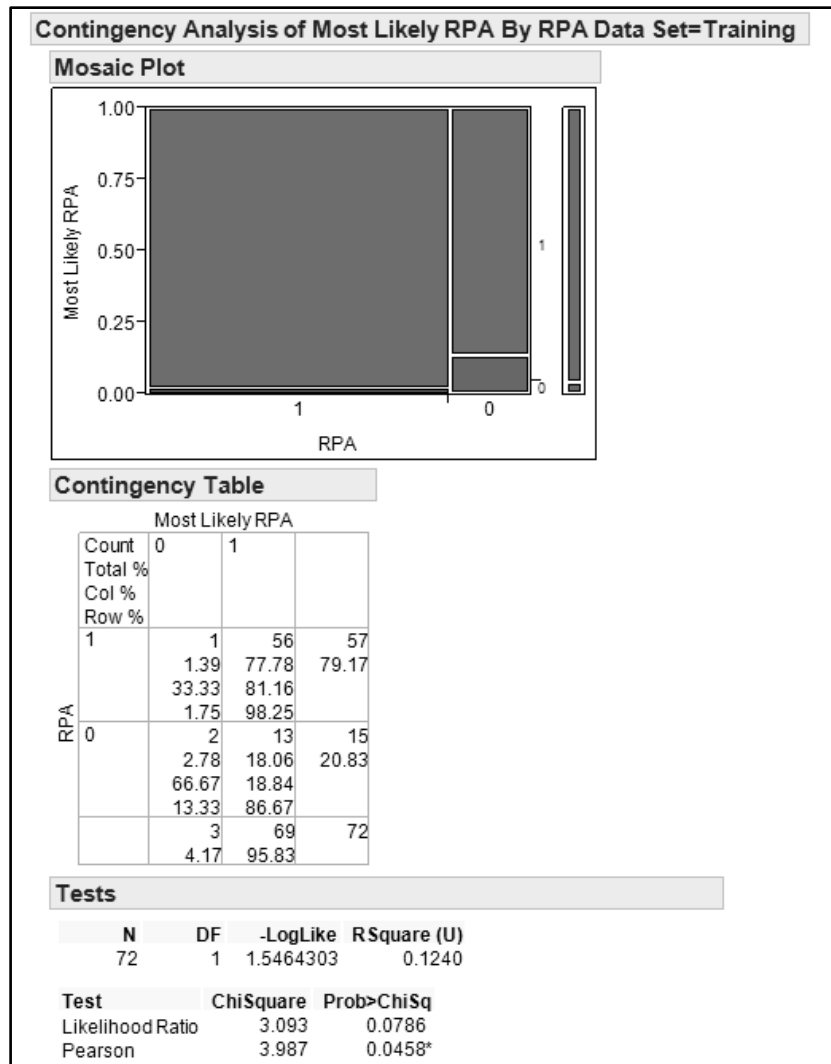


Figure 20. Level I Cross Validation Training Set Results

The test set (Figure 21) produced similar results by misclassifying three of 16 mishaps (18.8 percent). The MQ-1 was accurately predicted 12 out of 12 times (100 percent) and the MQ-9 was accurately predicted one out of four times (25.0 percent). The similar misclassification rates indicate agreement between the test and training sets. Additionally, it can be noted that the model is much more efficient at accurately predicting MQ-1 mishaps relative to MQ-9 mishaps.

