Runway Incursion Severity Risk Analysis:

September 14, 2012

FINAL Report

Produced for:

Runway Safety
Office of Safety
Air Traffic Organization
Federal Aviation Administration
Washington, D.C.

Produced by:

Lee Biernbaum and Garrett Hagemann
U.S. Department of Transportation
Research and Innovative Technology Administration
Volpe National Transportation Systems Center
Cambridge, MA

DOCUMENT APPROVAL

Document Number of Final Version	Date of Document mm/dd/yyyy	Approver Initials	Date of Approval
Draft	07/31/2012		
Final	09/14/2012		

DOCUMENT CHANGE HISTORY

Document Number	Version Number	Draft or Final	Date of Document mm/dd/yyyy	Author's Initials	Author's Org	Description of Change

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. **DOT-VNTSC-FAA-12-13**	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle **Runway Incursion Severity Risk Analysis**		5. Report Date **September 14, 2012** 6. Performing Organization Code
7. Author(s) **Lee Biernbaum, Garrett Hagemann**		8. Performing Organization Report No. **DOT-VNTSC-FAA-12-13**
9. Performing Organization Name and Address **U.S. Department of Transportation Volpe National Transportation Systems Center 55 Broadway Cambridge, MA 02142**		10. Work Unit No. (TRAIS) 11. Contract or Grant No. **FA6AB3, FA01B1**
12. Sponsoring Agency Name and Address **U.S. Department of Transportation Federal Aviation Administration Air Traffic Organization, Office of Safety and Technical Training 800 Independence Avenue, SW Washington, DC 20591**		13. Type of Report and Period Covered **Final Report, September 2010 – September 2012** 14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

Runway incursions are defined as the unauthorized presence of a vehicle, pedestrian, or aircraft on a runway. Identifying situations or conditions in which runway incursions are more likely to be severe can suggest policy implications and areas for future safety research. Previous work in this area focused on a narrative approach. This study seeks to examine runway incursions from a statistical perspective and provide insights into the broad trends underlying severity.

This report analyzes 10 years of runway incursion event information. A variety of FAA data sources were used to provide information on the event itself, airport characteristics, and airport operations at the time of the incident. Weather information was also incorporated using automated weather readings from airports. The culmination of the analysis is a series of discrete choice models focusing on different sets of incident characteristics.

As this represents the first regression-based analysis of these data, the results are suggestive rather than definitive. For example, controller incidents appear to be more severe on average. The results also suggest some areas for further investigation: specifically a need for understanding the frequency of incursions and improvements to the severity measure.

17. Key Words **Runway incursion, air traffic control, incident severity, safety analysis, discrete choice, multinomial logit, ordinal logit, airport safety, runway safety**	18. Distribution Statement **No restrictions**		
19. Security Classif. (of this report) **Unclassified**	20. Security Classif. (of this page) **Unclassified**	21. No. of Pages **271**	22. Price

EXECUTIVE SUMMARY

Runway incursions are used to identify pre-collision behavior. Understanding those factors that increase the severity of a runway incursion may help identify situations that are more dangerous and potentially mitigate that danger. A runway incursion is defined as the unauthorized presence of a vehicle, pedestrian or aircraft on a runway. Runway incursions are rated according to severity: category D represents the least severe incidents (generally one aircraft) while category A represents the most severe (up to and including a collision). Incidents are also identified by who is "responsible" for the incursion: a controller, a pilot, or a vehicle.

The purpose of this research is to examine the underlying factors that contribute to the severity of runway incursions. The research detailed in this report does not seek to explain the causes of *particular events*, but rather focuses on *broader trends* in incursion severity. Understanding those broader patterns can provide insight into policy-making and identify areas for future research.

Prior to examining any data, a literature review was undertaken to identify hypotheses potentially relevant explanatory variables. However, little quantitative research has been done on runway incursions. Much of the research that has been done has been qualitative in nature. Some identified trends, but generally focus on individual events rather than broad factors that may influence severity. Thus, to the best knowledge of the authors, the research in this report is *the first systematic statistical analysis of runway incursions*.

The analysis focused on the set of all runway incursions that occurred from 2001 to 2010. The FAA curated this dataset, which contains basic information about the incursion and related aircraft. One of the Volpe Center's innovations was to combine multiple FAA and non-FAA data sources to incorporate information not available in the base dataset. These additional sources included the FAA's Air Traffic Quality Assurance (ATQA) database and Operational Network (OPSNET) database, while weather and information on airport layout were gathered from other parties.

A variety of statistical techniques were also used to examine the dataset. Due to the lack of previous research, much of the effort focused on cross tabulations of the data. This technique revealed interesting relationships among the variables both in terms of incident severity and incident type. A preliminary modeling effort was also undertaken. Some of the major conclusions drawn from the research are:

- Controller incidents are approximately three times more likely to be severe than other incident types.
- Incident type and severity distributions statistically significantly vary by region, indicating policy impacts will also vary by region.
- Evidence suggests controller age does not impact severity.
- Commercial carriers are 60% less likely to be involved in severe conflict incursions but are more likely to be involved in conflict incursions overall.

- Additional runway intersections increase the likelihood of a severe event, but more total runways decreases the likelihood of a severe event.
- Incidents during takeoff are 2.5 times more likely to be severe when compared with taxiing. Incidents during landing are 1.7 times as likely to be severe when compared with taxiing.

In addition to identifying factors that contribute to severity, this research effort identified areas for future research. Some of the research that could contribute most to an understanding of the risks related to runway incursions are:

- Estimating models of incursion frequency (rather than severity) to shed light on how other variables impact safety.
- Investigating the nature of the ordering (if any) of severity between C and D events.
- Understanding the relationship between incident type (OE/PD/VPD) and severity.
- Examining why LAHSO operations appear to have fewer than expected incursions despite being a riskier operation.
- Refining and clarifying traffic complexity measures.
- Investigating the relationship between time on shift and frequency of incursions.
- Disentangling the effects of various visibility-related measurements (i.e., visibility, ceiling, cloud coverage).

TABLE OF ACRONYMS

Acronym	Definition
AC/AT	Air Carrier / Air Transport
AIP	Airport Improvement Program
AMASS	Airport Movement Area Safety System (a predecessor to ASDE)
ARTS II	Automated Radar Terminal System, Version II
ARTS III	Automated Radar Terminal System, Version III
ASDE	Airport Surface Detection Equipment
ASDE-3	Airport Surface Detection Equipment, Version 3
ASDE-X (and ASDEX)	Airport Surface Detection Equipment, Model X
ASRA	Aviation System Reporting System
ATC	Air Traffic Control
ATQA	Air Traffic Quality Assurance
ETMSC	Enhanced Traffic Management System Counts
FAROS	Final Approach Occupancy Signal
GA	General Aviation
ICAO	International Civil Aviation Organization
IIA	Independence of Irrelevant Alternatives
LAHSO	Land and Hold Short Operation
METAR	From the French Mètéorologique Aviation Régulière. Hourly weather reports automatically generated
NAS	National Airspace System
OE	Operator Error
OEP	Operational Evolution Partnership
OLS	Ordinary Least Squares
OPSNET	Operations Network Database
PD	Pilot Deviation

Acronym	Definition
RI	**Runway Incursion**
STARS	**Standard Terminal Automation Replacement System**
TIPH	**Taxi Into Position and Hold**
V/PD or VPD	**Vehicle or Pedestrian Deviation**
VFR	**Visual Flight Rules**
VMC	**Visual Metrological Conditions**
VOD	**Vehicle Operation Deviations**

TABLE OF CONTENTS

TABLE OF TABLES

TABLE OF FIGURES

1. INTRODUCTION

The focus of this research is to examine the underlying factors that contribute to the severity of runway incursions. A runway incursion is an event in which a person, vehicle, or aircraft enters the runway safety area without authorization. From the perspective of the FAA, runway incursions represent dangerous pre-collision behavior. In accordance with standards established by the International Civil Aviation Organization (ICAO), runway incursions are ranked according to their severity, with category D being the least dangerous and category A being a narrowly avoided collision.[1] As it is believed that reducing the severity of incursions reduces the likelihood of having a collision, it is important to understand those factors that influence incursion severity.

Previous research has focused on qualitative examinations of incursion reports. Case studies were used to understand some trends and identify common causes. The research detailed in this report does not seek to explain the causes of particular events, but rather focuses on broader trends in incursion severity. Understanding those patterns can provide insight into policy-making and identify areas for future research.

1.1. Background

Runway incursions are classified in two main ways. Severity is ranked from category D to category A. In addition to that ranking is a classification of who was at fault for the incident: controller at fault, called operational errors (OEs); pilot at fault, called pilot deviations (PDs); and vehicle or pedestrian at fault, called vehicle or pedestrian deviations (V/PD). In fiscal year 2008, FAA adopted a new definition of runway incursions, conforming to ICAO standards.[2] When compared to previous years, the new definition produces more runway incursions, even with no change in underlying behavior or safety. Thus, any comparison with previous years needs to be done so in the light of the changing definitions. However, this definitional change does not reclassify any severe incursions (class A or B). Below is an overview of recent incursion trends to provide context for the research that follows.

Fiscal Year 2008

During FY 2008, there were 1,009 runway incursions.[3] Twenty-five of those incursions were classified as severe (category A or B), resulting in a rate of approximately 0.43 severe incursions per million operations (takeoff or landing). The overall rate of runway incursions was 17.2 incursions per million operations.

---

1 The FAA also ranks collisions as Category A events. This practice deviates from the ICAO standard, which does not consider collision events as Category A events.

2 Please see Appendix A: Runway Incursion Definition for a complete definition including severity classifications. Note that the appendix uses the new definition of a runway incursion.

3Federal Aviation Administration (2010).

Fiscal Year 2009

During FY2009, there were 951 runway incursions.[4] Of those incursions, 12 were categorized as severe, representing a rate of 0.23 severe incidents per million operations. The overall rate of runway incursions was 18 incursions per million operations.

Fiscal Year 2010

There were 966 incursions during fiscal year 2010, representing a rate of approximately 18.9 incursions per million operations.[5] Of those, 6 were categorized as severe, representing a rate of 0.12 severe incidents per million operations.[6]

Fiscal Year 2011

There were 954 incursions during fiscal year 2011, representing a rate of approximately 18.8 incursions per million operations. Of those, 7 were categorized as severe, representing a rate of 0.14 severe incidents per million operations.

Runway Incursion Trends

Compared to the previous year, FY2010 saw an increase in both the number of incursions and the rate of runway incursions. However, the number of incursions in FY2010 is still below the FY2008 total of 1009 incursions. The incursion rate has increased in general from 2008 (a rate of 17.2 incursions per million operations) to 18.9 incursions per million operations in FY2010. FY2011 saw the rate remain almost constant, though there was a slight drop in total number of incursions.

FY2010 saw a continued decrease in the overall number of severe incursions. The rate of severe incursions has also declined. This is contrary to the trend in the overall rate and count of runway incursions. FY2011 saw an increase in both the rate and total number of severe incursions. However, both were very slight and likely not representative of an increasing trend.

FAA Response to Runway Incursions

FAA has recently placed a renewed focus on runway safety, starting with a Call to Action in August 2007. A plan “focused on changes in cockpit procedures, airport signage and markings, air traffic procedures, and technology” was developed.[7] Further deployment of systems such as ASDE-3/AMASS and ASDE-X will increase controller awareness of movement areas. FAA has also begun deployment of runway status lights at 23 airports. The new light system “gives pilots a visible warning when runways are not safe to

---

4 Ibid.

5 Ibid.

6 For FY2010 and FY2011, incursion statistics are taken from http://asias.faa.gov while operations statistics are taken from the OPSNET database. The statistics from that database are current as of 8/1/2012.

7 Federal Aviation Administration (2008).

enter, cross, or depart on".[8] The first lights are already online with the full set expected to be in service by 2016.[9] Yet another effort to reduce runway incursions is the deployment of the Final Approach Occupancy Signal (FAROS) system. The FAROS system "activates a flashing light visible to aircraft on approach as a warning to pilots when a runway is occupied and hazardous for landing" – essentially the arrival counterpart to runway status lights.[10]

These interventions by FAA are an attempt to control some of the causes and impacts of runway incursions. FAROS and runway status lights are designed to give pilots more information so that they can avoid a runway incursion (by performing a go around or stopping at the hold short line, for example). The ground surveillance technologies (ASDE-3/AMASS and ASDE-X) help improve situational awareness for controllers and provide controllers with early warnings of potential collisions. Both are human factors improvements meant to mitigate runway incursion risk.

1.2. Method of Investigation

The goal of this research was to use statistical methods to identify trends in runway incursion severity. The basis of this research was the set of all incursions that occurred between January 1, 2001 and September 30, 2010. During this time period, there were approximately 8,800 incursions. The methodology focused on analyzing these 8,800 incursions and detecting patterns in airport, aircraft, controller, and pilot characteristics. Additional information on weather was included where feasible.

The analysis was effectively split into two parts. The first part was focused on one-way and two-way descriptive statistics and analyzing cross tabulations of variables. As much of the information describing the incursion was categorical in nature, this provided an effective means of analyzing these variables. Additionally, this allowed a wider array of variables to be tested. These results focus on comparing variables pairwise, so are less able to account for interactions.

To counteract some of the limitations of the cross tabulation approach, a modeling effort was undertaken. This allowed multiple variables to be included at once and their interactions to be understood. However, as this was a more time-intensive process, the sample had to be limited. It was decided to focus on controller incursions. Thus, the cross tabulations examine many more variables across and broader array of incursion types while the modeling effort attempts to delve deeper into the relationship between these variables and severity within a limited sample. Again, due to time and resource constraints, these models should be considered *preliminary only*; a more intensive modeling exercise would provide significant improvements to the understanding of runway incursion severity.

The modeling effort focused on discrete choice models. Due to the apparently ordered nature of the rankings, an ordered logit model was presumed to be the appropriate model. Evidence suggests,

8 Government Accountability Office (2008).

9 Federal Aviation Administration (2011).

10 Ibid.

however, that the assumptions of the ordered model did not hold. Therefore, multinomial logit models were employed to capture a more nuanced look at the impacts on severity.

1.3. Overview of the Document

This document is divided into three major sections followed by conclusions and appendices. The first section presents the results of a review of previous runway incursion literature as well as discrete choice modeling. The second section presents descriptive statistics and the results of the cross tabulations. This addresses some of the basic distributions of the data, serves as an introduction to the data involved in the modeling, and presents some basic results. The third section details the modeling effort, including the supporting methodology. These results supplement those seen in the second section and form the basis for any conclusions drawn.

Following the main body of the paper are a series of appendices. Appendix A addresses the definition of a runway incursion. Appendix B addresses additional data issues discovered during the research process. Appendix C provides additional detail on the statistical methodologies used in this report. Appendix D contains a list of identified future research needs.

2. LITERATURE REVIEW

The research reviewed in this section falls into two major categories. The first set of papers covers previous research on runway incursions. Understanding severity was not the main goal of these papers; rather, they focused on understanding the causes behind runway incursions. This first set of research papers provided insights into what variables or concepts might play a role in determining incident severity. These suggested variables can be further divided into policy variables – which can be directly affected to produce a change – and control variables – which are not directly affected by policy, but still play a role.

The second set of papers focus on discrete choice modeling. While not necessarily in the context of runway incursions, or even aviation, this research demonstrates relevant methodology. Section 4.1 of this paper, on methodology, was heavily influenced by these papers.

It is apparent from this literature review that a rigorous econometric model of runway incursion severity has not been previously developed. The previous research on runway incursions has been focused on the human factors elements that can cause runway incursions. There is also a wealth of information on modeling injury severity, mostly from the highway community. The combination of these two research traditions guided the development of a model of runway incursion severity.

2.1. Previous Runway Incursion Research

Previous research on runway incursion causes been mostly conducted in the human factors arena and divides the research into three areas: pilots, controllers, and other airport personnel. The papers outlined below represent the culmination of an extensive research process. The review began with some known sources and a broad search for literature related to the causes and severity of runway incursions. These sources provided additional citations that proved to be of interest to the review process. Ultimately, however, few papers focus specifically on the causes and severity of runway incursions. The following summary attempts to provide a fair representation of the state of the practice.

Cardosi and Yost produced an extensive literature review on the subject of human factors in runway incursions.[11] A summary of their findings is presented here.[12]

> Cardosi and Yost note that a common theme among the papers they reviewed was miscommunication or failure to coordinate between two controllers. In addition to that common theme, other factors such as losing track of an aircraft or forgetting its position were also cited as contributing to runway incursions. Another study (Kelly and Steinbacher 1993) focused on frequency congestion and found that many incidents were associated with blocked transmissions or incomplete messages. Lastly, Skaliotis (1991) found that the "number of incursions was not well correlated with the number of operations. It suggested that local factors

11 Cardosi and Yost (2001).

12 In addition to the literature review, Cardosi and Yost examined safety data. This analysis of both pilots and controllers and will be discussed in the relevant sections below.

at particular airports are more important than high operations at determining the risk of an accident/incident" during the time period studied. [13]

2.1.1. Pilots

DiFiore and Cardosi examined 231 reports filed by pilots or co-pilots from the Aviation System Reporting System (ASRA).[14] DiFiore and Cardosi found that, by far, communication factors were cited most often overall as contributing to runway incursions. Position awareness (i.e., the pilot being aware of his or her location in the airfield) was cited next most often. The analysis then focused on certain kinds of runway incursions: crossing the hold short line, crossing the runway without a clearance, taxi into position and hold (TIPH), and entering the runway without authorization. The authors offered the broad categorizations of human factors mentioned previously, but were also able to focus on specific issues (such as misunderstanding ATC phraseology).

Cardosi and Yost performed an analysis of safety data submitted by pilots. They examined 76 incident reports and found that unclear airport markings and controller-pilot miscommunication were the two most cited causes of incursions.

2.1.2. Controllers

In addition to their literature review and analysis of pilot related human factors, Cardosi and Yost looked at reports focusing on controller-related issues.[15] They found that the five most common contributing factors, in order, were (lack of) aircraft observation, coordination, communication errors, visual data, and ground operations. Following the analysis of reports, Cardosi and Yost examined the underlying report data to perform their own independent analysis. They found that the most common contributing factors were controllers forgetting about the status of a runway or an aircraft, controller-pilot communication errors, controller coordination errors, and supervisor/controller in charge working a control position simultaneously.

2.1.3. Other Airport Personnel

Scarborough, Bailey, and Pounds examined vehicle operation deviations (VODs) – where one party involved in a runway incursion is driving a ground vehicle (as opposed to an aircraft) – to attempt to find factors associated with this type of deviation.[16] They used logistic regression and found a statistically significant relationship between a driver not observing markings, signals, or lighting and the presence of

---

13 Ibid.

14 DiFiore and Cardosi (2006).

15 Cardosi and Yost (2001).

16 Scarborough*, et al.* (2008).

inclement weather. On the other hand, no relationship was found between construction outside the movement area and VODs.

The Airport Cooperative Research Program, part of the Transportation Research Board, sponsored a synthesis project focused on winter operations.[17] The report provides a thorough exploration of factors contributing to vehicle-aircraft incidents during winter operations. The report group factors into several broad categories, including:

- Communication,
- Environment,
- Human performance,
- Situational awareness,
- Time pressures,
- Personnel, vehicles, and equipment resources, and
- Operational factors.

The report cited poor communication (e.g., using the incorrect radio frequency, equipment mishaps, and frequency congestion), poor visibility, fatigue, time pressures (to clear the runway as quickly as possible to resume aircraft operations), and several operating factors as major causes of runway incursions during winter operation. While the report focused on winter operations, it provides insight into ground operations in general.

2.2. Severity Research on Other Modes

While research focusing on incursion severity seems to be lacking from the current runway incursion literature, the question of factors contributing to automobile crash severity has been examined extensively. This highway literature can provide important insight into how to approach modeling runway incursion severity. In addition, reviewing crash severity literature can illuminate those areas were runway incursions are similar to and diverge from the highway crash literature and will require careful consideration.

2.2.1. Safety Research

Schneider IV et al. examined the factors contributing to driver injury severity along horizontal curves in Texas.[18] A multinomial logit approach was used and separate models were developed for three different curve radii (small, medium and large). Some of their findings can be translated to a runway incursion framework while others are less easily translated. The authors found that not wearing a seatbelt greatly increased the chance of a fatality. The same is true for the presence of alcohol and drugs. Those factors have no clear analogues in the runway incursion framework. The authors also examined environmental factors and found that clear weather and daylight increase the chance of a less severe accident. Weather may also play a role in runway incursion severity. Another factor the authors considered was

17 Quilty (2008).

18 Schneider IV*, et al.* (2009).

vehicle type. Certain vehicle types (motorcycles) were associated with higher probabilities of more severe injuries while others (semi- and pickup trucks) were not. This translates rather directly into examining the impact of aircraft type on the runway incursion severity. However, the relationship between pilot experience and aircraft type would need to be carefully considered.

Kockelman and Kweon also examined the factors contributing to driver injury severity.[19] The authors used an ordered probit methodology and focused on different types of crashes: single versus two vehicle crashes. Again, the authors found a relationship between driver injury severity and vehicle type as well as alcohol. Interestingly, the authors did not find an effect for daylight (versus nighttime) on injury severity. The authors also found evidence of a non-linear relationship between injury severity and driver age. It is unclear how age may translate into a useful concept for runway incursions, but it speaks to the need to examine the included variables in a non-linear way as well. Lastly, the authors examined how the angle of the crash – head-on versus rear-end for example – contributes to driver injury severity. This suggests examining a similar notion of angle for runway incursions. For example, it may be that more severe incursions are associated with more certain relative angles between aircraft.[20]

Islam and Mannering provide another example of a multinomial logit approach.[21] The authors focused on differing gender-age group combinations (male and female, young, middle-aged, and elderly drivers). This paper examines automobile-specific factors that could have contributed to injury severity. However, coefficients are reported for only some of the models (and then only the statistically significant ones), and select elasticities are reported in the comparison tables. This makes it difficult for the reader to gain a full understanding of implications of the model and removes the context for the results. Additionally, findings that are not statistically significant are as important as those results which are statistically significant. Reporting even insignificant results is a critical step in the research process. This analysis does provide an interesting template for comparing different subgroups of a population. Lam provides another example of an ordered probit approach targeted at comparing different age groups in a graduated licensing system in Australia.[22]

2.2.2. Methodological Concerns

Xie et al. used a similar ordered probit model but the coefficients were estimated using a Bayesian approach.[23] They examined the outcome of using different priors on the coefficient estimates. They also compared the results of standard ordered probit to a Bayesian ordered probit on the complete and a

19 Kockelman and Kweon (2002).

20 The runway incursion dataset provided did not allow for this kind of analysis, but it remains an interesting question for future research.

21 Islam and Mannering (2006).

22 Lam (2003).

23 Xie*, et al.* (2009).

restricted sample to gauge the impact the differing methodologies had when compared on a small sample of data, a property of interest for statistical models. The restricted sample represents a random selection of 100 records from the complete set of 76,994 records. In the complete sample, they found results consistent with other studies: increased age and alcohol usage increase the injury severity. Both being male and certain vehicle types (vans and SUVs) reduce injury severity. The researchers found similar results between the standard ordered probit and Bayesian ordered probit in terms of coefficient magnitudes and standard errors for the full sample. When examining the restricted sample, the authors found that the Bayesian ordered probit provided answers more similar to those obtained on the full sample. This indicates that the Bayesian approach may be better suited to examining small datasets.

Abdel-Aty used an ordered probit approach and found similar results when looking at crashes at three different roadway types in Florida (roadway sections, signalized intersections, and toll plazas).[24] The author also tested these results against differing estimation procedures. Ordered logit models gave similar results, while a multinomial logit did not perform as well (as measured by how well the model predicted the known data and with fewer variables found to be significant). A nested logit procedure was also tested, but was found to be difficult to implement; the model also provided little improvement over the ordered probit in terms of model fit. The analysis provides insight into some methodological considerations but is not as informative for examining runway incursions. The variables used are specific to the road sections considered (such as whether or not an electronic toll tag was in use).

Perera and Dissanayake also used an ordered probit approach.[25] Their analysis focused on injury severity among older drivers. They developed two models, one for urban roads and one for rural. They found similar results as other studies, however the analysis is simplistic. For example, they used a series of binary variables to represent vehicle type. The general form of the variables is that they are equal to one if the vehicle was that type, and zero otherwise. They included binary variables for cars, vans, pick-ups, and SUVs. Note that these categories are by definition mutually exclusive: a car cannot be a van or a pickup or an SUV – knowing that one of the variables is equal to one reveals the value of the other vehicle variables. All coefficients for these variables are positive in the rural model. The authors report that the vehicles are associated with increased injury severity. However, without a reference case, the positive coefficients are inherently meaningless and must be compared amongst themselves. Pickups, with the lowest positive coefficient, thus reduce injury severity compared to other vehicle types rather than increase injury severity. The focus on older drivers and driver age renders this paper not very informative for runway incursions. However, it is illustrative of a methodological trap that needs to be avoided.

These papers present a summary of the types of methodologies that may be used to understand runway incursion severity. Yet, the papers have some flaws worth noting with the intention that the same flaws are avoided during the modeling process for the current research. Several of the papers suffered from reporting deficiencies, such as not reporting all coefficients. Other papers suffered from methodological

---

24 Abdel-Aty (2003).

25 Perera and Dissanayake (2010).

problems in their variable definitions or interpretation, such as the Perera and Dissanayake paper just described.

While this research is suggestive of methodologies and factors to consider for runway incursions, there is a subtle difference between crash injury severity and runway incursion severity. Crash injury severities are conditional on a crash having already occurred whereas runway incursions are attempting to classify the underlying risk associated with an incident.

It is important to keep these differences in mind when using injury severity literature to inform a study on runway incursions. While the underlying methodology will not change, the interpretation of the coefficients will be slightly different.

2.3. Conclusions

This literature review provided a starting point for developing a model for runway incursion. The research that was reviewed suggested several variables that warrant further examination:

- Policy Variables
    - The presence of technologies like ASDE-X
    - Runway configuration
- Control Variables
    - Weather conditions
    - Time of day
    - Presence of construction
    - Aircraft type
    - Pilot characteristics (if available)

Notably, most of the suggested variables are "control variables," and may not directly influence severity. While it is important that the control variables are present in the model, they provide little actionable information. However, the response of an airport to these control variables may be a policy lever that could be examined. Additionally, it would be valuable in future research to translate potential relevant policy decisions of airports into variables for evaluation.

3. DATA DESCRIPTION AND DESCRIPTIVE STATISTICS

3.1. Datasets

3.1.1. Runway Incursion Data

Source

The RI database is maintained by the FAA Runway Safety Office. It contains information on 10,408 runway incursions from January 2, 2001 through September 30, 2010. It is hand-populated based on reports filed in response to an incursion and contains information expected to be of use to the Runway Safety Office in responding to, and preventing further, incursions.[26] Recall that the FAA adopted a new definition of Runway Incursions in 2008. Incursions prior to 2008 were given a rank consistent with the new system ensuring that the most current definition is used for this analysis.

Contents

The Runway Incursion database contains basic information on each incursion (date, time, airport, type), aircraft, parties involved (e.g., private citizen, airport personnel), the type of error, current conditions at the airport, and the closest vertical and horizontal distance between the aircraft.

Data Issues

There are inconsistencies in how the data are coded from year to year that warranted additional data cleaning. For example, it appears to vary as to how a “no” is recorded; that is, sometimes a variable was left “missing” to signify “no” while in other cases (sometimes for the same variable), a “no” was specifically entered. In others, it appears that “unknown” was used as a valid response in some, but not all, years of the database.

Additionally, the database was provided without a detailed codebook; follow up with the Runway Safety Office was required to ascertain the meaning of some specific variables or codings. Full details on the various data problems encountered and their resolution are in Appendix B: Data Issues.

Sometimes, the database provides more detail than necessary for this analysis (e.g., aircraft type, which hold short line was crossed). This information was consolidated into categories that are more general for the purposes of this analysis.

3.1.2. ATQA OE

Source

The ATQA database contains the preliminary and final incident reports for (Air Traffic Controller) OEs both en-route and on the surface. The Runway Safety Office provided an extract of OE incidents related to surface events. This database contains 1,504 unique records. Fields that contain personally identifiable information or relevant only to airborne events were not included in the extract.

---

26 The Runway Incursion Database is no longer updated with new events. Runway Incursions are still noted in the ATQA database, but the more detailed process is no longer performed.

Contents

The database contains all of the information collected in the preliminary and final investigation (FAA Forms 7210-2 and 7210-3). The database contains information on the aircraft involved in the incident (see subsequent section), the controller and conditions in the tower, some descriptions of the event, and information about the facility (including radar and other equipment in use at the time of the incident).

The ATQA OE database also contains information on causal factors related to the incident. These data were deemed inappropriate for this analysis for several reasons. Firstly, the causal factors are related to the severity of the incident by definition in some instances. Thus, they are inappropriate for a modeling effort as they *determine* the outcome. Second, the causal factors are not conditioning factors; the causal factors, rather, indicate *how* an incident happened. Consider an incident where one of the causal factors is hear back/read back error. Reducing the number of hear back/read back errors would surely reduce the number of incursions, but provides little guidance on what conditions increase or decrease the likelihood of such errors. Finally, the data quality on these variables was also quite low. Thus, even if the causal factors were determined to be beneficial to this analysis, the data quality prevented their inclusion.

Data Issues

Many of the variables in this dataset are inconsistently coded over time. Others contain a large number of “missing values.” These missing values in some cases may be interpreted as a “no,” (i.e., the form instructed one to check the box if the answer is yes) but in other cases, the form presents options for both “yes” and “no,” but missing values are still prevalent. In these circumstances, it may not be possible to distinguish between a missing entry intended to be a “no” and those entries left missing because the true state of the variable is unknown. For variables with missing values that are not of the yes/no type (e.g., Current Shift Start Time), observations containing missing values will be excluded from some types of analysis.

Additionally, for incidents involving multiple controllers or aircraft, the database turned over to the Volpe Center does not distinguish between multiple involved aircraft or between multiple involved controllers. While FAA Forms 7210-2 and 7210-3 do allow for multiple aircraft and controllers, the data appear not to have been preserved in the database extract sent to the Volpe Center. It will be assumed that the aircraft or controller information provided will be for the primary aircraft or controller “at fault” or in the wrong location, though this may not be true in all cases.

Other variables, such as aircraft type, appear to have little standardization in the type of responses allowed on the form. In these cases, variables were cleaned by the Volpe Center before they were used for analysis.

3.1.3. ATQA PD

Source

The ATQA database contains information on PD's in addition to information on OE's. Like the information for OE's, the PD data covers both en-route and on the ground incidents. The Runway Safety Office provided an extract of PD incidents related to surface events. This database contains 6,434 unique records. Fields that contain personally identifiable information or relevant only to airborne events were not included in the extract.

Contents

The database contains all of the information collected in the preliminary and final investigation (FAA Forms 8020-17 and 8020-17). The database contains information about the pilot certifications, pilot actions, other pilot characteristics, and some information about the incident (such as aircraft type and some aircraft equipment).

Data Issues

As with the ATQA OE data, variables are inconsistently coded over time. The same issues regarding missing values are present in the PD data: in some cases, it is impossible to distinguish between missing values that are "no," missing values that mean "not applicable," and unknown values. This is doubly complicated for variables where "unknown" is a valid answer on the form. As in the OE data, there are some variables, such as Duty Time in Last 24 hours, which contain missing values indicating those observations had to be excluded from certain analyses.

Similar to the ATQA OE data, the observations in this database are for one aircraft only. In cases where two pilots were involved, the information for the second pilot appears to not have been preserved. It is assumed that the data presented pertain to the pilot and aircraft at fault only.

Finally, some variables required standardization in terms of nomenclature. This is a similar problem to those noted in the ATQA OE database. For example, there are a large number of pilot certification fields. In some cases, respondents selected "other" but provided a description that matches one of the available options. A simple text matching process was developed to locate those records that matched an already existing category. In some cases, such as a common response to "other," additional categories were created.

3.1.4. Weather Information

Source

METAR, from the French Mètéorologique Aviation Régulière, "is the international standard code format for hourly surface weather observations."[27] Hourly METAR weather readings at airports are archived by Plymouth State University in New Hampshire.[28] These METAR readings represent a standardized set of

27 Source: http://www.ncdc.noaa.gov/oa/wdc/metar/

information automatically collected by weather stations. Plymouth State University was able to provide weather readings for a large fraction of the location-hour pairs in the RI dataset.

Contents

The hourly readings contain information about temperature, humidity, wind conditions, visibility conditions, and information about active weather such as storms. In addition, some readings contain summary amounts of precipitation for the past 6 or 24 hours.

Data Issues

Approximately 122 events did not receive weather data, representing 64 different facilities.

Readings of average precipitation over the previous 6 or 24 hours are not reported in every METAR record. Consequently, these data are missing from a substantial portion of WX database entries. These variables were deemed impossible to use. A more sophisticated look at the weather data may be able to incorporate the precipitation measures into an analysis.

3.1.5. Airport Characteristics

Source

Airport characteristic data were gathered by a research team at the University of Virginia Center for Risk Management and Engineering Systems and provided to the FAA for a related study on safety risks at airports. These tables (one for each region) contain information on 498 airports.

Information on runways was gathered from FAA Form 5010 submissions. The Volpe Center pulled all 5010 facility and runway data as of July 2011. A summary of grants distributed by the Airport Improvement Program (AIP) provided information on funded runway construction projects that is used to back out information on runways that opened between an incursion and the present 5010 filing.

Contents

For each airport, the airport characteristics file contains information about the overall characteristics, average weather, geometric layout, number of incursions by severity, and average operations.

The 5010 report contains detailed information on each runway and the location of the facility as a whole. The vast majority of this information was discarded, as it was not useful to this project. The data kept, however, indicate the number of runways at each airport, the length of the shortest and longest runways, and if are Land and Hold Short Operations (LAHSOs) procedures on any runway at the airport.

Data Issues

The variables contained in the excel spreadsheets have plain-text names which are easily human-readable. However, their spreadsheets do not contain additional information on how each data element

---

28 Website: http://vortex.plymouth.edu/

was gathered or recorded. For example, for average "rainy days," it is neither clear what makes a day "rainy," nor how many years over which the data were averaged.[29]

Data entry and display from region to region are inconsistent. The number of columns on the summary of inputs page varies, data are sometimes inappropriately rounded (e.g., percentages to 100% or 0%), data are rounded to a different number of digits, or inaccurate column headings are applied to data on some sheets.

Moreover, other data appear unrealistic. In some cases, clusters of airports report identical weather data, which may be reasonable. However, six airports across Massachusetts report the same weather data, despite being 130 miles apart. Notably, two of these are on Cape Cod, which has significantly different weather from Western or Northern Massachusetts, the location of the other four.

3.1.6. Operations Data

Source

Hourly operations data are available from FAA through the Enhanced Traffic Management System Counts (ETMSC) system. Larger time aggregations, such as daily or yearly operations, are also available through OPSNET.

Contents

The sample data contained hourly readings for approximately 515 airports. For each hour, counts of commercial air carrier, air taxi, general aviation (GA), and military traffic are given. The counts provided by ETMSC are allocations of the daily operations (as reported by OPSNET) at that airport to specific hours. The allocation is done proportionally based on flights with flight plans within a given hour. Thus, if only one flight filed a flight plan that day, total daily operations would be allocated to the hour in which that one operation occurred. Because GA and military flights do not file flight plans as frequently, it is possible that their distribution across the day is unaccounted for.

Data Issues

The main concern with this dataset is the systematic undercounting of GA and military flights. This may present a problem for modeling if the non-flight planned operations are at systematically different times of day than those that file a flight plan, resulting in an allocation of daily operations that does not reflect reality. Ultimately, the correlation between daily, yearly, and hourly operations is fairly high. Therefore, due to the high correlation and higher reliability of daily and yearly data, hourly operations are not used in the modeling effort.

---

29 Through communication with the researchers at University of Virginia, it was determined that the weather information came from http://weatherbase.com. It appears that the information presented on weatherbase.com is derived from historical National Weather Service Records (of varying length per data element). Particularly for "rainy days" no definition is provided on weatherbase.com.

3.2. Merged Data Set

The above datasets were aggregated through a variety of processes that resulted in one overall dataset. The processes used to combine datasets fall into two major categories: event-specific data and more general data. The event-specific data are contained in the Runway Incursion database and the two ATQA databases. The more general information constitutes the airport characteristics, and operations data. The weather data required special treatment before they could be combined with the Runway Incursion database.

3.2.1. Merging Disparate Datasets

General Information

Matching the more general data to the Runway Incursion database was simple. Using the incident location (airport code), date, and time the general data could be easily matched. There is no need to differentiate between multiple incidents at the same airport for certain variables, such as number of runways at an airport.[30] Thus, adding these variables to the underlying Runway Incursion database was simple.

Event-Specific Data

Conversely, for event-specific data, there is a need to distinguish between multiple incidents at the same place and time. For the ATQA OE data, this was accomplished using a unique event identifier. Approximately 249 records in the Runway Incursion database did not have matching records in the ATQA OE dataset. The process for combining the Runway Incursion database with the ATQA PD data was more complicated. The ATQA PD database did not contain a unique record identifier that matched any identifier in the Runway Incursion database. A sequential matching procedure was employed to pair records from the ATQA PD database with the Runway Incursion database.

The first step involved matching records that were unique by date and location. That is, records in each database that were the only one at that airport on that date were considered to be matches. A spot check of those matches indicates that they describe the same incident (e.g., aircraft involved, type of incident). The second step in the sequential match involved hand pairing records that were not already matched. Records were considered matches if they were identical on an increasingly looser set of criteria. For example, the exact times of the incidents were compared. If this did not result in a match, a comparison of information such as the aircraft involved and the hour of the incident followed. This process resulted in 4,193 records that were in both databases and 1,547 records only in the RI database.

---

30 Specifically in the case of number of runways at an airport, the information may change over time. A list of runways built during the time period covered by the data was assembled. Subsequently, the airport characteristics were updated by year to ensure that the number of runways was accurate and any related variables were changed appropriately (e.g., intersecting runways, parallel runways).

Weather Data

As mentioned previously, the weather data is reported hourly, representing point estimates of the conditions at that time. The Runway Incursion database contains the time of the event down to the minute. Because weather data did not necessarily align with the timing of the incursion event, a way to interpolate the weather at the time of the event was developed. Two methods were developed: one for variables that change continuously (like temperature) and one for variables that change discretely (such as precipitation).

The method for continuous variables relied on linear interpolation. The two weather readings on either side of the incident were used as the basis for the interpolation. The method for variables that changed discretely relied on picking the observation closest to the time of the incident. The weather readings occur roughly hourly so the closest reading is, in general, less than 30 minutes away. This method was used for the variables including the weather code (indicating precipitation, fog, smoke, haze, etc.). The remainder of the variables (temperature, cloud cover, etc.) were all subject to the linear interpolation method. The combination of these two methods provided a set of data that could be matched exactly to the Runway Incursion database, making the matching trivial after the interpolation steps.

3.2.2. Summary Statistics

Before examining specific sets of variables, some general characteristics of the merged dataset are worth presenting. It is important to keep these facts in mind when examining specific variables, as the context in terms of the larger dataset is important.

As mentioned previously, incursion events are categorized along two major axes: incident severity and incident type. Table 1 presents the cross tabulation of these two categories and the results of Pearson's Chi-Squared test (Chi-Squared for short). Additionally, Table 2 presents the expected frequency.

The expected frequencies represent the hypothetical distribution of observations across the two categories if the two variables were unrelated. That is, the expected distribution holds the row totals constant but divides observations proportionally among the columns. Deviation from that expected distribution is taken as indication that the rows and columns are not unrelated. For more information please see Appendix C.1.

Table 1 – Observed Incident Type Distribution by Severity

	OE	PD	V/PD	Total
A	53	63	16	132
B	45	77	23	145
C	943	1,822	543	3,308
D	227	3,340	1,660	5,227
Total	1,268	5,302	2,242	8,812

Chi2 score: 1146.89

Degrees of Freedom: 6

P-value: 0.00

Table 2 – Expected Incident Type Distribution by Severity

	OE	PD	V/PD	Total
A	19	79	34	132
B	21	87	37	145
C	476	1,990	842	3,308
D	752	3,145	1,330	5,227
Total	1,268	5,302	2,242	8,812[31]

The first thing worth noticing is the frequencies across the various cells. Interestingly, across the time period covered by our sample, PD incidents occur about twice as often as V/PD incidents and four times as often as OE incidents. The predominance of PD incidents is also true for the different severity categories. The overall frequency is not the only metric of importance, however. Note that OE incidents are the least frequent overall but are the second most frequent for categories A, B, and C. In fact, category A OE incidents occur approximately four times as often as category A PD incidents (giving OEs the highest rate of category A incidents). Thus, while overall frequency is interesting, it is also important to understand the relative frequency of each category. For example, a policy intervention directed at reducing PD incursions (as they are the most common) would do less to reduce category A incursions than an intervention targeted at OE incursions.

31 Tables contain rounded numbers for convenience, consequently row and column totals may not be the same as the sum of the displayed cells. The totals are accurate.

The difference between relative frequency and overall frequency raises the need to test for differences in the two. This is where a Chi-Squared test can be useful.[32]

As reported in Table 1, the Chi-Squared statistic is extraordinarily high and associated with a p-value of approximately zero. This indicates that the distribution of incursion severity is not uniform across the different incident types. Because this is a joint test, it is unable to distinguish which categories are over or under represented. That is, this test indicates that there is *some* relationship between incident type and severity, but cannot shed light on what that relationship might be. A cursory look at the observed and expected numbers reveals that OE incidents appear to be over represented in categories A, B, and C while being underrepresented in category D incursions. The opposite is true for PD and V/PD incidents, which are underrepresented in categories A, B, and C and overrepresented in category D.

This pattern may be the result of one or more underlying processes. Firstly, the increased severity among OE incidents might merely be a function of the nature of OE incidents; in other words, OE incidents are naturally more dangerous. An alternative explanation is that controllers have been successfully trained to avoid category D incidents.[33] If controllers were trained to avoid category D incidents (i.e., relatively minor incidents) the remaining incidents would be the more severe incidents. Under this scenario, the rate of OE category A, B, and C incursions is natural, but the rate of OE category D incursions is artificially low. This would be consistent with the observations in Table 1. Yet a third possibility is that controller actions are always double-checked by the pilot. That is, each command given by a controller must be enacted by a pilot. That pilot has the ability to error check those commands and perhaps forestall the least dangerous situations (such as turning onto a closed runway). This is in contrast to pilots who are able to take actions without someone double-checking them, such as rolling over a hold short line or turning onto a runway without contacting the tower.

While some of the causes suggested above might be more or less likely, it is important to note that there may be multiple explanations. The results presented in Table 1 indicate that additional research is required to understand the true nature of the relationship between incident type and incident severity. Results presented later in this paper may help focus research on why OE incidents may be more severe than other incident types.

Table 1 indicated that there is a relationship between severity and incident type. Table 3 further explores this focusing on OE events. The results presented in Table 3 represent the impact of an incident being categorized as OE on severity. As with all regression results, it is important to note that these results represent correlation rather than causation.

Table 3 – Logit Estimate of Impact on Severity, OE Incident

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB
OE Incident	**3.45**	**.446**	**0.00**	**2.67**	**4.44**

32 See Appendix C.1 for more information on calculating chi-squared statistics.

33 Given that much of the focus on runway incursions is centered on the idea that preventing small mistakes will cascade into prevention of larger mistakes, training focusing on Category D-type incidents may have been a reasonable practice.

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB

Table 3 presents the results of the logit output in terms of odds ratios.[34] As described in Table 3 the odds of a severe incident are approximately 3.4 times as high for OE events as for non-OE events. This is in accordance with the results seen in Table 1, but is a more precise measure of how much more likely OEs are to be severe.

Table 4 presents the same information, but restricted to only conflict events. Here the alternative to "severe" is category C rather than both categories C and D. The effect of being an OE still persists, though in reduced magnitude.

Table 4 – Logit Estimate of Impact on Severity, OE Incident, Conflict Only

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB
OE Incident	**1.37**	**.180**	**0.02**	**1.06**	**1.78**

The pattern of incursions across regions is also informative. Table 5 and Table 6 present the breakdown of incident type by region while Table 7 and Table 8 present the breakdown of incident severity by region. While the above results presented in Table 1 indicate that there is a relationship between incident type and severity, it is difficult to control for such relationships in a two-way table.

Table 5 – Observed Incident Type Distribution by Region

	AAL Alaska	ACE Central	AEA Eastern	AGL Great Lakes	ANE New England	ANM Northwest Mountain	ASO Southern	ASW Southwest	AWP Western Pacific	Total
OE	**27**	**35**	**194**	**248**	**50**	**111**	**250**	**130**	**223**	**1,268**
PD	**174**	**265**	**429**	**775**	**204**	**495**	**998**	**545**	**1,417**	**5,302**
V/PD	**200**	**76**	**218**	**426**	**53**	**167**	**335**	**245**	**522**	**2,242**
Total	**401**	**376**	**841**	**1,449**	**307**	**773**	**1,583**	**920**	**2,162**	**8,812**

34 This is a direct transformation of the raw logit coefficients to aid interpretation. The odds of an event are defined as the ratio of that event happening to that event not happening. For example, the odds of seeing heads on a coin toss are 1:1 (or just 1). If an event has a probability of happening of 25% the odds are 1:3 (or 1/3). Conversely, if an event has a probability of happening of 75% the odds are 3:1 (or 3). The odds ratio as it is presented in Table 3 is just a measure of how the odds change when that dependent variable changes. In this case, as the dependent variable is either 0 (not an OE) or 1 (OE event) it is merely the ratio of odds of being severe between non-OE and OE events. The 95% CI LB and 95% CI UB cells represent the lower and upper bounds of the 95-percent confidence interval surrounding the estimated odds ratio.

Chi2 score: 300.01

Degrees of Freedom: 16

P-value: 0.00

Table 6 – Expected Incident Type Distribution by Region

	AAL Alaska	ACE Central	AEA Eastern	AGL Great Lakes	ANE New England	ANM Northwest Mountain	ASO Southern	ASW Southwest	AWP Western Pacific	Total
OE	58	54	121	209	44	111	228	132	311	1,268
PD	241	226	506	872	185	465	952	554	1,301	5,302
V/PD	102	96	214	369	78	197	403	234	550	2,242
Total	401	376	841	1,449	307	773	1,583	920	2,162	8,812

Table 7 – Observed Severity Distribution by Region

	AAL Alaska	ACE Central	AEA Eastern	AGL Great Lakes	ANE New England	ANM Northwest Mountain	ASO Southern	ASW Southwest	AWP Western Pacific	Total
A	1	2	18	23	1	13	35	7	32	132
B	2	4	20	19	6	8	29	12	45	145
C	105	123	344	522	123	276	619	337	859	3,308
D	293	247	459	885	177	476	900	564	1,226	5,227
Total	401	376	841	1,449	307	773	1,583	920	2,162	8,812

Chi2 score: 83.30

Degrees of Freedom: 24

P-value: 0.00

Table 8 – Expected Severity Distribution by Region

	AAL Alaska	ACE Central	AEA Eastern	AGL Great Lakes	ANE New England	ANM Northwest Mountain	ASO Southern	ASW Southwest	AWP Western Pacific	Total
A	6	6	13	22	5	12	24	14	32	13
B	7	6	14	24	5	13	26	15	36	14
C	151	141	316	544	115	290	594	345	812	3,30
D	238	223	499	860	182	459	939	546	1,282	5,22
Total	401	376	841	1,449	307	773	1,583	920	2,162	8,81

The most striking feature of these tables is the Chi-Squared statistics rather than the individual cells. The test statistics indicate that the distribution by region is not uniform for incident type or incident severity. This is not surprising given the result of the relationship between severity and incident type noted above. There are likely a variety of causes of this discrepancy, such as varying traffic patterns between regions and the prevalence of general aviation in each region. The overarching point is that any policy intervention will have differing impacts across regions.

3.3. Descriptive Statistics

The following section focuses on the analysis of various groups of variables. The groups of variables to be discussed include aircraft information, pilot information, controller information, weather information, and other variables. It is important to keep the overall distributions noted in the previous section in mind when examining these subsets of the data.

While formally a regression model, the logistic regressions (logits) presented in this section serve a role similar to the descriptive statistics above, a way to explore, rather than explain, the data. The results presented in this section focus on single variables with some examples of two or three variables at a time. The drawback of using these logit models is that the dependent variable must be dichotomized – destroying some information inherent to the rankings. It was chosen to examine severe (categories A and B) versus non-severe (categories C and D) events. To reiterate, these results serve more as data exploration and as a way to being to quantify the effect of various variables rather than as a formal modeling exercise. More formal modeling results are presented in Section 4.3.

3.3.1. Aircraft Information

Aircraft information originates from both the Runway Incursion and ATQA OE databases. All of the variables are of a categorical nature. These variables cover information about what the aircraft was doing at the time of the incident.

Intersecting Runway Departure or Arrival

(Runway Incursion Database)

The Runway Incursion database contains information on whether there was a departure or arrival on an intersecting runway. Figure 1 presents the distribution of this variable. Table 9 and Table 10 contain the cross tabulation of this variable by incident severity. Note that category D incursions were excluded (as by definition an event would not be a D if this variable was yes). This table includes the results of Fisher's Exact test. This is a similar test to the Chi-Squared test indicated above, and tests the same hypothesis (independence of row and column categories), but is applicable when some cells have very small values and the assumptions of the Chi-Squared test do not apply.[35]

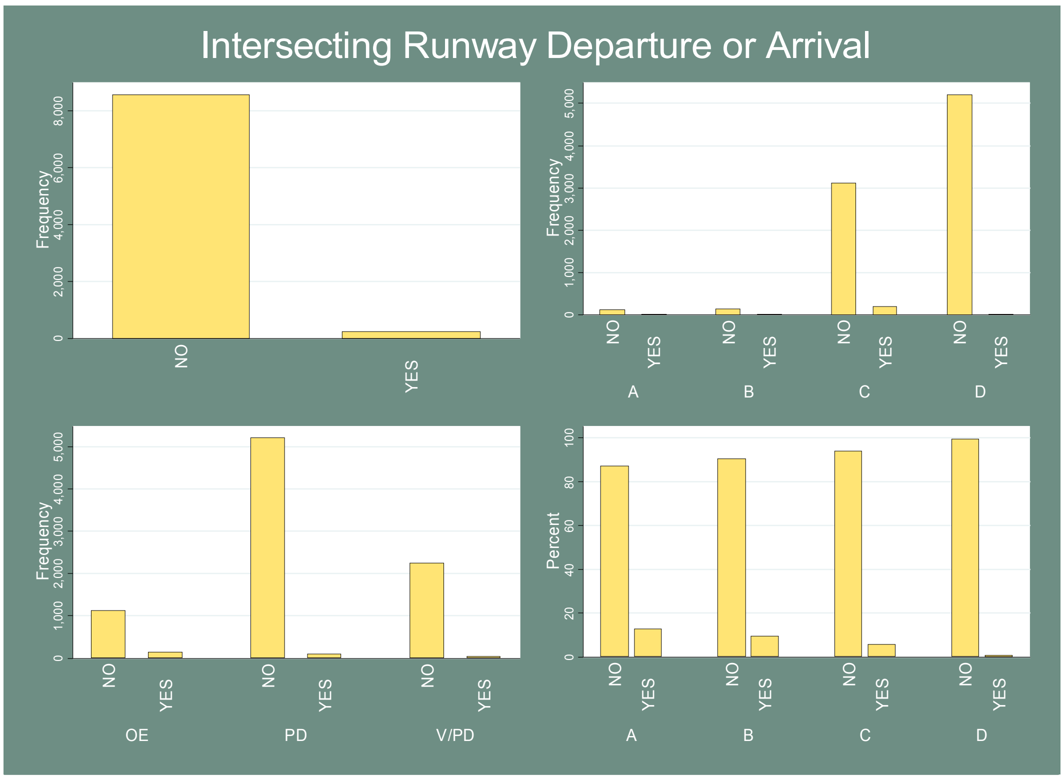

Figure 1 – Distribution of Intersecting Runway Departure or Arrival

Table 9 – Observed Distribution of Intersecting Runway Departure or Arrival by Severity

	A	B	C	Total
NO	**115**	**131**	**3,113**	**3,359**

35 While Fisher's Exact test and the Chi-Squared test are similar, they are best used in different situations. The Chi-Squared test relies on asymptotic assumptions to calculate the p-value while Fisher's Exact test calculates the p-value exactly; i.e. Fisher's Exact test is the non-parametric analogue to the Chi-Squared test. In this analysis, the Chi-Squared test was the preferred test. However, when the asymptotic assumptions seemed impractical (read: low expected values in a large fraction of table cells), Fisher's Exact test was employed. Further details on the calculation for Fisher's Exact test can be found in Rice (2007).

	A	B	C	Total
YES	**17**	**14**	**195**	**226**
Total	**132**	**145**	**3,308**	**3,585**

P-value: 0.00[36]

Table 10 – Expected Distribution of Intersecting Runway Departure or Arrival by Severity

	A	B	C	Total
NO	**124**	**136**	**3,099**	**3,359**
YES	**8**	**9**	**209**	**226**
Total	**132**	**145**	**3,308**	**3,585**

This table indicates that there is a relationship between these two variables. Examining the observed versus expected values indicates that incidents with departure or arrivals on intersecting runways occur more frequently than expected among category A and B incursions than among category C incursions. In some sense, this is not surprising given the definition of incursion severity – if there is an arrival or departure on an intersecting runway it is more likely that the two planes will come into conflict. Given that, this result indicates that these events are more severe than other conflict events.

Table 9 indicated that there was a relationship between this variable and severity. Table 11 presents the results in terms of odd ratios. Again, category D incursions are excluded for definitional reasons.

Table 11 – Logit Estimate of Impact on Severity, Intersecting Runway or Departure

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB
Intersecting Runway Departure or Arrival	**2.01**	**.411**	**0.00**	**1.35**	**3.00**

The results suggest that the odds of being severe for incidents with an operation on an intersecting runway are approximately twice as large as those without. Again, this is consistent with Table 9, but is a more quantitative look at this relationship.

36 In general, the tables presented in this section follow the convention of presenting only the P-value when Fisher's Exact was performed. When a Chi Squared test was performed, the Chi Squared test as well as its statistic will be presented.

Table 12 presents the results of a logit where the dependent variable is a flag for an OE incident or not. Again, category D incursions were excluded for definitional reasons. Note that the alternative here is "not OE"; that is, both V/PD and PD incidents are included in the alternative. The odds of an incident being an OE are approximately 4.4 times as high if there is an operation on an intersecting runway. Recall that no V/PD incidents were coded as "yes" for this variable. This should temper the effect somewhat, as seen in Table 13.

Table 12 – Logit Estimate of Impact on Incident Type, Intersecting Runway or Departure

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB
Intersecting Runway Departure or Arrival	**4.44**	**.633**	**0.00**	**3.36**	**5.87**

Table 13 – Logit Estimate of Impact on Incident Type, Intersecting Runway or Departure, OE and PD Only

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB
Intersecting Runway Departure or Arrival	**3.40**	**.484**	**0.00**	**2.56**	**4.48**

Table 14 and Table 15 contain a cross tabulation of the same variable by incident type. Category D incursions are still excluded. Interestingly, intersecting runway departure or arrivals occur most frequently for OE incidents. The Chi-Squared statistic supports the conclusion that there is a relationship between these two variables. The observed values reveal two things. First, there is only one V/PD incident where this is variable is coded as yes. This suggests that airport vehicles are effectively never in a situation where this could be coded yes. Secondly, OEs are over represented while PDs are underrepresented. This relationship holds even when V/PDs are excluded from the analysis, as seen in Table 16 and Table 17. This indicates that intersecting runway departures or arrivals are proportionally less a problem for pilots than controllers. Further research is required in this area to detail why that is the case.

Table 14 – Observed Distribution of Intersecting Runway Departure or Arrival by Incident Type

	OE	PD	V/PD	Total
NO	**901**	**1,876**	**582**	**3,359**
YES	**140**	**86**	**0**	**226**
Total	**1,041**	**1,962**	**582**	**3,585**

Chi2 score: 141.38

Degrees of Freedom: 2

P-value: 0.00

Table 15 – Expected Distribution of Intersecting Runway Departure or Arrival by Severity

	OE	PD	V/PD	Total
NO	**975**	**1,838**	**545**	**3,359**
YES	**66**	**124**	**37**	**226**
Total	**1,041**	**1,962**	**582**	**3,585**

Table 16 – Observed Distribution of Intersecting Runway Departure or Arrival by Incident Type, OE & PD

	OE	PD	Total
NO	**901**	**1,876**	**2,777**
YES	**140**	**86**	**226**
Total	**1,041**	**1,962**	**3,003**

Chi2 score: 80.31

Degrees of Freedom: 1

P-value: 0.00

Table 17 – Expected Distribution of Intersecting Runway Departure or Arrival by Severity, OE & PD

	OE	PD	Total
NO	**963**	**1,814**	**2,777**
YES	**78**	**148**	**226**
Total	**1,041**	**1,962**	**3,003**

Landed or Departed on Closed Taxiway or Runway

(Runway Incursion Database)

Figure 2 presents the overall distribution of this variable. Table 18 and Table 19 present a cross tabulation of this variable by severity. This table excludes incidents that were classified as V/PD incidents. The definition of this variable is meaningless in the context of a V/PD, as vehicles cannot land or takeoff; additionally, only one V/PD was coded as yes on this variable. Additionally, Table 18 contains the output of Fisher's Exact test.

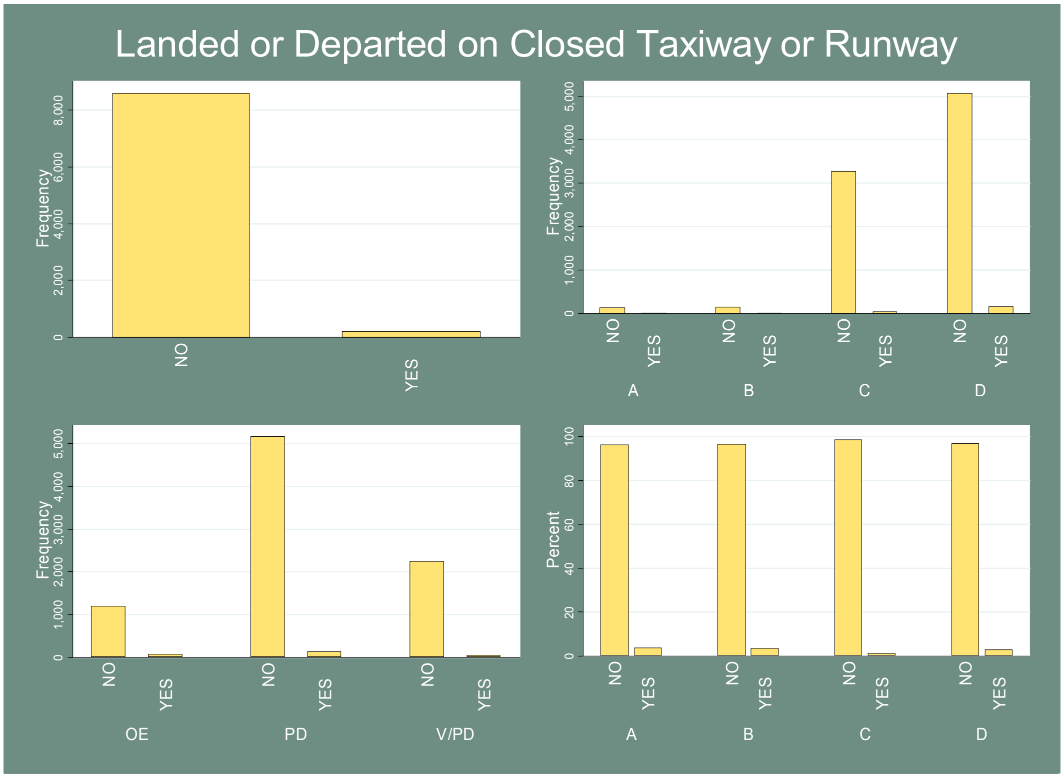

Figure 2 – Distribution of Landed or Departed on Closed Taxiway or Runway

Table 18 – Observed Distribution of Landed or Departed on Closed Taxiway or Runway by Severity

	A	B	C	D	Total
NO	**111**	**117**	**2,725**	**3,412**	**6,365**
YES	**5**	**5**	**40**	**155**	**205**
Total	**116**	**122**	**2,765**	**3,567**	**6,570**

P-value: 0.00

Table 19 – Expected Distribution of Landed or Departed on Closed Taxiway or Runway by Severity

	A	B	C	D	Total
NO	**112**	**118**	**2,679**	**3,456**	**6,365**
YES	**4**	**4**	**86**	**111**	**205**

	A	B	C	D	Total
Total	**116**	**122**	**2,765**	**3,567**	**6,570**

The results clearly indicate a relationship between this variable and severity. The expected values indicate that categories A, B, and D are overrepresented while category C is underrepresented. A possible interpretation of this split is that, while landing or departing on a closed taxiway or runway is a dangerous action, the definition of category D precludes a higher rating if there is no other aircraft around. That is, landing or departing on a closed taxiway or runway is inherently quite dangerous. When another aircraft is nearby, this becomes a severe conflict event (category A or B). If no other plane is nearby, the event is rated a D, despite the inherent danger of the action. This is only one possible explanation; further testing is required to rule out or confirm this hypothesis.

Table 20 and Table 21 present the breakdown of this variable by incident type. Note that again V/PDs have been excluded for the reasons noted above. Table 20 also includes the results of a Chi-Squared test. The test indicates that there is a relationship between this variable and the type of incident. OE incidents are observed more frequently than one would expect.

Table 20 – Observed Distribution of Landed or Departed on Closed Taxiway or Runway by Incident Type

	OE	PD	Total
NO	**1,200**	**5,165**	**6,365**
YES	**68**	**137**	**205**
Total	**1,268**	**5,302**	**6,570**

Chi2 score: 26.14

Degrees of Freedom: 1

P-value: 0.00

Table 21 – Expected Distribution of Landed or Departed on Closed Taxiway or Runway by Incident Type

	OE	PD	Total
NO	**1,228**	**5,137**	**6,365**
YES	**40**	**165**	**205**
Total	**1,268**	**5,302**	**6,570**

This variable brings up another important issue. While OE incidents occur twice as often, proportionally, the baseline for comparison is important. Table 20 presents the universe of PD and OE runway incursions. Comparing the *observed* rate to the total number of events indicates if this is a larger fraction of observed events for one group or another, but says little about the *error* rate. In terms of this variable, pilot incursions occur roughly twice as often as controller errors. However, that comparison is conditional on all the incursions that have occurred. There is no information available about how often pilots or controllers are presented with an opportunity to commit this error, which may be the more appropriate basis for comparison rather than number of incursions. One possible way to address this issue is to identify the number of operations per individual. Throughout the day, pilots are presented with far fewer opportunities to land an aircraft on a closed runway than a controller might be, and further research needs to account for this.

Landed or Departed Without Clearance Communication

(Runway Incursion Database)

Figure 3 presents the overall distribution of this variable. Table 22 and Table 23 present a cross tabulation of this variable by severity. V/PD incidents are again excluded from the analysis as this variable makes little sense in that context. For reference, zero V/PD incidents were coded yes on this variable.

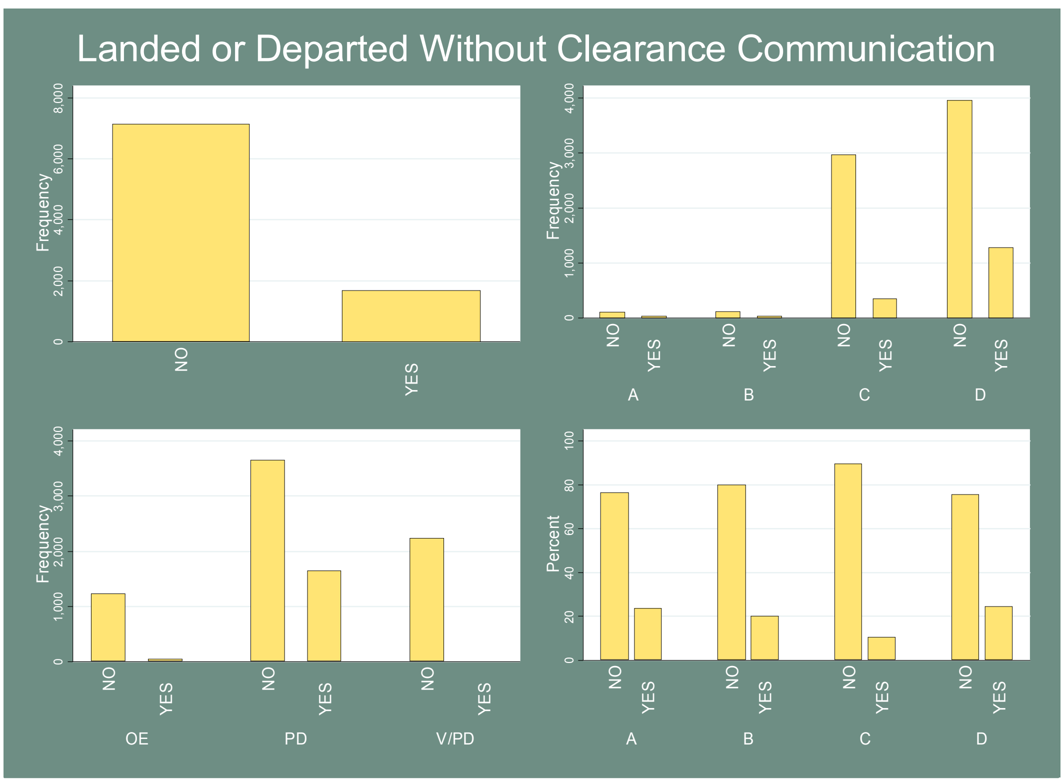

Figure 3 – Distribution of Landed or Departed Without Clearance Communication

Table 22 – Observed Distribution of Landed or Departed Without Clearance Communication by Severity

	A	B	C	D	Total
NO	85	93	2,417	2,286	4,881
YES	31	29	348	1,281	1,689
Total	116	122	2,765	3,567	6,570

Chi2 score: 444.07

Degrees of Freedom: 3

P-value: 0.00

Table 23 – Observed Distribution of Landed or Departed Without Clearance Communication by Severity

	A	B	C	D	Total
NO	86	91	2,054	2,650	4,881
YES	30	31	711	917	1,689
Total	116	122	2,765	3,567	6,570

Again, the test statistics indicate that there is a relationship between severity and this variable. A similar pattern to that seen for landing or departing on a closed runway or taxiway is seen: category D is observed more frequently than expected while the opposite is true for category C. A similar explanation of the pattern can be hypothesized for this variable as well. Table 24 and Table 25 presents the same cross tab, but examine conflict events only.

Table 24 – Observed Distribution of Landed or Departed Without Clearance Communication by Severity, Conflict Only

	A	B	C	Total
NO	101.0	116.0	2,960.0	3,177.0
YES	31.0	29.0	348.0	408.0
Total	132.0	145.0	3,308.0	3,585.0

Chi2 score: 32.29

Degrees of Freedom: 2

P-value: 0.00

Table 25 – Expected Distribution of Landed or Departed without Clearance Communication by Severity, Conflict Only

	A	B	C	Total
NO	**117**	**128**	**2,932**	**3,177**
YES	**15**	**17**	**376**	**408**
Total	**132**	**145**	**3,308**	**3,585**

Excluding Ds from the analysis removes the conflict/non-conflict event dynamic: Categories A, B, and C are all conflict events. The Chi-Squared test again indicates a relationship between these variables. Given the expected values, it appears that this variable may increase severity, once the presence of a second aircraft is controlled for.

Table 26 presents the estimate of the odds ratio with respect to severity for this variable.

Table 26 – Logit Estimate of Impact on Severity, Landed or Departed without Clearance Communication

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB
Landed or Departed Without Clearance Communication	**.973**	**.148**	**0.86**	**.722**	**1.31**

Contrary to the results presented in Table 22, there is no increase in the likelihood of a severe event given that an aircraft landed or departed without clearance. This is likely due to the loss of information from consolidating the severity categories. Table 27 presents the same regression, excluding category D events (i.e., removing the conflict/non-conflict dynamic). The relationship seen in Table 22 is now clearly visible, indicating that incidents where an aircraft landed or departed without clearance have odds approximately 2.3 times larger of being severe.

Table 27 – Logit Estimate of Impact on Severity, Landed or Departed without Clearance Communication, Conflict Only

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB
Landed or Departed Without Clearance Communication	**2.34**	**.374**	**0.00**	**1.71**	**3.20**

Table 28 presents the results of a logit where the dependent variable is whether or not the incident was an OE. AS V/PDs were excluded, the alternative here is PD; thus, the odds ratio indicates the increase (or decrease) in the likelihood of being an OE compared to a PD. The results indicate that incidents where an aircraft landed or departed without clearance are dramatically less likely to be OEs. This is not surprising given the nature of the error.

Table 28 – Logit Estimate of Impact on Incident Type, Landed or Departed Without Clearance Communication

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB
Landed or Departed Without Clearance Communication	**.068**	**.011**	**0.00**	**.049**	**.095**

Taxiing Out for Departure

(Runway Incursion Database)

This variable indicates whether the primary aircraft was taxiing out for departure or not. Observations coded no may be in any other phase of flight. Figure 4 presents the overall distribution of this variable. Table 29 and Table 30 present the breakdown of this variable by severity.

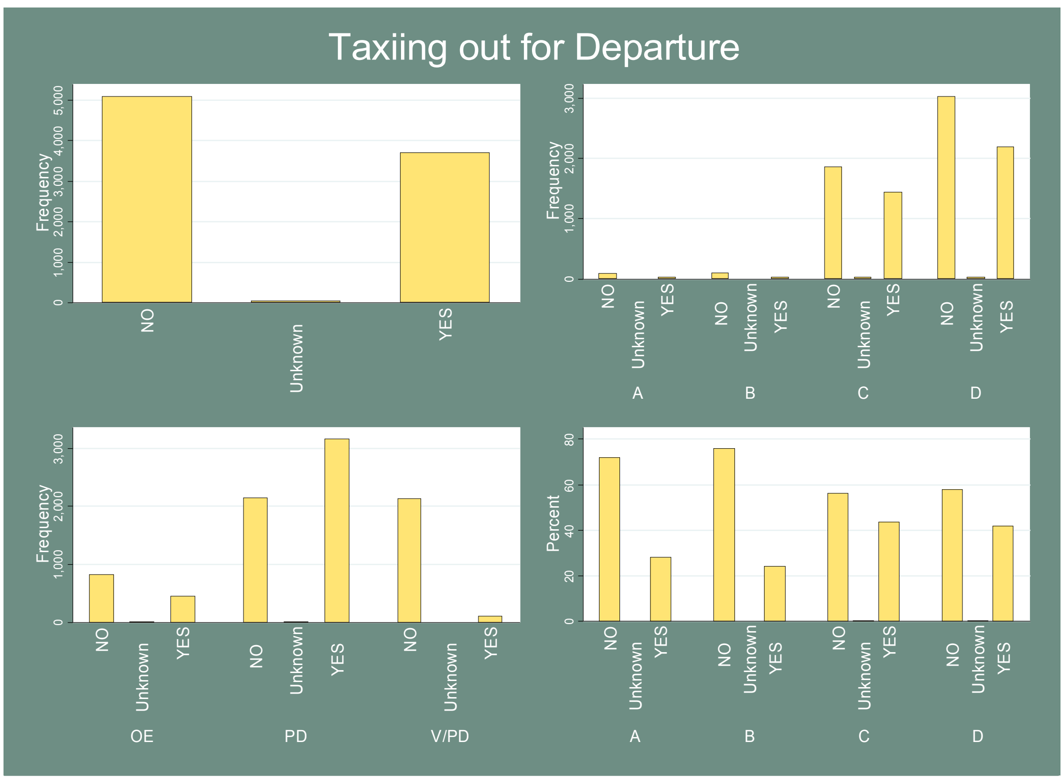

Figure 4 – Distribution of Taxiing Out for Departure

Table 29 – Observed Distribution of Taxiing Out for Departure by Severity

	A	B	C	D	Total
NO	95	110	1,863	3,031	5,099
YES	37	35	1,443	2,195	3,710
Total	132	145	3,306	5,226	8,809

Chi2 score: 33.18

Degrees of Freedom: 3

P-value: 0.00

Table 30 – Expected Distribution of Taxiing Out for Departure by Severity

	A	B	C	D	Total
NO	76	84	1,914	3,025	5,099
YES	56	61	1,392	2,201	3,710
Total	132	145	3,306	5,226	8,809

The Chi-Squared statistic indicates that there is a relationship between this variable and severity. The expected values indicate that conflict events are underrepresented while category D events are observed more often than expected. This may be indicative of the kind of behavioral errors with which this variable is associated. For example, if taxiing aircraft rarely interact with other aircraft on a runway (i.e. only when the taxiing aircraft is crossing the runway), any given error is more likely to be a D than any other category.[37]

Table 31 and Table 32 present the breakdown of this variable by incident type. V/PDs are dramatically underrepresented when compared with the expected value. This is likely an indication that vehicles on aircraft grounds are rarely near aircraft that are taxiing out for departure. This variable is coded yes more frequently (both in relative and absolute terms) for PD incidents than OE incidents. Again, without the proper baseline (total taxi operations by group) it is hard to tell if one group is committing the error more than the other; however, given that there is an error, this appears to be more associated with pilots than controllers.

37 As a reminder, this research examines only runway incursions. If an incident occurred while taxiing, but did not involve a runway (such as a collision on taxiways or crossing a hold short line at a taxiway intersection), it would not be reported in the dataset used for this analysis.

Table 31 – Observed Distribution of Taxiing Out for Departure by Incident Type

	OE	PD	V/PD	Total
NO	821	2,143	2,135	5,099
YES	446	3,157	107	3,710
Total	1,267	5,300	2,242	8,809

Chi2 score: 1969.36

Degrees of Freedom: 2

P-value: 0.00

Table 32 – Expected Distribution of Taxiing Out for Departure by Incident Type

	OE	PD	V/PD	Total
NO	733	3,068	1,298	4,366
YES	534	2,232	944	3,176
Total	1,267	5,300	2,242	7,542

Land and Hold Short

(Runway Incursion Database)

This variables codes for whether or not there was a land and hold short operation in effect for one of the aircraft involved in the incident. It is important to keep in mind the overall low frequency of errors involving LAHSO, there are only 17 such incursions. Consequently, it is difficult to draw any strong conclusions regarding incident severity; however, that no category A or B incidents occurred during a LAHSO. All 17 incidents were category C or D (16 Cs and 1 D). Figure 5 presents the overall distribution of this variable. Table 33 and Table 34 present the frequency of this variable by incident type. The test statistic indicates that there is a relationship between these variables, and OEs appear to be overrepresented. Without information on the number of LAHSOs that do not result in a runway incursion, it is difficult to determine the appropriate baseline rate of comparison.

Figure 5 – Distribution of Land and Hold Short

Table 33 – Observed Distribution of Land and Hold Short by Incident Type

	OE	PD	Total
NO	1,258	5,295	6,553
YES	10	7	17
Total	1,268	5,302	6,570

P-value: 0.00

Table 34 – Expected Distribution of Land and Hold Short by Incident Type

	OE	PD	Total
NO	1,265	5,288	6,553
YES	3	14	17
Total	1,268	5,302	6,570

Table 35 provides an estimate of the increase in the odds of being an OE if an event occurs during a LAHSO. Note that V/PDs were excluded from this regression to be consistent with Table 33. While the effect is fairly large in magnitude, it is also imprecise given the lower frequency of LAHSOs in the dataset.

Table 35 – Logit Estimate of Impact on Incident Type, Land and Hold Short Operation

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB
LAHSO	**6.01**	**2.97**	**0.00**	**2.28**	**15.8**

Evasive Action Taken

(ATQA OE)

This variable codes for whether or not the aircraft took evasive action.[38] This variable only applies when the incursion involves two aircraft (or an aircraft and a vehicle) so the relevant set is only category A, B, and C incursions[39]. Figure 6 presents the distribution of this variable. Table 36 and Table 37 present the breakdown of this variable by severity. This variable originates in the ATQA OE dataset, so is relevant only to OE incidents.

---

38 A definition of evasive action is not provided, either in the database or on the reporting form. Thus, it is unclear what the threshold for this variable to be coded as “yes” is.

39 This limitation is not just based on intuition. There are no incidents coded yes on this variable and as Category D.

Figure 6 – Distribution of Evasive Action Taken

Table 36 – Observed Distribution of Evasive Action Take by Severity

	A	B	C	Total
No	**28**	**27**	**687**	**742**
Unknown	**7**	**3**	**66**	**76**
Yes	**13**	**9**	**91**	**113**
Total	**48**	**39**	**844**	**931**

P-value: 0.00

Table 37 – Expected Distribution of Evasive Action Taken by Severity

	A	B	C	Total
No	**38**	**31**	**673**	**742**
Unknown	**4**	**3**	**69**	**76**

	A	B	C	Total
Yes	**6**	**5**	**102**	**113**
Total	**48**	**39**	**844**	**931**

Categories A and B appear to be observed more frequently than statistically expected. Intuition suggests that aircraft that have to take evasive action are in more dangerous situations. There is a possibility that evasive action may be taken into account with the definitions of categories A and B. Table 38 and Table 39 present the breakdown among category A and B only. The test statistic indicates that there is no relationship between the variable and severity. Combining this with the results from Table 36 indicate that evasive action helps distinguish between category C and the remaining two categories, rather than uniformly increasing severity.

Table 38 – Observed Distribution of Evasive Action Taken by Severity, A and B Only

	A	B	Total
No	**28**	**27**	**55**
Unknown	**7**	**3**	**10**
Yes	**13**	**9**	**22**
Total	**48**	**39**	**87**

Table 39 – Expected Distribution of Evasive Action Taken by Severity, A and B Only

	A	B	Total
No	**30**	**25**	**55**
Unknown	**6**	**4**	**10**
Yes	**12**	**10**	**22**
Total	**48**	**39**	**87**

Phase of Flight

(Runway Incursion Database)

The Runway Incursion database contains information on the phase of flight of the primary aircraft involved at the time of the incident. The three possibilities are taxiing, takeoff, and landing. Table 40 presents the results of a simple logit looking at the impact on the odds of being severe.

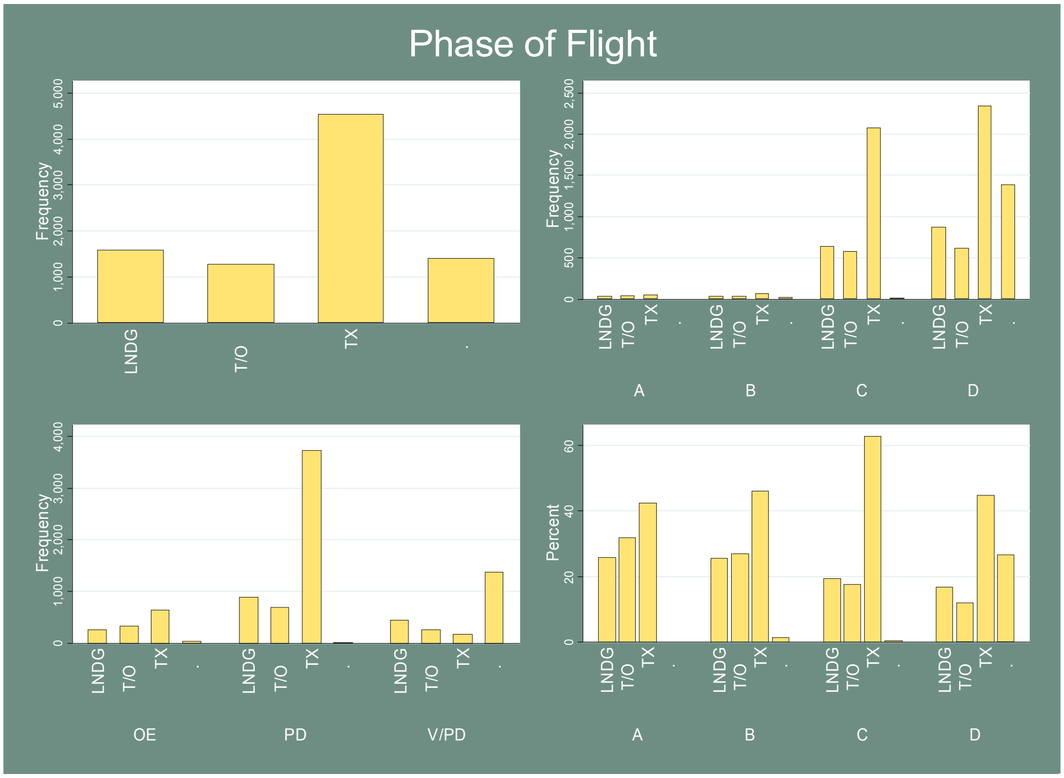

Figure 7 – Distribution of Phase of Flight

Table 40 – Logit Estimate of Impact on Severity, Phase of Flight

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB
Landing	**1.68**	**.255**	**0.00**	**1.25**	**2.26**
Takeoff	**2.43**	**.357**	**0.00**	**1.82**	**3.24**

The baseline for comparison is a taxiing aircraft. It appears that both takeoff and landing tend to be more severe than taxiing aircraft. While not an inherently surprising result, the disparity between takeoff and landing is interesting. Takeoff appears to be the more dangerous of the two situations compared to taxiing. Perhaps this has to do with acceleration versus deceleration of the aircraft (i.e., aircraft taking off are in general moving faster towards a potential collision while landing aircraft are already breaking as part of the landing procedure).

Table 41 presents the results for the odds of being an OE incident. Again, both takeoff and landing have higher odds than taxiing. But the magnitude of the impact is not as important as the disparity between takeoff and landing. It appears that an incident involving an aircraft taking off is more likely to be a controller error than an incident involving a landing aircraft. Naively, one might have assumed that the impacts would be the same. This may have implications for controller processes or training, pending the results of a more in depth study of this issue.