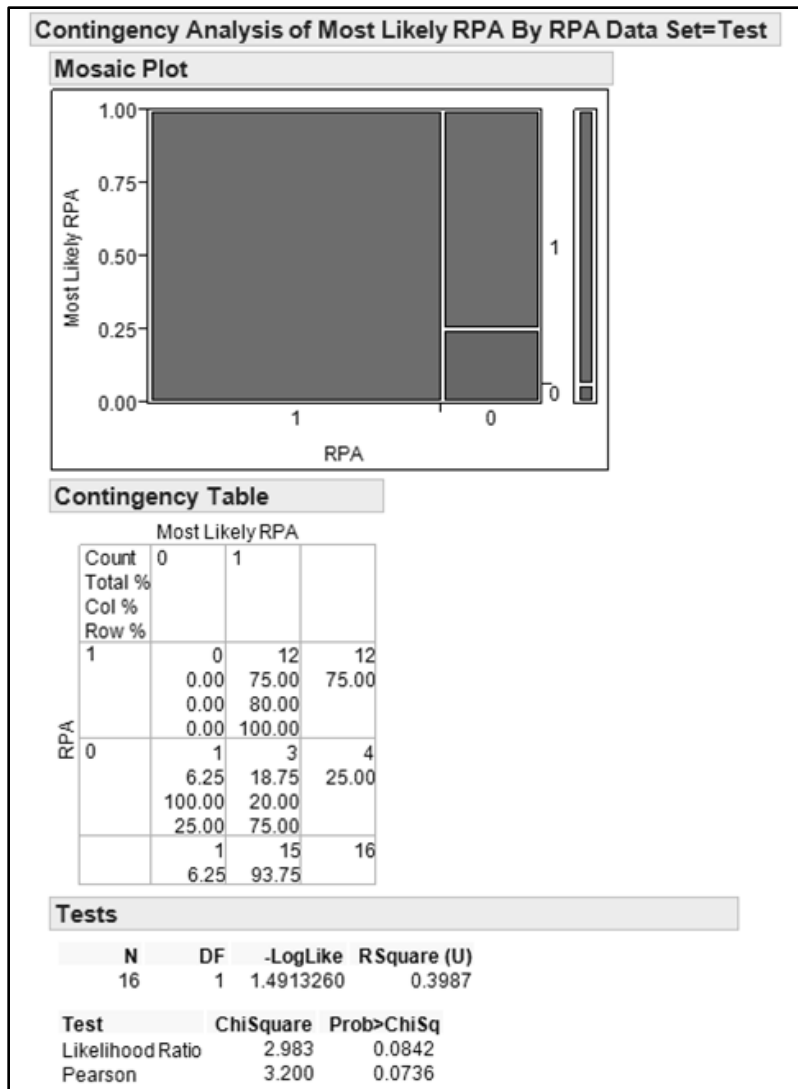


Figure 21. Level I Cross Validation Test Set Results

4. Level II Analysis

Of the 20 DoD HFACS categories at Level II, 17 were tested for homogeneity with regard to RPA type. The categories AV (Violations), PE1 (Physical Environment), and PP2 (Self-Imposed Stress) were removed from the logistic regression due to small sample size and associated numerical instability. AE1 (Skill-Based Errors), AE3 (Perception Errors), PC1 (Cognitive Factors), OC (Organizational Climate), and OP (Organizational Processes) met the defined threshold. AE1 ($p = 0.28$), was the one exception which was included due to its approximate value of 0.25.

Table 16. Summary of Level II Mishap Distribution

Level II Category	χ^2	df	<i>p</i>-value
AE1	1.164	1	0.281
AE2	0.673	1	0.412
AE3	2.141	1	0.143
AV	N/A	N/A	N/A
PE1	N/A	N/A	N/A
PE2	0.143	1	0.705
PC1	3.480	1	0.062
PC2	0.079	1	0.779
PC3	0.017	1	0.895
PC4	0.475	1	0.491
PC5	0.000	1	0.984
PP1	0.539	1	0.463
PP2	N/A	N/A	N/A
SI	0.108	1	0.743
SF	0.228	1	0.633
SP	0.759	1	0.384
SV	0.059	1	0.809
OR	0.555	1	0.456
OC	2.290	1	0.130
OP	2.121	1	0.145

Factors found to be significant in the stepwise regression were included in the logistic analysis: AE1 (Skill-Based Errors), AE3 (Perception Errors), PC1 (Cognitive Factors), OC (Organizational Climate), and OP (Organizational Processes). The baselines for Domain and Phase randomly chosen were Operations and Landing respectively. The stopping rule for this fit was defined by minimum AIC. The model identified OC (Organizational Climate), Domain (Logistics/Maintenance), Phase (Ground Ops), and Phase (Enroute) as parameters to include in the logistic model (Figure 22).

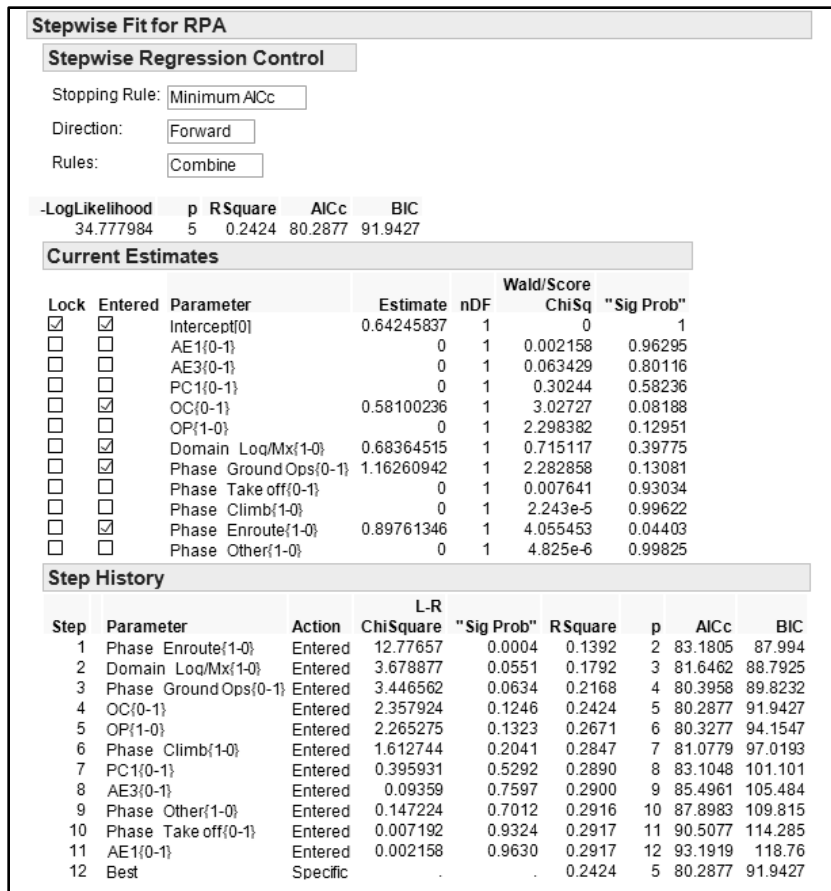


Figure 22. Level II Stepwise Fit Results for RPA

A nominal logistic model was fit to the data identified by the stepwise regression. The Whole Model Test reveals that there is statistically significant evidence to suggest that the model is useful in differentiating between RPA type ($\chi^2_{(4)} = 22.3, p = 0.0002$). The Lack of Fit test suggests there is little evidence to support a lack of fit with the selected model ($p = .19$). Phase (Enroute) was identified as the most statistically significant parameter ($p = .04$). The resulting model for Level II is:

$$\text{logit}(\hat{p}) = 0.642 - 0.581(OC) + 0.684(Log / Mx) - 1.163(GroundOps) + 0.898(Enroute)$$

This model implies that Logistics/Maintenance, and Enroute related RPA mishaps are associated with MQ-1 mishaps, whereas Ground Operations and Organizational Climate related mishaps are associated with MQ-9 mishaps.