Table 41 – Logit Estimate of Impact on Incident Type, Phase of Flight

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB
Landing	**1.18**	**.095**	**0.04**	**1.01**	**1.38**
Takeoff	**2.13**	**.163**	**0.00**	**1.83**	**2.47**

Finally, Table 42 presents a crude model that controls for the effect of phase of flight and its interaction with incident type. Phase of flight and incident type appear to have independent effects on severity. The magnitude of the effects appears consistent with that seen in their separate estimates, though the exact values have shifted slightly. Lastly, there is no interaction between phase of flight and incident type; they are merely the sum of their parts.[40]

Table 42 – Logit Estimate of Impact on Severity, Incident Type and Phase of Flight

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB
Landing	**1.56**	**.290**	**0.02**	**1.09**	**2.25**
Takeoff	**2.14**	**.403**	**0.00**	**1.48**	**3.09**
OE Incident	**2.59**	**.526**	**0.00**	**1.74**	**3.86**
Landing and OE Incident	**1.15**	**.377**	**0.68**	**.602**	**2.18**
Takeoff and OE Incident	**.987**	**.305**	**0.97**	**.539**	**1.81**

Commercial Carrier

(Runway Incursion Database, ATQA)

Given the more stringent requirements for pilots on commercial carriers, it may be the case that they are less likely to be involved in serious incidents. Additionally, commercial carrier pilots are flying into different airports than the majority of GA pilots. For the purposes of this analysis, a commercial carrier is any carrier not flying under GA regulations (part 91), military regulations, or conducting on demand operations (part 135). This essentially divides the population into scheduled carriers (domestic and

40 Note that odds ratios are multiplicative. In this case, the combined impact on the odds of a severe incident of an OE involving an aircraft taking off is approximately 5.5.

foreign) and other carriers.[41] Table 43 presents the distribution of this variable. Table 44 presents the impact of this categorization on the odds of a severe event.

Table 43 – Distribution of Commercial Carrier Status

	A	B	C	D	Total
NO	**106**	**112**	**2,305**	**3,141**	**5,664**
YES	**26**	**30**	**955**	**611**	**1,622**
Total	**132**	**142**	**3,260**	**3,752**	**7,286**

Table 44 – Logit Estimate of Impact on Severity, Commercial Carrier Status

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB
Commercial Carrier	**.893**	**.136**	**0.46**	**.6612**	**1.20**

When considering all incident categories, there is no impact from being a commercial carrier. When considering only conflict events, as shown in Table 45, the relationship becomes more pronounced. This disparity between conflict and non-conflict events is not unusual, likely indicating that commercials carriers (as defined above) are in situations where category D events can occur less frequently. Conflict versus non-conflict aside, commercial carriers still appear to be involved in less severe incidents, reducing the odds of a severe incursion by almost 40%. This may be due to pilot experience, as noted above. A focused research effort examining issues such as pilot training, pilot experience, familiarity with the airport, total pilot hours, and other factors could help explain the origin of this fairly large effect.

Table 45 – Logit Estimate of Impact on Severity, Commercial Carrier Status, Conflict Only

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB
Commercial Carrier	**.620**	**.096**	**0.00**	**.458**	**.840**

Finally, Table 46 presents the interaction between commercial carrier and incident type on severity. Interestingly, the impact of commercial carrier flag is not a good indicator of severity, once incident type is accounted for. Some of the impact of incident type on severity is also diminished. The interaction between incident type and commercial carrier status is also interesting. This simplistic model suggests that while OE incidents in general are more severe, OE incidents involving commercial carriers tend to be less severe. This is possibly capturing some of the same factors as the OEP 35 flag (e.g., pilot experience, pilot familiarity) but it is interesting that the interaction exists for OE incidents but not PD

41 Note that some carriers under part 135 do in fact fly scheduled service. However, it is impossible to distinguish those part 135 aircraft that are scheduled from those that are not for the purposes of this analysis.

incidents. The mechanism through which commercial status interacts with OE incidents should be investigated further, but these results suggest it should be included in an OE focused model.

Table 46 – Logit Estimate of Impact on Severity, Commercial Carrier Status and Incident Type

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB
Commercial Carrier	**1.15**	**.408**	**0.70**	**.571**	**2.30**
OE Incident	**2.38**	**.585**	**0.00**	**1.47**	**3.85**
PD Incident	**.583**	**.136**	**0.02**	**.369**	**.919**
Commercial Carrier & OE	**.371**	**.160**	**0.02**	**.159**	**.866**
Commercial Carrier & PD	**.733**	**.316**	**0.47**	**.314**	**1.71**

Number of Aircraft Involved

(ATQA OE)

This variable measures the number of aircraft involved in an incident.[42] This variable is only available for OE incidents. Table 47 and Table 48 present the observed and expected frequencies of this variable. Note that category D incursions were excluded from this analysis. Recall that an incident is category D by definition if there is only one entity involved. Thus, there are no observed values of this variable other than one for category D incidents.

Table 47 – Observed Distribution of Number of Aircraft Involved by Severity

	A	B	C	Total
0	**0**	**0**	**1**	**1**
1	**4**	**7**	**124**	**135**
2	**43**	**31**	**704**	**778**
3	**1**	**1**	**14**	**16**
4	**0**	**0**	**1**	**1**
Total	**48**	**39**	**844**	**931**

P-value: 0.59

42 This variable measures aircraft specifically, though an incursion can be committed by an aircraft, another vehicle, a person, or an animal. Thus, this variable can take on values of zero or one and still be a conflict event due to the presence of non-aircraft entities.

Table 48 – Expected Distribution of Number of Aircraft Involved by Severity

	A	B	C	Total
0	0	0	1	1
1	7	6	122	135
2	40	33	705	778
3	1	1	15	16
4	0	0	1	1
Total	48	39	844	931

Note that the majority of incidents involve one or two aircraft. However, there does not appear to be a relationship between severity and the number of aircraft involved (except in category D).

3.3.2. Pilot Information

This information describes the pilot involved in the incident. This information comes from the Runway Incursion and ATQA PD databases. Some variables may only pertain to PD incidents, which are noted in the variable specific discussion. Some variables are categorical while others are continuous.

Foreign Aircraft or Pilot

(Runway Incursion Database)

This variable indicates whether or not the pilot or aircraft involved were of foreign origin. Table 49 and Table 50 provide the breakdown of this variable by severity. Figure 8 presents the overall distribution of this variable.

Figure 8 – Distribution of Foreign Pilot Status

The distribution is weighted towards the conflict categories of A, B and C. The test statistic indicates that there is a relationship between these two variables. An underlying cause may be that most foreign pilots or aircraft entering the United States are commercial. Commercial pilots are (generally) at busier airports, and so are less likely to be in a category D due to the increased traffic at the airport. Because of the strong relationship between foreign pilot status and commercial carrier status, it is difficult to draw strong conclusions about the effect of foreign pilot status on severity.

Table 49 – Observed Distribution of Foreign Aircraft or Pilot by Severity

	A	B	C	D	Total
NO	129	141	3,208	5,119	8,597
YES	3	4	100	108	215
Total	132	145	3,308	5,227	8,812

Table 50 – Expected Distribution of Foreign Aircraft or Pilot by Severity

	A	B	C	D	Total
NO	**129**	**141**	**3,227**	**5,099**	**8,597**
YES	**3**	**4**	**81**	**128**	**215**
Total	**132**	**145**	**3,308**	**5,227**	**8,812**

Pilot Lost

(ATQA PD)

This variable indicates whether the investigation determined if the pilot was lost at the time of the incident. Figure 9 presents the overall distribution of this variable. Table 51 and Table 52 present a tabulation of this variable by severity. Fisher's Exact test indicates that there is no relationship between these variables. While not entirely surprising, this does indicate that, at least at a cursory level, pilots being lost on the airfield does not increase the severity of an ensuing incident. It may, however, increase the likelihood of an incident occurring at all; this cannot be tested without "normal operations" data for all non-incident operations.

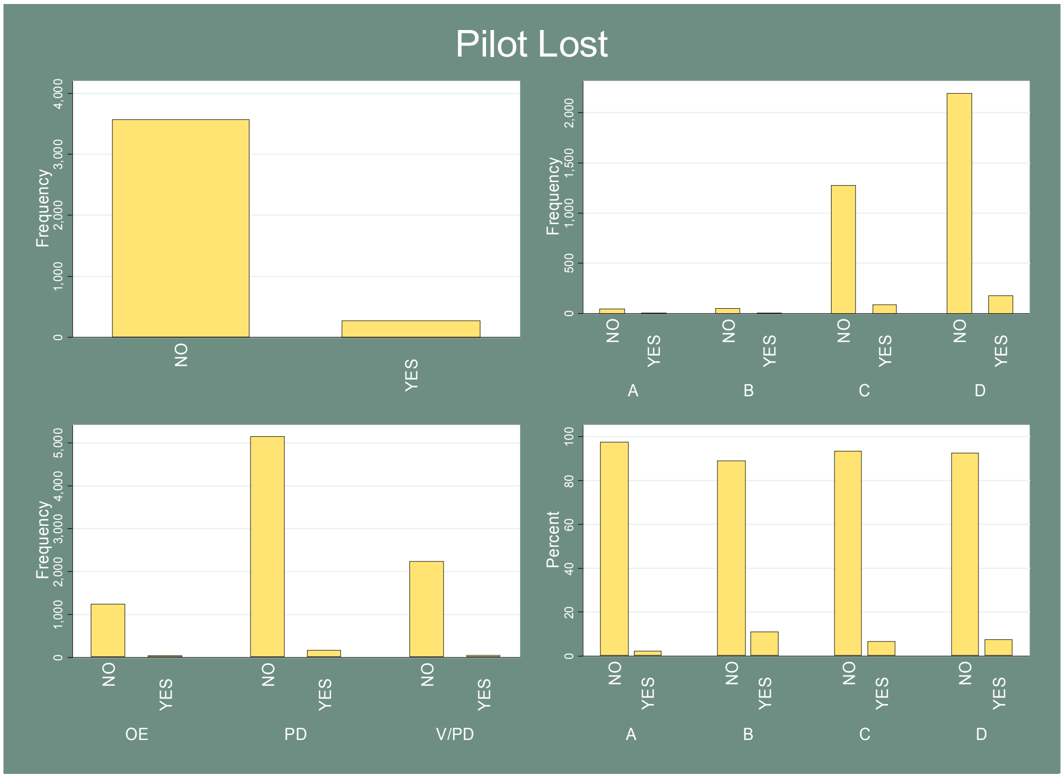

Figure 9 – Distribution of Pilot Lost

Table 51 – Observed Distribution of Pilot Lost by Severity

	A	B	C	D	Total
NO	41	49	1,277	2,196	3,563
YES	1	6	89	177	273
Total	42	55	1,366	2,373	3,836

P-value: 0.30

Table 52 – Expected Distribution of Pilot Lost by Severity

	A	B	C	D	Total
NO	39	51	1,269	2,204	3,563
YES	3	4	97	169	273
Total	42	55	1,366	2,373	3,836

Pilot Ratings

(ATQA PD)

The ATQA PD database contains information on pilot ratings. These ratings include: single engine sea, single engine land, multiengine sea, multiengine land, rotorcraft, glider, lighter than air, and other. The sea and land ratings for multiengine and single engine categories were combined due to the low number of sea plane ratings in the dataset. The distribution of response after the combination of sea and land ratings can be seen in Figure 10. Chi-Squared tests were run for each category separately; the majority of the categories have no significant relationship with severity. The only category that presented a marginally significant result was the multiengine rating category. The results of this test are presented in Table 53 and Table 54. The pattern of expected versus observed is unclear. The major contribution to the test statistic appears to be from the overrepresentation of category C incursions. This may be an artifact of the distribution of multiengine rating in the population; that is, pilots with multiengine ratings may fly into busier airports and thus would be more likely to be in a conflict event.

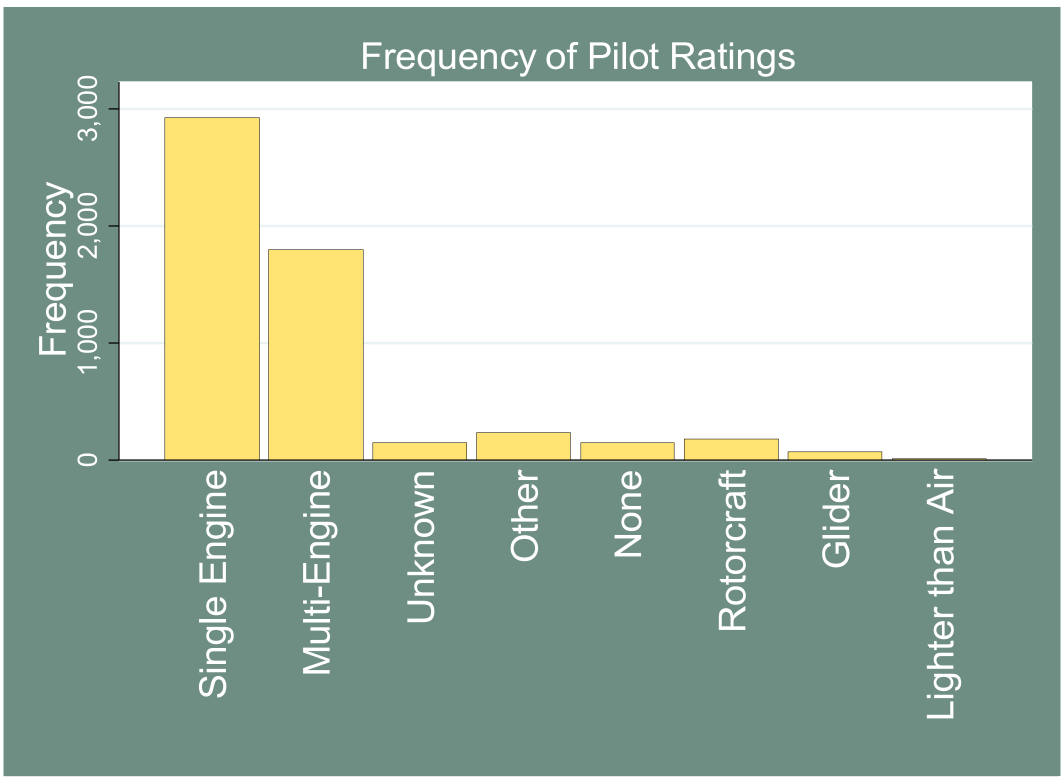

Figure 10 – Frequency of Pilot Ratings by Rating Category

Table 53 – Observed Distribution of Multiengine Rating by Severity

	A	B	C	D	Total
NO	22	36	693	1,287	2,038

	A	B	C	D	Total
YES	**20**	**19**	**673**	**1,086**	**1,798**
Total	**42**	**55**	**1,366**	**2,373**	**3,836**

Chi2 score: 7.68

Degrees of Freedom: 3

P-value: 0.05

Table 54 – Expected Distribution of Multiengine Rating by Severity

	A	B	C	D	Total
NO	**22**	**29**	**726**	**1,261**	**2,038**
YES	**20**	**26**	**640**	**1,112**	**1,798**
Total	**42**	**55**	**1,366**	**2,373**	**3,836**

Entered Runway without Clearance

(Runway Incursion Database)

If the primary aircraft in the event entered a runway without clearance, this variable is coded yes. The Chi-Squared statistic, contained in Table 55 and Table 56, indicates that there is a relationship between this variable and severity. Category C is underrepresented while all other categories are over represented.

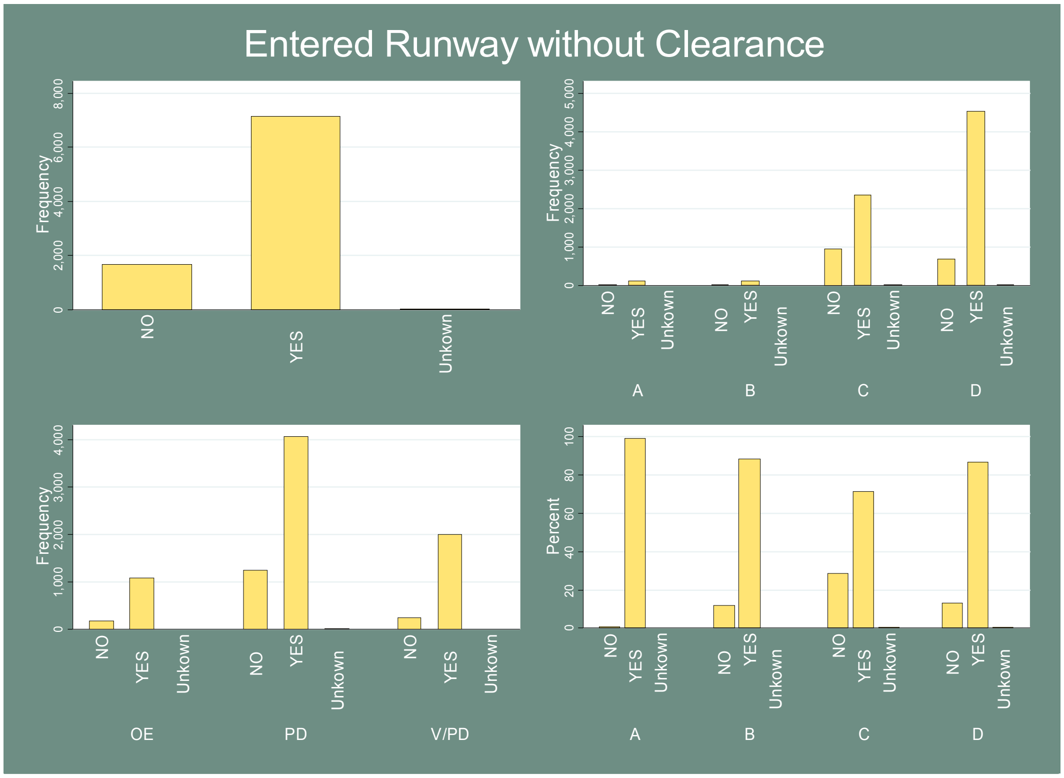

Figure 11 – Distribution of Entered Runway without Clearance

Table 55 – Observed Distribution of Entered Taxiway or Runway without Clearance by Severity

	A	B	C	D	Total
NO	1	17	950	696	1,664
YES	131	128	2,357	4,530	7,146
Total	132	145	3,307	5,226	8,810

Chi2 score: 347.97

Degrees of Freedom: 3

P-value: 0.00

Table 56 – Expected Distribution of Entered Taxiway or Runway without Clearance by Severity

	A	B	C	D	Total
NO	25	27	625	987	1,664

	A	B	C	D	Total
YES	**107**	**118**	**2,682**	**4,240**	**7,146**
Total	**132**	**145**	**3,307**	**5,226**	**8,810**

Pilot Instrument Rating

(ATQA PD)

This variable indicates the instrument rating of the pilot involved in the database. Interestingly, the coding on this variable contains information on if the pilot was rated previously, but is not currently. Figure 12 presents the overall distribution of this variable. Table 57 and Table 58 present a breakdown of this variable. Note that unknown ratings were excluded, as they provide little insight into the impact of this variable.

Figure 12 – Distribution of Pilot Instrument Rating

Table 57 – Observed Distribution of Pilot Instrument Rating by Severity

	A	B	C	D	Total
Current	**15**	**21**	**686**	**1,106**	**1,828**
Not Current	**1**	**2**	**74**	**155**	**232**

	A	B	C	D	Total
No Rating	**20**	**23**	**351**	**680**	**2,060**
Total	**36**	**46**	**1,111**	**1,941**	**3,134**

Table 58 – Expected Distribution of Pilot Instrument Rating by Severity

	A	B	C	D	Total
Current	**21**	**27**	**648**	**1,132**	**1,828**
Not Current	**3**	**3**	**82**	**144**	**232**
No Rating	**12**	**16**	**381**	**665**	**2,060**
Total	**36**	**46**	**1,111**	**1,941**	**3,134**

Table 59 – Difference between Observed and Expected of Pilot Instrument Rating by Severity

	A	B	C	D
Current	**-6**	**-6**	**+38**	**-26**
Not Current	**-2**	**-1**	**-8**	**+11**
No Rating	**+8**	**+7**	**-30**	**+15**

The Chi-Squared test statistic indicates that there is a relationship between severity and instrument rating. Current and Not Current are underrepresented for categories A and B, see Table 59 for the deviations between observed and expected. The opposite is true for No Rating. For categories C and D, Not current and No Rating are underrepresented for category C and overrepresented for category D, while Current is observed more than expected for category C and less than expected for category D. When restricted to only conflict events (Table 60 and Table 61), Current and Not Current follow a similar pattern in terms of observed compared to expected values and the mitigating impact on severity is still present. However, the impact of having a non-current rating may be non-linear. These data suggest that having ever been rated is associated with lower incident severity. Without additional statistical and human factors study, it is unclear if these pilots get into fewer severe situations, better recover from mistakes, or if this can be explained by other factors, including the possibility of a spurious correlation.

Specifically, this variable is easily conflated with commercial carrier status (as all commercial carriers are instrument rated while not all GA pilots are instrument rated).

Table 60 – Observed Distribution of Pilot Instrument Rating by Severity, Conflict Only

	A	B	C	Total
Current	15	21	686	722
Not Current	1	2	74	77
No Rating	20	23	351	799
Total	36	46	1,111	1,193

Table 61 – Expected Distribution of Pilot Instrument Rating by Severity, Conflict Only

	A	B	C	Total
Current	22	28	672	722
Not Current	2	3	72	77
No Rating	12	15	367	799
Total	36	46	1,111	1,193

Finally, Table 62 presents an estimate of the impact on severity. The baseline here is considered to be a pilot with no instrument rating. Thus, the odds ratios represent the reduction in the likelihood of being severe if the pilot were either currently rated or rated in the past, but not currently. These estimates support the conclusion that pilots with either current rating or a past rating are less likely to be involved in a severe incursion. Interestingly, the confidence intervals for the two estimates overlap, indicating that the magnitude of the two estimates cannot be considered statistically different. That is, this preliminary estimate suggests that pilots with past ratings are as safe as pilots with current ratings. Further research into pilot instrument ratings should account for the three rating groups (current, past, and never rated) and further investigate whether current and past ratings have the same impact on severity.

Table 62 – Logit Estimate of Impact on Severity, Pilot Instrument Rating

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB
Current Rating	**0.48**	**0.11**	**0.00**	**0.31**	**0.75**
Rated, but not Current	**0.31**	**0.19**	**0.05**	**0.10**	**1.02**

Pilot Hours in Make and Model

(ATQA PD)

For each PD incident, pilots are required to report hours in the make and model of aircraft in which the incident occurred. Table 63 presents the percentiles of this distribution. While the overall distribution is interesting, the distribution of pilot hours by severity level is more pertinent to the question at hand. Figure 13 presents this distribution.

Table 63 – Percentiles of Pilot Hours in Make and Model

	10th	25th	50th	75th	90th
A	**24**	**74**	**200.5**	**508**	**1400**
B	**25.5**	**87.5**	**160**	**680**	**1925**
C	**38**	**100**	**400**	**1400**	**3500**
D	**31**	**100**	**350**	**1199**	**3000**
Overall	**35**	**100**	**356**	**1200**	**3100**

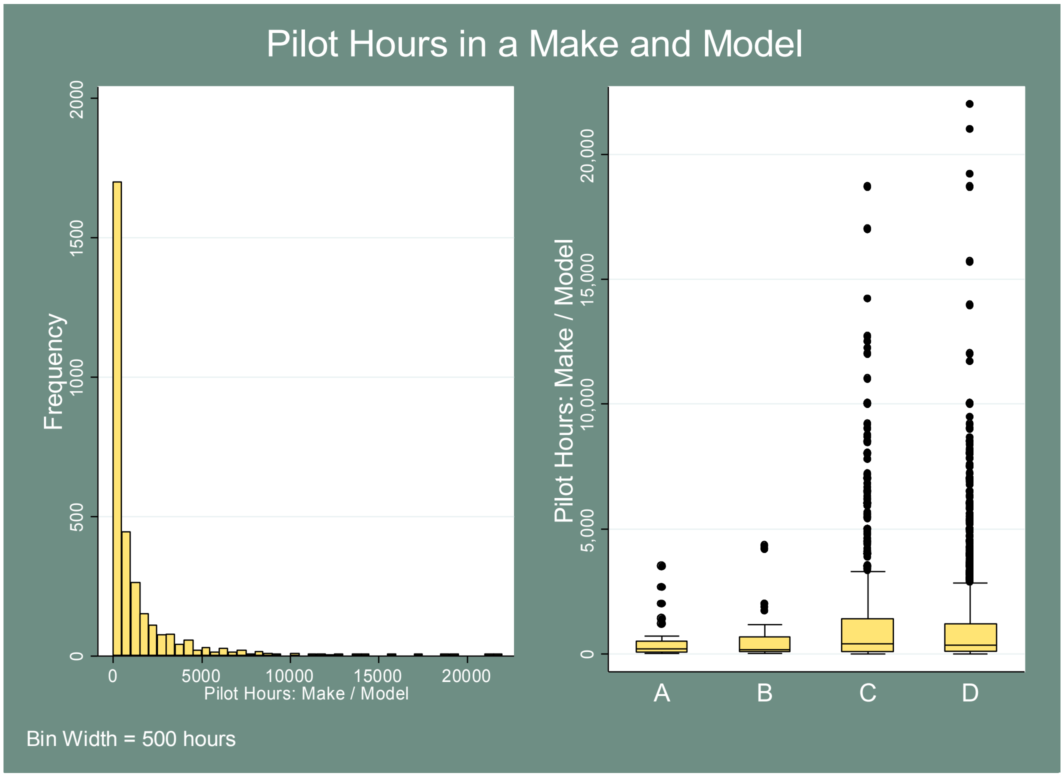

Figure 13 – Distribution of Pilot Hours in a Make and Model

Figure 13 presents two pieces of information. A histogram representing the distribution detailed in Table 63 is on the left. The graph on the right presents the distribution by severity in terms of a boxplot (or box and whisker plot).[43]

Figure 13 reveals that no pilot with more than 5,000 hours in the make and model involved in the incident has committed a severe incursion (category A or B). Figure 13 also reveals that the distribution of hours is weighted heavily towards zero for all severities. It appears, however, that the median value for categories A and B are lower than those of C and D, indicating a leftward shift in the distribution. In other words, pilots involved in category C and category D incursions tend to have more hours in that make and model. There are two possible explanations that come to mind.

Appendix A: The most obvious is that there is an effect of experience. As pilots spend more hours in a make and model they are less likely to commit serious incursions.

Appendix B: An alternative explanation is that bad pilots do not ever get many hours in a make and model. Under this hypothesis, error rates are fairly constant across experience levels but pilots that commit many serious errors stop being pilots (e.g., they do not enjoy it, cannot get licensed). This would lead to lower hour pilots being concentrated in categories A and B rather than in C or D.

43 See Appendix C.2 for more information about the Box and Whisker Plot.

Further statistical testing is required to distinguish between these two hypotheses. The two hypotheses also suggest different policy responses. One implies a policy to encourage pilots to get more hours more quickly. The other hypothesis implies that there needs to be a better way to identify poor pilots and remove them from the population.

The medians of a continuous variable separated by groups can be compared using what is called the Kruskal-Wallis rank test.[44]

Table 64 reports the results of a Kruskal-Wallis test on pilot hours in the involved make and model. The test statistic indicates that the four severity categories are jointly different from each other. However, the pairwise comparisons indicate that no two groups can be considered different from each other (note that a stricter criterion has been used to determine significance given the multiple comparison issue noted in Appendix C.3).

Table 64 – Kruskal-Wallis Test Results for Pilot Hours in Make and Model

	A	B	C	D
Number of Observations	**34**	**40**	**1,107**	**1,914**
Mean Rank	**1,229.07**	**1,310.53**	**1,591.20**	**1,534.64**

In Figure 13 categories A and B appear similar as do categories C and D. Grouping the categories in this manner produce a dichotomized variable, which can be easier to analyze; some of the techniques in Section 4 rely on this dichotomous nature. On the other hand, grouping categories together causes a loss of information. In this case, it is no longer possible to distinguish between conflict and non-conflict events. Thus, while Table 65 presents the results of such a dichotomous grouping, further investigation into the differences between categories (especially categories C and D) is warranted.

Table 65: Kruskal-Wallis Test Results for Pilot Hours in Make and Model, Severe versus Non-Severe

	Not Severe	Severe
Number of Observations	**3021**	**1554.73**
Mean Rank	**74**	**1273.10**

44 See Appendix C.3 for more information about Kruskal-Wallis tests.

3.3.3. Airport Characteristics

This information describes the characteristics of the airport at which the incident occurred. In general, this information will stay the same from incident to incident at the same airport[45] so most of the interesting variation in these variables is between airports. The conclusions are therefore more useful when comparing different types of airports.

Special Procedures

(ATQA OE)

This variable indicates if special procedures were in effect at the time of the incident. Figure 14 presents the distribution of this variable.

45 Some variables do vary at an incident level or across time and will be noted accordingly.

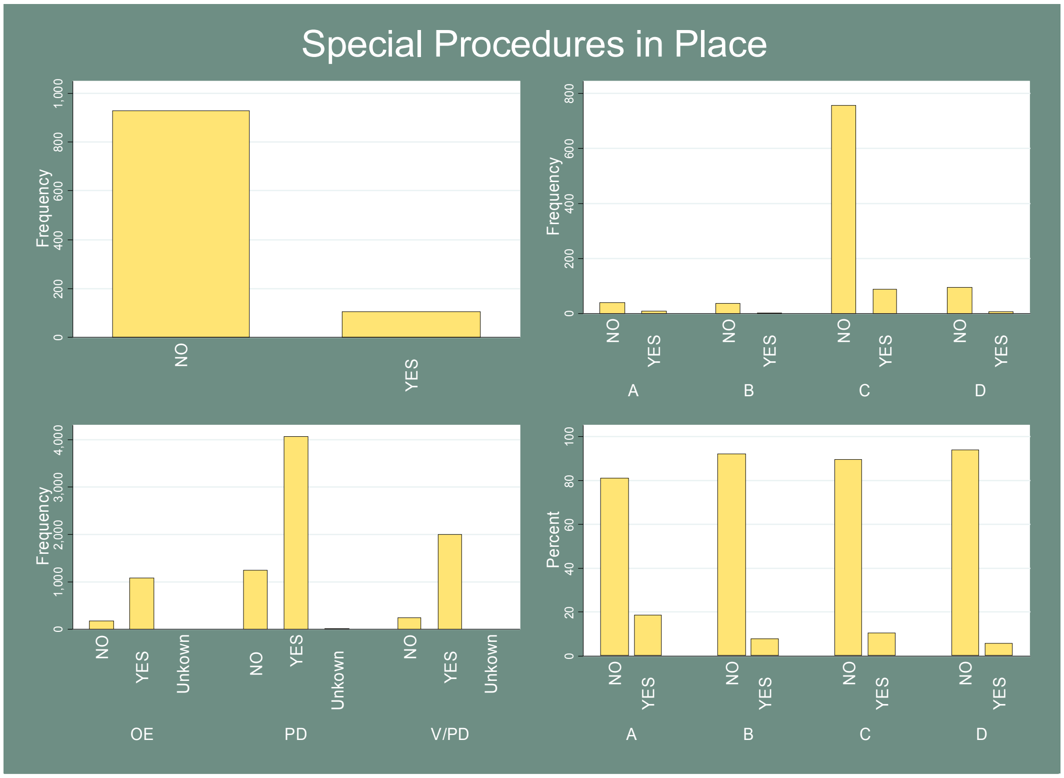

Figure 14 – Distribution of Special Procedures in Place

Table 66 and Table 67 reports the breakdown of this variable by severity and the results of Fisher's Exact test. Note that this variable can only be examined among OE incidents. The test statistic indicates that special procedures have no effect on the severity of an incident. While there is no impact on severity, no information can be gleaned about the impact on frequency of incursions while special procedures are in effect.

Table 66 – Observed Distribution of Special Procedures in Place by Severity

	A	B	C	D	Total
No	**39**	**36**	**75 6**	**96**	**927**
Yes	**9**	**3**	**88**	**6**	**106**
Total	**48**	**39**	**84 4**	**102**	**1,033**

Table 67 – Expected Distribution of Special Procedures in Place by Severity

	A	B	C	D	Total
No	**43**	**35**	**757**	**92**	**927**

	A	B	C	D	Total
Yes	**5**	**4**	**87**	**10**	**106**
Total	**48**	**39**	**844**	**102**	**1,033**

Traffic Complexity Code

(ATQA OE)

This variable indicates the complexity of traffic at the time of the incident on a five-point scale. This variable originates from the ATQA OE database and only applies to OE incidents. Figure 15 presents the distribution of hourly ops by complexity code. Higher complexity is associated with higher hourly operations. Recall that hourly operations are not entirely accurate and the extreme outliers likely represent data problems rather than actual observations. Regardless, the graph shows a distinct trend in median operations by complexity level. However, the degree to which the distributions overlap in the middle categories suggests that the definition of complexity may not be entirely clear in that region (or at least not entirely defined by hourly operations).

The positive correlation between complexity code and hourly operations is not visible for OEP 35 airports. There is a slight trend in median hourly operations, however the overlap between categories is much more pronounced. It is also helpful to keep these values in mind when examining the results presented below.

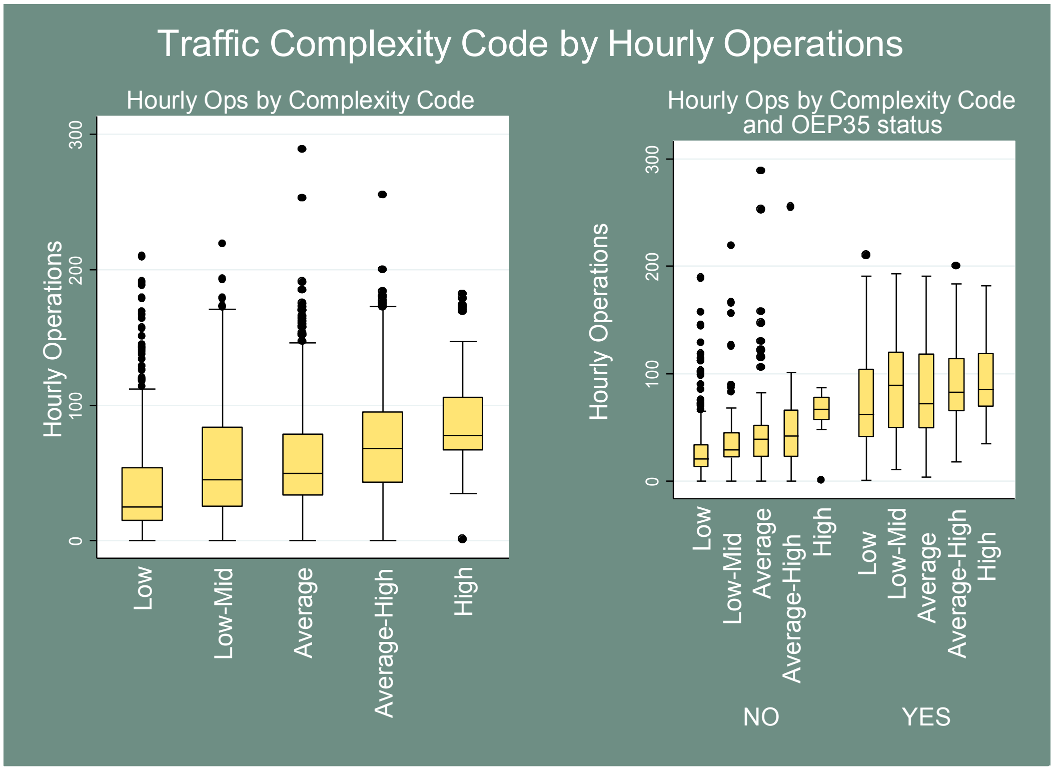

Figure 15 – Distribution of Hourly Operations by Complexity Code, OEP 35 versus Non-OEP 35

Figure 16 – Distribution of Traffic Complexity Code

Figure 16 presents the overall distribution for complexity code. Table 68 and Table 69 present the distribution of responses by severity category and the results of a Chi-Squared test.

Table 68 – Observed Distribution of Traffic Complexity by Severity

	A	B	C	D	Total
Low	**11**	**8**	**250**	**65**	**334**
Low-Mid	**3**	**8**	**160**	**17**	**188**
Average	**22**	**14**	**248**	**17**	**301**
Average-High	**7**	**7**	**144**	**2**	**160**
High	**5**	**2**	**42**	**1**	**50**
Total	**48**	**39**	**844**	**102**	**1,033**

Chi2 score: 70.79

Degrees of Freedom: 12

P-value: 0.00

Table 69 – Expected Distribution of Traffic Complexity by Severity

	A	B	C	D	Total
Low	**16**	**13**	**273**	**33**	**334**
Low-Mid	**9**	**7**	**154**	**19**	**188**
Average	**14**	**11**	**246**	**30**	**301**
Average-High	**7**	**6**	**131**	**16**	**160**
High	**2**	**2**	**41**	**5**	**50**
Total	**48**	**39**	**844**	**102**	**1,033**

It appears that category D incidents are observed more than expected for low complexity, while the conflict events are observed more frequently than expected for average complexity. Category C incursions appear more often than expected for all levels of complexity except the lowest. This suggests that increased complexity is associated with increased severity.

There is a variety of other problems associated with the interpretation of this variable. First and foremost, this is a purely subjective measure. The reporting form offers no guidance on what constitutes “average” or “high” complexity so interpretations of “high” complexity may differ person-to-person or day-to-day. Secondly, due to the lack of guidance, the measure is poorly calibrated. For example, “average complexity” may refer to what is average for a given tower, average for a time of day, or average across the entire NAS. Thus, even though someone may report “average” complexity, it is difficult to tell what the comparison is (i.e., “average” relative to what?). Thirdly, without normal operations it is difficult to discern the true impact of this variable; that is, it is possible that incursions themselves are more likely in high complexity times even if it does not affect the distribution of severity. It is possible that high complexity occurs twice as often for incursion events as for normal operations. Nevertheless, the results in Table 68 indicate that there is a relationship between complexity and incident severity.

Factors leading to Traffic Complexity

(ATQA OE)

This is a set of variables describing the factors leading to the traffic complexity ranking given in the previous variable. As with Traffic Complexity Code, it only applies to OE incidents. Figure 17 provides the frequency of yes and no by each factor, as well as the number of missing values.

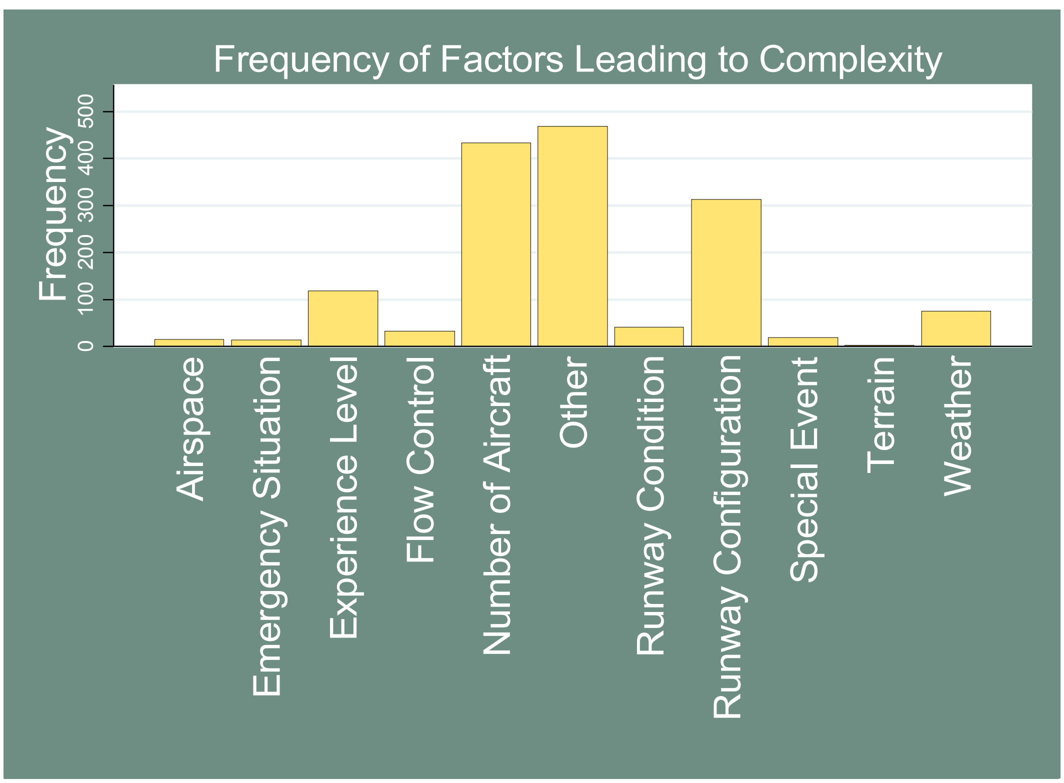

Figure 17 – Frequency of Factors Leading to Traffic Complexity

In many of the cases, it is unclear how these factors may interact with traffic complexity, let alone severity. For example, “Experience” may indicate a lack of experience or that the controller’s higher level of experience reduced the complexity. Additionally, the quality of the data is called into question as the flag for "N/A" is indicated alongside other factors. No test statistics are reported for these variables and *any* interpretation of them is likely erroneous. They are reported here to bring to light the problems in the data that prevent additional analysis.

Part 139 Airport Status

(Runway Incursion Database)

This variable indicates whether the airport at which the incursion happened is categorized as a Part 139 airport.[46] Table 70 and Table 71 present the distribution of this variable by severity. Note that a significant Chi-Squared statistic is also reported indicating some relationship between the severity of the event and Part 139 statistics. This is likely due to the higher traffic at Part 139 airports in general compared to non-Part 139 airports. Figure 18 presents the overall distribution of this variable.

Table 72 and Table 73 report the same results, but limited to only conflict events (categories A through C). After removing category D events from the comparison, the significant relationship is no longer detected, indicating that the result seen in Table 70 is likely driven by the disparity between conflict and non-conflict events, which is itself based on the activity level of the airport, rather than on a real relationship with severity.

46Part 139 status indicates that the airport serves scheduled and unscheduled service with more than 9 passenger seats on a regular basis. Source: http://www.faa.gov/airports/airport_safety/part139_cert/?p1=what

Figure 18 – Distribution of Part 139 Status

Table 70 – Observed Distribution of Part 139 Status by Severity

	A	B	C	D	Total
No	37	37	737	1,348	2,159
Yes	95	108	2,571	3,879	6,653
Total	132	145	3,308	5,227	8,812

Table 71 – Expected Distribution of Part 139 Status by Severity

	A	B	C	D	Total
No	32	36	810	1,281	2,159
Yes	100	109	2,498	3,946	6,653

	A	B	C	D	Total
Total	**132**	**145**	**3,308**	**5,227**	**8,812**

Table 72 – Observed Distribution of Part 139 Status by Severity, Conflict Only

	A	B	C	Total
No	**37**	**37**	**737**	**811**
Yes	**95**	**108**	**2,571**	**2,774**
Total	**132**	**145**	**3,308**	**3,585**

Table 73 – Expected Distribution of Part 139 Status by Severity, Conflict Only

	A	B	C	Total
No	**30**	**33**	**748**	**811**
Yes	**102**	**112**	**2,560**	**2,774**
Total	**132**	**145**	**3,308**	**3,585**

Table 74 and Table 75 reports the distribution of incident type by Part 139 status. The Chi-Squared statistic indicates that there is also a relationship between incident type and Part 139 status. The expected values indicate that this is likely due to an overrepresentation of OE and PD incidents and a corresponding underrepresentation of V/PD incidents among Part 139 airports. Table 76 and Table 77 reports the same information, excluding V/PDs. The reported Chi-Squared statistic indicates that the relationship detected in Table 74 is observed again. Here, it appears that PDs are observed less frequently than expected at Part 139 Airports and the opposite is true for OE incidents. It is unclear why this disparity exists among incident types; further research in the prevalence of different incident types by Part 139 status is required to understand what is reported in Table 74 and Table 76.

Table 74 – Observed Distribution of Part 139 Status by Incident Type

	OE	PD	V/PD	Total
No	186	1,197	776	2,159
Yes	1,082	4,105	1,466	6,653
Total	1,268	5,302	2,242	8,812

Table 75 – Expected Distribution of Part 139 Status by Incident Type

	OE	PD	V/PD	Total
No	311	1,299	549	2,159
Yes	957	4,003	1,693	6,653
Total	1,268	5,302	2,242	8,812

Table 76 – Observed Distribution of Part 139 Status by Incident Type, OE & PD

	OE	PD	Total
No	186	1,197	1,383
Yes	1,082	4,105	5,187
Total	1,268	5,302	6,570

50

edom: 1

Table 77 – Expected Distribution of Part 139 Status by Incident Type, OE & PD

	OE	PD	Total
No	267	1,116	1,383

	OE	PD	Total
Yes	**1,001**	**4,186**	**5,187**
Total	**1,268**	**5,302**	**6,570**

Table 70 and Table 72 addressed the issue of severity and Part 139 status. The results presented in those two tables indicate that any relationship between Part 139 status and severity is a product of the conflict/non-conflict event dynamic. Therefore, due to the loss of information from combining the categories, it is unlikely that an effect would be detected related to severity in the logit framework. Table 74 and Table 76 addressed the issue of incident type and Part 139 status. The results presented in Table 78 indicate that there is a relationship with incident type and that incidents at Part 139 airports have twice the odds of being an OE as non-Part 139 airports. All incursions, and thus airports, included in this analysis are controlled. It is possible that the disparity between Part 139 and non-Part 139 airports may be related to the differing pilot populations between airport types. As noted earlier, further research into why OE incidents are more common at Part 139 airports is warranted.

Table 78 – Logit Estimate of Impact on Incident Type, Part 139 Status

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB
Part 139 Status	**2.06**	**.172**	**0.00**	**1.75**	**2.43**

OEP 35 Airport Status

(Runway Incursion Database)

This variable indicates whether or not the airport at which the incursion occurred is considered part of the OEP 35, the 35 busiest airports in the country in 2000. Though OEP 35 is used in this analysis, the same results hold for the Core 30, the 30 airports of interest to the FAA in 2011, a designation the FAA is using going forward. Figure 19 presents the overall distribution of OEP 35 status.

Figure 19 – Distribution of OEP 35 Status

Table 79 presents the estimated effect on the odds of being severe if an incident occurs at an OEP 35 airport.

Table 79 – Logit Estimate of Impact on Severity, OEP 35 Status

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB
OEP 35	**1.40**	**.189**	**0.01**	**1.07**	**1.82**

The increase in the odds of a severe incident is moderate compared to some of the other variables examined. Given that OEP 35 airports are extremely busy, it is possible that this relationship is merely a product of the higher likelihood of conflict events at a busy airport. Table 80 presents the same estimate, excluding category D incursions. Not surprisingly, the previous relationship is now not detected, indicating that OEP 35 status is likely a better indicator of conflict versus non-conflict rather than severity.

Table 80 – Logit Estimate of Impact on Severity, OEP 35 Status, Conflict Only

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB
OEP 35	**.880**	**.122**	**0.36**	**.671**	**1.15**

Table 81 presents a look at the impact on the odds of being an OE. Interestingly, the impact on OEs is fairly strong, increasing the odds by around 170%. Given the relationship between OE incidents and severity, it is prudent to check if the impact on severity is an independent effect. That is, given that OEP 35 incidents are more likely to be OEs and that OEs are also likely to be more severe, it is not surprising that OEP 35 incidents are more severe. Table 82 presents a multivariate logit that controls for this relationship and examines the impact on severity.

Table 81 – Logit Estimate of Impact on Incident Type, OEP 35 Status

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB
OEP 35	2.72	.175	0.00	2.40	3.09

Table 82 – Logit Estimate of Impact on Severity, OEP 35 Status and Incident Type

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB
OEP 35	1.53	.262	0.01	1.09	2.14
OE Incident	4.36	.678	0.00	3.21	5.91
OEP 35 & OE Incident	.446	.126	0.00	.256	.777

The results indicate that not only is there an independent impact on severity, there is an interaction between OEP 35 status and incident type. The results indicate that incidents at OEP 35 airports tend to be more severe, OE incidents tend to be more severe, but OE incidents at OEP 35 airports are less severe than the combination would suggest – there is a mitigating factor in the interaction of OEP 35 status and incident type. Table 83 presents the same results but excludes category D incidents. Here, the independent OEP 35 impact is no longer detected, but the interaction is still detected – though just barely. It is possible that this mitigating factor is related to controller experience or skill (broadly defined). That hypothesis would indicate that only the most skilled controllers are at the OEP 35 airports and they make less severe mistakes than their non-OEP 35 counterparts. That is only one hypothesis and is difficult to test. A deeper understanding of the differences in controllers between OEP 35 airports and non-OEP 35 airports is required to formulate better hypotheses and to test them adequately.

Table 83 – Logit Estimate of Impact on Severity, OEP 35 Status and Incident Type, Conflict Only

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB
OEP 35	1.05	.183	0.79	.743	1.48
OE Incident	1.71	.271	0.00	1.25	2.33
OEP 35 & OE Incident	.553	.158	0.04	.315	.969

Land and Hold Short Capability at Airport

(Airport Database)

This variable indicates if an airport is capable of LAHSO operations. This is in contrast to the variable described previously in Table 33, which indicates if one of the aircraft involved was performing a LAHSO. Figure 20 contains the overall distribution for this variable.

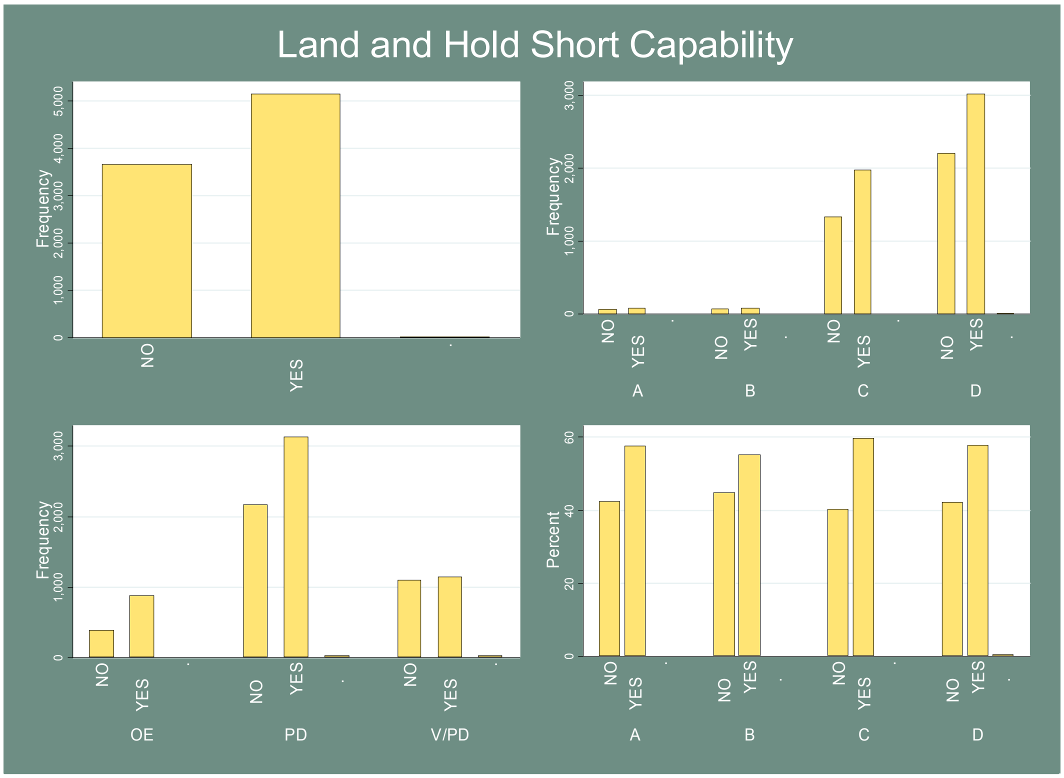

Figure 20 – Distribution of LAHSO Capability at Airport

Table 84 and Table 85 present the distribution by incident type. Table 84 indicates that OE and PD incidents are observed more frequently than expected with a corresponding underrepresentation of V/PD incidents. This relationship is found to be statistically significant. It may be possible that LAHSO capability is correlated with other factors such as Part 139 status and overall traffic levels. Thus, it may be that LAHSO capability is correlated with something that creates this disparity in incident types, rather than LAHSO capability being the driving factor behind the disparity.

Table 84 – Observed Distribution of LAHSO Capability by Incident Type

	OE	PD	V/PD	Total
No	**391**	**2,170**	**1,094**	**3,655**
Yes	**877**	**3,128**	**1,145**	**5,150**
Total	**1,268**	**5,298**	**2,239**	**8,805**

Table 85 – Expected Distribution of LAHSO Capability by Incident Type

	OE	PD	V/PD	Total
No	**526**	**2,199**	**929**	**3,655**
Yes	**742**	**3,099**	**1,310**	**5,150**
Total	**1,268**	**5,298**	**2,239**	**8,805**

Table 86 – Correlation of LAHSO Capability and Other Airport Characteristics

	Correlation
Part 139 Status	**0.5278**
OEP 35 Status	**0.2291**
Daily Operations	**0.1151**

Table 86 presents the correlation of LAHSO Capability with other relevant airport characteristics. Interestingly, it is not highly correlated with any of these factors. The relationships seen in Table 84 and Table 85 cannot be attributed to that correlation. Further research is warranted to better understand how LAHSO capability is correlated with incident type. Table 84 indicated a significant relationship between incident type and LAHSO capability at an airport, while Table 87 provides an estimate of the impact LAHSO capability has on the odds of being an OE (i.e., an OE has odds 71% higher under LAHSO capability).

Table 87 – Logit Estimate of Impact on Incident Type, Land and Hold Short Capability at Airport

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB
LAHSO Capability	**1.71**	**.111**	**0.00**	**1.51**	**1.95**

Table 88 and Table 89 present the distribution by severity. There is no relationship between severity and LAHSO capability, as indicated by the insignificant Chi-Squared statistic. Combining this result with previous results raises some interesting questions. The logic chain is as follows:

1. The results in Table 84 indicate that OEs are more common than expected at LAHSO capable airports.

2. As seen in Table 1 there is a relationship between incident type and severity, with OEs tending to be more severe.
3. These two results combined might indicate that LAHSO capable airports should be more severe, but that does not seem to be the case.

Further research into severity, incident type, and LAHSO capability might help clarify this surprising (lack of) relationship.

Table 88 – Observed Distribution of LAHSO Capability by Severity

	A	B	C	D	Total
No	**56**	**65**	**1,333**	**2,201**	**3,655**
Yes	**76**	**80**	**1,975**	**3,019**	**5,150**
Total	**132**	**145**	**3,308**	**5,220**	**8,805**

Table 89 – Expected Distribution of LAHSO Capability by Severity

	A	B	C	D	Total
No	**55**	**60**	**1,373**	**2,167**	**3,655**
Yes	**77**	**85**	**1,935**	**3,053**	**5,150**
Total	**132**	**145**	**3,308**	**5,220**	**8,805**

Daily Operations

(OPSNET)

As noted in Section 3.1.6, operations are available on a variety of time scales: hourly, daily, and annually. The ideal operations measure is both granular and accurate. The hourly counts provided by ETMSC are the most granular option available, but due to the way VFR operations are allocated to hours of the day, the accuracy of the data is questionable, at best. In fact, the allocation procedure indicates that some incursions happened in hours with zero operations, which is extremely unlikely. Yearly operations are much more stable, but do not offer the granularity that may be important as operations vary throughout the year. Daily operations offer a good mix of granularity and accuracy. Figure 21 presents the distribution of this variable overall, and by severity. Table 90 presents the median daily operations by severity while Table 91 presents the results of a Kruskal-Wallis test.

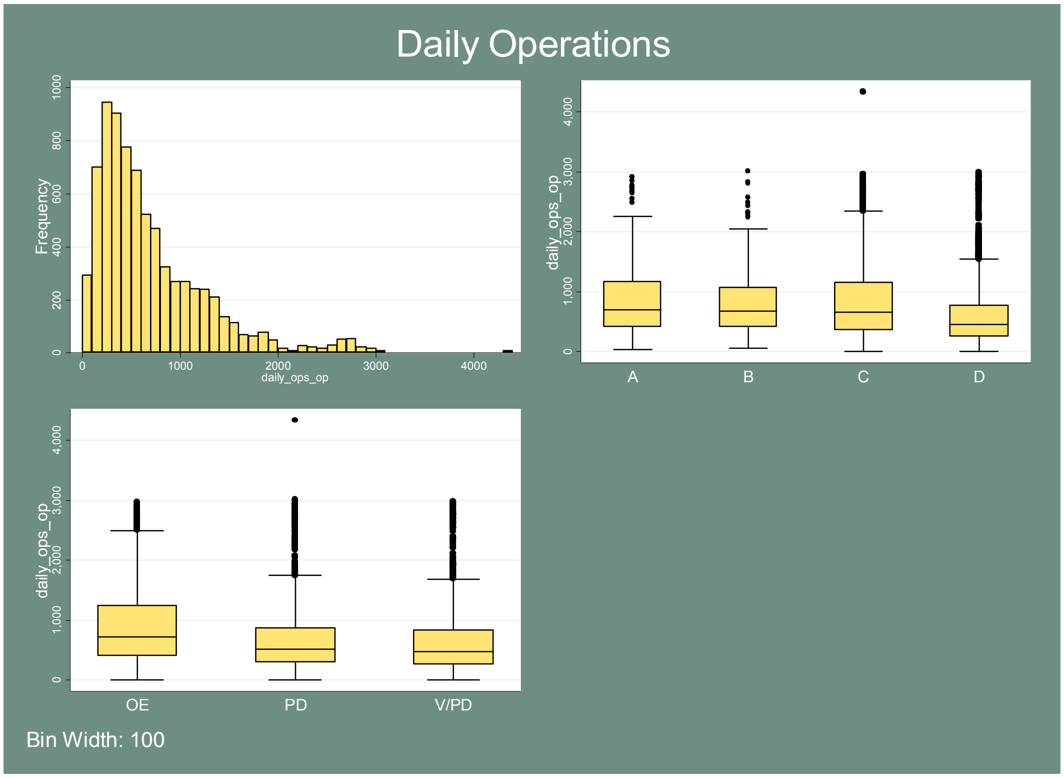

Figure 21 – Distribution of Daily Operations

Table 90 – Percentiles of Daily Operations

	10th	25th	50th	75th	90th
A	**202**	**408**	**695.5**	**1,169**	**1,743**
B	**295.5**	**410**	**673.5**	**1,071.5**	**1,889.5**
C	**230**	**371**	**654**	**1,161**	**1,636**
D	**145**	**257**	**451**	**770**	**1,225**
Overall	**170**	**298**	**530.5**	**936**	**1,412**

Table 91 – Kruskal-Wallis Test Results for Daily Operations

	A	B	C	D
Number of Observations	**114**	**120**	**2978**	**4422**
Mean Rank	**4485.79**	**4573.25**	**4385.44**	**3397.29**

The results of the Kruskal-Wallis indicate that daily operations jointly differ across severity categories. Category D appears to have many fewer median daily operations than any of the other categories. The pairwise comparison tests indicate that categories A, B, and C can all be distinguished from D statistically. However, categories A, B, and C are pairwise indistinguishable. This indicates that daily operations are likely a better determinant of conflict versus non-conflict event rather than contributing to severity.

Percent of Operations that are Air Carrier / Air Transport

(Airport Database)

This variable indicates the average percent of traffic at an airport that is categorized as Air Carrier or Air Transport.[47] Figure 22 presents the distribution of this variable by severity, while Table 92 presents the percentiles of the distribution. Table 93 reports the results of a Kruskal-Wallis test by severity.

47 This data element was contained in the database of airport characteristics the Volpe Center received from the University of Virginia (via FAA). It appears that the values are derived from OPSNET, however it is unclear over what time span this average is calculated.

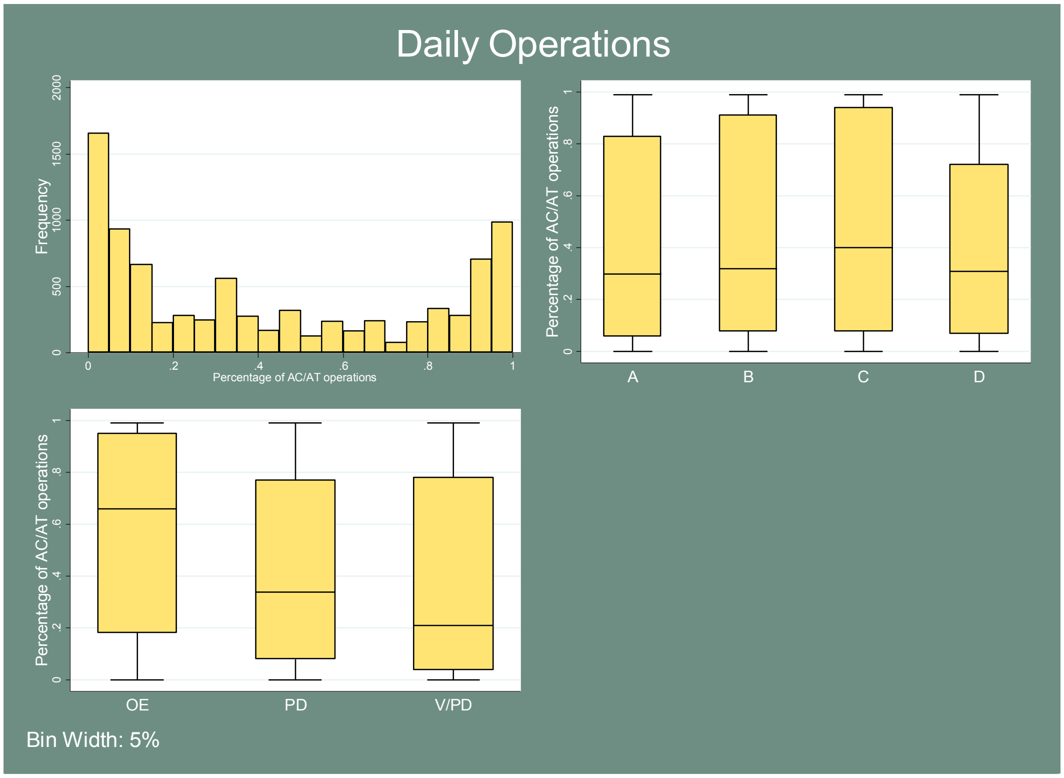

Figure 22 – Distribution of AC/AT Percent of Operations

Table 92 – Percentiles of AC/AT Percent of Operations by Severity

	10th	25th	50th	75th	90th
A	.01	.06	.30	.83	.985
B	.02	.08	.32	.91	.96
C	.02	.08	.40	.94	.98
D	.02	.07	.31	.72	.95
Overall	.02	.08	.34	.83	.96

Table 93 – Kruskal-Wallis Test Results for AC/AT Percent of Operations

	A	B	C	D
Number of Observations	130	144	3299	5178
Mean Rank	4261.07	4419.64	4725.50	4155.00

Interestingly, all severity levels appear to have similar medians, with the values for category C tending to be a bit higher. Additionally, the interquartile range for category D incursion appears to be smaller, indicating a more narrowly distributed variable (especially given the overwhelming prevalence of category D). The result of the Kruskal-Wallis test supports the conclusion that the categories are jointly different, but offers little information for the pairwise comparisons. Category C can be distinguished from category D, but no other pairs are significantly different. This may indicate that high percentage AC/AT airports are also very busy and are thus unlikely to commit an error in the absence of another aircraft. Further exploration will need to control for the operations at the given airport to disentangle the two effects.

Table 94 and Table 95 examine the percent of operations that are AC/AT by incident type. All three incident types appear to have very different distributions. OE incidents have a higher median percentage while V/PD incidents have the lowest. The results of the Kruskal-Wallis test corroborate this, indicating that the three incident types are jointly different as well as all pairwise different from each other. This suggests that policy interventions need to account for traffic mix at an airport. That is, any policy intervention targeted predominately at one kind of airport will have differing impacts on severity and incident types across airports.

Table 94 – Percentiles of AC/AT Percent of Operations by Incident Type

	10th	25th	50th	75th	90th
OE	.04	.18	.66	.95	.99
PD	.02	.08	.34	.77	.95
V/PD	.01	.04	.21	.78	.96
Overall	.02	.08	.34	.83	.96

Table 95 – Results of Kruskal-Wallis Test for AC/AT Percent of Operations by Incident Type

	OE	PD	V/PD
Number of Observations	1267	5250	2234
Mean Rank	5393.27	4307.19	3960.76

68

dom: 2

Number of Runway Intersections

(Airport Database)

This variable measures the number of runway intersections at the airport where the incursion occurred. Figure 23 and Table 96 gives the distribution of this variable. Table 97 gives the results of a Kruskal-Wallis test by severity.

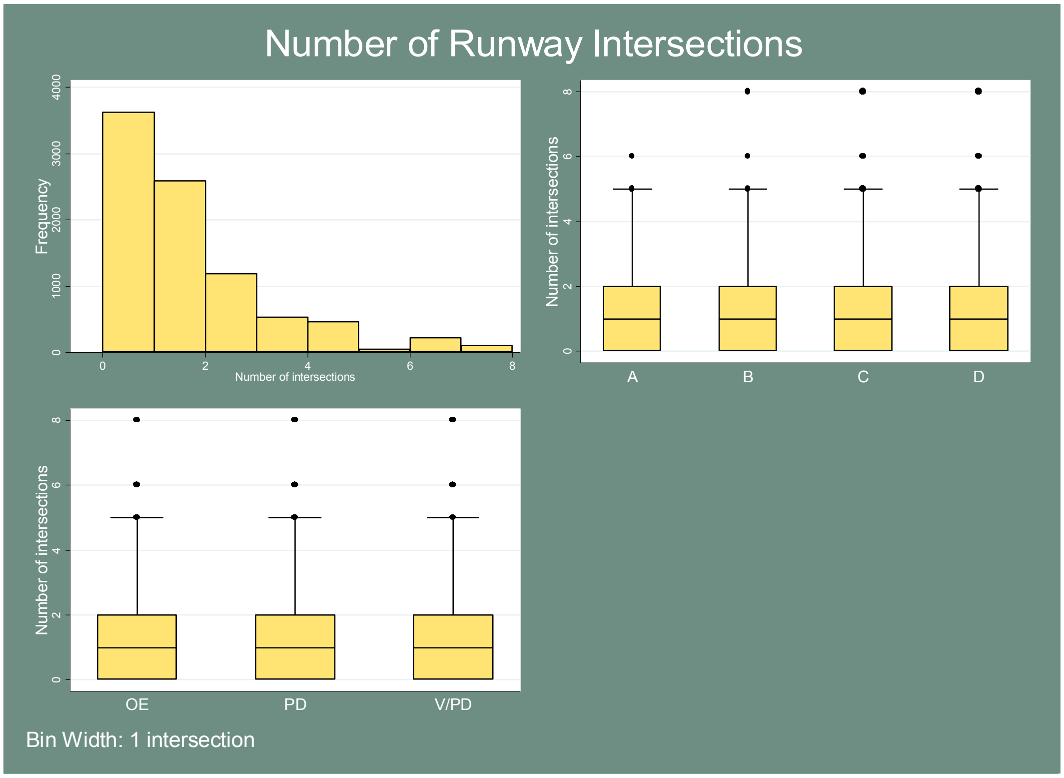

Figure 23 – Distribution of Number of Runway Intersections

Table 96 – Percentiles of Number of Runway Intersections by Severity

	10th	25th	50th	75th	90th
A	0	0	1	2	4
B	0	0	1	2	4
C	0	0	1	2	4
D	0	0	1	2	3
Overall	0	0	1	2	3

Table 97 – Kruskal-Wallis Test Results for Number of Runway Intersections

	A	B	C	D
Number of Observations	**132**	**145**	**3308**	**5226**
Mean Rank	**4725.56**	**4836.27**	**4563.58**	**4286.24**

On a pairwise basis, only categories C and D can be considered different. Table 98 presents the results of a Kruskal-Wallis test, examining conflict only events. The three severity categories can no longer be considered jointly different. This indicates that number of runway intersections is helpful for identifying conflict or non-conflict events but not severity among conflict events.

Table 98 – Kruskal-Wallis Test Results for Number of Runway Intersections, Conflict Only

	A	B	C
Number of Observations	**132**	**145**	**3308**
Mean Rank	**1850.91**	**1892.81**	**1786.31**

Number of Runways

(Airport Database)

This variable indicates the total number of runways at the airport where the incident occurred. Note that this is not the number of runways in operation at the time of the incident. A measure of the number of operating runways was unavailable and future research may want to explore how that impacts severity. Figure 24 and Table 99 present the distribution of the number of runways. Table 100 presents the results of a Kruskal-Wallis test by severity.

The results of the Kruskal-Wallis test indicate that there is a difference in number of runways between the severity categories. Examining the distribution indicates that category D appears the most different in terms of percentiles. Additionally, only categories C and D can be considered pairwise different. It is likely that the observed relationship will be no longer significant once category D incursions are excluded from the analysis (controlling for the conflict versus non-conflict dynamic).

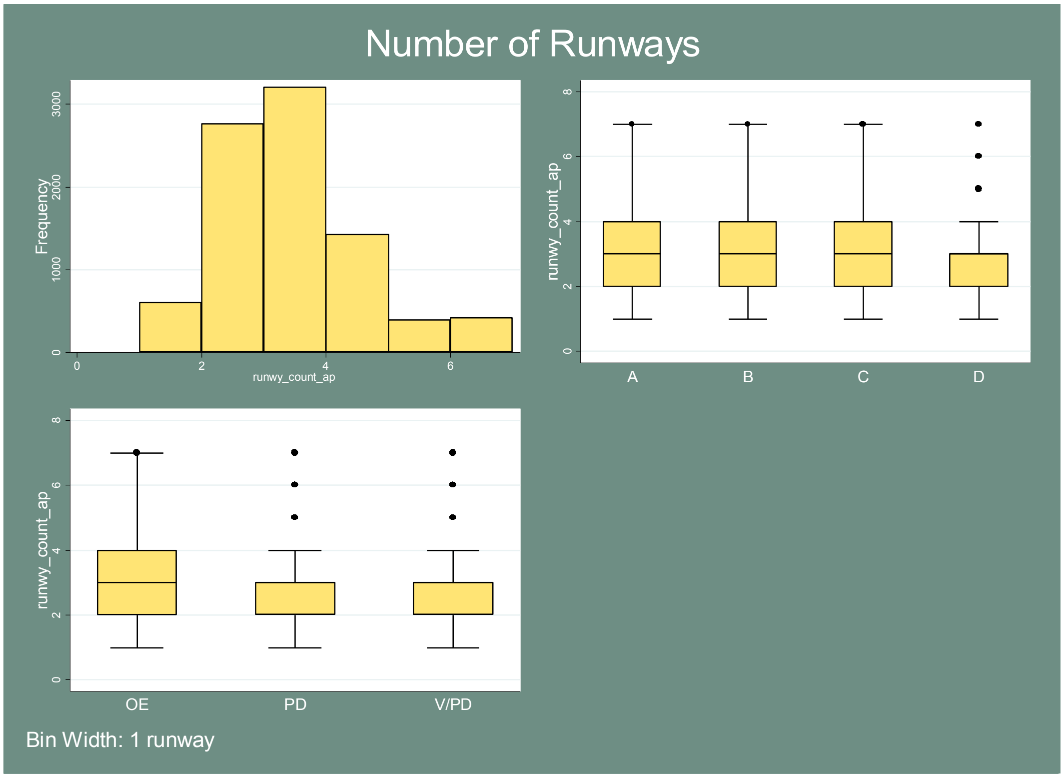

Figure 24 – Distribution of Number of Runways

Table 99 – Percentiles of Number of Runways by Severity

	10th	25th	50th	75th	90th
A	2	2	3	4	5
B	2	2	3	4	5
C	2	2	3	4	5
D	2	2	3	3	4
Overall	2	2	3	4	4

Table 100 – Kruskal-Wallis Test Results for Number of Runways

	A	B	C	D
Number of Observations	132	145	3308	5226

	A	B	C	D
Mean Rank	**4498.24**	**4637.70**	**7442.31**	**4166.26**

Table 101 presents the results of a Kruskal-Wallis test examining conflict events only. As expected, the relationship between number of runways and severity is no longer significant. It is likely that number of runways is a proxy for overall traffic levels and likelihood of two planes conflicting. A similar argument may hold for the number of runway intersections.

Table 101 – Kruskal-Wallis Test Results for Number of Runways, Conflict Only

	A	B	C
Number of Observations	**132**	**145**	**3308**
Mean Rank	**1691.51**	**1742.27**	**1799.27**

Number of Hotspots

(Airport Database)

This variable indicates the number of hotspots identified at an airport. A hotspot is defined as "a location on an airport movement area with a history of potential risk of collision or runway incursion, and where heightened attention by pilots and drivers is necessary."[48] Table 102 and Figure 25 present the distribution of this variable while Table 103 presents the results of a Kruskal-Wallis test.

The severity categories are jointly different while only categories C and D can be considered pairwise different. As with total runways and runway intersections, it is instructive to examine conflict events only. The evidence is weaker for conflict only events, as seen in Table 104. However, the change is not as dramatic as for number of runways or number of runway intersections. Thus, number of hotspots

48 Federal Aviation Administration (2012). http://www.faa.gov/airports/runway_safety/hotspots/hotspots_list/

appears to be most useful in identifying conflict versus non-conflict events but may also provide some additional information regarding severity for conflict events.

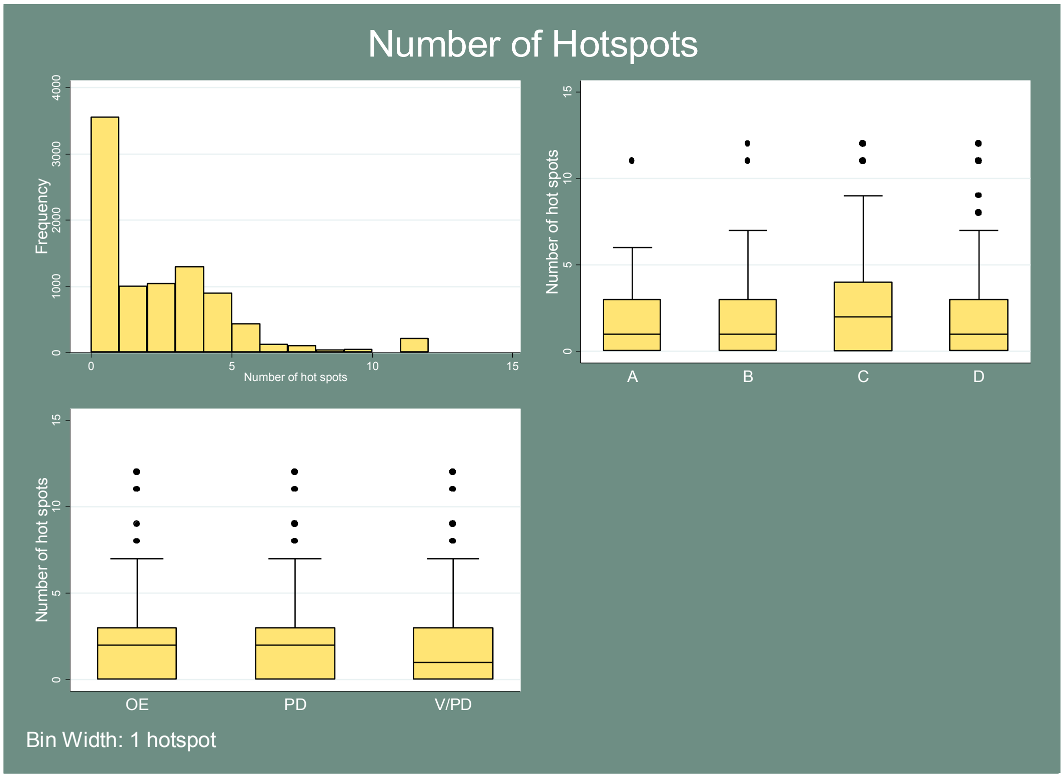

Figure 25 – Distribution of Number of Hotspots

Table 102 – Percentiles of Number Hotspots by Severity

	10th	25th	50th	75th	90th
A	0	0	1	3	4
B	0	0	1	3	5
C	0	0	2	4	5
D	0	0	1	3	4
Overall	0	0	1	3	5

Table 103 – Kruskal-Wallis Test Results for Number of Hotspots

	A	B	C	D
Number of Observations	**132**	**145**	**3308**	**5226**
Mean Rank	**4445.25**	**4354.09**	**4746.78**	**4190.74**

Table 104 – Kruskal-Wallis Test Results for Number of Hotspots, Conflict Only

	A	B	C
Number of Observations	**132**	**145**	**3308**
Mean Rank	**1679.18**	**1640.72**	**1804.22**

3.3.4. Radar

These variables are derived from the ATQA OE dataset. Correspondingly, they can only be analyzed for OE incidents. They describe the radar systems available to the controller at the airport where the incident took place. In some instances the radar variables have been combined to cover multiple similar versions of a system. Where this occurs, the specific systems included will be noted. Brief definitions of the different radar systems examined follow:[49]

- STARS: STARS (standard terminal automation replacement system) "is a new terminal air traffic control system that uses modern, commercial, open architecture computing equipment to replace existing [ARTS] systems."
- ASDE: ASDE (airport surface detection equipment) is a radar system that tracks ground based vehicles and aircraft. A variety of ASDE systems have been installed throughout the years. ASDE-X, the latest iteration, attempts to uses a slightly different set of hardware to achieve a similar effect to that of previous ASDE systems.
- ARTS: ARTS (Automated Radar Terminal System) encompasses several versions of a similar system. At its core, ARTS is a radar processing system to associate data with specific radar

49 Nolan (2011).

tracks. ARTS-III actually represents an older version of the technology. ARTS-II represents an attempt to produce a lower cost version of the ARTS-III system.

STARS

(ATQA OE)

Table 105 and Table 106 present the observed and expected distributions of STARS by severity. Fisher's exact test indicates that there is a relationship between severity and the availability of the STARS radar system. Categories A, B and D appear underrepresented while Category C is over represented.

Table 105 – Observed Distribution of STARS by Severity

	A	B	C	D	Total
No	**39**	**34**	**617**	**61**	**751**
Yes	**9**	**5**	**227**	**41**	**282**
Total	**48**	**39**	**844**	**102**	**1,033**

Table 106 – Expected Distribution of STARS by Severity

	A	B	C	D	Total
No	**35**	**28**	**614**	**74**	**751**
Yes	**13**	**11**	**230**	**28**	**282**
Total	**48**	**39**	**844**	**102**	**1,033**

To better understand how this variable impacts severity, category D incursions are excluded from the following tables (Table 107 and Table 108). This eliminates the conflict versus non-conflict dynamic that appears in Table 105. As with the entire range of severity, categories A and B are underrepresented while category C is overrepresented. However, the relationship between STARS and severity is weaker. This indicates that some of the relationship seen in Table 105 can be attributed to discriminating between conflict and non-conflict. This may be a product of where STARS is deployed; that is, STARS may be deployed where the baseline rate for conflict events is higher regardless of its impact on severity. Nevertheless there appears to be weak evidence suggesting that the presence of STARS is associated with lower severity incidents.

Table 107 – Observed Distribution of STARS by Severity, Conflict Only

	A	B	C	Total
No	**39**	**34**	**617**	**690**
Yes	**9**	**5**	**227**	**241**

	A	B	C	Total
Total	**48**	**39**	**844**	**931**

Table 108 – Expected Distribution of STARS by Severity, Conflict Only

	A	B	C	Total
No	**36**	**29**	**626**	**690**
Yes	**12**	**10**	**218**	**241**
Total	**48**	**39**	**844**	**931**

ASDE

(ATQA OE)

It is important to acknowledge that this ASDE variable does not discriminate between different versions of the ASDE system. That is, this variable indicates the presence of ASDE-3 or ASDE-X. This is due to how the information was ended in the Runway Incursion Database. Regardless, Table 109 and Table 110 present the distribution of this variable. Interestingly, there appears to be a strong relationship between severity and the presence of ASDE. Categories A, B, and D are underrepresented while category C is overrepresented. This is likely a product of how the ASDE systems were deployed. ASDE is deployed at major airports, where a non-conflict event (category D) is less likely. Therefore, it is instructive to look at the conflict only distribution as presented in Table 111 and Table 112.

Table 109 – Observed Distribution of ASDE by Severity

	A	B	C	D	Total
No	**35**	**32**	**535**	**83**	**685**
Yes	**13**	**7**	**309**	**19**	**348**
Total	**48**	**39**	**844**	**102**	**1,033**

Table 110 – Expected Distribution of ASDE by Severity

	A	B	C	D	Total
No	**32**	**26**	**560**	**68**	**685**

	A	B	C	D	Total
Yes	**16**	**13**	**284**	**34**	**348**
Total	**48**	**39**	**844**	**102**	**1,033**

The conflict only distribution indicates a similar pattern to the overall distribution. There is some evidence that ASDE is associated with lower severity events (in this case, category C incursions). This indicates that the lower than expected number of D incursions seen in Table 109 is likely a product of the distribution of ASDE systems with respect to airports.

Table 111 – Observed Distribution of ASDE by Severity, Conflict Only

	A	B	C	Total
No	**35**	**32**	**535**	**602**
Yes	**13**	**7**	**309**	**329**
Total	**48**	**39**	**844**	**931**

Table 112 – Expected Distribution of ASDE by Severity, Conflict Only

	A	B	C	Total
No	**31**	**25**	**546**	**602**
Yes	**17**	**14**	**298**	**329**
Total	**48**	**39**	**844**	**931**

Given that both ASDE and STARS appear to reduce the severity of runway incursions, it would be interesting to investigate whether or not there is any synergy between STARS and ASDE. The logit results presented in Table 113 indicate that STARS and ASDE are both associated with lower severity incidents, but there is no synergy between the systems. That is, the effect of STARS and ASDE is exactly the sum of its parts. Note, however, that the odds ratios for STARS and ASDE in isolation are not precisely estimated; this is likely a product of including the interaction term in the estimation. Though the evidence for the isolated impact of ASDE or STARS is weaker in this logit model, combining these results with those from the Fisher's Exact test indicates that there is evidence that these radar systems reduce severity.