Nominal Logistic Fit for RPA				
Converged in Gradient, 6 iterations				
Whole Model Test				
Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	11.129966	4	22.25993	0.0002*
Full	34.777984			
Reduced	45.907950			
RSquare (U)	0.2424			
AICc	80.2877			
BIC	91.9427			
Observations (or Sum Wqts)	88			
Measure	Training	Definition		
Entropy RSquare	0.2424	$1 - \text{Loqlike}(\text{model}) / \text{Loqlike}(0)$		
Generalized RSquare	0.3450	$(1 - (L(0) / L(\text{model}))^{2/n}) / (1 - L(0)^{2/n})$		
Mean -Loq p	0.3952	$\sum -\text{Loq}(p_{ij}) / n$		
RMSE	0.3558	$\sqrt{\sum (y_{ij} - p_{ij})^2 / n}$		
Mean Abs Dev	0.2538	$\sum y_{ij} - p_{ij} / n$		
Misclassification Rate	0.1818	$\sum (p_{ij} \neq p_{Max}) / n$		
N	88	n		
Lack Of Fit				
Source	DF	-LogLikelihood	ChiSquare	Prob>ChiSq
Lack Of Fit	4	3.043924	6.087848	
Saturated	8	31.734060		
Fitted	4	34.777984		0.1927
Parameter Estimates				
Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	0.64245845	0.7673584	0.70	0.4025
OC[1]	-0.5810024	0.378877	2.35	0.1252
Domain Loq/Mx[1]	0.68364521	0.4491739	2.32	0.1280
Phase Ground Ops[1]	-1.1626095	0.6383123	3.32	0.0685
Phase Enroute[1]	0.89761351	0.4304794	4.35	0.0371*
For log odds of 1/0				
Effect Likelihood Ratio Tests				
Source	Nparm	DF	L-R ChiSquare	Prob>ChiSq
OC	1	1	2.35792406	0.1246
Domain LoqMx	1	1	2.77200761	0.0959
Phase Ground Ops	1	1	4.04088801	0.0444*
Phase Enroute	1	1	5.5592173	0.0184*

Figure 23. Level II Nominal Logistic Fit for RPA

The Odds Ratios (Figure 24) summarize the effect size and to describe the strength of association between the data. The Odds Ratio for OC (Organizational Culture) is 3.20. Organizational Culture related RPA mishaps are associated with greater likelihood of an MQ-9 mishap. Domain (Logistics/Maintenance) is 3.92. Logistics/Maintenance related RPA mishaps are associated with greater likelihood of an MQ-1 mishap relative to an MQ-9 mishap. The Odds Ratio for Phase (Ground Operations) is 10.23. Ground Operations related mishaps are associated with greater

likelihood of an MQ-9 mishap relative to an MQ-1 mishap. The Odds Ratio for Phase (Enroute) is 6.02. Enroute related mishaps are associated with greater likelihood of an MQ-1 mishap relative to an MQ-9 mishap.

Nominal Logistic Fit for RPA					
Converged in Gradient, 6 iterations					
Odds Ratios					
For RPA odds of 1 versus 0					
Tests and confidence intervals on odds ratios are likelihood ratio based.					
Odds Ratios for OC					
Level1	/Level2	Odds Ratio	Prob>Chisq	Lower 95%	Upper 95%
0	1	3.1963346	0.1246	0.7203896	15.058155
1	0	0.3128584	0.1246	0.0664092	1.3881378
Odds Ratios for Domain_Log/Mx					
Level1	/Level2	Odds Ratio	Prob>Chisq	Lower 95%	Upper 95%
0	1	0.2547964	0.0959	0.0312473	1.2516472
1	0	3.924702	0.0959	0.7989472	32.002751
Odds Ratios for Phase_Ground Ops					
Level1	/Level2	Odds Ratio	Prob>Chisq	Lower 95%	Upper 95%
0	1	10.228919	0.0444*	1.0579243	246.09468
1	0	0.097762	0.0444*	0.0040635	0.9452473
Odds Ratios for Phase_Enroute					
Level1	/Level2	Odds Ratio	Prob>Chisq	Lower 95%	Upper 95%
0	1	0.1660897	0.0184*	0.0224284	0.7562226
1	0	6.0208414	0.0184*	1.3223619	44.586353

Figure 24. Level II Odds Ratio Results

The ROC curve identified the tradeoffs between successfully identifying True Positive values and mistakenly identifying False Positives. The resulting ROC curve value for the dataset at Level II was .823 (Figure 25). The value indicates that the model may have some trouble with misclassification.