Table 113 – Logit Estimates of Impact on Severity, ASDE and STARS

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB
STARS	**0.59**	**0.18**	**0.08**	**0.33**	**1.06**
ASDE	**0.50**	**0.18**	**0.06**	**0.24**	**1.03**
STARS & ASDE	**1.15**	**0.74**	**0.83**	**0.32**	**4.07**

ARTS II

(ATQA OE)

As mentioned previously, ARTS II represents a lower cost version of the ARTS III system. This variable indicates if any version of ARTS II was available to the controller at the time of the incident. Table 114 and Table 115 present the observed and expected distribution. There is no indication of any relationship between the presents of ARTS II and severity. As the ARTS systems are focused on airborne traffic, this is not an unexpected result.

Table 114 – Observed Distribution of ARTS II by Severity

	A	B	C	D	Total
No	**40**	**33**	**733**	**88**	**894**
Yes	**8**	**6**	**111**	**14**	**139**
Total	**48**	**39**	**844**	**102**	**1,033**

Table 115 – Expected Distribution of ARTS II by Severity

	A	B	C	D	Total
No	**6**	**5**	**114**	**14**	**139**
Yes	**48**	**39**	**844**	**102**	**1,033**
Total	**42**	**34**	**730**	**88**	**894**

ARTS III

(ATQA OE)

The ARTS III system is the more feature rich and expensive version of the ARTS systems under consideration. This variable indicates whether ARTS III was available to the controller at the time of the incident. Table 116 and Table 117 present the observed and expected distribution of this variable.

Table 116 – Observed Distribution of ARTS III by Severity

	A	B	C	D	Total
No	**27**	**25**	**527**	**78**	**657**
Yes	**21**	**14**	**317**	**24**	**376**
Total	**48**	**39**	**844**	**102**	**1,033**

Table 117 – Expected Distribution of ARTS III by Severity

	A	B	C	D	Total
No	**31**	**25**	**537**	**65**	**657**
Yes	**17**	**14**	**307**	**37**	**376**
Total	**48**	**39**	**844**	**102**	**1,033**

Categories A and C appear over represented while categories B and D appear under represented. It is important to reiterate that the Fisher's Exact test indicates that there is some relationship between the two variables (i.e., there is systematic relationship between rows and columns in the table). It does not test for any particular direction or even if that relationship is consistent. For a better understanding of how ARTS III may impact severity, category D incursions can be excluded, removing the conflict versus non-conflict dynamic.

Table 118 and Table 119 examine ARTS III in terms of conflict events only. The relationship seen in Table 116 is no longer present. As with ASDE it is possible that this relationship is due to how ARTS III is deployed – busier airports received the expensive ARTS III system.

Table 118 – Observed Distribution of ARTS III by Severity, Conflict Only

	A	B	C	Total
No	**27**	**25**	**527**	**579**
Yes	**21**	**14**	**317**	**352**
Total	**48**	**39**	**844**	**931**

Table 119 – Expected Distribution of ARTS III by Severity, Conflict Only

	A	B	C	Total
No	**30**	**24**	**525**	**579**
Yes	**18**	**15**	**319**	**352**
Total	**48**	**39**	**844**	**931**

3.3.5. Controller Variables

These variables originate from the ATQA OE database and therefore only pertain to OE incidents. The variables in this section describe the controller or controller's situation at the time of the incident.

Employee Alerted to Incident By

(ATQA OE)

This variable indicates who alerted the controller to the incident. Recall that this is coded only for OE incidents; so in all cases the controller was at fault, though the incident may be first identified by a different party. The overall frequency of each response is presented in Figure 26. Table 120 and Table 121 present the distribution as well as the results of a Chi-Squared test.

Figure 26 – Frequency of Categories of Employee Alerted to Incident By

Table 120 – Observed Distribution of Employee Alerted to Incident By, by Severity

	A	B	C	D	Total
Conflict Alert	0	0	3	0	3
MSAW_EMSAW	0	0	1	0	1
Self-identified	12	10	296	33	351
Facility personnel	8	12	284	56	360
Pilot	22	11	158	1	192
Other	6	6	96	12	120
Total	48	39	838	102	1,027

Table 121 – Expected Distribution of Employee Alerted to Incident By, by Severity

	A	B	C	D	Total
Conflict Alert	0	0	2	0	3
MSAW_EMSAW	0	0	1	0	1
Self-identified	16	13	286	35	351
Facility personnel	17	14	294	36	360
Pilot	9	7	157	19	192
Other	6	5	98	12	120
Total	48	39	838	102	1,027

The majority of incidents appear to be identified by persons other than the controller. Additionally, incidents identified by pilots tend to be more severe than expected. All categories except category D incidents are higher than expected (with category A being twice as high as expected). The opposite pattern holds for incidents identified by other facility personnel. The pattern is less clear for self-identified incidents, where categories A, B and D are lower than expected and category C is observed more frequently than expected. The deviations from the expected values are much higher for pilot identified incidents than for either self-identified or those identified by other personnel.

Table 122 presents the results of a simple logit focusing on OE incidents identified by pilots.

Table 122 – Logit Estimate of Impact on Severity, Employee Alerted to Incident By, Conflict Only

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB
Employee Alerted to Incident By Pilot	3.00	.713	0.00	1.88	4.78

The results indicate that the odds of an OE incident being severe if it is identified by a pilot are 3 times higher than incidents not identified by pilots. This is consistent with the information contained in Table 120.

One possible explanation for this pattern is that, due to their proximity, pilots are able to identify the most serious incidents. This would cause the increase in pilot-reported serious OE incidents. This trend may not be unique to OE incidents, but there is no counterpart variable describing PD incidents. Further research is warranted to better understand how severity and who identifies the incident are related.

Controller Time on Shift

(ATQA OE)

This variable tracks the time the controller was on shift before the incident occurred. Again, this is only available for OE incidents. Figure 27 and Table 123 present the distribution of this variable while Table 124 presents the results of Kruskal-Wallis test by severity category.

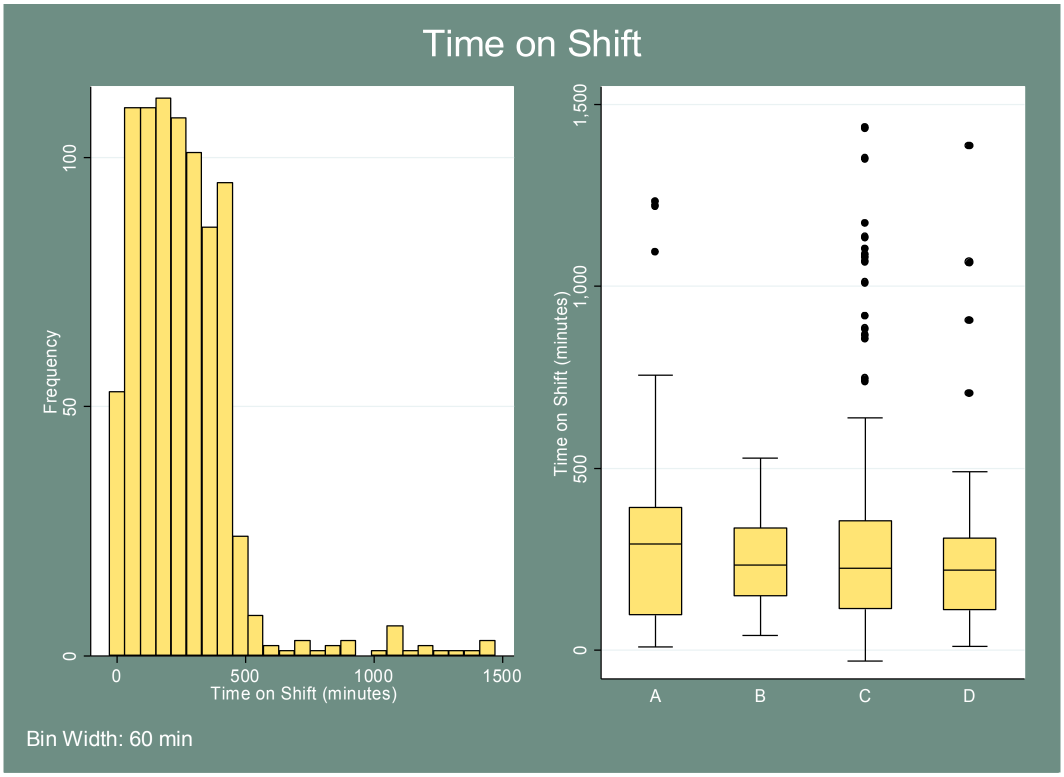

Figure 27 – Distribution of Time on Shift

Table 123 – Percentiles of Time on Shift

	10th	25th	50th	75th	90th
A	36	96	293	392	462
B	68	150	234	337	424
C	46	113	226	355	427
D	48	109	220	308	431
Overall	46	115	227	354	427

Table 124 – Kruskal-Wallis Test Results for Time on Shift

	A	B	C	D
Number of Observations	43	37	685	70
Mean Rank	456.26	437.08	415.96	404.35

	A	B	C	D

The overall distribution is confined mostly before 500 minutes. This is not entirely surprising, as shift length is regulated. However, it is worth noticing the observations above approximately 500 minutes. These observations are certainly outliers and may be misreported. However, the number is not large enough to distort the distribution and, without further information, the values are certainly *possible* if unlikely and so should not be excluded.

The distributions by severity level look fairly similar. This observation is borne out by the results of the Kruskal-Wallis test that indicate no joint difference between the groups. The most obvious explanation for this is that time on shift does not influence severity of the incident. It is possible that the frequency of incidents might go up as time on shift goes up.[50] It is important to note that no information on controller shifts without incursions is available – the vast majority of shifts have no incursions. Further investigation into the relationship between time on shift and frequency of incursion is warranted.

Controller Age

(ATQA OE)

This variable indicates the controller age in years. As this variable is derived from ATQA, it is only available for OE incidents. Table 125 and Figure 28 present the distribution of controller age while Table 126 gives the results of a Kruskal-Wallis test by severity.

50 As a side note, there appears to be some evidence that time on shift *may* impact event frequency. The distribution of time on shift is fairly flat for times under approximately 500 minutes. If the probability of an incursion happening is independent of time on shift, one would expect a distribution that decreases as time increases as not all shifts are the same length and controllers “drop out” of the distribution as shifts end.

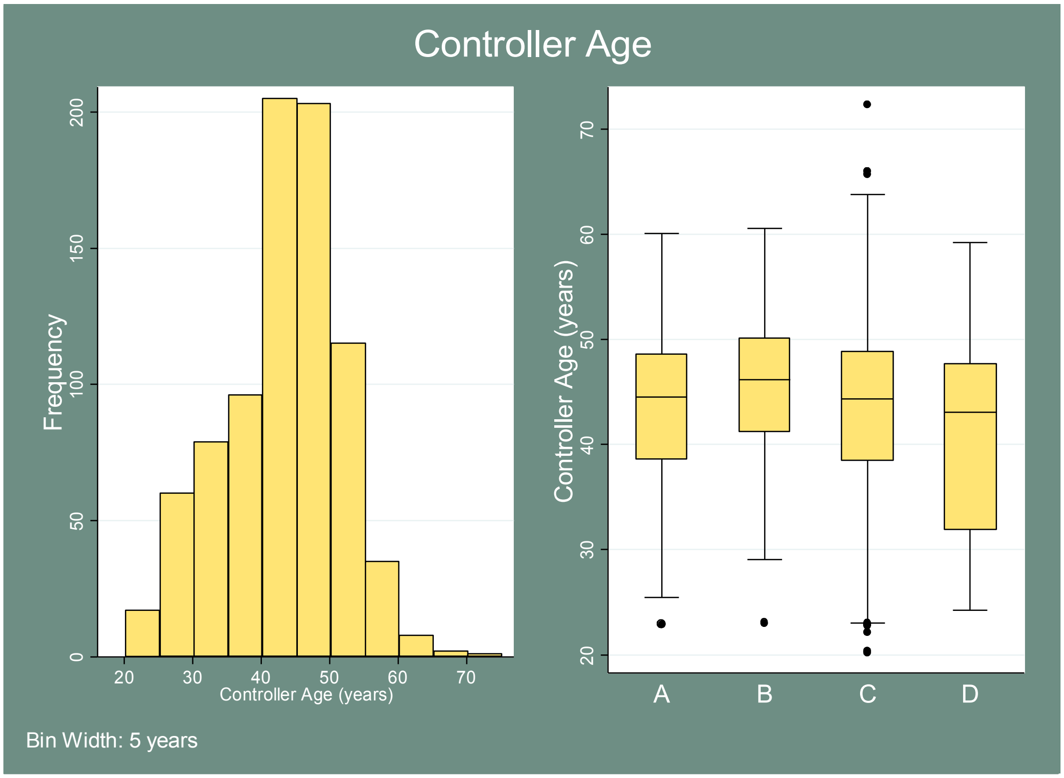

Figure 28 – Distribution of Controller Age

Table 125 – Percentiles of Controller Age

	10th	25th	50th	75th	90th
A	31	39	45	49	52
B	33	41	46	50	58
C	31	39	44	49	53
D	27	32	43	48	54
Overall	31	38	44	50	53

Table 126 – Kruskal-Wallis Test Results for Controller Age

	A	B	C	D
Number of Observations	41	37	673	70
Mean Rank	404.22	476.26	412.76	363.54

There does not appear to be a relationship between controller age and incident severity. Controller age is a weak proxy for controller experience. A more focused look at controller experience may reveal a different pattern. Additionally, it is important to note that these results are in terms of severity and nothing can be said about the frequency with which controllers of a given age commit errors.

Relevant Training in the Last Year

(ATQA OE)

This variable indicates whether the controller was involved in "relevant" training in the last year. Note that this is a self-reported variable on the controller incident reporting form. Additionally, no guidance is given on what constitutes relevant training. At a minimum it is assumed to be training broadly related to runway incursions.

Table 127 – Observed Distribution of Relevant Training in Last Year by Severity

	A	B	C	D	Total
No	5	7	114	12	138
Yes	39	32	592	59	722
Total	44	39	706	71	860

Table 128 – Expected Distribution of Relevant Training in Last Year by Severity

	A	B	C	D	Total
No	7	6	113	11	138
Yes	37	33	593	60	722
Total	44	39	706	71	860

There does not appear to be any relationship between receiving training and severity. It is possible that training may affect the frequency with which errors occur, but no conclusion regarding frequency can be drawn from these results.

Controller Workload

(ATQA OE)

Controller workload measures the number of aircraft the controller was responsible for at the time of the incident. This is a self-reported variable on the controller error reporting form.

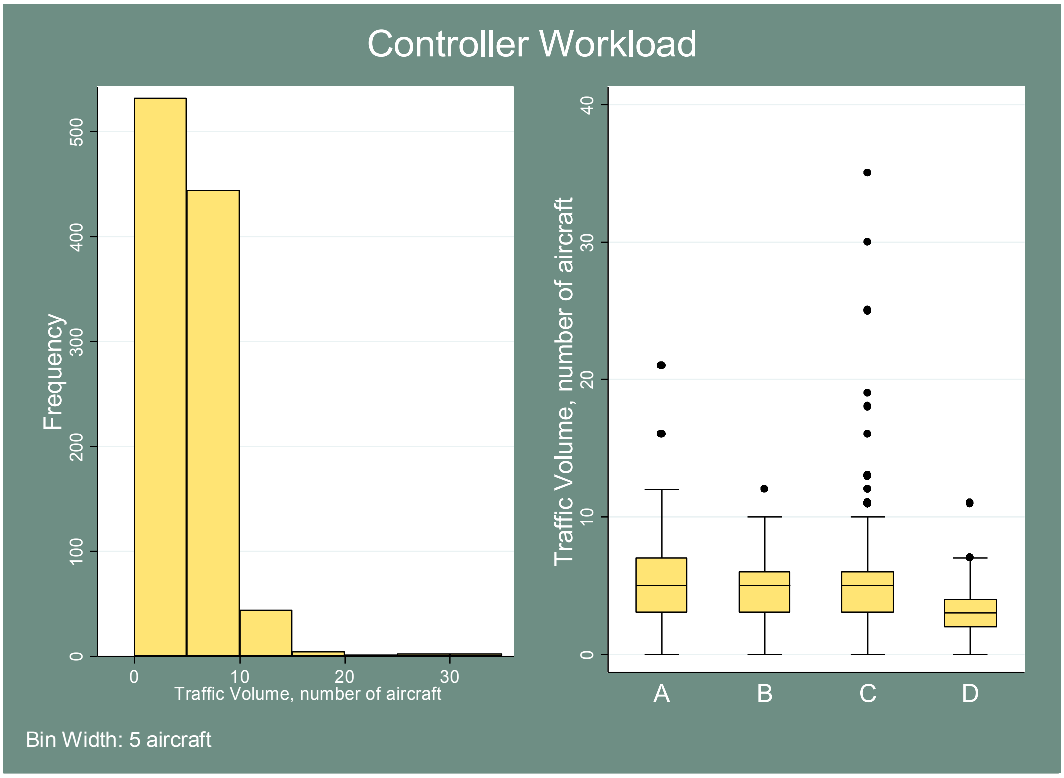

Figure 29 – Distribution of Controller Workload

Table 129 – Percentiles of Controller Workload

	10th	25th	50th	75th	90th
A	2	3	5	7	8
B	1	3	5	6	10
C	2	3	5	6	8
D	1	2	3	4	5
Overall	2	3	4	6	8

Table 130 – Kruskal-Wallis Test Results for Controller Workload

	A	B	C	D
Number of Observations	48	38	841	102

	A	B	C	D
Mean Rank	**549.94**	**559.32**	**537.02**	**300.46**

The test results indicate that the severity categories are jointly different in terms of controller workload. Further, all categories can be considered pairwise different from category D (no other pairwise comparisons are significantly different). Table 131 presents the results of a Kruskal-Wallis test for conflict events only. Once the conflict versus non-conflict dynamic has been eliminated, controller workload does not appear to have a different distribution by severity. Controller workload may serve as a proxy for the overall traffic level at an airport, rather than directly impacting severity. A more focused look at extreme controller workload levels may also reveal a different pattern (given that the overall distributions are fairly narrow).

Table 131 – Kruskal-Wallis Test Results for Controller Workload, Conflict Only

	A	B	C
Number of Observations	**48**	**38**	**841**
Mean Rank	**477.02**	**483.99**	**462.35**

3.3.6. Weather Variables

These variables capture the weather conditions surrounding the incident. As described previously, the weather data originates from the METAR data archived by Plymouth University. It is then interpolated to represent a best approximation of the conditions at the time of the incident.

Temperature

(Weather Database)

The temperature at the time of the incident is interpolated from the closest hourly readings. Figure 30 presents the overall distribution of temperature, the distribution by severity, and the distribution by incident type. The percentiles of the distribution, conditional on severity and incident type, are presented in Table 132 and Table 134.

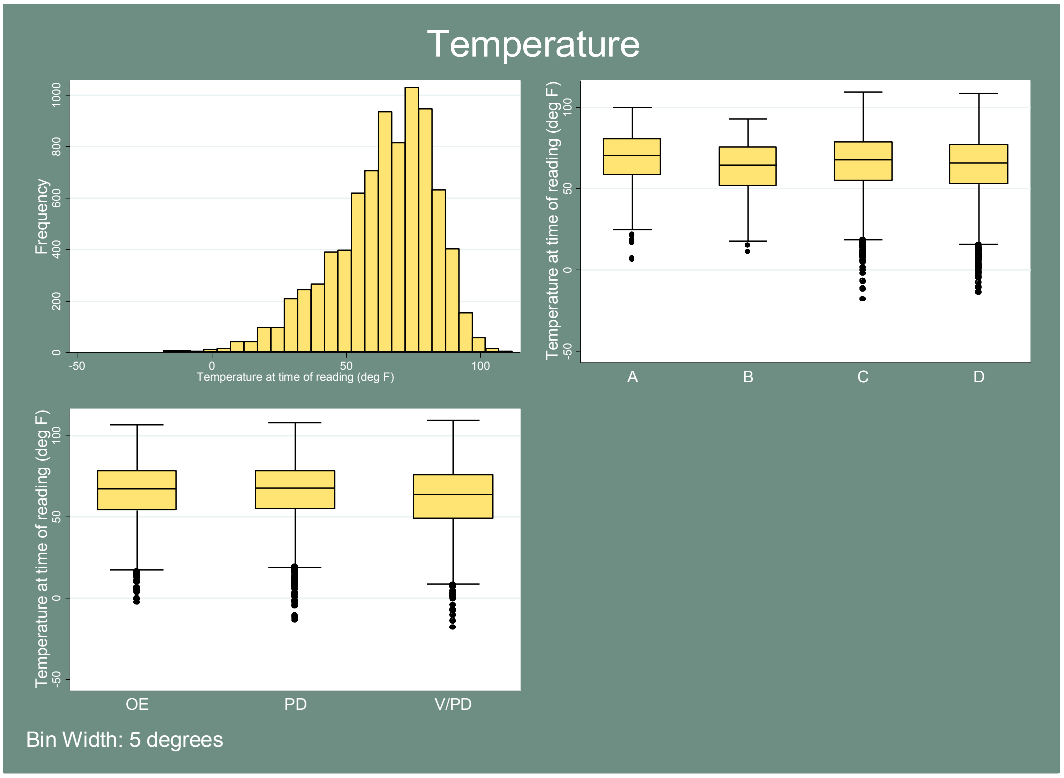

Figure 30 – Distribution of Temperature

The overall distribution is unsurprising. This data covers approximately ten years and the 50 States, the District of Columbia, and U.S. Territories; thus, the range seems reasonable. The overall distribution is skewed slightly left, but not dramatically so. The distribution by severity appears fairly similar. Category A and C incursions appear to have slightly higher median temperatures. One might anticipate that ice (and thus cold temperatures) would have a disproportionate effect on severity, but that does not seem to be the case. The distributions by severity also appear quite similar, with V/PD incidents having a slightly lower median temperature. This may be indicative of the involvement of snow removal vehicles in V/PD incidents.[51] To further test these apparent differences by severity and incident type, two Kruskal-Wallis tests were performed, the results of which are presented in Table 133 and Table 135.

Table 132 – Percentiles of Temperature by Severity

	10th	25th	50th	75th	90th
A	**36**	**58.35**	**70.7833**	**81**	**88**

51 There are 81 incidents involving snow removal vehicles in the database, 63 of which are V/PD incidents, constituting approximately 3% of V/PD incidents.

	10th	25th	50th	75th	90th
			3		
B	38.225	52	64.825	75.8	81.58333
C	38.31667	54.8	67.93333	79	86
D	37	52.75	66	77.4	84.65
Overall	37	53.76667	66.76667	78	85.13333

Table 133 – Kruskal-Wallis Test Results for Temperature by Severity

	A	B	C	D
Number of Observations	122	130	3110	4787
Mean Rank	4529.80	3798.59	4187.81	3997.63

Table 134 – Percentiles of Temperature by Incident Type

	10th	25th	50th	75th	90th
OE	37.93333	54	67.51667	78.46667	85.2
PD	40	54.83333	68	78.63333	85.55
V/PD	32.66667	48.96667	64	75.96667	84
Overall	37	53.76667	66.76667	78	85.13333

Table 135 – Kruskal-Wallis Test Results for Temperature by Incident Type

	OE	PD	V/PD
Number of Observations	1222	4945	1982
Mean Rank	4121.64	4197.31	3741.08

2

The results for severity indicate that, while jointly different, few of the categories can be declared different from each other. Categories C and D are the only two categories that can be declared different. This is partially due to the smaller sample of A and B incidents, leading to less precise estimates of their distributions. There does not seem to be a trend with severity and temperature. It is unclear how temperature alone might impact severity, but temperature may be a proxy for more specific weather phenomena, such as snow and ice. While snow and ice may impact severity, it is possible current operational practices (such as reducing traffic volume) already compensate for the increased risk of a severe incident. Further research, focusing on these particular phenomena (icy runways and snow) may disentangle the operational effects from the weather effects.

The test by incident type indicates that the three incident types are jointly different and that V/PD incidents are distinct from both OE and PD incidents (OE and PD incidents are not able to be distinguished). This supports the conclusion drawn from the distributional graph, but provides no further indication as to why V/PDs might have a different distribution of temperature. There is a broad range of factors that could influence V/PD incidents to occur at lower temperatures, including: the national geographic distribution of V/PD incidents, the prevalence of snow removal equipment in V/PD incidents, and changes in airport vehicle driver behavior due to cold weather. It is unlikely that temperature causes V/PDs; investigating factors related to cold weather that may cause V/PDs may be helpful in understanding this distribution and its policy implications.

Dew Point

(Weather Database)

This variable provides an estimate of the dew point at the time of the incident. The dew point indicates the temperature at which water vapor in the air condenses into liquid water. Higher dew points are associated with more humid air and severe weather.[52] As with the many of the weather variables, it is unlikely that a higher or lower dew point causes increased or decreased severity. However, factors

52 National Weather Service Weather Forecast Office (2012).

related to dew point (such as haziness or approaching weather) may contribute to increased or decreased severity. Figure 31 presents the distribution of this variable overall, by severity, and by incident type.

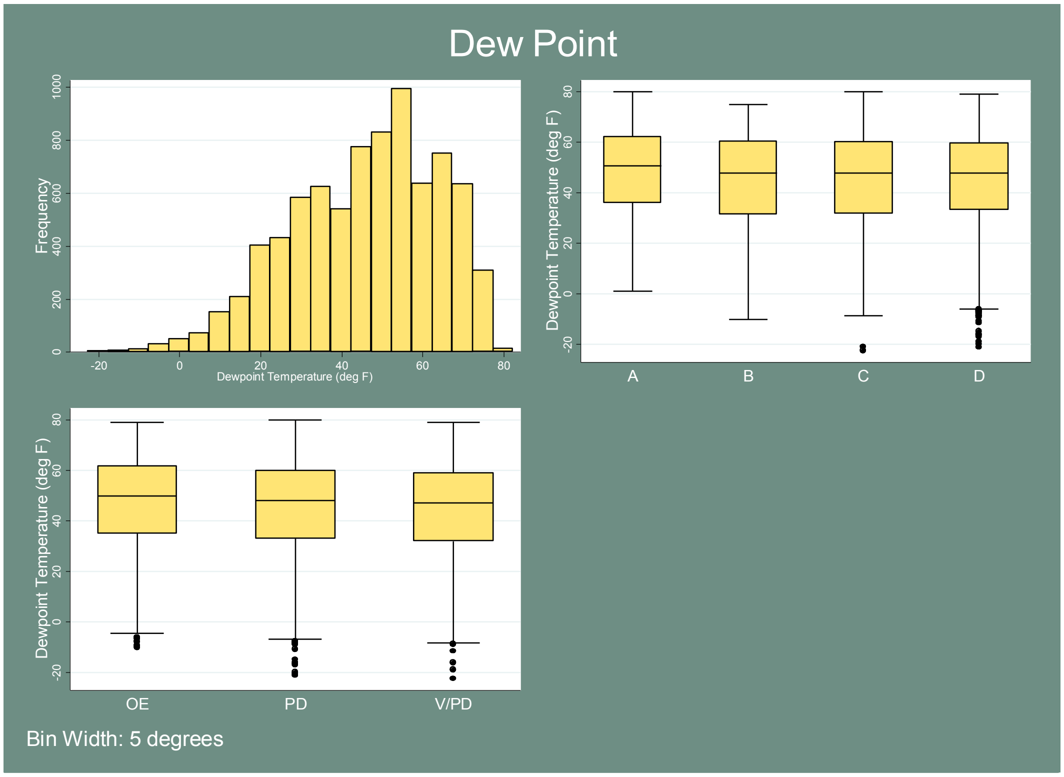

Figure 31 – Distribution of Dew Point

The distributions across severity types appear fairly similar. Category A incursions appear to have a slightly higher median dew point than the other three categories. Similarly, OE incursions appear to have a higher median dew point than either PD or V/PD incursions. Table 136 and Table 138 present the percentiles of the distribution by severity and incident type. Table 137 and Table 139 present the results of Kruskal-Wallis tests by severity and incident type.

Table 136 – Percentile of Dew Point by Severity

	10th	25th	50th	75th	90th
A	23	36	50.55	62.2166 7	70
B	16	31.75	48	60.5666 7	70
C	19.9	32	48	60.2666 7	68.9333 3
D	21	33.3666	48	59.8833	68.2666

	10th	25th	50th	75th	90th
		6		4	7
Overall	21	33	48	60	68.5833 4

Table 137 – Kruskal-Wallis Test Results for Dew Point by Severity

	A	B	C	D
Number of Observations	120	128	3104	4743
Mean Rank	4290.4 4	4078.6 4	4023.1 1	4057.3 3

Chi2 score 1.74

Degrees of Freedom: 3

P-value: 0.63

Table 138 – Percentile of Dew Point by Incident Type

	10th	25th	50th	75th	90th
OE	21.8	34.95	50	61.8666 6	69.5666 7
PD	21	33	48	60	68.2666 7
V/PD	19.4666 7	32	47.0333 3	59	68.1333 3
Overall	21	33	48	60	68.5833 4

Table 139 – Kruskal-Wallis Test Results for Dew Point by Incident Type

	OE	PD	V/PD
Number of Observations	1219	4925	1951
Mean Rank	4290.4 4	4078.6 4	4023.1 1

5

dom: 2

The results by severity indicate that the severity categories are indistinguishable jointly. That is, it appears that dew point does not vary systematically by severity category. This is not entirely surprising, given that there is no strong hypothesis for how or why dew point would impact severity. If dew point were a proxy for another underlying cause (such as haziness), it is not a strong enough proxy to show up in these results. A more focused examination of other weather phenomena may provide additional insight.

The results by incident type do indicate differences among groups. While the three incident types are jointly different, only OE incidents can be distinguished from any other group (PD and V/PD incidents are indistinguishable). It is unclear why OEs have a higher median dew point. It is likely that there is some underlying cause associated with dew point that is manifesting in this test statistic. A more focused study may reveal that underlying cause (or causes) or indicate that this is a spurious correlation.

Temperature-Dew Point Difference

(Weather Database)

Continuing with the examination of temperature measures, this variable examines the difference between temperature and the dew point. When the dew point and temperature are closer, fog and precipitation are more likely.[53] Figure 32 presents the distribution of this variable overall, by severity, and by incident type.

53 Ibid.

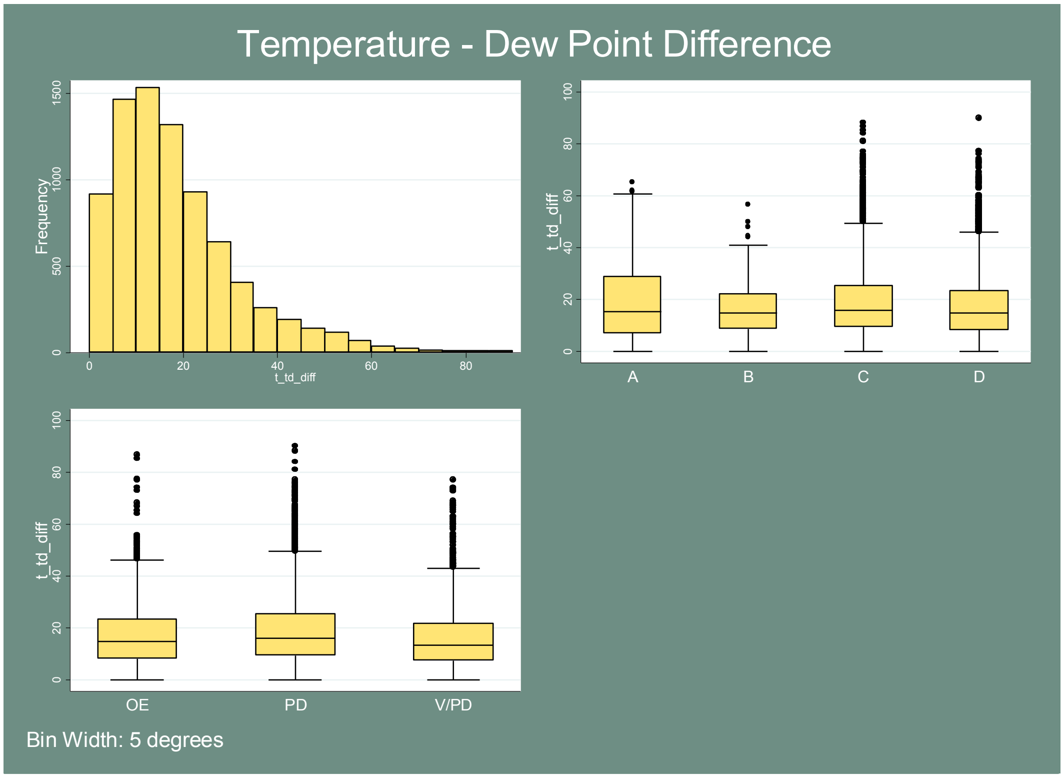

Figure 32 – Distribution of Temperature – Dew Point Difference

Firstly, there are no negative values. This is due to an intrinsic relationship between dew point and temperature. Secondly, the differences between temperature and dew point can be quite large, though most of the distribution is contained below the twenty-degree difference mark. The distribution by severity appears fairly similar among all four categories. By incident type, the distributions also appear similar, but PD incidents may have a slightly higher median difference. The percentiles of the distributions by severity and incident type are contained in Table 140 and Table 142. The results of Kruskal-Wallis tests by severity and incident type are contained in Table 141 and Table 143.

Table 140 – Percentiles of Temperature – Dew Point Difference by Severity

	10th	25th	50th	75th	90th
A	**2.399999**	**7.150002**	**15.4**	**28.95**	**39.48333**
B	**3.233334**	**8.833336**	**14.89166**	**22.275**	**30.4**
C	**5**	**9.391668**	**16**	**25.51666**	**38.06667**
D	**4.166668**	**8.5**	**14.95**	**23.53333**	**34.7**
Overall	**4.36666**	**8.94999**	**15.3333**	**24.4**	**36**

	10th	25th	50th	75th	90th
	5	7	4		

Table 141 – Kruskal-Wallis Test Results for Temperature – Dew Point Difference by Severity

	A	B	C	D
Number of Observations	120	128	3104	4743
Mean Rank	4060.2 6	3810.8 9	4209.3 2	3948.5 2

Table 142 – Percentiles of Temperature – Dew Point Difference by Incident Type

	10th	25th	50th	75th	90th
OE	4	8.25	14.9	23.5333 3	34.2
PD	5	9.44999 9	16.2	25.5333 3	38.8333 3
V/PD	3.96666 7	7.53333 3	13.45	21.8	32.0333 3
Overall	4.36666 5	8.94999 7	15.3333 4	24.4	36

Table 143 – Kruskal-Wallis Test Results for Temperature – Dew Point Difference by Incident Type

	OE	PD	V/PD
Number of Observations	1219	4925	1951
Mean Rank	3891.1 7	4239.5 5	3662.4 5

3

dom: 2

The results in Table 141 indicate that the severity levels differ jointly, but only categories C and D can be distinguished from each other. It is possible that with more observations, Categories A and B might also be able to be distinguished. It appears that category C has a slightly higher median difference than category D. Further research is required to understand if this is indicative of a true impact on severity or an artifact of the data.

The results of the test by incident type indicate that the three incident types are not only jointly significant, but all pairwise different from each other. The source of the observed differences is unclear. PD incidents have the largest median difference while V/PD incidents have the smallest. The difference between temperature and dew point is related to the chance of precipitation, and it is possible that pilot behavior is responding to this. That is, if fewer pilots (presumably GA) fly when the chance of precipitation is higher; this may drive the median difference higher. Explanations for the variation in OE and V/PD incidents are less forthcoming. Factors related to the difference of temperature and dew point (notably precipitation) and how those factors impact incidents of various types should be investigated further.

Cloud Ceiling

(Weather Database)

This variable measures the height of the cloud ceiling at the time of the incident. It was interpolated in a similar fashion to the other weather variables. Figure 33 presents the distribution of this variable.

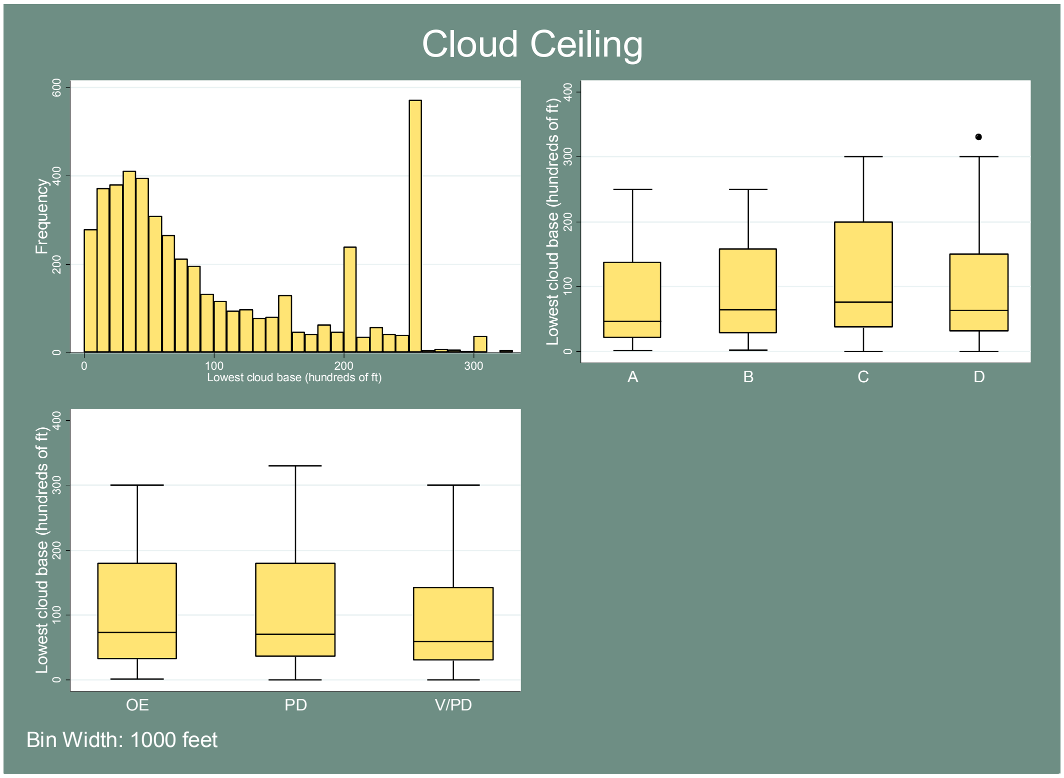

Figure 33 – Distribution of Cloud Ceiling

Note that there is a large amount of peaking at certain values (approximately 15, 20, 25, and 30 thousand feet). This is likely due to rounding by those reporting the incident. The distribution by severity indicates that the median cloud level increases as severity decreases from A to C. Category D breaks this pattern. As noted previously, it is possible that category D incidents are a product of a different process than conflict incidents – this may be yet another supporting indication. Cloud ceiling looks fairly similar between OE and PD incidents while V/PD incidents appear to have a slightly lower median. Table 144 and Table 146 present the percentiles of the distribution by severity and incident type. Table 145 and Table 147 present the results of Kruskal-Wallis tests by severity and incident type.

Table 144 – Percentiles of Cloud Ceiling by Severity

	10th	25th	50th	75th	90th
A	8	22.6	46.68333	137.2	250
B	17	29.3	65	158.3333	250
C	17.5	37.06667	76.25	200	250
D	14.03333	32	64	150	250

	10th	25th	50th	75th	90th
Overall	**15**	**34.08333**	**68.95833**	**168.5357**	**250**

Table 145 – Kruskal-Wallis Test Results for Cloud Ceiling by Severity

	A	B	C	D
Number of Observations	**58**	**82**	**1893**	**2755**
Mean Rank	**2032.20**	**2278.98**	**2531.74**	**2311.27**

Table 146 – Percentiles of Cloud Ceiling by Incident Type

	10th	25th	50th	75th	90th
OE	**15**	**32**	**73.15**	**180**	**250**
PD	**16.2**	**36.2**	**70**	**180**	**250**
V/PD	**12**	**29.60833**	**60**	**142.4333**	**250**
Overall	**15**	**34.08333**	**68.95833**	**168.5357**	**250**

Table 147 – Kruskal-Wallis Test Results for Cloud Ceiling by Incident Type

	OE	PD	V/PD
Number of Observations	**822**	**2798**	**1168**
Mean Rank	**2450.44**	**2442.40**	**2240.37**

The results indicate that cloud ceiling differs significantly by incident type and severity. Both OE and PD incidents are distinguishable from V/PD incidents while OE and PD incidents are not pairwise different. This supports the observation noted above and warrants further investigation as it is not clear why cloud cover should impact incidents where a vehicle or pedestrian was at fault (or even runway incursions in general, though it may impact visibility for pilots and controllers).

The patterns by severity are less clear. While jointly different, only categories A and C and C and D can be considered pairwise different. All other combinations are not significantly different. This is similar to the pattern noted in the distribution – that A, B, and C incursions have a trend in median ceiling level while category D appears similar to B, breaking the pattern – but there is not strong evidence to support it. Thus, it is possible that there is an impact of cloud ceiling height on event severity, but the effects are not clear. The mechanism through which cloud ceiling would impact runway incursion severity is also not clear. If the factor at play is really visibility, a more direct measurement of visibility would offer improved explanatory power.

Cloud Coverage

(Weather Database)

This variable indicates how much of the sky was covered with clouds. The original rating is presented as a series of increasing fractions from Clear (0/8ths of the sky covered) to Overcast (8/8ths of the sky covered). Due to the sequential nature of these categories (and their approximation to fractions), it was decided to turn this variable into a numeric variable describing how many eighths of the sky is covered. Thus, the variable ranges from 0 to 8. Table 148 presents the mapping from the original categories to the numeric values. As the original categories covered a range of values, the midpoint of each range was used.[54]

Table 148 – Mapping of Cloud Coverage Categories to Numeric Values

Original Category	Numeric Value
Clear (0/8)	**0**
Few (between 0/8 and 2/8)	**1**
Scattered (between 2/8 and 4/8)	**3**

54 The categories presented in Table 129 present an interesting problem. First, the categories are of differing widths. Clear and Overcast only cover one value while Few, Scattered, and Broken represent ranges. Additionally, some categories overlap, while others are adjacent. Clear indicates 0/8 parts of the sky is covered. The next category, Few, indicates that between 0 and 2 out of 8 parts are covered. This category picks up exactly where clear left off. Scattered begins at 2 where Few left off and ends at 4. Broken, however, begins at 5 – one unit more than where Scattered ends. Overall this is likely a minor quirk in the definition, but it may create artifacts in the data and ends up making the top part of the scale more spaced out than the bottom half.

Original Category	Numeric Value
Broken (between 5/8 and 7/8)	**6**
Overcast (8/8)	**8**

After conversions to a 0 to 8 scale, values were interpolated between the two points and then rounded. This was an attempt to more accurately represent the precision of the information in the data. The original data did not contain the high level of decimal precision implied by the interpolation process, thus the data was rounded to the nearest half. The final data measures the number of eighths of the sky covered from 0 to 8, measured in steps of 0.5. While the units may seem odd, the variable can still be interpreted as the fraction of the sky covered with clouds.

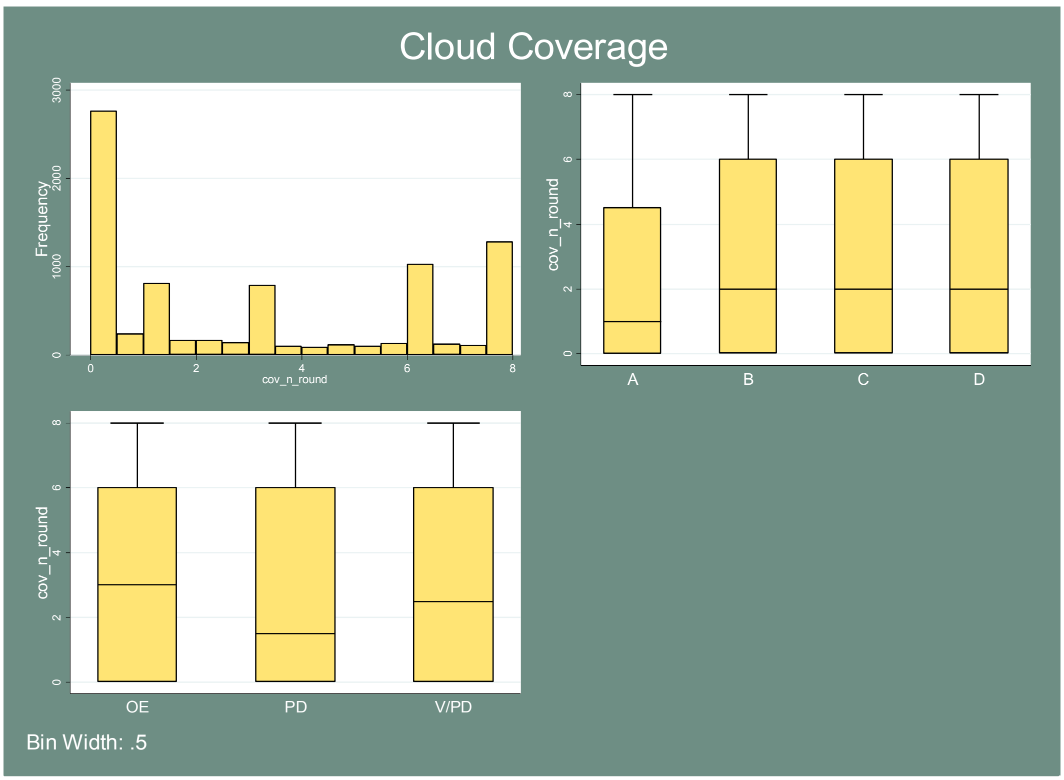

Figure 34 – Distribution of Cloud Coverage, Rounded

The rounding procedure mentioned above has a distinct effect on the distribution of this variable, as seen in Figure 34. Note that in addition to the rounding to the nearest half, there are also distinct spikes are certain values – such as 1, 3, 6, and 8. These values are the midpoints of the original categories, as indicated in Table 148. There are still a fair amount of observations in between these values, as generated by interpolation, but this clumping is important to be aware of when considering the impact this variable may have.

When considered by severity, all the categories appear similar aside from category A, which has lower median cloud coverage. This is surprising, as a naïve hypothesis is that increased cloud coverage would increase severity. Further testing is required to determine if this difference is significant or an artifact of the data. Similarly for incident type, while the different types appear to have different median cloud cover values, further testing is required to see if the difference is significant. Table 149 and Table 151 present the percentiles of the distribution by severity and incident type. Table 150 and Table 152 present the results of a Kruskal-Wallis test by severity and incident type to examine these issues.

Table 149 – Percentiles of Cloud Coverage by Severity

	10th	25th	50th	75th	90th
A	**0**	**0**	**1**	**4.5**	**8**
B	**0**	**0**	**2**	**6**	**8**
C	**0**	**0**	**2**	**6**	**8**
D	**0**	**0**	**2**	**6**	**8**
Overall	**0**	**0**	**2**	**6**	**8**

Table 150 – Kruskal-Wallis Test Results for Cloud Coverage by Severity

	A	B	C	D
Number of Observations	**58**	**82**	**1893**	**2755**
Mean Rank	**2032.20**	**2278.98**	**2531.74**	**2311.27**

Table 151 – Percentiles of Cloud Coverage by Incident Type

	10th	25th	50th	75th	90th
OE	**0**	**0**	**3**	**6**	**8**
PD	**0**	**0**	**1.5**	**6**	**8**
V/PD	**0**	**0**	**2.5**	**6**	**8**
Overall	**0**	**0**	**2**	**6**	**8**

Table 152 – Kruskal-Wallis Test Results for Cloud Coverage by Incident Type

	OE	PD	V/PD
Number of Observations	**1223**	**4961**	**2013**
Mean Rank	**4467.49**	**3965.44**	**4204.29**

The results by severity indicate that the categories are not jointly different. This indicates that the lower median coverage observed in Figure 34 is an artifact of the data rather than a substantial difference. The results by incident type are more interesting. All incident types are jointly different as well as pairwise different. Pilots appear to have the lower median than the other two incident types indicating that pilot incidents tend to happen with less of the sky covered by clouds. VFR are also more likely when there are fewer clouds – increasing the number of pilots flying, and thus potentially involved in a runway incursion. It is likely that cloud coverage, like cloud ceiling, is related to visibility. Cloud coverage should be investigated as part of a broader study on weather impacts, although the main influence appears to be on incident type rather than on severity.

Visibility

(Weather Database)

While the previous two variables dealt with visibility indirectly, this variable measures visibility directly. This variable measures the distance one can see (approximately) in miles. Figure 35 and Figure 36 present the same information, but figure twelve focuses on reports of visibility less than 10 miles.

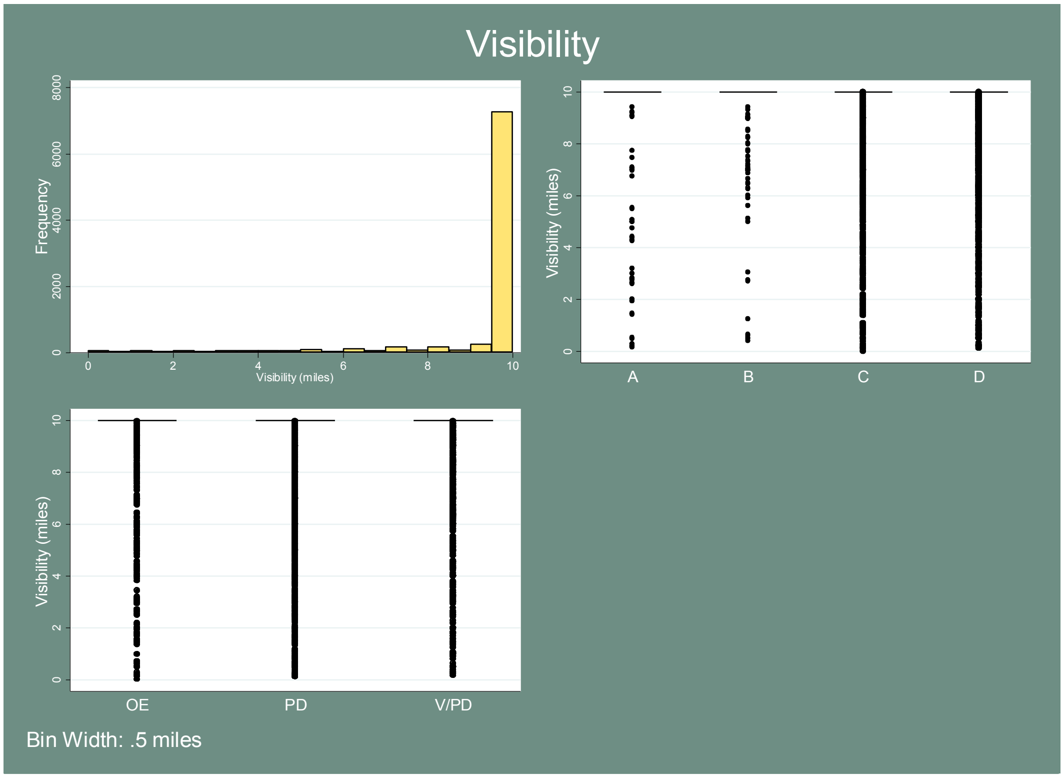

Figure 35 – Distribution of Visibility

As Figure 35 indicates there is extreme bunching of visibility readings at 10 miles, which is effectively a coding for unlimited. The bunching is so dramatic that when broken down by severity or incident type, all parts of the box plot are coded as 10 miles – i.e. the upper and lower whiskers, 25th, 50th, and 75th percentiles are all 10 miles. Figure 36 focuses on the distribution of readings less than 10 miles (i.e., times with less than unlimited visibility), to enable a clearer analysis of the distribution of visibility.

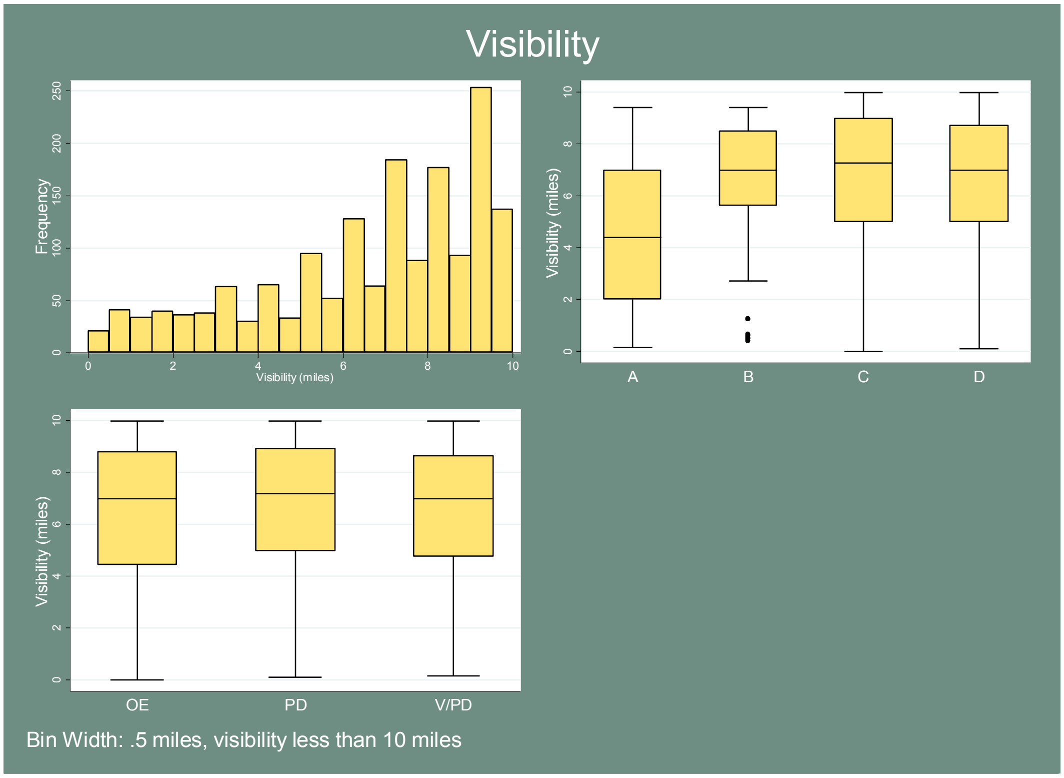

Figure 36 – Distribution of Visibility, Visibility Less than 10 miles

The distribution appears to be leftward skewed, with more readings occurring at higher readings (though less than 10 miles). This is likely indicative of a larger trend in behavior of less traffic when the conditions are low visibility. This may be due to changes in flight rules (visual versus instrument) or due to pilots simply choosing to stay on the ground. There also appears to be bunching near whole values, indicating some rounding taking place among those generating METAR readings.

The distribution by severity hints that category A incursions may occur with a lower median visibility, though the interquartile range is fairly large, as seen in Table 153. The remaining categories of B, C, and D all appear to have similar median visibilities. Category B also has smaller whiskers. While all other categories cover almost the entire range, category B's whiskers are much smaller, covering slightly more than half the range. This is indicative that the distribution of visibility among category B incidents is narrower than other categories. The distribution across incident types appears almost identical in terms of median, interquartile range and whiskers. The percentiles by incident type are given in Table 155.

Table 154 and Table 156 present the results of Kruskal-Wallis tests by severity and incident type.

Table 153 – Percentiles of Visibility by Severity

	10th	25th	50th	75th	90th
A	.2733333	2	4.4	7	9.05
B	1.25	5.61666	7	8.48333	9

	10th	25th	50th	75th	90th
		7		3	
C	**2.141667**	**5**	**7.266667**	**9**	**9.466667**
D	**2.516667**	**5**	**7**	**8.720833**	**9.366667**
Overall	**2.416667**	**5**	**7**	**8.8**	**9.4**

Table 154 – Kruskal-Wallis Test Results for Visibility by Severity

	A	B	C	D
Number of Observations	**27**	**35**	**603**	**1008**
Mean Rank	**480.02**	**799.57**	**861.42**	**833.25**

Table 155 – Percentiles of Visibility by Incident Type

	10th	25th	50th	75th	90th
OE	**1.96**	**4.416667**	**7**	**8.8**	**9.4**
PD	**2.516667**	**5**	**7.183333**	**8.9**	**9.433333**
V/PD	**2.183333**	**4.772358**	**7**	**8.65**	**9.366667**
Overall	**2.416667**	**5**	**7**	**8.8**	**9.4**

Table 156 – Kruskal-Wallis Test Results by Incident Type

	OE	PD	V/PD
Number of Observations	**257**	**942**	**474**
Mean Rank	**799.48**	**858.27**	**815.08**

	OE	PD	V/PD

The results indicate that the severity categories are jointly different while the incident types are not. Category A incursions can be distinguished from categories C and D. After the correction for multiple comparisons, categories A and B are considered not significantly different, albeit barely. With more observations in each category, it is likely that categories A and B could be distinguished. This suggests that the lower median visibility for category A is significant. Note that these are all conditional on visibility being less than 10 miles. Without that constraint, the categories are indistinguishable.

Properly controlling for the relation among visibility, ceiling, and cloud cover might reveal the nature of the interaction. Indeed, many weather phenomena (such as precipitation) might impact severity through reduced visibility. This research hints at the impact weather may have, but a more thorough undertaking with precise weather data would illuminate some of these issues.

Visual Meteorological Conditions (VMC)

(Runway Incursion Database)

This variable indicates (broadly) the weather conditions at the time of the incident. This is not to be confused with visual (or instrument) flight rules which indicate the operating procedure at that time. VMC indicates that the weather was good enough for visual flight. The overall frequency of this variable is noted in Figure 37.

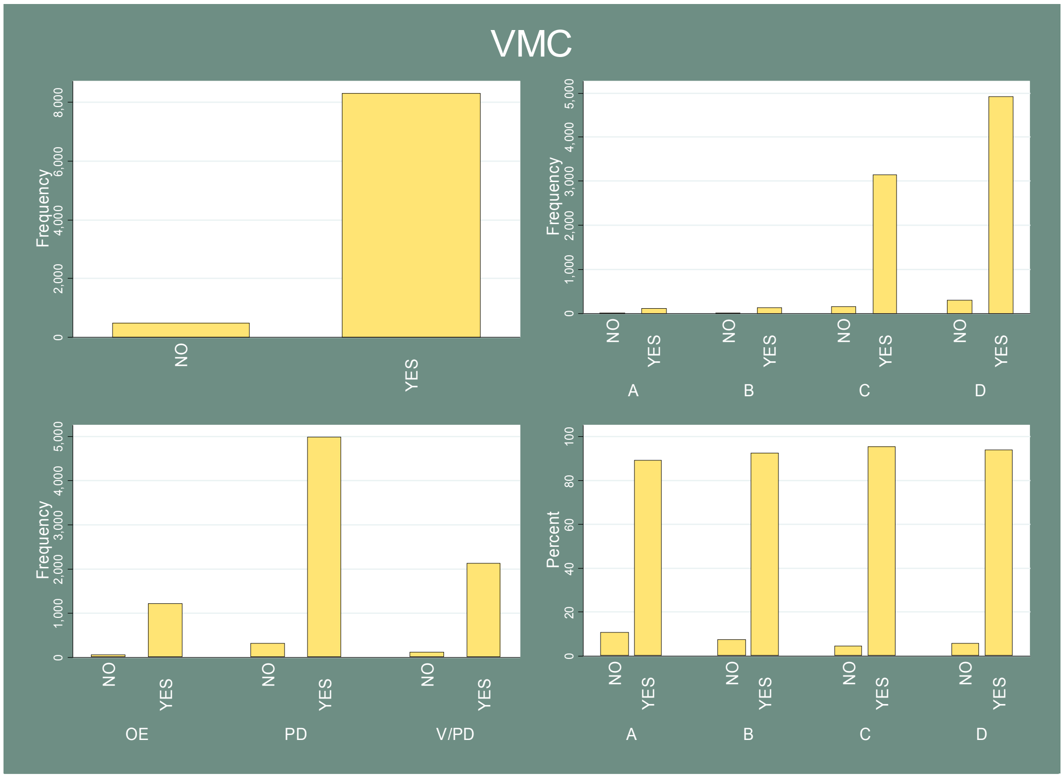

Figure 37 – Overall Distribution of VMC

Table 157 indicates the impact of VMC on the odds of being severe.

Table 157 – Logit Estimate of Impact on Severity, VMC

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB
VMC	**.578**	**.124**	**0.01**	**.379**	**.881**

Not surprisingly, VMC are associated with less severe incidents. The magnitude is also quite large, almost halving the odds. This impact is likely related to visibility and, perhaps, reduced complexity of operations. Because this variable (possibly) conflates many different effects, it is less attractive as a modeling variable.

Sea Level Pressure Deviation

(Weather Database)

This variable indicates the air pressure at the time of the incident, normalized to sea pressure. Pressure varies with altitude, thus it is important to normalize to a standard altitude (in this case, sea level). Thus, it is most helpful to examine this variable in terms of deviation from standard pressure (1013.25 mb). Figure 38 presents this distribution. The percentiles of the distribution by incident type and severity are presented in Table 158 and Table 159 while the results of a Kruskal-Wallis test by severity and incident type are presented in Table 160 and Table 161, respectively.

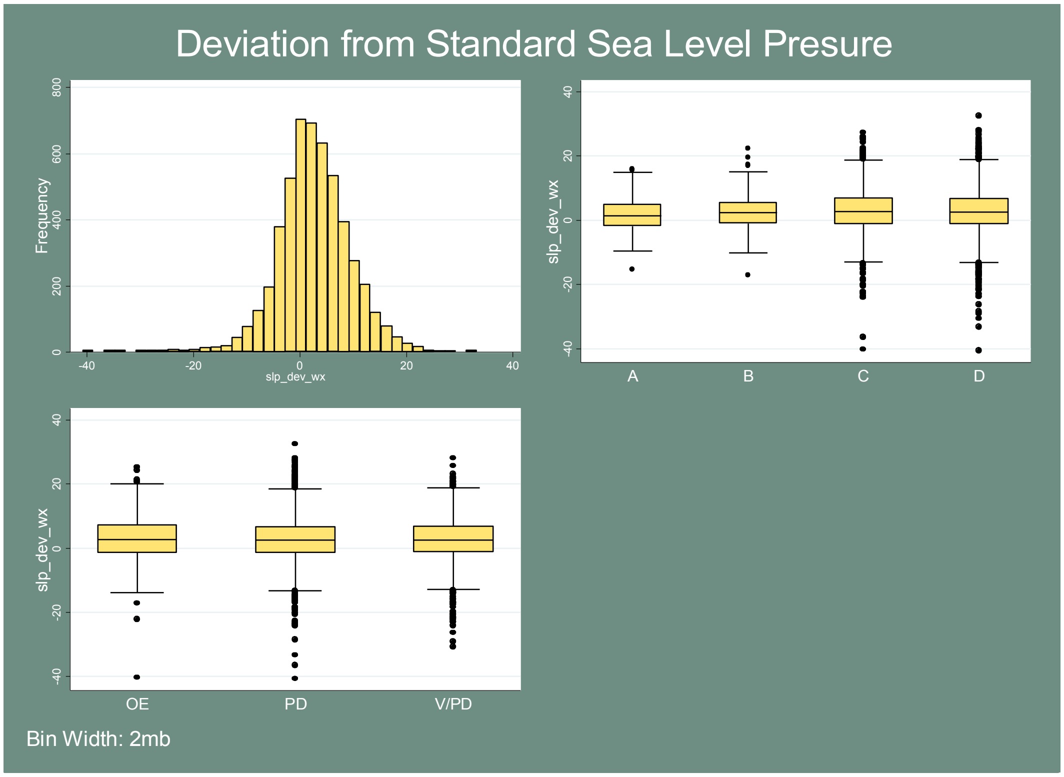

Figure 38 – Distribution of Deviation of Sea Level Pressure from Standard Pressure

Table 158 – Percentiles of Deviation of Sea Level Pressure by Severity

	10th	25th	50th	75th	90th
A	-4.063353	-1.870853	1.272512	4.997483	8.01001
B	-3.731671	-1.024994	2.349976	5.449982	12.93162
C	-4.8	-1.160004	2.783358	6.829978	11.36664
D	-5.076664	-1.288324	2.493335	6.773335	11.09669
Overall	-4.91667	-1.256665	2.572498	6.77001	11.14005

Table 159 – Percentiles of Deviation of Sea Level Pressure by Incident Type

	10th	25th	50th	75th	90th
OE	-4.650024	-1.244983	2.821649	7.310004	11.65002
PD	-4.666676	-1.276668	2.510009	6.663354	10.97664
V/PD	-5.516679	-1.150024	2.549988	6.849976	11.41001
Overall	-4.91667	-1.256665	2.572498	6.77001	11.14005

There does not appear to be any relationship between this variable and severity, nor between this variable and incident type. This conclusion is supported by the results of the Kruskal-Wallis tests outlined below.

Table 160 – Kruskal-Wallis Test Results for Deviation of Sea Level Pressure by Severity

	A	B	C	D
Number of Observations	72	55	1997	3062
Mean Rank	2327.37	2473.89	2626.74	2580.23

Table 161 – Kruskal-Wallis Test Results for Deviation of Sea Level Pressure by Incident Type

	OE	PD	V/PD
Number of Observations	795	3170	1221
Mean Rank	2659.30	2580.94	2583.27

Weather Phenomena

(Weather Database)

In addition to basic weather information, the METAR reports contain information regarding any weather phenomena at the measurement time. The majority of these phenomena encompass different kinds of precipitation. In addition to the various kinds of precipitation, haze, fog, and smoke are also indicated. As Figure 39 indicates, the majority of incursions occur when there are no weather phenomena. This is not surprising, given that many amateur pilots may not be able to fly in less than pristine meteorological conditions. Figure 40 presents the same distribution but excludes cases of “No Weather.” Overall, the distribution is dominated by “haze,” “light rain,” and “light snow.”

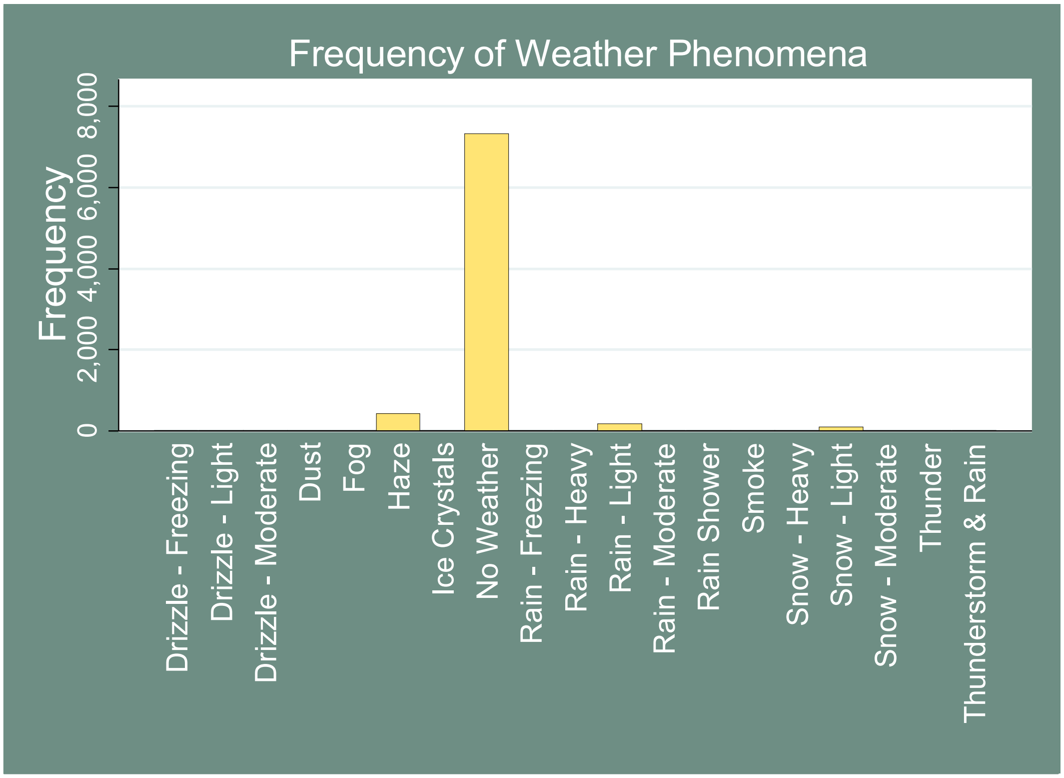

Figure 39 – Distribution of Weather Phenomena

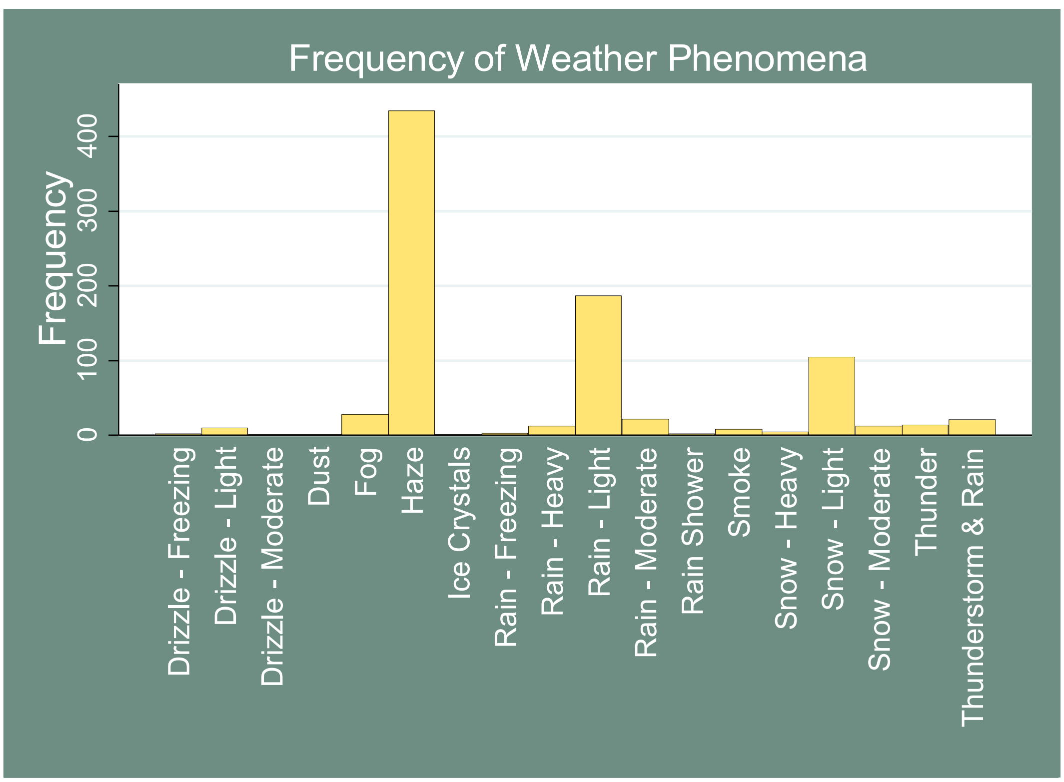

Figure 40 – Distribution of Weather Phenomena, excludes "No Weather"

To simplify the analysis, the various weather phenomena codes have been collapsed into a dichotomized variable indicating if there was *any* weather at the time of the incident. Table 162 and Table 163 present the observed and expected distributions of this indicator.

Table 162 – Observed Distribution of No Weather Indicator by Severity

	A	B	C	D	Total
Weather Present	**21**	**17**	**296**	**535**	**869**
No Weather	**101**	**114**	**2,817**	**4,291**	**7,323**
Total	**122**	**131**	**3,113**	**4,826**	**8,192**

Table 163 – Expected Distribution of No Weather Indicator by Severity

	A	B	C	D	Total
Weather Present	**13**	**14**	**330**	**512**	**869**
No Weather	**109**	**117**	**2,783**	**4,314**	**7,323**
Total	**122**	**131**	**3,113**	**4,826**	**8,192**

The test results indicate that there is a relationship between this variable and severity. Categories A, B, and D appear underrepresented, although categories A and B are barely lower than the expected values. It appears that the relationship is driven primarily by the observed and expected results from categories C and D. After excluding non-conflict events the results are similar. Categories A and B are underrepresented, while category C incursions are observed more than expected. This indicates that incursions tend to be less severe when there are no weather phenomena.

Table 164 – Observed Distribution of No Weather Indicator by Severity, Conflict Only

	A	B	C	Total
Weather Present	**21**	**17**	**296**	**334**
No Weather	**101**	**114**	**2,817**	**3,032**
Total	**122**	**131**	**3,113**	**3,366**

Table 165 – Expected Distribution of No Weather Indicator by Severity, Conflict Only

	A	B	C	Total
Weather Present	**12**	**13**	**309**	**334**
No Weather	**110**	**118**	**2,804**	**3,032**
Total	**122**	**131**	**3,113**	**3,366**

Wind Speed

(Weather Database)

This variable measures the wind speed at the time of the incident (in knots). Figure 41 and Table 166 present the distribution of wind speed. Table 167 contains the results of a Kruskal-Wallis test by severity. There does not appear to be a significant relationship between wind speed and severity.

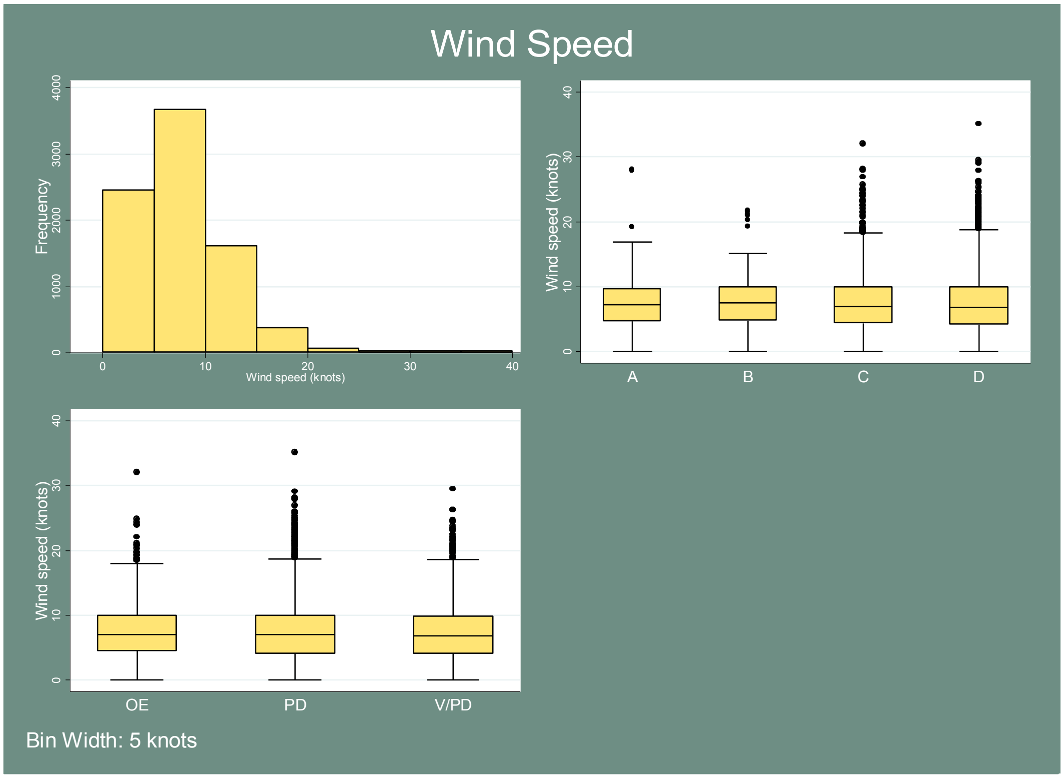

Figure 41 – Distribution of Wind Speed

Table 166 – Percentiles of Wind Speed by Severity

	10th	25th	50th	75th	90th
A	2.566667	4.65	7.25	9.716666	12.83333
B	1.4	4.8	7.541667	10	11.93333
C	1.9	4.416667	7	10	13.2
D	1.4	4.15	6.85	10	13.06667
Overall	1.6	4.266667	6.966667	10	13.1

Table 167 – Kruskal-Wallis Test Results for Wind Speed

	A	B	C	D
Number of Observations	122	132	3128	4829
Mean Rank	4167.0	4229.4	4166.4	4061.9

	A	B	C	D
	1	5	8	0

3.3.7. Other Variables

These variables do not necessarily fall into the other categories above.

Snow Removal Vehicle Involved

(Runway Incursion Database)

This variable indicates whether a snow removal vehicle was involved in the event. This variable incorporates many effects under one umbrella: decreased visibility to snow, special operating procedures to accommodate snow removal and weather, and unfamiliar drivers with access to runways. It is not possible to disentangle these without more accurate measures of the component factors, such as driver experience or (especially) weather / visibility. Figure 42 presents the overall distribution of this variable.

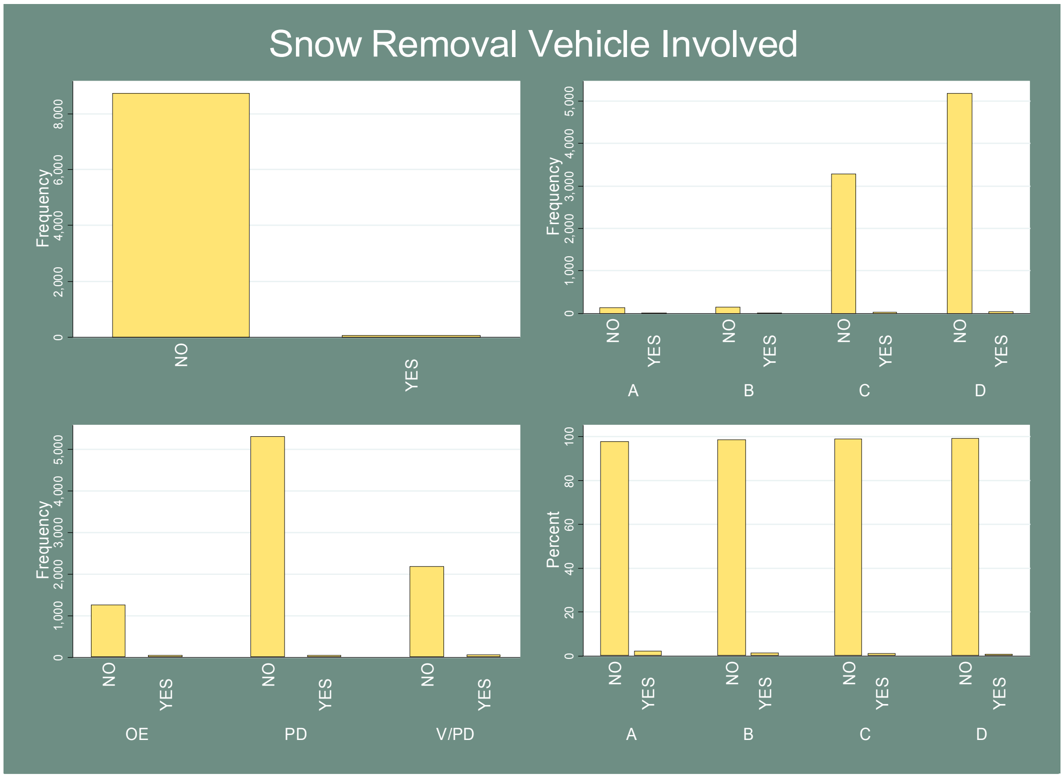

Figure 42 – Distribution of Snow Removal Vehicle Involved

Table 168 and Table 169 present the distribution of this variable by severity, while Table 170 and Table 171 present the distribution by incident type.

Table 168 – Observed Distribution of Snow Removal Vehicle Involved by Severity

	A	B	C	D	Total
No	129	143	3,275	5,184	8,731
Yes	3	2	33	43	81
Total	132	145	3,308	5,227	8,812

Table 169 – Expected Distribution of Snow Removal Vehicle Involved by Severity

	A	B	C	D	Total
No	131	144	3,278	5,179	8,731
Yes	1	1	30	48	81

	A	B	C	D	Total
Total	**132**	**145**	**3,308**	**5,227**	**8,812**

Table 170 – Observed Distribution of Snow Removal Vehicle Involved by Incident Type

	OE	PD	V/PD	Total
No	**1,257**	**5,295**	**2,179**	**8,731**
Yes	**11**	**7**	**63**	**81**
Total	**1,268**	**5,302**	**2,242**	**8,812**

Table 171 – Expected Distribution of Snow Removal Vehicle Involved by Incident Type

	OE	PD	V/PD	Total
No	**1,256**	**5,253**	**2,221**	**8,731**
Yes	**12**	**49**	**21**	**81**
Total	**1,268**	**5,302**	**2,242**	**8,812**

The distribution by severity, and its associated Fisher's Exact test statistic, indicates no relationship between severity and snow removal vehicles. While there are a relatively low number of observations, no dramatic trend by severity presents itself. This could be due to the fact that current operational changes when snow removal vehicles are present already compensate for the increased risk introduced.

The distribution by incident type is more interesting. Firstly, the Chi-Squared statistic indicates that there is a relationship between the presence of snow removal vehicles and type. There are approximately 3 times as many observed V/PD incidents than expected. PD incidents are dramatically under-represented while OE incidents are close to their expected value. The large number of V/PD incidents is interesting, indicating that when snow removal vehicles are involved in an incident, they are disproportionally at fault.

Given the high concentration of V/PD incidents, it is instructive to examine the severity of those incidents more closely. Recall that Table 168 indicated no relationship between severity and the presence of snow removal vehicles. That test statistic was calculated for all incident types, whereas

Table 172 and Table 173 present the same information, distribution by severity, but only for V/PD incidents.