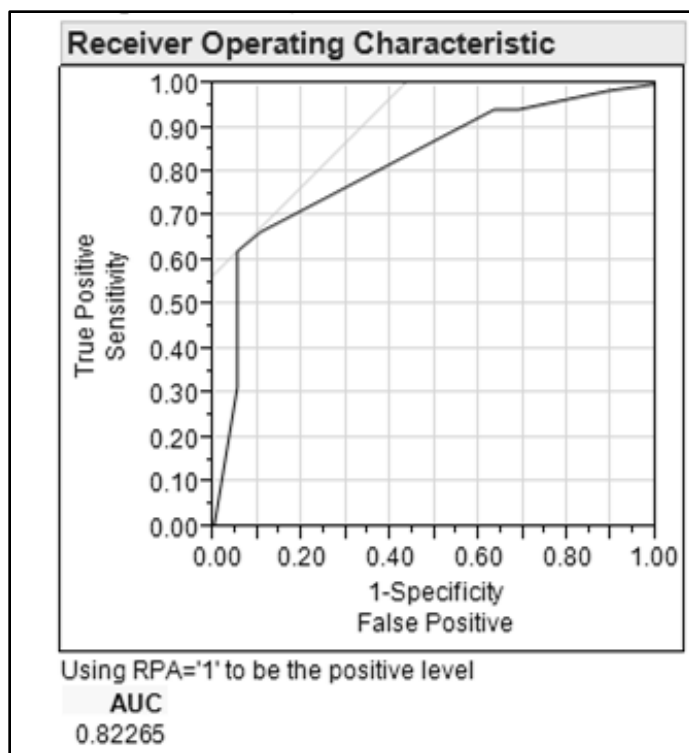


Figure 25. Level II ROC Curve

5. *Level II Cross Validation*

Cross-Validation was performed to assess how well the Level II model classifies records outside of the data. Cross validation was executed by assigning a training set (n = 75) and a test set (n = 13) from the data. The analysis from the training set (Figure 26) shows that the model misclassified 13 of the 72 mishaps (18.1 percent); however, the results further indicate the fit of the model is strong for predicting MQ-1s (93.3 percent) and relatively weak for predicting MQ-9s (40.0 percent).

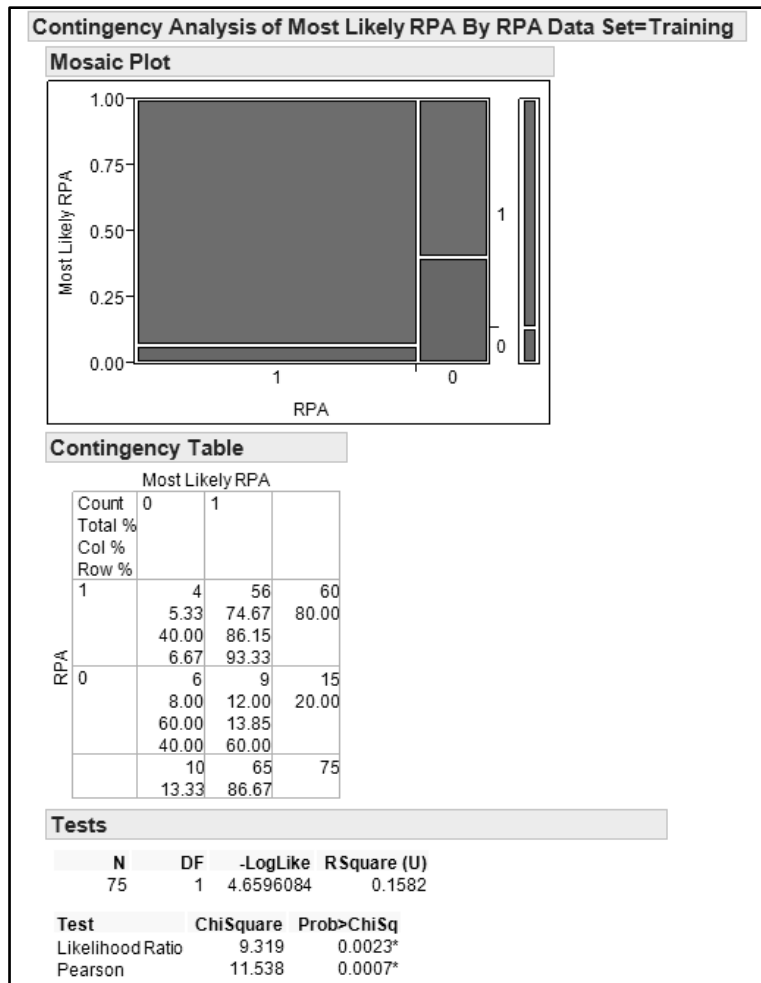


Figure 26. Level II Cross Validation Training Set Results

The test set (Figure 27) produced similar results by misclassifying three of 13 mishaps (23.7 percent). The MQ-1 was accurately predicted nine out of nine times (100 percent) and the MQ-9 was accurately predicted only once out of four times (25 percent). The similar misclassification rates indicate agreement between the test and training sets. Additionally, it can be noted that the model is much more efficient at accurately predicting MQ-1 mishaps relative to MQ-9 mishaps. These results are consistent with the Level I Cross Validation.