Table 172 – Observed Distribution of Snow Removal Vehicle Involved by Severity, V/PD Only

	A	B	C	D	Total
No	**14**	**22**	**520**	**1,623**	**2,179**
Yes	**2**	**1**	**23**	**37**	**63**
Total	**16**	**23**	**543**	**1,660**	**2,242**

Table 173 – Expected Distribution of Snow Removal Vehicle Involved by Severity, V/PD Only

	A	B	C	D	Total
No	**16**	**22**	**528**	**1,613**	**2,179**
Yes	**0**	**1**	**15**	**47**	**63**
Total	**16**	**23**	**543**	**1,660**	**2,242**

Here, the test statistic indicates that there is a relationship among severity. Category D appears to be underrepresented while the conflict categories are all overrepresented. This indicates that V/PD incidents involving snow removal vehicles tend to be more severe than V/PD incidents not involving snow removal vehicles. The trend among conflict incidents is less clear (partly due to sample size issues). It is possible that snow removal vehicles are more likely to conflict with aircraft than other types of vehicles due to the nature of their operations. A better examination of the involvement in snow removal vehicles would account for the fact that snow removal vehicles are some of the few vehicles operating on runways. Further investigation is necessary to determine if snow removal vehicles are actually more risky or their over representation in conflict events is a product of their unique activities.

Day/Night Indicator

(Runway Incursion Database)

This variable indicates if the event occurred during the daytime. As the hours of daylight shift throughout the year, this is perhaps a better (though slightly subjective) measure than the hour the incident occurred. This variable originates from the Runway Incursion database and is thus available for a large number of incidents. Figure 43 presents the overall frequency of this day/night indicator (note that a coding of "yes" indicates daytime).

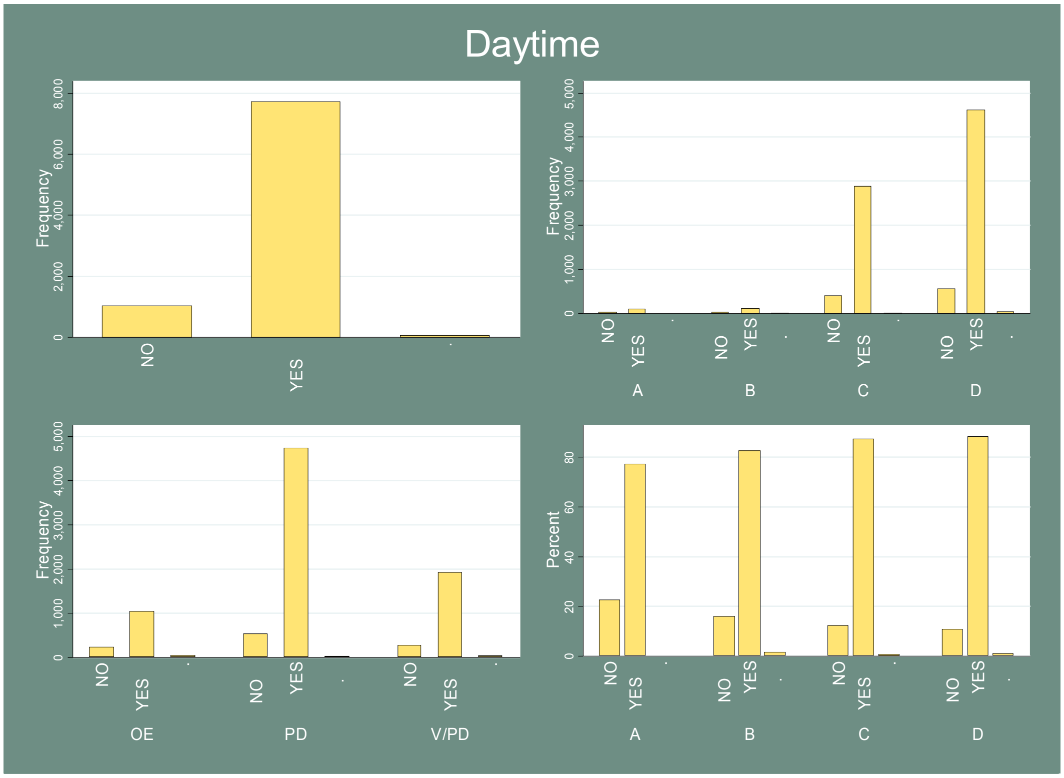

Figure 43 – Overall Frequency of Day/Night Indicator

Table 174 and Table 175 present the distribution of this variable by incident severity.

Table 174 – Observed Distribution of Day/Night by Severity

	A	B	C	D	Total
No	30	23	408	568	1,029
Yes	102	120	2,891	4,611	7,724
Total	132	143	3,299	5,179	8,753

Table 175 – Expected Distribution of Day/Night by Severity

	A	B	C	D	Total
No	16	17	388	609	1,029

	A	B	C	D	Total
Yes	**116**	**126**	**2,911**	**4,570**	**7,724**
Total	**132**	**143**	**3,299**	**5,179**	**8,753**

As daytime and nighttime are opposites, it may be more instructive to examine the "No" row above; that is, observations coded as "No" for daytime must, by definition, have occurred after dark. The test statistic indicates that there is indeed a relationship between daytime/nighttime and severity. Examining the expected values indicates that categories A, B, and C are overrepresented at night while category D is underrepresented. This suggests that conflict incidents are more likely to occur at night.

Table 176 and Table 177 present the distribution by incident type. Again, there is a significant relationship between these two variables. OE and V/PD incidents occur more often than expected at night, while PD incidents occur less frequently than expected at night. This may be due to macroscopic patterns in pilot behavior throughout the day. Less experienced pilots may not be (or be allowed to be) flying at night and thus are unable to commit errors. Given the strong relationship between incident type and severity, it is possible that the severity relationship seen in Table 176 is a product of the relationship of incident type. Further research into the relationship between day/night and severity should account for incident type explicitly. Additionally, it is unclear why day/night would impact the three incident types differently. An examination of into these differing impacts and how they may contribute to severity would help better understand the impact of day/night on runway incursions.

Table 176 – Observed Distribution of Day/Night by Incident Type

	OE	PD	V/PD	Total
No	**220**	**533**	**276**	**1,029**
Yes	**1,047**	**4,749**	**1,928**	**7,724**
Total	**1,267**	**5,282**	**2,204**	**8,753**

Table 177 – Expected Distribution of Day/Night by Incident Type

	OE	PD	V/PD	Total
No	**149**	**621**	**259**	**1,029**
Yes	**1,118**	**4,661**	**1,945**	**7,724**

	OE	PD	V/PD	Total
Total	**1,267**	**5,282**	**2,204**	**8,753**

Events occurring at night have odds of being severe of approximately 83% higher than those occurring in daytime. As indicated in Table 178, this result is fairly precise. A similar result holds for the impact on the likelihood of being an OE (compared to either PD or V/PD), as seen in Table 179. Interestingly, the effects are approximately the same size. It is possible this similarity is driven by the underlying relationship between incident type and severity. The results presented in Table 180 attempts to correct for this.

Table 178 – Logit Estimate of Impact on Severity, Night

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB
Night	**1.83**	**.287**	**0.00**	**1.35**	**2.49**

Table 179 – Logit Estimate of Impact on Incident Type, Night

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB
Night	**1.75**	**.145**	**0.00**	**1.49**	**2.06**

Table 180 – Logit Estimate of Impact on Night[55]

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB
OE Incident	**1.70**	**.142**	**0.00**	**1.44**	**2.00**
Severe	**1.62**	**.257**	**0.00**	**1.19**	**2.21**

Interestingly, the effects persist. That is, night impacts severity, even when accounting for incident type, and night also impacts incident type even when accounting for severity. These results also tell us two further things. Firstly, the size of the impacts is indistinguishable (the difference of the coefficients is not significantly different from zero). Secondly, there is no interaction effect. That is, night makes an incident more likely to be severe and more likely to be an OE, but only as the sum of its parts. Another way to think about it is that the odds ratios are multiplicative: night increases the odds of a severe OE by

55 Strictly speaking, this logit is constructed such that an indicator for night is the dependent variable. As regressions only estimate correlation, the calculation of the coefficients is indifferent to whether night is the dependent or independent variable. Thus the regression was structured in this way to enable the appropriate comparison: the impact of night on OE status and the impact of night on severity. This is only possible because all three variables are binary flags.

approximately 2.9 (1.7 * 1.7 = 2.9). That the effects are relatively constant in size over multiple model specifications and are precisely estimated indicates that this is likely a robust impact. Further research into the exact mechanism through which night impacts severity and controller actions may yield results that could improve operations.

Collision

(Runway Incursion Database)

Collisions between aircraft are also tracked in the Runway Incursions database, provided they occur on a runway. While exceedingly rare (only 7 appear in the 10 years covered by the dataset), it may be helpful to examine these incidents. Note that all collisions are considered a category A incursion, so no analysis of severity is possible.

Table 181 – Logit Estimate of Impact on Likelihood of Collision, OE Incident

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB
OE Incident	**14.9**	**12.5**	**0.00**	**2.89**	**77.0**

Table 181 indicates the increase in the odds of a collision, given that the event is an OE (the alternative being PD or V/PD). While the increase is quite dramatic (almost 15 times as high as non-OE incidents), the confidence interval is also quite large. It is important to consider the variance in the estimate as well as the magnitude of the estimate. There is little doubt that an OE incident has higher odds of being a collision, but the odds may increase anyway from approximately 2 to 77 times. Due to the extreme rarity of collision events, it will be difficult to get a more precise estimate without much more data, which is, in this case, not a desired event. This result further supports the claim that OE incidents tend to be more severe, but more research into why OE incidents are more severe is still required.

4. MODELING METHODS AND RESULTS

4.1. Methodology Background

While analysts use a variety of modeling methods, the purpose of this research is to engage in statistical analysis using regression models. Within regression models, though, a wide range of specifications are possible; selecting an appropriate model (or series of appropriate models) requires an understanding of the different assumptions underlying each model. These underlying assumptions can also impact the interpretation of model results, which can in turn affect policy recommendations. This section will review basic regressions as well as discrete choice models.

4.1.1. Regression as a Concept

The most basic regression framework is ordinary least squares (OLS) regression. Given a dependent variable **Y** and a set of independent variables **X**, the basic structure can be described as:

$$Y = \beta X + \varepsilon$$

where β is a set of coefficients that can be estimated that captures the effects of variables, and ε is a random disturbance term that includes "unobserved variables," that are not captured in X. In this framework, β represents the marginal impact of an increase in X on Y. If β is positive, then increased X is associated with increased Y; if β is negative, then increased X is associated with decreased Y. It is also important to note that this framework merely describes the relationship between X and Y and says nothing of causation in either direction.

In the context of regression analysis, OLS regression is applicable to a wide range of situations. For example, it can be used to explore the relationship between income and demographic factors or the health impacts of various policy decisions. It allows the researcher to decompose the effects of exogenous variables, controlling for their differing impacts on the dependent variable. OLS regression is extremely flexible in terms of the relationships between variables that can be captured. The X described above can include just a few variables, or many with interactions between them. OLS regression is also simple to implement.

Despite its many advantages, OLS regression has some serious shortfalls when trying to describe data such as runway incursion severity. By definition, the severity of a runway incursion falls into one of several categories: A through D. The convention in this case is to number the categories 1 through 4, with A being the highest number (thus positive β suggest increasing severity). However, it becomes quickly apparent that OLS does not bound the estimation in any way. That is, given the right confluence of negative βs, OLS may predict a score less than one (or perhaps even a negative score).

Consider a more concrete example: suppose OLS regression is used to model the optimal runway choice at a hypothetical airport based on factors such as aircraft size, weather, and destination. This hypothetical airport has three runways: 1-19, 9-27, and 15-33. Given the description of a new hypothetical flight, the model predicts an optimal runway choice of 4.73. Firstly, runway 4.73 is not a valid choice at any airport. Worse still, there is no particular rounding rule that could be assured of providing correct results.

Figure 44 below presents this distinction graphically. The figure depicts a hypothetical sample of heights and weights and plots the relationship between them. Notice that various intermediate values of height are shown and that the values of height are not restricted in any fashion. These data are appropriate for analyzing with OLS regression.

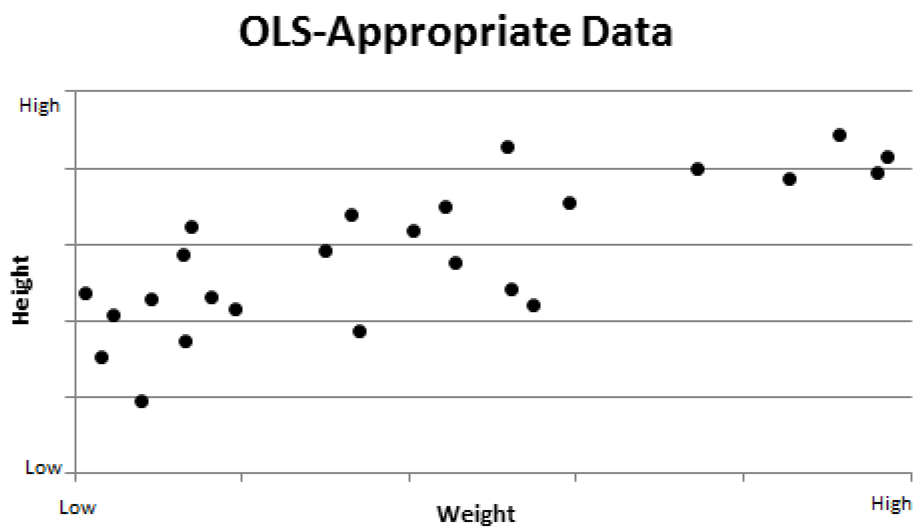

Figure 44 - Example OLS Data

The following figure, Figure 45, depicts data that is not appropriate for analyzing with OLS and is categorical in nature. Notice that the heart attack risk group outcome is restricted to only three values: low medium high and intermediate values are not possible.

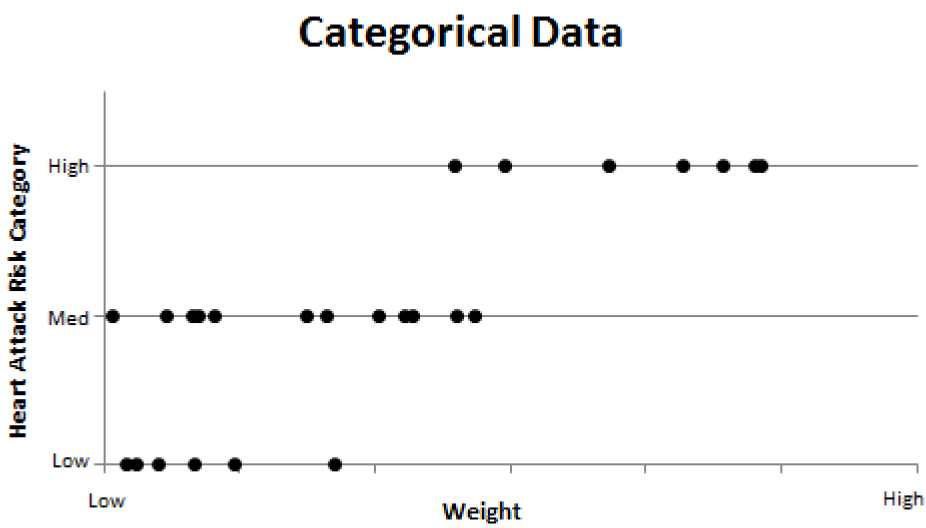

Figure 45 - Example Categorical Data

In addition to the problems relating to boundedness and integer values mentioned above, OLS has an additional, and perhaps more important, failing in relation to incursion severity data. Incursion severity data has the property that it is merely ordinal, not cardinal. That is, incursion severity data has some sort of ranking (A is more severe than B, etc.) but the ranking does not describe the distance between ranks. An incursion of severity level B is more severe than a C-level incursion, which is in turn more severe than a D-level incursion. However, a category B incursion may be *much* more severe than a C compared to the difference between a category C incursion and a category D incursion. While there is logic to assigning severity ratings of A-D on a scale of 1-4 (with A being the highest), this decision is entirely arbitrary. In fact, given the substantial effort invested in preventing category A and B incursions, one could suggest that the proper scale should be 2, 3, 6, and 12 (for severity D, C, B, and A). Using this scale, one could argue that Category D and C E incursions are progressively more severe at a constant rate, but that Category B incursions are twice as severe as Category C incursions. Moreover, in this case, a Category A incursion is twice as severe as a Category B, 4 times more severe than a Category C, and 6 times more severe than a Category D. This would certainly be in line with the specific concern for A and B-level incursions, but without some sort of specific analytical and numeric rationale, this categorizing system is just as arbitrary as using 1-4. That is, how can we be sure the real ranks are not 2, 3, 6, and 11.5? Consequently, one needs a form of regression that can provide accurate and useful results in the absence of a perfectly defined scale.

OLS regression does not acknowledge this aspect of the data. OLS treats the change between any two categories as equal and makes it a suboptimal choice for analyzing data such as runway incursions.

4.1.2. Alternatives to Linear Regression

Data like runway incursion severity falls into a category that can be described as "discrete choice" data. The data points are placed into distinct categories, often of a qualitative nature. An entire class of models has been developed to analyze discrete choice data and overcome the limitations of OLS regression discussed above.

Discrete choice models have been developed to look at binary choice, such as whether or not to participate in the labor market and to analyze sets with more than two choices. These multi-choice models come in a variety of flavors such as ordered (which recognizes an inherent ordering in the categories) and multinomial (which do not recognize any ranking among choices). There are additional extensions to the multinomial model framework that seek to relax several of the constraints imposed by the standard multinomial model; for more information, see Appendix C.6.

A significant portion of the safety and severity literature utilizes regressions models utilizing a somewhat different framework than traditional "frequentist" statistics. The basis for these alternative Bayesian models is described in Appendix C.4. These models remain an interesting alternative modeling methodology for future research, but due to the lack of previous statistical studies in this field, it was deemed most useful to utilize the frequentist models as they are less computationally intensive, easier to understand for readers new to the topic, and should provide similar (if not identical) results to the Bayesian models.[56]

Beyond the world of OLS and its extensions, the basis for (frequentist) econometrics is maximum likelihood estimation (MLE). MLE can be used to estimate a plethora of different model types and all of the models discussed later in this report are estimated using MLE techniques. The focus of MLE is the likelihood function, L:[57]

$$L(\beta \mid y_1, \ldots, y_n) \equiv f(y \mid \beta)$$

for a sample of n observations, each with a value of y, noted as $y_1 \ldots y_n$. This equation represents the likelihood of observing the data, y, given parameters β. For this particular application, the likelihood function, f or L, represents the distribution of runway incursion severities. This formulation can be extended to include other conditioning variables X:[58]

$$L(\beta \mid y_1, \ldots, y_n, X) \equiv f(y \mid X, \beta)$$

On the above equation, Greene notes:

> the likelihood function is written in this fashion to highlight our interest in the parameters and the information about them that is contained in the observed data. However, it is understood that the likelihood function is not meant to represent a probability density…, the parameters are assumed to be fixed constants which we hope to learn about from the data[59]

---

56 The results of the two frameworks converge to the same results due to the lack of any informed priors adding additional information/usefulness to the Bayesian models.

57 Greene (2003).

58 Ibid.

This likelihood function can be thought of as the data generation process. Suppose y is the probability of rain today. Then X will be variables that may influence that, such as temperature, humidity, and atmospheric pressure. β characterizes the impact of those variables on y. The likelihood can also be thought of as the probability of observing that set of y, given X and β. Maximum likelihood estimation, true to its name, seeks to choose a β to maximize the above expression (the probability of observing that set of y given X and β.)

β is of fundamental interest to the econometrician and policy-maker. β captures the effects of the various exogenous variables X on the dependent variable y. It is from this information that informed policy decisions can be made.

4.1.3. Discrete Choice Models

The Problem

As noted earlier, runway incursion severity rankings fall into a category known as discrete choice data. A variety of models have been developed to analyze these types of data. Each of the potential models has underlying assumptions and characteristics that may influence the applicability of that model to the analysis of runway incursion severity.

To clarify the discussion about which model to use, the various competing models can be separated along two axes: logit versus probit, and multinomial versus ordered. Logit and probit refer to assumed distributions of the random disturbance terms. This can have impacts on the assumptions underlying each kind of model. Ordered and multinomial refer to how the model interprets the various choices (i.e., alternative levels of the dependent variable). Both kinds of models deal with choice sets with three or more alternatives. However, the ordered models recognize an inherent ordering in the choices while multinomial models assume there is no underlying order to the choices. Table 182 illustrates this breakdown.

Table 182 – Discrete Choice Models under Consideration

	Logit	**Probit**
Ordered	**Ordered logit**	**Ordered probit**
Multinomial	**Multinomial logit (conditional logit) and extensions**	**Multinomial probit**

At a simple level, the decision is between one of these four possibilities. The criteria governing this decision include tractability, precision, and how well the model reflects reality. Additionally, there is value in comparing different models. The comparison may provide additional insight into the relationship among variables as well as serve as a sensitivity analysis to the assumptions of the model.

59 Ibid., p. 468-469.

Comparisons across rows and across columns are valuable in the sense that they hold fixed one set of assumptions. For example, ordered logits are best compared to ordered probits (holding the ordering assumption fixed, but changing the distributional assumption) and multinomial logits (holding the distributional assumption fixed and relaxing the ordering assumption). Thus, the preferred model is one whose neighbors are also favorable in terms of the decision criteria above.

Logits versus Probits

There are some general comments that pertain to the columns of Table 182 that are true regardless of the row chosen. The major distinction between logit and probit models are the distribution of the random disturbance term (ε, which captures the impact of unobserved variables). In general, probit models assume a normal distribution for at least some component of ε, while logistic models assume a logistic distribution.[60]

In practical terms, the distinction between logit and probit models appears to be minute. Horowitz examines this issue by comparing a known multinomial probit function to its logit approximation. He finds that several thousand observations are required to distinguish between the two models, depending on the correlation between the random disturbances for each choice.[61] Dow and Endersby seek to compare multinomial probit and logit models in a more applied setting, examining vote data and finding similar conclusions to Horowitz. The predicted probabilities are similar between the two models and the authors note that a sample size of 1500 is not enough to distinguish between the two models.[62] Greene also suggests that ordered logit and probit models provide similar results in practice.[63] This claim is corroborated in a study by O'Donnell and Connor.[64] Consequently, if one finds significantly different results between the two models (in terms of variable significance and predicted probabilities), further investigation would be required.

It is important to note that the interpretation of the models does not depend on the distributional assumption. The difference in implementation is important from a theoretical perspective, but is largely transparent to the reader.

---

60 The implications the different assumptions have for the model are relevant, but a thorough discussion of the differences in the distributions (and the properties of those distributions) is outside of the scope of this paper. For a more in-depth discussion of the assumptions underlying these models, both in regards to the random disturbance term and other properties, please see: Greene (2003), Washington*, et al.* (2011).

61 Horowitz (1980).

62 Dow and Endersby (2004).

63 Greene (2003), p. 737.

64 O'Donnell and Connor (1996).

Multinomial versus Ordered

As noted earlier, both ordered and multinomial models address choice sets with multiple alternatives. However, the main difference is that ordered models recognize an inherent ordering of the choices while multinomial models do not.

Of course, situations such as runway incursion severity are clearly ordered by intention, but multinomial models can also be used to examine ordered data, providing some potential benefits as well as drawbacks. Ordered models place a strong constraint on the estimated coefficients. Washington et al. provide an example: consider accident severity data that has severity rankings of property damage only, injury, and fatality. Additionally, suppose the effect of airbag deployment was of interest. An ordered model constrains the coefficient to either "increase the probability of a fatality (and decrease the probability of property damage only) or decrease the probability of fatality (and increase the probability of property damage only)."[65] This may not be the case in reality. Airbag deployment may reduce the probability of a fatality and of property damage only, due to an increase in probability of an injury. A multinomial specification allows the flexibility for such effects.[66]

While ordered models do not allow for this sort of complexity, they do provide more intuitive coefficient interpretation. If the coefficient is positive, increasing the value of the explanatory variable unambiguously increases the probability of being in the highest category and the probability of being in the lowest category decreases, though intermediate categories have a more subtle relationship.[67] Thus, a tradeoff must be made between accounting for additional accuracy in modeling complex relationships between severity levels and providing results that are useful and practical to policy-makers. Moreover, this distinction only exists in the event that the effect of an explanatory variable is not the same across severity levels.

Similarly, Washington et al. note that "if an unordered model (such as the multinomial logit model [MNL]) is used to model ordered data, the model parameter estimates remain consistent but there is a loss of efficiency."[68] In other words, the multinomial estimates are less precise than an ordered model, but are unbiased estimates of the effects. There is an essential "trade off … between recognizing the ordering of the responses and losing the flexibility in specification offered by unordered outcome models."[69]

---

65 Washington*, et al.* (2011), p. 358.

66 Ibid.

67 Greene (2003), p. 738.

68 Washington*, et al.* (2011), p. 345.

69 Ibid., p. 359.

Specific Model Discussion and Examples

In addition to the more general properties mentioned above, some of the specific models have additional properties that may make them desirable or undesirable. In addition to specifics of the various models, examples of the models in applied settings will be provided.

Ordered Logit and Ordered Probit

The above sections outline the basic differences between various discrete choice models. Ordered logit and ordered probit models vary only in their choice of distributional assumption. For reference, ordered logit models assume a logistic distribution on the random disturbance term, while ordered probit models assume a normal distribution. There is a slight preference for the ordered probit model due to the normality assumption, which, barring evidence that it is invalid, is convenient, but there is no inherent theoretical basis for that preference and the practical differences are likely small. The models have no additional specific properties that require additional discussion.

O'Donnell and Connor provide an example of using an ordered logit to examine injury severity.[70] Their study focuses on comparing the results to that of an ordered probit model on the same data. As noted earlier, the theoretical prediction of similar results is validated, though there are aspects of the modeling methodology in this paper that should not be replicated. Specifically, the authors use a measure of model fit (the Schwarz Bayesian Information Criterion (SBIC)) to aid in selecting variables for inclusion in the model. Starting with a large set of variables, variables were removed algorithmically as determined by the SBIC formula. Thus, the models presented in this paper may be prone to overfit, reducing the actual usefulness of the model outside of its specific dataset.

Kockelman and Kweon provide a good example of an ordered probit in practice, again examining injury severity.[71] Lauer examines educational attainment in France and Germany using an ordered probit framework.[72] Xie et al. provide a good example of an ordered probit model implemented in a Bayesian framework (in addition to a frequentist framework).[73]

Multinomial Logit

Multinomial logits (MNL) are the most studied version of the multinomial models. The multinomial logit has several features that make it distinct from the multinomial probit. In terms of model comparison, MNL models are best compared to ordered logits and multinomial probits.

The first feature of a MNL model that distinguishes it from a multinomial probit is the distribution of the random disturbance terms. In the MNL framework, the random disturbances for different choices are

---

70 O'Donnell and Connor (1996).

71 Kockelman and Kweon (2002).

72 Lauer (2003).

73 Xie et al. (2009).

assumed to be uncorrelated.[74] In other words, the unobserved variables that influence the probability of choice A are entirely unrelated to the unobserved variables that influence the probability of choice B. This property may not hold in reality, resulting in faulty estimates from the model.

A direct result of the assumption regarding the correlation of the random disturbances is what is called the independence of irrelevant alternatives (IIA) property. Specifically, the ratio of any two choice probabilities is independent of the probabilities of any other possible choices.[75] This is often characterized in the red bus-blue bus problem:

> "…consider the estimation of a model of choice of travel mode to work where the alternatives are to take a personal vehicle, a red transit bus, or a blue transit bus. The red and blue transit buses clearly share unobserved effects that will appear in their disturbance terms and they will have exactly the same functions [choice probabilities] if the only difference in their observable characteristics is their color. For illustrative purposes, assume that, for a sample commuter, all three modes have the same value [from the model]…(the red and blue bus will, and assume that costs, time, and other factors that determine the likelihood of the personal vehicle being chosen works out to the same value as the buses). The predicted probabilities yield each mode with a 33% chance of behind selected. This outcome is unrealistic since the correct answer is a 50/50 chance of taking a personal vehicle and a 50/50 chance of taking a bus (both red and blue bus combined) and not 33.33% and 66.67%, respectively, as the [multinomial logit] would predict. The consequences of an IIA violation are incorrect probability estimates."[76]

The MNL also has another undesirable property in regards to parameter estimation. Specifically, "estimable parameters relating to variables that do not vary across outcome alternatives can, at most, be estimated in *I*-1 of the functions determining the discrete outcome (*I* is the total number of discrete outcomes)."[77] For example, suppose gender were a relevant variable to a model of mode choice. If there were three choices (e.g., bus, train, or automobile), the model could only estimate the effect of being male on the two out of three choices. This is a fairly severe limitation of the multinomial logit model if there are a large number of effects that are of interest, but do not vary across categories. One potential way to address this is to normalize the coefficients for one outcome (the "base" outcome). Thus, parameters for variables that do not vary across categories can be estimated for the remaining categories. The coefficients are then interpreted as a change relative to the base outcome.

As noted above, MNL models see extensive use in practice (especially in comparison to multinomial probit models). Islam and Mannering provide a good example of a multinomial logit being used to examine injury severity.[78] Dow and Endersby provide an example of a multinomial logit looking at voter

---

74 Greene (2003)., p. 724

75 Ibid., p. 724

76 Washington*, et al.* (2011)., p. 326

77 Ibid., p. 318.

78 Islam and Mannering (2006).

behavior in comparison to a multinomial probit model.[79] Finally, Schneider IV et al. also examine injury severity using a multinomial logit framework.[80] Additional discussion of the theoretical aspects of the multinomial logit specifications can be found in Washington et al. and Greene.[81] There are extensions to the MNL model that seek to relax some of these restrictions, such as IIA. Two of the most common extensions are nested logit and random parameter models. A brief discussion of these extensions can be found in Appendix C.6.

Multinomial Probit

Like multinomial logit models, the multinomial probit is an unordered discrete choice specification. It is not commonly used due to be "the difficulty in computing the multivariate normal probabilities…."[82] However, with the challenges of estimation come some benefits.

The major benefit of a multinomial probit as compared to a multinomial logit is the lack of correlation structure on the random disturbances. Recall that in a multinomial logit, the random disturbance terms were assumed to be uncorrelated for different alternatives. Multinomial probits have no such restriction on the correlation and allow a freer set of correlations between disturbance terms.[83] This translates directly into another benefit: multinomial probit models do not have the IIA property. Further discussion of the multinomial probit model can be found in Greene and Washington et al. In general, the multinomial probit specification appears to be preferable to the multinomial logit due to the less stringent assumptions of the multinomial probit model. However, computational difficulty remains a major challenge and is the major disadvantage of a multinomial probit framework. The likelihood for the multinomial probit specifications contains the standard normal cumulative distribution function (CDF), which has no closed form solution. Thus, the likelihood function has no closed form solution.[84] Due to the multiple integrals required for multinomial models, evaluating these expressions can be extremely computational intensive compared to the logit specification.[85]

An example of a multinomial probit in applied work can be found in Dow and Endersby.[86] In addition to implementing the model, the authors provide some additional insight into the comparison between

---

79 Dow and Endersby (2004).

80 Schneider IV*, et al.* (2009).

81 Washington*, et al.* (2011)., Greene (2003).

82 Greene (2003), p. 728.

83 Ibid., p. 728.

84 Washington*, et al.* (2011), p. 312.

85 Greene (2003)., p. 728

86 Dow and Endersby (2004).

multinomial logit and probit models. Horowitz (1980) provides another comparison of multinomial logit and probit models.[87] Further commentary on why a multinomial probit specification may be preferable can be found in Horowitz (1991).[88]

4.2. Methods chosen

Given the discussion above, it is clear that each model has some pros and cons associated with it. As noted earlier, it is important for not only the chosen model to be desirable, but also the comparison models. Recall that the decision criteria for a model to be desirable included tractability, precision, and how well it reflects reality. The specific nature of the runway incursion data does not suggest any particular model choice. Though the data does have some sense of ordering to the categories, multinomial models provide some advantages in terms of analysis, especially as the ordering present in the data may be the result of multiple processes.

Due to the nature of the data (i.e., severity ratings from A to D), it was initially desired to focus on the analysis on the ordered family of models. However, as discussed below, the assumptions of the ordered model were not satisfied in many cases.[89] This led to the use of multinomial models to relax the ordering constraint. Logit models were chosen over probit models due to computational simplicity, similarity in results when a subset of models were compared head-to-head, and evidence that that the assumption of IIA is not violated with these data.

4.3. Models

The models presented below do not contain all of the variables presented in the previous chapter. The results of the tests presented in that section helped inform the modeling process. Though the ordering models' assumptions may not be satisfied, the results are presented for comparison and completeness. Also note that these results are restricted to only OE incidents, limiting the sample size and the variables that could be included. Finally, the dependent variable is the severity of the incident, with category A being considered most severe and given a rank of four.

4.3.1. Aircraft

This model contains variables relating to the aircraft involved at the time of the incident. The results of the ordered model are presented in Table 183. Category D incursions were excluded due to the inclusion of the variable measuring the number of aircraft involved. As Category D incursions involve only one aircraft by definition, including category D incursions obscures the true impact of the number of aircraft on severity.

---

87 Horowitz (1980).

88 Horowitz (1991).

89 As an aside, ordered probit models were also run. They gave very similar results, leading to the conclusion that the distributional differences between logit and probit models are of little consequence for this data.

Table 183 – Ordered Logit Model for Aircraft Variables

Variable	Coefficient	Standard Error	P-Value	95% CI LB	95% CI UB
# of Aircraft Involved	**.556**	**.312**	**0.08**	**-.056**	**1.17**
Commercial Carrier	**-1.11**	**.319**	**0.00**	**-1.73**	**-.482**
Landing	**.380**	**.321**	**0.24**	**-.248**	**1.01**
Takeoff	**.762**	**.282**	**0.01**	**.209**	**1.31**
Daily Operations[90]	**.001**	**.002**	**0.61**	**-.003**	**.005**

N = 866	LR Chi-Squared Stat: 23.91
LL = -293.21923	LR P-value: 0.00
LL_0 = -305.17656	Ordered Test P-value: 0.67

The sign of the coefficients can be interpreted just as they would for simple logits: positive values increase the likelihood of the incursion being rated as a category A while negative values decrease that likelihood.[91] The opposite impact is had for category C incursions – positive indicates less chance of being a category C while negative increases the probability of being category C. The impact on category B is ambiguous and requires further calculations to determine.

The signs of variables are consistent with many of the conclusions drawn in Section 3.3. Takeoff is still more dangerous than taxiing. The impact of landing is not statistically different from zero (i.e., the model cannot distinguish if there is a change in probability due to landing or not). Commercial carrier status reduces the likelihood of a category A incident. The daily operations at an airport appear not to impact the likelihood of a category A incident, once these other variables are controlled for. Finally, additional aircraft increase the likelihood of a category A incident.

Interestingly, this model satisfies the constraints on the ordered logit model.[92] This is likely due to the exclusion of the category D incursions. To be consistent with the other categories, the results of binary

---

90 The units on Daily Operations are actually tens of daily operations. Thus the coefficient represents the marginal impact of an additional 10 operations per day.

91 For more information on interpreting the results of regression output, please see Appendix C.4.

92 This can be determined from the "Ordered Test P-value" reported in the footer of Table 158. The ordering test tests the hypothesis that the effects of the model variables are consistent across all category types. The insignificant test statistic (0.67) thus indicates that the impacts of the variables are consistent across the three severity categories.

logit – similar to those presented in Section 3.3 – and of a multinomial logit are presented below. They are broadly consistent with the ordered model, though the multinomial model provides a more nuanced look at the relationship among these variables. The consistency is not surprising given that the assumptions of the ordered logit are satisfied.

Table 184 – Binary Logit of Aircraft Variables

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB
# of Aircraft Involved	1.75	.549	0.08	.943	3.23
Commercial Carrier	.335	.107	0.00	.180	.626
Landing	1.47	.471	0.23	.781	2.75
Takeoff	2.08	.589	0.01	1.20	3.63
Daily Operations	1.00	.002	0.64	.997	1.01

N = 866	LR Chi-Squared Stat: 22.96
LL = -243.10764	LR P-value: 0.00
LL_0 = -254.58853	

The binary logit results are almost identical to the ordered results – though they are presented in odds ratio form. Recall that category D incursions are excluded, making the alternative category C only. The stability of the relationships indicates that collapsing categories A and B had little impact on the estimate of the effect of these variables. The multinomial results presented below offer a slightly more in-depth look at these effects.

Table 185 – Multinomial Logit of Aircraft Variables

Variable	Coefficient	Standard Error	P-Value	95% CI LB	95% CI UB
B: # of Aircraft Involved	0.4036986	0.4884421	0.41	-0.5536302	1.3610270
B: Commercial Carrier	-1.0894810	0.5035568	0.03	-2.0764340	-0.1025275
B: Landing	0.4045899	0.4603749	0.38	-0.4977282	1.3069080
B: Takeoff	0.1697855	0.4759628	0.72	-0.7630844	1.1026550
B: Daily Operations	0.0003766	0.0034865	0.91	-0.0064568	0.0072099
A: # of Aircraft Involved	0.650676	0.3879667	0.09	-0.1097248	1.411077
A: Commercial Carrier	-1.113953	0.3994795	0.01	-1.896918	-0.3309872
A: Landing	0.3683587	0.4276108	0.39	-0.469743	1.20646
A: Takeoff	1.0542	0.3463956	0.00	0.3752774	1.733123
A: Daily Operations	0.0014655	0.0027573	0.60	-0.0039387	0.0068696

N = 866	LR Chi-Squared Stat: 26.84
LL = -291.75439	LR P-value: 0.00
LL_0 = -305.17656	

With the ability to distinguish between category A and B, some additional insights arise. It is important to note that the total change in probably across categories must equal zero as the total probability across categories is constrained (i.e., you must be in one of these categories, so a reduction in the probability of one category must be countered by an increase in the probability of another). For example, commercial carrier status reduces the probability of a category B incursion by approximately .03 (from approximately p = .047 to approximately p = .017)[93]. The likelihood of a category A incursions is reduced by approximately .045. Therefore, commercial carrier status increases the likelihood of a category C incursion is increased by approximately .075.

Another lesson to take from this is that, although the variable had a similar estimated coefficient between categories, the impact in terms of probability can be different. This is a function of the formulation of the multinomial logit model. Thus, all coefficients must be interpreted in terms of changes in probability within their category, rather than directly compared across categories. The results of the categorical variables (in this case commercial carrier status and aircraft phase of flight) are presented in Table 186. The figures following that table provide the impact of the continuous variables on each category. In both the table and figures, the variables not changing are held at their mean.

Table 186 – Change in Probability of Severity Categories for Categorical Variables

	Category C	Category B	Category A
Commercial Carrier Status	.07	-.03	-.05[94]
Takeoff	-.03	.01	.02
Landing	-.06	.00	.06

93 In many cases, the probability of being a category A or category B event for these multinomial models is quite low. This is partially due to the fact that severe incidents are rare and the overwhelming majority of incidents are category C. Thus, while the absolute value of the change may be small, it may be large in percentage terms.