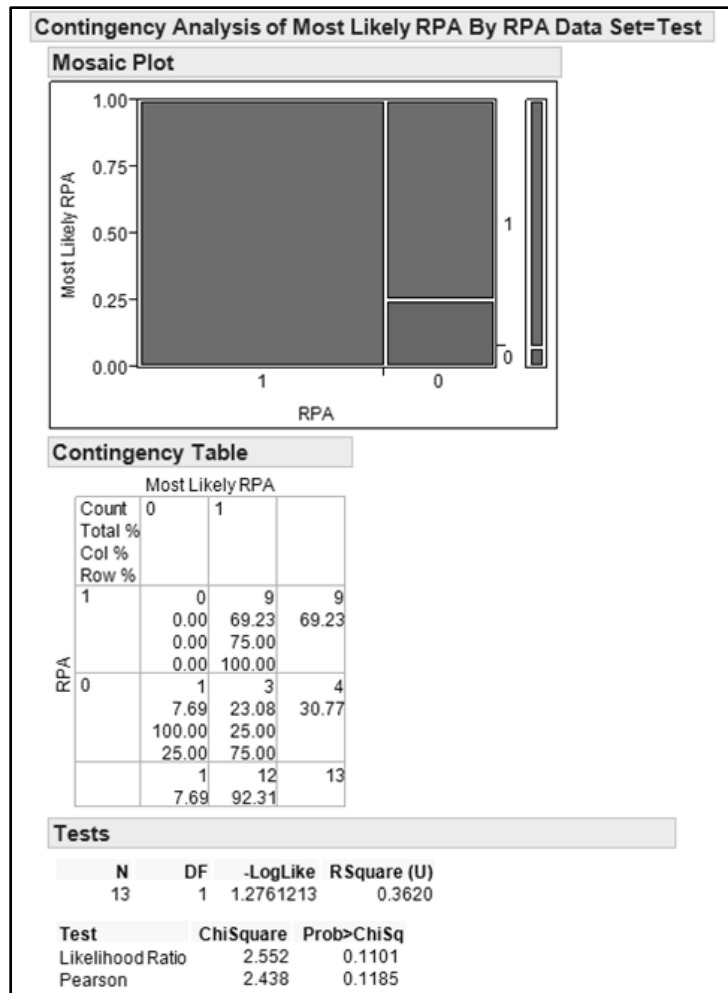


Figure 27. Level II Cross Validation Test Set Results

A second Cross-Validation was performed to further assess how well the model classifies records outside of the data at Level II. Cross validation was executed by assigning a training set (n = 76) and a test set (n = 12) from the data. The analysis from the second training set (Figure 28) shows that the model misclassified 15 of the 76 mishaps (19.7 percent); however, the results further indicate the fit of the model is strong for predicting MQ-1s (98.4 percent) and relatively weak for predicting MQ-9s (6.67 percent).

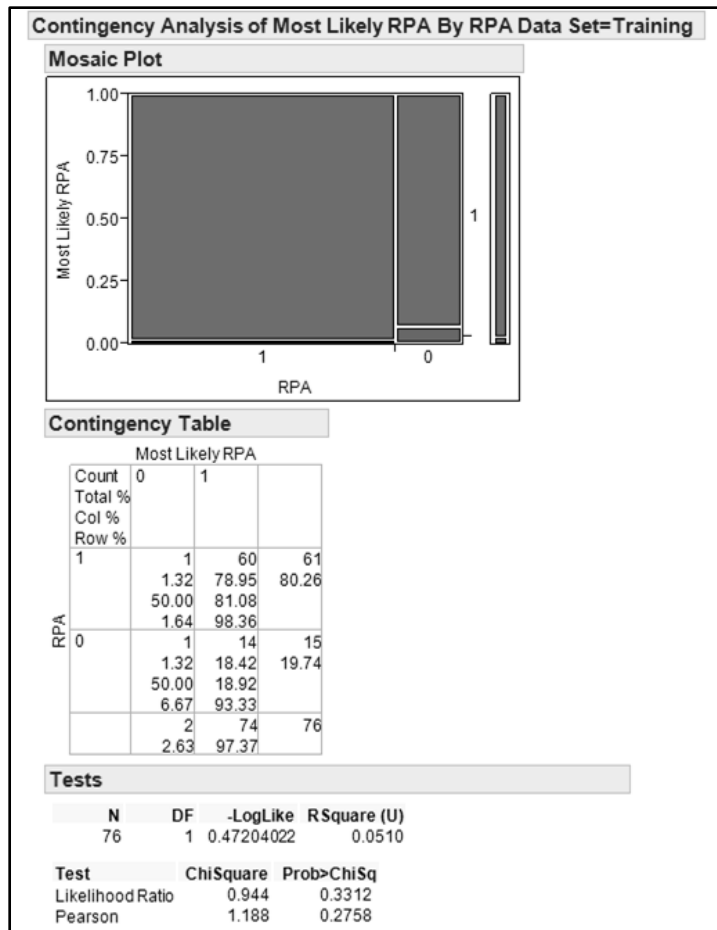


Figure 28. Level II Cross Validation Training Set Results (Second Iteration)

The second test set (Figure 29) produced similar results by misclassifying three of 12 mishaps for a rate of 25.0 percent. The MQ-1 was accurately predicted eight out of eight times (100 percent) and the MQ-9 was accurately predicted one out of four times (25.0 percent). The similar misclassification rates indicate agreement between the test and training sets. Additionally, it can be noted again that the model is much more efficient at accurately predicting MQ-1 mishaps relative to MQ-9 mishaps. These results are consistent with the Level I Cross Validation and the first Level II Cross Validation.

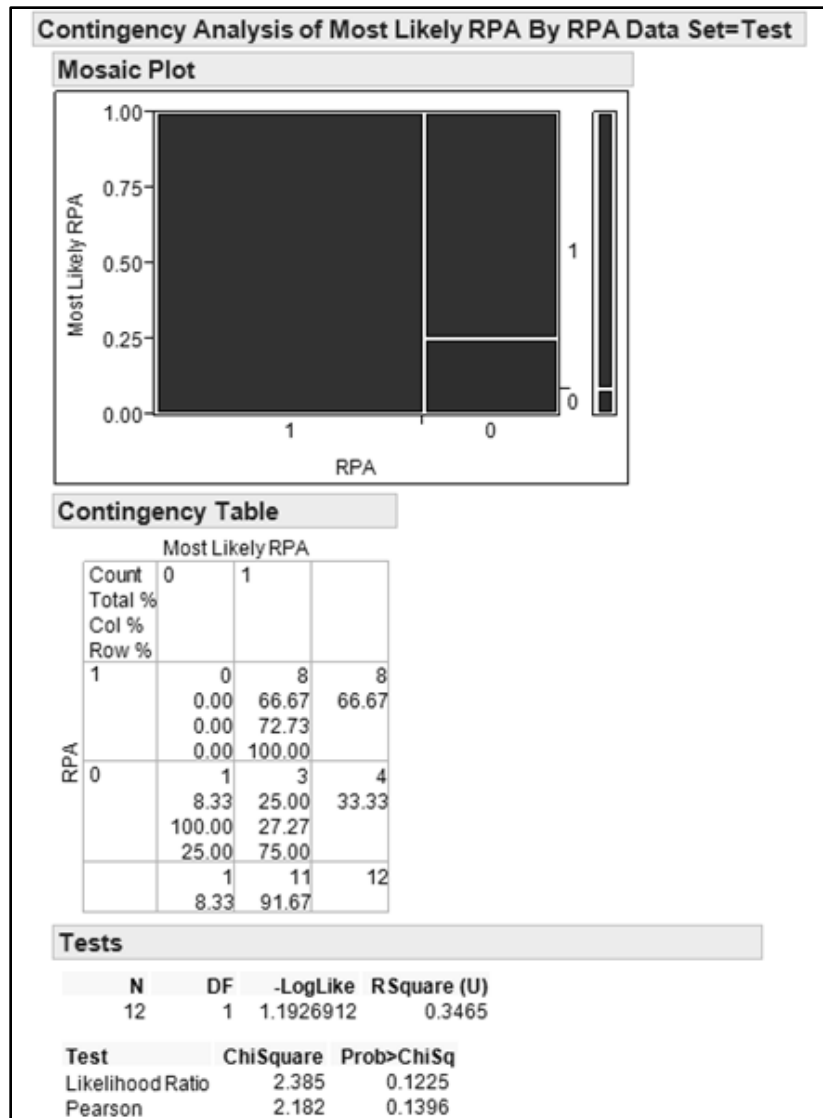


Figure 29. Level II Cross Validation Test Set Results (Second Iteration)

E. SUMMARY

The application of a chi-square analysis to evaluate the observed and expected frequencies at both Levels for the MQ-1 and MQ-9 provided statistical rationale for selecting nanocodes and covariates for inclusion in the logistic regression.

The HFACS Level I results of the logistic regression included only the two covariates, Domain and Phase, as qualified parameters in the construction of the model to predict RPA type. The analysis at Level I did not identify any latent or active failures, as

defined in DoD HFACS, for inclusion in the model. The analysis at this level suggests that the binary response variable (RPA type) was not associated with human error (DoD HFACS). The analyses fail to reject the null hypothesis that there is not an effect on RPA type on human performance concerns while operating RPA systems with the same GCS.

The Level II results of the logistic regression are consistent with the results from Level I. The model included only one nanocode group, Organizational Culture, and the same two Level I covariates, Domain and Phase, to predict mishap RPA type. The analysis only identified one latent failure and no active failures, as defined in DoD HFACS, for inclusion in the model. The hypothesis that there is not an effect on RPA type on human performance concerns while operating RPA systems with the same GCS cannot be rejected.

The near exclusion of the DoD HFACS nanocodes as variables in either model indicates that there is not sufficient human error evidence in this dataset to suggest that there is a relative difference in probability favoring the MQ-1 or MQ-9 mishap predictability based on the use of the same GCS.

THIS PAGE INTENTIONALLY LEFT BLANK

V. DISCUSSION

A. OVERVIEW

As the effort to demonstrate the viability and effectiveness of RPA systems continues, there is an increasing demand for improved total system performance; specifically reduced mishap rates. Based on the dramatic increase in Combatant Commander's requests for these mission critical systems during the last decade, in addition to the rapidly growing civilian RPA sector, it is evident these systems are becoming an integral component to our national defense and numerous civil aeronautics sectors. Results from a recent study of 221 DoD RPA mishaps spanning a 10-year period found that 79 percent of USAF RPA mishaps were human error-related (Tvaryanas, 2006). The analysis and understanding of where human error can be attributed in this realm is lacking in the current literature. In an effort to improve the understanding of RPA mishap epidemiology, an analysis was completed on USAF MQ-1 and MQ-9 RPA mishaps from 2006-2011. The dataset included 88 human error-related mishaps that were coded using DoD HFACS, an evolution of Reason's complex linear accident model, known as the Swiss Cheese Model (Reason, 1990).

The human error coding assigned by the mishap investigators was validated by conducting inter-rater reliability analyses. The moderate to good agreement identified between Rater 1 (original mishap investigator) and Raters 2 (aerospace medicine specialist) and 3 (aerospace physiologist) provided sufficient evidence to support validation of the study dataset.

The initial exploration of the data involved the organization of the data into two levels of the DoD HFACS hierarchy, Level I (Acts, Preconditions, Supervision, Organization) and Level II (20 subcategories of Level I). Covariates evaluated in the dataset included Phase of Flight (Ground Operations, Takeoff, Climb, Enroute, Landing, and Other), Mishap Domain (Operations, Logistics/Maintenance, and Miscellaneous), and Mishap Class (A, B, and C) by RPA type. The application of a chi-square analysis at

both Levels for the MQ-1 and MQ-9 identified Mishap Domain and Phase of flight to be statistically significant for inclusion as covariates in the logistic regression analysis.

The subsequent analysis applied a series of chi-square tests to identify statistical differences among the HFACS categories (at both levels) by RPA type. The application of the chi-square analysis to evaluate the observed and expected frequencies at both Levels for the MQ-1 and MQ-9 provided statistical rationale for selecting nanocodes and covariates for inclusion in the logistic regression. The resulting statistically significant (p-value < 0.25) categories and covariates were further analyzed by applying logistic regression techniques to the data. The resulting logistic regression models are designed to predict aircraft type within the mishap dataset. The models were assessed using ROC curves for accuracy and were cross validated using test sets from the study dataset. The models intend to provide quantitative data to inform RPA certification standards and to complement existing efforts to improve future system designs.

The Level I results of the logistic regression included only the two covariates, Domain and Phase, as qualified parameters in the construction of the model to predict RPA type. The analysis at Level I did not identify any latent or active failures, as defined in DoD HFACS, for inclusion in the model. The analysis at this level suggests that the binary response variable (RPA type) was not associated with human error (DoD HFACS). The analyses fail to reject the null hypothesis that there is not an effect on RPA type on human performance concerns while operating RPA systems with the same GCS.

The Level II results of the logistic regression are consistent with the results from Level I. The model included only one DoD HFACS category, Organizational Climate, and the same two Level I covariates, Domain and Phase, to predict mishap RPA type. The hypothesis that there is not an effect on RPA type on human performance concerns while operating RPA systems with the same GCS cannot be rejected.

The near exclusion of the DoD HFACS nanocodes as variables in either model indicates that there may not be sufficient human error evidence in this dataset to suggest that there is a relative difference in MQ-1 or MQ-9 mishap predictability based on the use of the same GCS.

The adequate incorporation of Human Systems Integration early in the system acquisition phases is dependent on quantitative and relevant data to serve as forcing functions in designing and building smart human-centered systems. The models derived in this study support performance improvement by quantifying mishap patterns and how those patterns resemble or differ between the MQ-1 and MQ-9 when operated with the same GCS.

B. RESEARCH QUESTION

This research was driven by the need to improve the understanding of human error patterns in the RPA operations realm. The specific research question was: Do the types of active failures (unsafe acts) and latent failures (preconditions, unsafe supervision, and organizational influences) differ between the MQ-1 and MQ-9 when operated with the same GCS? The single inclusion of Organizational Climate (organizational influence) in the Level II model suggests that there is not a statistically significant difference in RPA type mishaps with regard to human error.

The research analyzed the archive of HFACS data in addition to covariates such as Mishap Class, Mishap Domain, and Mishap Phase of Flight that are unique to each aircraft and those that are shared by both to identify potential human error patterns. It developed logistic regression models to predict aircraft type given the mishap dataset.

The Level I Model is defined as:

$$\text{logit}(\hat{p}) = 1.08 + .80(\text{Log} / \text{Mx}) - 1.08(\text{GroundOps}) + .81(\text{Enroute})$$

This model predicts that the specific Domain of the mishap in addition to the Phase of Flight in which the mishap occurred accurately predicts RPA type approximately 79 percent of the time within the dataset. Specifically, Logistics/Maintenance related RPA mishaps are associated with greater likelihood of an MQ-1 mishap relative to an MQ-9 mishap. Ground Operations related mishaps are associated with greater likelihood of an MQ-9 mishap relative to an MQ-1 mishap. Enroute related mishaps are associated with greater likelihood of an MQ-1 mishap relative to an MQ-9 mishap. There were no Level I (Acts, Preconditions, Supervision,

Organization) DoD HFACS identified in the analysis that were considered statistically different by chi-square and logistic regression for inclusion in the model. These results suggest that human performance requirements need to be closely coupled to the GCS and not necessarily RPA type.

The Level II Model is defined as:

$$\text{logit}(\hat{p}) = 0.642 - 0.581(OC) + 0.684(Log / Mx) - 1.163(GroundOps) + 0.898(Enroute)$$

This model predicts that the citing of the Level II DoD HFACS category (latent failure), Organizational Climate, is more strongly associated with MQ-9 mishaps. Additionally, the specific Domain of the mishap in addition to the Phase of Flight in which the mishap occurred accurately predicts RPA type approximately 82 percent of the time within the dataset. The covariate results are consistent with the Level I model. There was only one Level II DoD HFACS category, Organizational Climate (latent failure), identified in the analysis that was considered sufficiently diagnostic by chi-square and logistic regression for inclusion in the model. These results provide additional evidence that human performance requirements need to be closely coupled to the GCS and not necessarily to the RPA type.

C. IMPLICATIONS FOR SYSTEM DESIGN

RPA provide a unique challenge to developers of certification standards (e.g., FAA, DoD) because the GCS and the aircraft are separate and it is theoretically possible to mix and match GCSs and aircraft. The stated research question was, “what matters in terms of human performance: the GCS or the aircraft?” The dataset provided the opportunity to gain insight into this question as a natural experiment in which the cockpit (GCS) is controlled and the aircraft was varied. The study results suggest that the GCS is what matters in terms of human performance, not the aircraft. The unique patterns, or lack thereof, of human performance failures provide evidence supporting the development of GCS standards used in RPA systems. The author recognizes that further exploration and analysis must be accomplished to transition to a more comprehensive

understanding of RPA mishap patterns. The efforts presented in this study have contributed to the understanding this relatively new realm in aviation history, the RPA.

D. CONCLUSION

This study explored the potential human error patterns in the USAF MQ-1 and MQ-9 communities, and recommends a solution aimed at proactive mishap prevention. Only a single RPA-specific human error pattern was identified to be significant enough for inclusion in the models, organizational climate (latent failure). The identified covariates in the models provide valuable data supporting further exploration into improved safety approaches with the potential to reduce costly RPA accidents. The study hypothesis that there is an effect of aircraft type on the human performance challenges when operating an RPA system from the same GCS was rejected. Current and future RPA systems should consider and prioritize the impact of GCS design with regard to RPA total system performance.

The USAF should consider additional human error research on current and future weapon systems currently in the acquisitions process. The suggested research should not be limited to historical mishap data, but should include areas where latent conditions can be quantified as both positive and negative drivers in total system performance. These areas should focus on the design of the GCS. The scope of this study did not include the specific issues with regard to the GCS, nor did it investigate the characteristics of the GCS and any potential influence on human error-related RPA mishaps. It is therefore recommended that future research and development efforts focus on the specific parameters surrounding the design and function of the GCS. The data analysis at the beginning of Chapter 4 is a recommended starting point for potential human error analysis as related to the GCS. The statistically significant differences prevalent among both levels of DoD HFACS categories (Table 13) may provide a starting point for further analysis that was outside the scope of this project. By using the analysis in this research, the USAF may be able to develop effective system design strategies with the objective to reduce the growing cost of human error RPA mishaps.

THIS PAGE LEFT INTENTIONALLY BLANK

APPENDIX A. MQ-1 AND MQ-9 SYSTEM CHARACTERISTICS

Characteristic	MQ-1	MQ-9
Primary Function	Armed reconnaissance, airborne surveillance and target acquisition	Remotely piloted hunter/killer weapon system
Contractor	General Atomics Aeronautical Systems Inc.	General Atomics Aeronautical Systems, Inc.
Power Plant	Rotax 914F four cylinder engine	Honeywell TPE331-10GD turboprop engine
Thrust	115 horsepower	900 shaft horsepower maximum
Wingspan	55 feet (16.8 meters)	66 feet (20.1 meters)
Length	27 feet (8.22 meters)	36 feet (11 meters)
Height	6.9 feet (2.1 meters)	12.5 feet (3.8 meters)
Weight	1,130 pounds (512 kilograms) empty	4,900 pounds (2,223 kilograms) empty
Maximum takeoff weight	2,250 pounds (1,020 kilograms)	10,500 pounds (4,760 kilograms)
Fuel Capacity	665 pounds (100 gallons)	4,000 pounds (602 gallons)
Payload	450 pounds (204 kilograms)	3,750 pounds (1,701 kilograms)
Speed	Cruise speed around 84 mph (70 knots), up to 135 mph	Cruise speed around 230 miles per hour (200 knots)
Range	Up to 770 miles (675 nautical miles)	1,150 miles (1,000 nautical miles)
Ceiling	Up to 25,000 feet (7,620 meters)	Up to 50,000 feet (15,240 meters)
Armament	Two laser-guided AGM-114 Hellfire missiles	Combination of AGM-114 Hellfire missiles, GBU-12 Paveway II and GBU-38 Joint Direct Attack Munitions
Crew (remote)	Two (pilot and sensor operator)	Two (pilot and sensor operator)
Initial operational capability	Mar-05	Oct-07
Unit Cost	\$20 million (FY09\$M) (includes four aircraft, a GCS and a Primary Satellite Link)	\$53.5 million (includes four aircraft with sensors) (fiscal 2006 dollars)

THIS PAGE LEFT INTENTIONALLY BLANK

THIS PAGE LEFT INTENTIONALLY BLANK

LIST OF REFERENCES

- Greenwood, M. & Woods, H.M. (1919). *The incidence of industrial accidents upon individuals with special reference to multiple accidents*. (British Industrial Fatigue Research Board, Report No. 4). London: HMS.
- Hollnagel, E., Woods, D. D., & Leveson, N. (2006). *Resilience Engineering*. Burlington: Ashgate .
- Rasmussen, J. (1986). *Information processing and human-machine interaction: An approach to cognitive engineering*. New York: Elsevier Science.
- Reason, J. (1990). *Human error*. Cambridge: Cambridge University Press.
- Reason, J. (1997). *Managing the risks of organizational accidents*. Aldershot, UK: Ashgate
- Salmon, P., Regan, M., & Johnston, I. (2005). *Human error and road transport*. Monash University: Accident Research Centre.
- Shortcuts, rush to field are key factors in UAV accidents, report claims. (2005). *Defense Daily*, 227(2), 1. Retrieved from <http://search.proquest.com.libproxy.nps.edu/docview/234060717?accountid=12702>
- Taranto, M. T. (2012). *USAF RPA Training*. Monterey: Naval Postgraduate School.
- Tvaryanas, A. P. (2006). Human systems integration in remotely piloted aircraft operations. *Aviation, Space, and Environmental Medicine*, 77(12), 1278–1282. Retrieved from <http://search.proquest.com.libproxy.nps.edu/docview/29518480?accountid=12702>
- Tvaryanas, A. P., Thompson, W. T., Constable, S. H. (2006, July). Human factors in remotely piloted aircraft operations: HFACS analysis of 221 mishaps over 10 years. *Aviation, Space, and Environmental Medicine*, 77(7), 724–732.
- U.S. Air Force. (2012). Air Force Safety Automated System Retrieved March 18, 2013, from <https://afsas.kirtland.af.mil>. Restricted-access data.
- U.S. Air Force. (2012). Air Force Safety Center. Retrieved March 18, 2013, from <https://afsas.kirtland.af.mil>. Restricted-access data.
- U.S. Air Force. (2012a). MQ-1 Specifications. Retrieved March 21, 2013, from U.S. Air Force: <http://www.af.mil/information/factsheets/factsheet.asp?fsID=122>

U.S. Air Force. (2012b). MQ-9 Specifications. Retrieved March 21, 2013, from U.S. Air Force: <http://www.af.mil/information/factsheets/factsheet.asp?fsID=6405>

USAF/SEF. (2008). Safety Investigations and Reports. USAF.Vermont: Ashgate.

Webster, N., White, D., & Wurmstein, T. (2005). DoD HFACS. Retrieved March 18, 2013, from <https://afsas.kirtland.af.mil>. Restricted-access data.

Wiegmann, D. A., & Shappell, S. A. (2003). A human error approach to aviation accident analysis: The human factors analysis and classification system. Burlington, VT: Ashgate.

INITIAL DISTRIBUTION LIST

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