94 Note that changes in probability may not add one due to rounding.

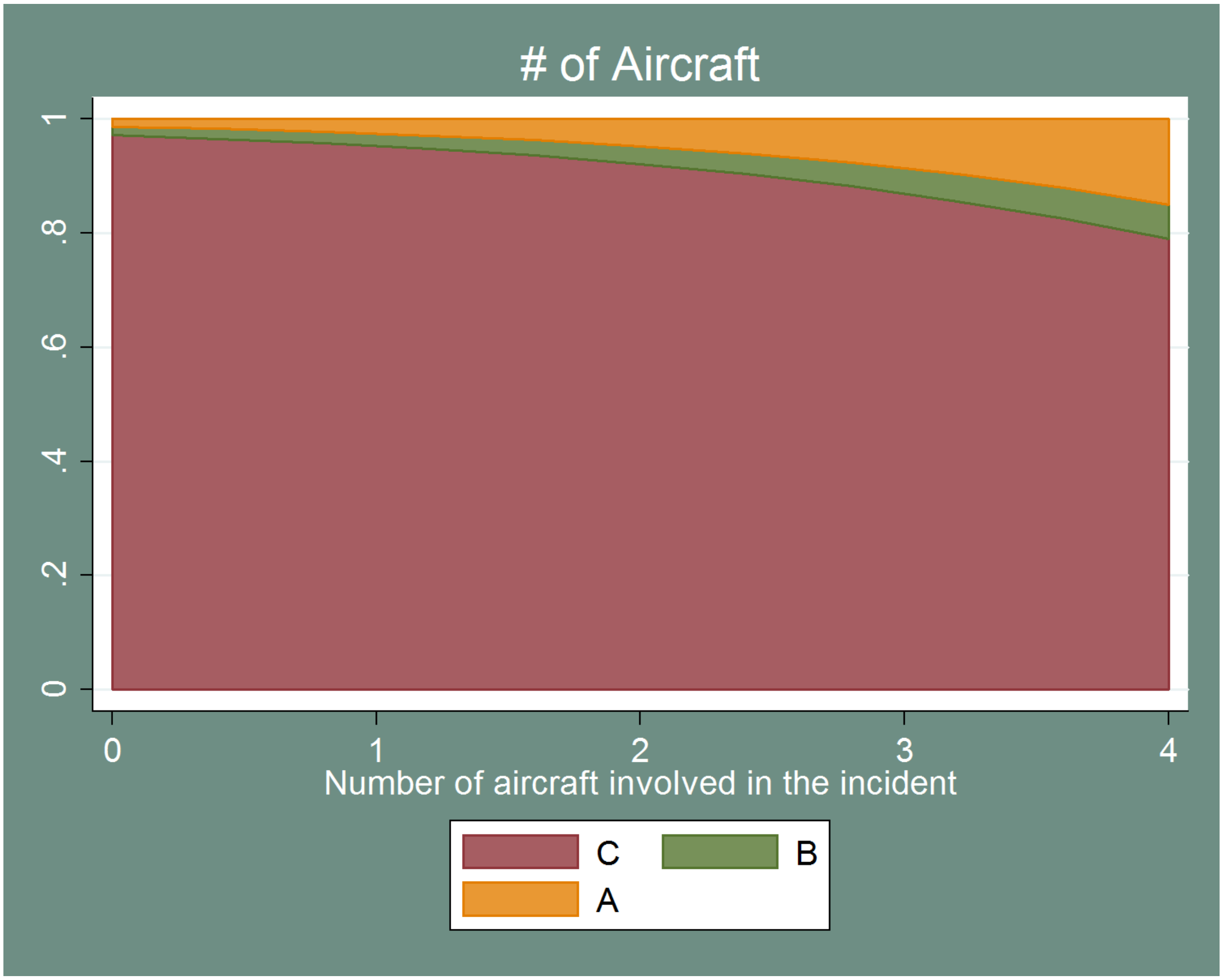

Figure 46 – Impact on Probability of Severity Categories of Number of Aircraft

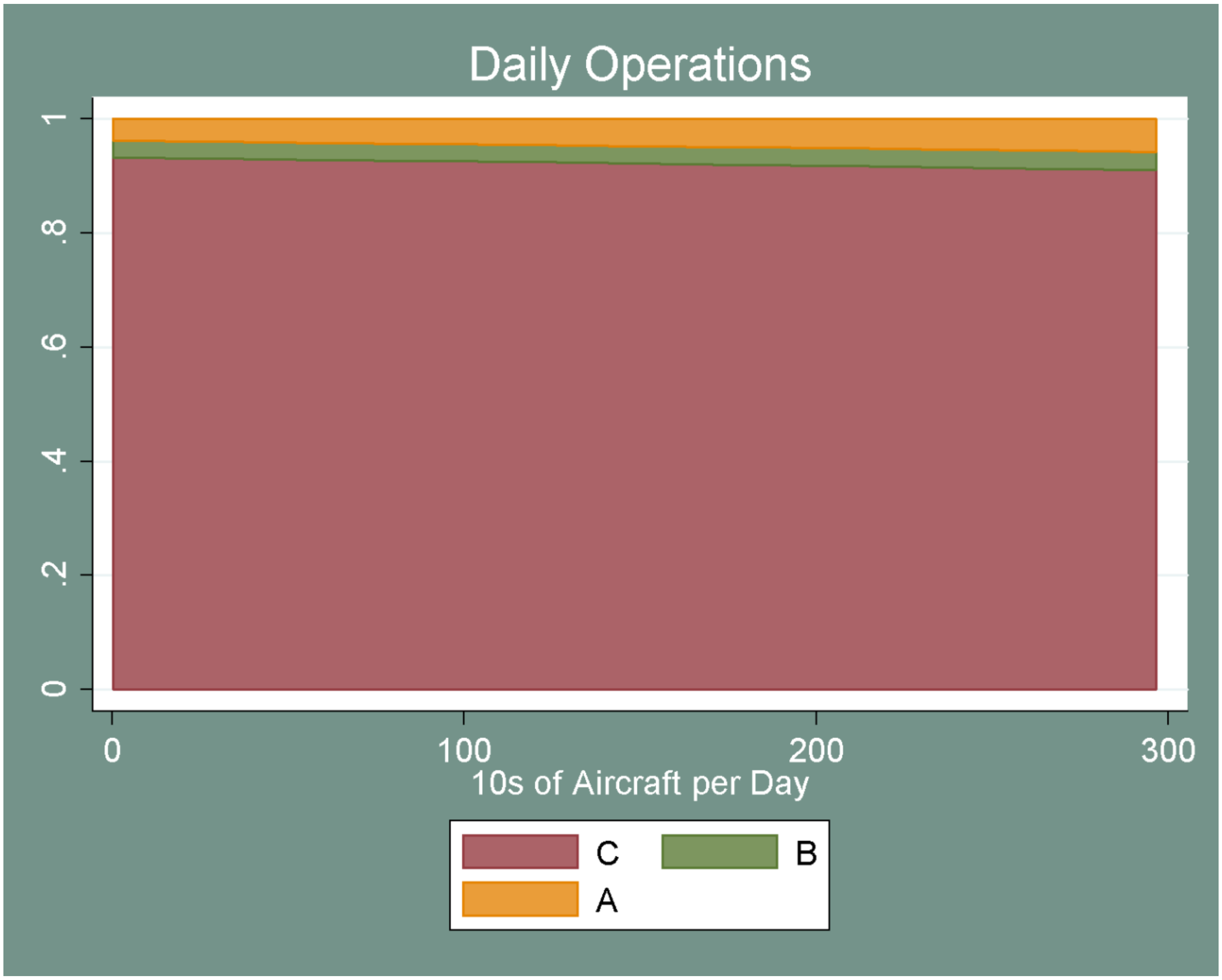

Figure 47 – Impact on Probability of Severity Categories of Daily Operations, Aircraft

The impact of the number of daily operations is fairly slight (not surprising given that the coefficients are not statistically significant). Number of aircraft, on the other hand, appears to increase the probability of category A fairly dramatically as the number of aircraft involved increases.

As noted above, the disparity between categories A and B are of interest. The model does not appear to describe the underlying process of category B incursions very well. The variables that appear significant in the ordered model appear to maintain significance only for category A (and only moderately for the number of aircraft). Thus, it appears that the impact of number of aircraft and aircraft phase of flight are localized to category A incursions rather than category B.

Finally, it is important to check that the assumptions underlying the multinomial logit model are met. As noted earlier, the major assumption for a multinomial logit model is that of IIA. Testing for violation of IIA revolves around estimating models excluding one alternative at a time and comparing coefficients. While the test statistics and associated p-values are presented, research suggests that these tests are not particularly useful for testing for violations of the IIA assumption.[95,96] While the test for violation of IIA is not particularly powerful, it represents the best available test. Additionally, information from the test can be combined with prior knowledge of the categorization (i.e., ranking) system for a better understanding of the IIA issue. The following table presents the results of a test for IIA in this model.[97] Insignificant test statistics suggest that the IIA assumption is valid in this case. For this model, the test statistics are insignificant regardless of which outcome is removed.

This model provides some interesting insights. First, it appears that amount of daily traffic at an airport does not have an impact on incident severity in the presence of these other variables. This is in contrast to models presented in subsequent section and is likely due to the exclusion of category D incursions. Second, phase of flight (specifically takeoff) appears to impact category A incursions, rather than both categories A and B. This is possibly a definitional effect, rather than a true relationship with severity. Similarly, number of aircraft involved appears to only increase the likelihood of category A incursions rather than both severe categories (although the coefficient is barely significant at a wider 10% criterion; also recall the earlier caution about multiple comparisons). Commercial carrier status appears to reduce the likelihood of both severe categories. This may be related to pilot experience, but it is surprising that that effect would show up for OE incidents as well. This further supports the idea that commercial carriers and GA pilots must be considered separately, even from a controller's perspective.

Table 187 – Results of IIA Test for Aircraft Variables

Omitted Outcome	Chi-Squared Stat	Degrees of Freedom	P-Value
C	2.65	6	0.85
B	0.93	6	0.99
A	3.86	6	0.70

4.3.2. Airport

This set of models examines the physical characteristics of the airport at which the incursion occurred. It is important to note that the variables in these models do not vary by incursion (in general). This introduces a problem into the model in that the errors (in a statistical sense) are possibly correlated

95 Long and Freese (2006).

96 Despite the lack of a sufficient test for the IIA assumption, it does not appear to be a problem with this data. The severity categories represent a mutually exclusive set of categories that describe the entirety of the severity spectrum. Thus, the dismissal of IIA as a problem is based both on the evidence of the (albeit weak) tests and theoretical ground.

97 The tests presented in this section are derived from the Stata package called SPost. The package performs the Hausman-McFadden tests for IIA using Stata's built in command suest. The test focuses on comparing the coefficients for a model containing all alternatives to models removing one alternative at a time. More information can be found in: J. Scott Long and Jeremy Freese (2005) Regression Models for Categorical Outcomes Using Stata. Second Edition. College Station, TX: Stata Press.

between observations. This affects the standard errors estimated from the model. It is unlikely to cause a major shift in standard errors, given that there are a large number of airports involved. While there are repeated observations at the same airport, they are not so common (relatively) as to dominate the estimation sample. Future research into airport models could attempt to account for repeated observations at the same airport via clustering or another method.

The results of the ordered model are presented below. This model does not satisfy the assumptions of the ordered model (as seen by the Ordered Test P-value in Table 188). However, when category D incursions are excluded (as seen in Table 189), the model does conform to the assumptions of the ordered model. This supports the idea that category D incursions follow a separate process from categories A through C and may not be part of the same continuous ordering.

Table 188 – Ordered Logit Results for Airport Variables

Variable	Coefficient	Standard Error	P-Value	95% CI LB	95% CI UB
# of Runway Intersections	0.1138214	0.0653972	0.08	-0.0143546	0.241998
# of Runways	-0.3065381	0.0961104	0.00	-0.4949109	-0.11817
# of Hotspots	-0.0728477	0.0390897	0.06	-0.1494621	0.003767
Difference of AC/AT and GA Percents	0.3109389	0.2970045	0.30	-0.2711793	0.893057
AC/AT Percent of Traffic	-0.4287666	0.2952404	0.15	-1.0074270	0.149894
Daily Operations	0.0102631	0.0021114	0.00	0.0061248	0.014401

N = 969	LR Chi-Squared Stat: 28.09
LL = -608.22534	LR P-value: 0.00
LL_0 = -622.2712	Ordered Test P-value: 0.00

Table 189 – Ordered Logit Results for Airport Variables, Conflict Only

Variable	Coefficient	Standard Error	P-Value	95% CI LB	95% CI UB
# of Runway Intersections	0.23436	0.1002157	0.019	0.037941	0.430779
# of Runways	-0.319587	0.164589	0.052	-0.64218	0.003002
# of Hotspots	-0.0972646	0.0650232	0.135	-0.22471	0.030178
Difference of AC/AT and GA Percents	0.3864231	0.4578904	0.3990000	-0.51103	1.283872
AC/AT Percent of Traffic	-0.6208724	0.3972058	0.1180000	-1.39938	0.157637
Daily Operations	0.0045543	0.0032610	0.1630000	-0.00184	0.010946

N = 870	LR Chi-Squared Stat: 14.50
LL = -295.42478	LR P-value: 0.02
LL_0 = -302.67675	Ordered Test P-value: 0.13

Although the overall model is invalid because of the ordering assumption, it is still worth noting some of the results. First, number of runway intersections plays a role. When excluding category D, this variable's coefficient is both larger and considered more significant (but is less precisely estimated). This same situation can be seen for overall runway count, although the effect is in the opposite direction, reducing severity. The number of hotspots at an airport is only (marginally) significant when category D incursions are included. The expectation is that this variable may help explain category D in the multinomial model, and no other categories. A similar expectation is held for daily operations, which serves as an overall control on the frequency of incursions (i.e., incursions are more likely with more traffic, even if the rate of incursions per operations is constant), yet is no longer significant when category D incursions are excluded. Thus, daily operations may help explain category D but not the other categories.

As discussed above, a simpler alternative to the multinomial model is to combine categories C and D and categories A and B. While ultimately a loss of detail, these models are simpler to interpret and focus the discussion on the impact on severe incursions – the categories of most interest for preventing crashes. The results of this binary logit are presented in Table 190.

Table 190 – Binary Logit Results for Airport Variables

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB
# of Runway Intersections	1.26122	0.1256319	0.02	1.037532	1.533134
# of Runways	0.7033451	0.1159668	0.03	0.509124	0.971659
# of Hotspots	0.8994699	0.058366	0.10	0.79205	1.021458
Difference of AC/AT and GA Percents	1.4833310	0.6772210	0.39	0.606204	3.629591
AC/AT Percent of Traffic	0.5566192	0.2192933	0.14	0.257162	1.204784
Daily Operations	1.0058800	0.0032209	0.07	0.999587	1.012212

N = 969	LR Chi-Squared Stat: 14.52
LL = -254.1839	LR P-value: 0.02
LL_0 = -261.44236	

The results for the binary logit are not dissimilar to those for the ordered model with all four severity categories. As in the ordered model, number of runway intersections increases the likelihood of a severe event. The impact is actually comparable in size to the impact in the ordered model, though these are expressed as odds ratios: for each additional runway intersection, the odds of a severe incursion are increased by approximately 25%. Counteracting this is the impact of having additional runways, which reduces the odds of a severe incursion by approximately 30% for each additional runway. Exposure (i.e., total operations) also plays a role in increasing severity, as seen in the ordered model; however, its impact is marginal at best.

The results from the multinomial logit support many of the conclusions drawn above. There are no categorical variables, so the impacts of all variables are depicted in the following charts. As in the ordered and binary models, increasing numbers of runway intersections are associated with increased severity. This change in probability appears to result from a decrease in the probability of category C incursions. This may suggest that runway intersections are associated with conflict events rather than category D incursions.

Table 191 – Multinomial Logit Results for Airport Variables

Variable	Coefficient	Standard Error	P-Value	95% CI LB	95% CI UB
D: # of Runway Intersections	-0.0298008	0.0789456	0.71	-0.18453	0.12493
D: # of Runways	0.3194948	0.1353774	0.02	0.05416	0.58483
D: # of Hotspots	0.0281814	0.0526785	0.59	-0.07507	0.131429
D: Difference of AC/AT and GA Percents	0.4109714	0.4275559	0.34	-0.42702	1.248966
D: AC/AT Percent of Traffic	0.3618836	0.3922496	0.36	-0.40691	1.130679
D: Daily Operations	-0.0200377	0.0039995	0.00	-0.02788	-0.0122
B: # of Runway Intersections	0.2856436	0.1629479	0.08	-0.03373	0.605016
B: # of Runways	-0.3129901	0.2331212	0.18	-0.7699	0.143919
B: # of Hotspots	-0.3321268	0.1353336	0.01	-0.59738	-0.06688
B: Difference of AC/AT and GA Percents	0.5351369	0.7505647	0.48	-0.93594	2.006217
B: AC/AT Percent of Traffic	-0.0665126	0.6102332	0.91	-1.26255	1.129523
B: Daily Operations	0.0046776	0.0054149	0.39	-0.00594	0.015291
A: # of Runway Intersections	0.2200878	0.1245765	0.08	-0.02408	0.464253
A: # of Runways	-0.3580728	0.225181	0.11	-0.79942	0.083274
A: # of Hotspots	-0.0091251	0.0725845	0.90	-0.15139	0.133138

Variable	Coefficient	Standard Error	P-Value	95% CI LB	95% CI UB
A: Difference of AC/AT and GA Percents	0.2301513	0.5724773	0.69	-0.89188	1.352186
A: AC/AT Percent of Traffic	-0.8456367	0.5073891	0.10	-1.8401	0.148828
A: Daily Operations	0.0046211	0.0039781	0.25	-0.00318	0.012418

N = 969	LR Chi-Squared Stat: 67.16
LL = -588.69072	LR P-value: 0.00
LL_0 = -622.2712	

Table 192 – Results of IIA Test for Airport Variables

Omitted Outcome	Chi-Squared Statistic	Degrees of Freedom	P-Value
D	6.71	14	0.95
C	10.40	14	0.73
B	8.53	14	0.86
A	9.03	14	0.83

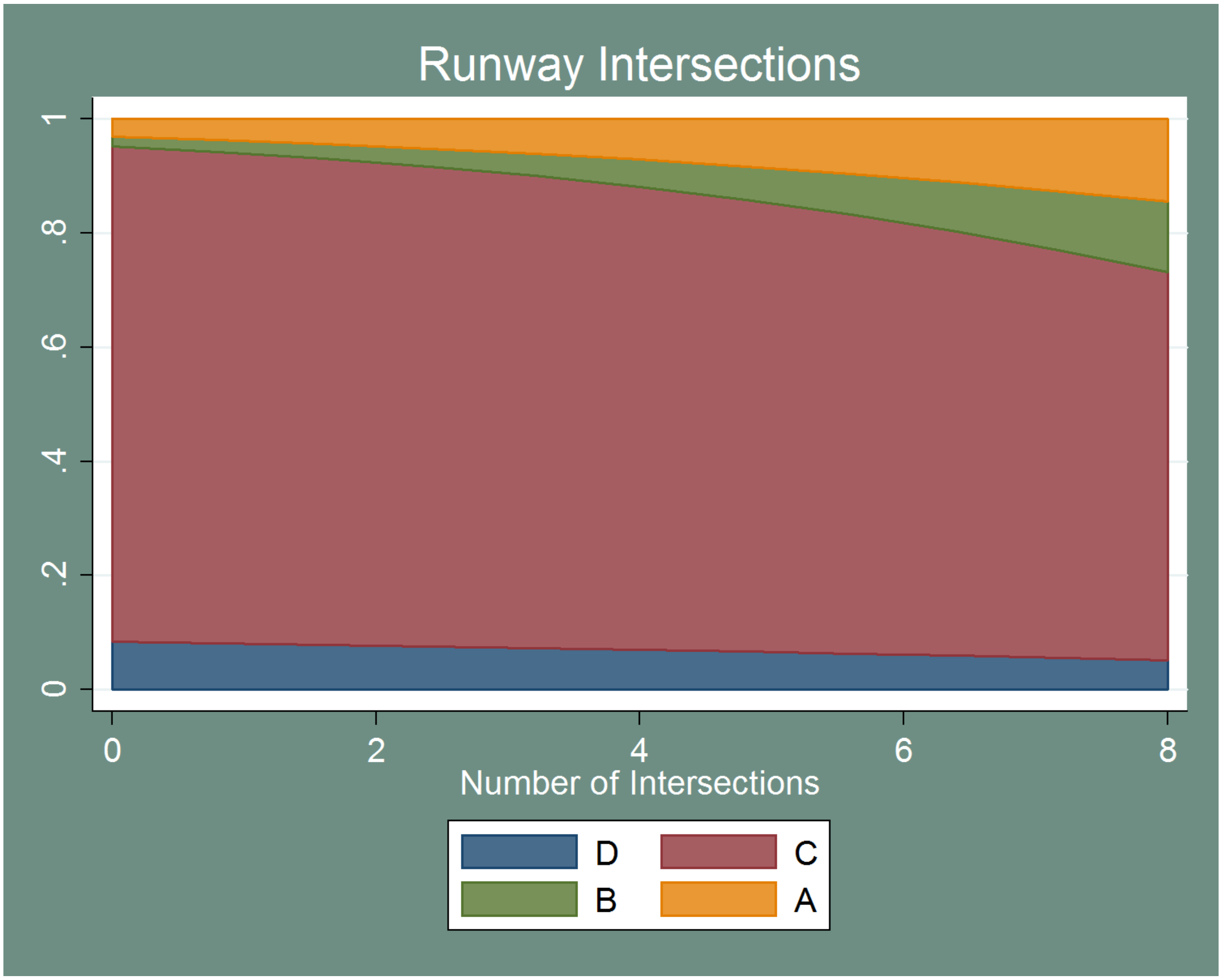

Figure 48 – Impact on Probability of Severity Categories of Number of Runway Intersections

The effect of number of runways appears, on the other hand, to be mostly a shift from the severe categories to category D. One potential explanation is that increased alternative runways can reduce the number of operations that could conceivably conflict. The impact of this variable is also fairly dramatic across the range seen in the dataset.

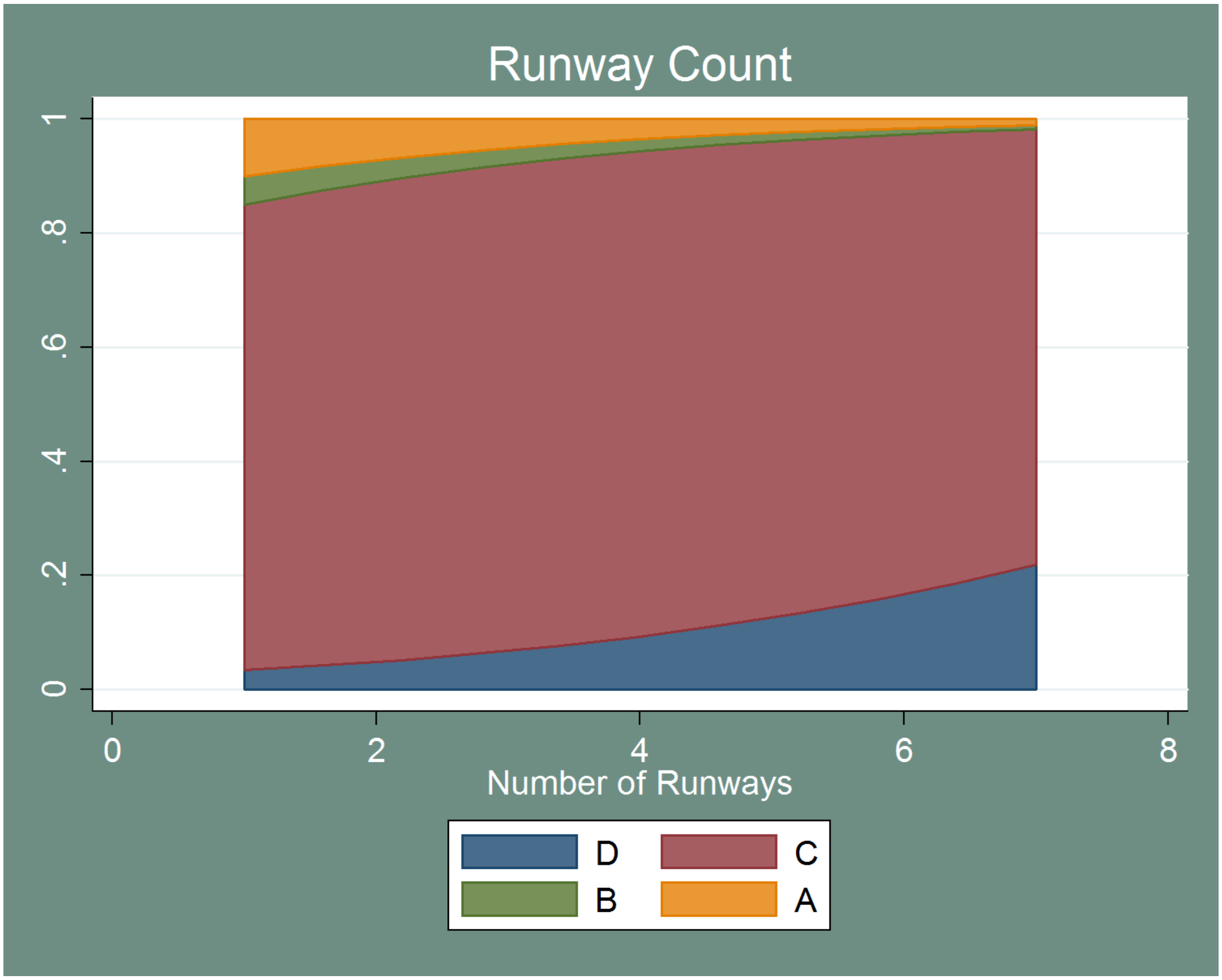

Figure 49 – Impact on Probability of Severity Categories of Number of Runways

Number of hotspots presents an interesting effect. The only severity category that appears to change over the range of this variable is category B. Overall, the impact of this variable appears to be to reduce severity – both categories C and D to increase in area on the chart. However, the impact on category B is still surprising.

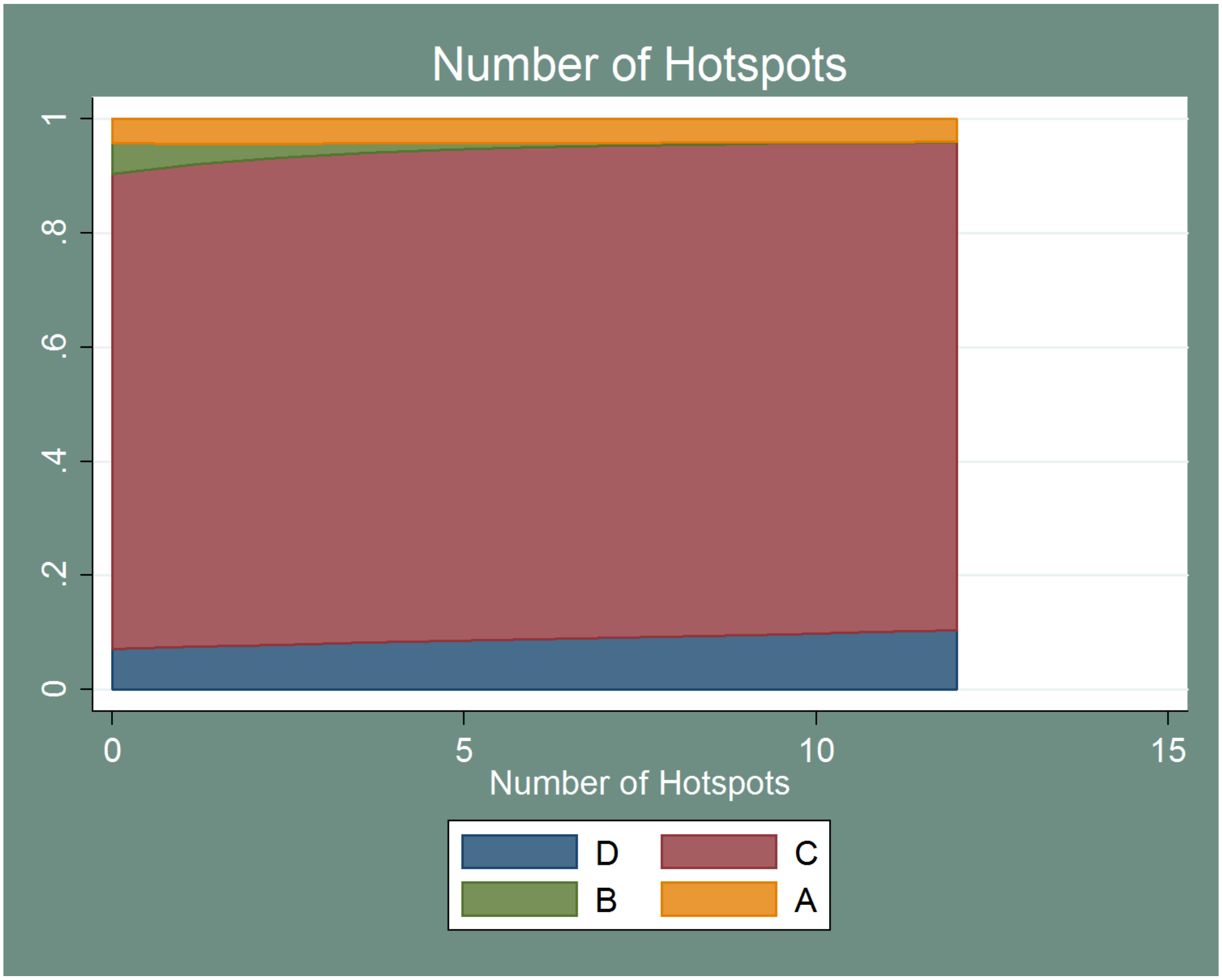

Figure 50 – Impact on Probability of Severity Categories of Number of Hotspots

Daily operations also have a fairly strong impact. The impact is consistent with that seen in the ordered and binary models as well as models containing other sets of variables. The increased severity is likely explained by the increased probability of a conflict event, although the relative probability of category A incursions increases over the range of the variable.

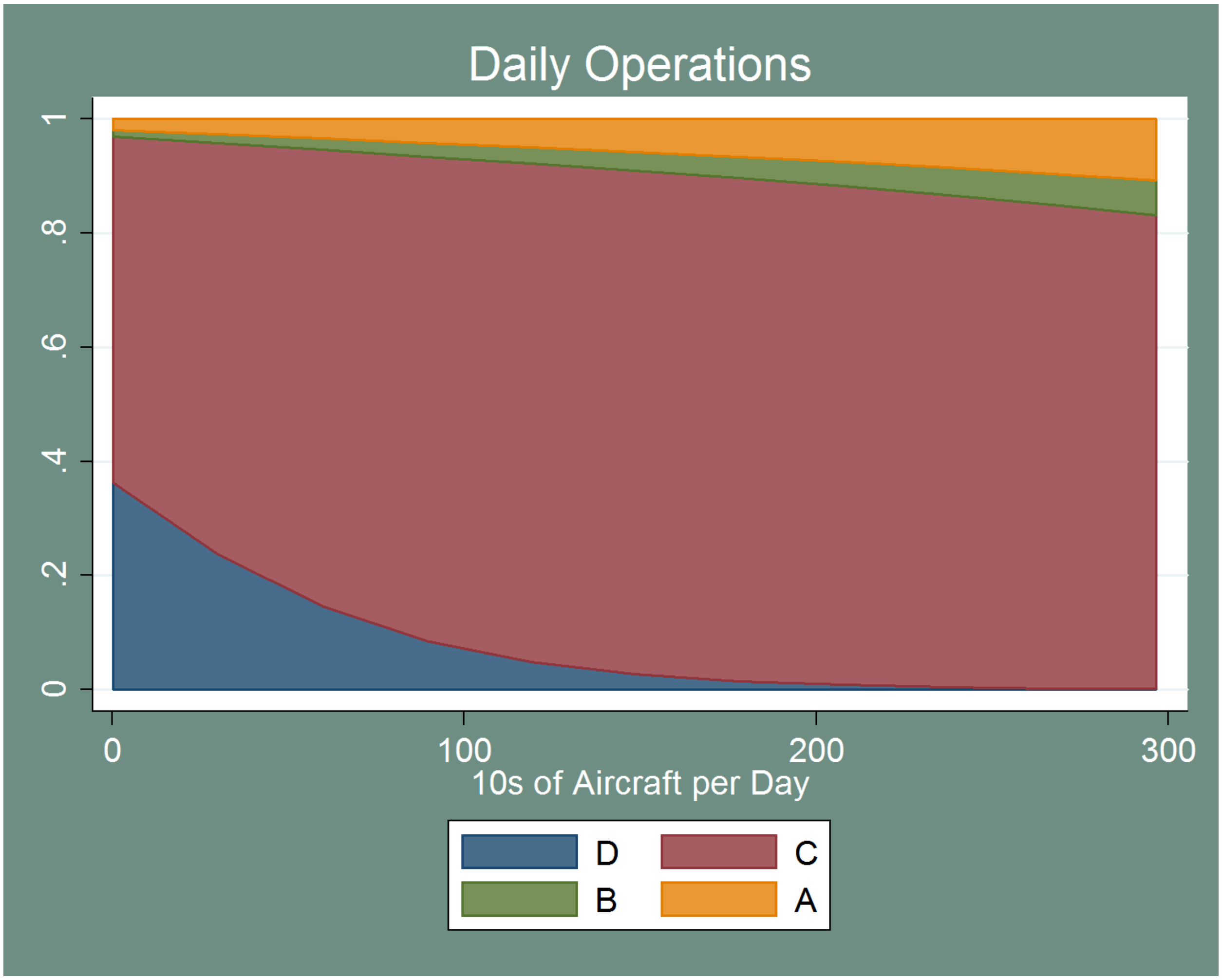

Figure 51 – Impact on Probability of Severity Categories of Daily Operations, Airports

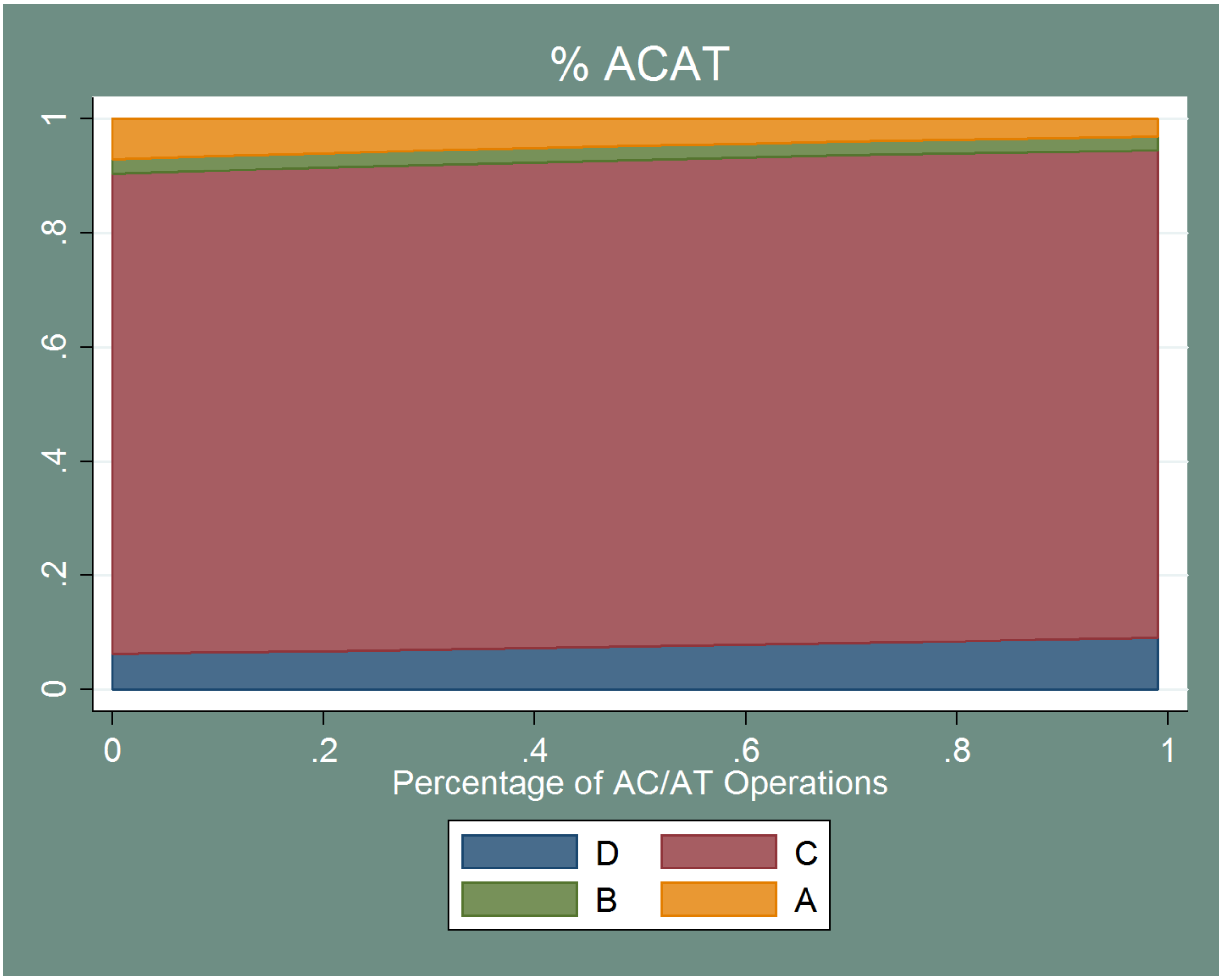

Figure 52 – Impact on Probability of Severity Categories of Percent of AC/AT Traffic

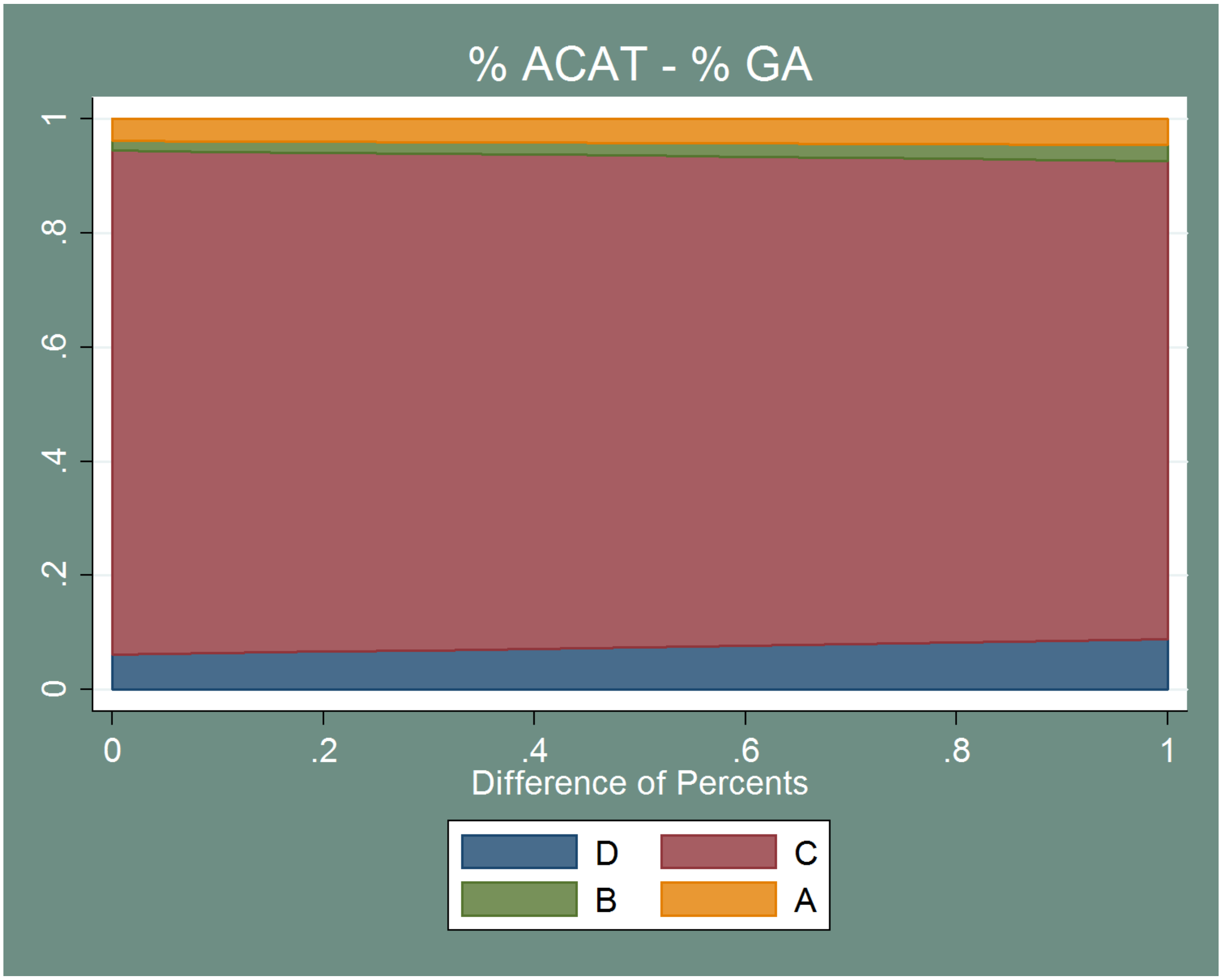

Figure 53 – Impact on Probability of Severity Categories of Difference between Percent AC/AT and GA Traffic

4.3.3. Radar

These models encompass the various radar technologies available in the dataset. This allows for a comparison between the various systems and their impacts on severity.

It is important to note that the ASDE flag in this mode represents any form of ASDE; no distinction was made in the Runway Incursion dataset between ASDE-3 (or earlier) and ASDE-X. Additionally, ARTS was into simplified into variables representing their major version numbers (II or III).

Table 193 – Ordered Logit Results for Radar Variables

Variable	Coefficient	Standard Error	P-Value	95% CI LB	95% CI UB
STARS	-0.9624492	0.2781872	0.00	-1.50769	-0.41721
ASDE	-0.3037916	0.2797329	0.28	-0.85206	0.244475
STARS & ASDE	0.7448619	0.3988561	0.06	-0.03688	1.526606

Variable	Coefficient	Standard Error	P-Value	95% CI LB	95% CI UB
ARTS II	-0.0827244	0.3026437	0.79	-0.6759	0.510446
ARTS III	-0.0530018	0.2529053	0.83	-0.54869	0.442684
Daily Operations	0.0044514	0.0016586	0.01	0.001201	0.007702

N = 970	LR Chi-Squared Stat: 25.61
LL = -612.52423	LR P-value: 0.00
LL_0 = -625.33029	Ordered Test P-Value: 0.00

As these are a series of binary flags, it is important to remember that the alternative to these variables is that the respective system is not in place. Neither ARTS nor ASDE appears to reduce incident severity for OE incidents. STARS, on the other hand, appears to provide some benefit in terms of reducing severity. Interestingly, the interaction between STARS and ASDE is significant at approximately the 6% level. This is inconsistent with the results seen in Table 113, which indicated the interaction effect was insignificant. Additionally, the evidence of the benefit of ASDE seen in Table 111 is no longer observed, likely due to the inclusion of daily operations, which is highly correlated with ASDE (correlation = 0.58). Overall, the model with four severity alternatives does not satisfy the ordering constraint, indicating that these results are not indicative of the true relationship between these variables and severity. When excluding category D events, the ordering constraint is met, but no variable is significant.

Table 194 – Ordered Logit Results for Radar Variables, Conflict Only

Variable	Coefficient	Standard Error	P-Value	95% CI LB	95% CI UB
STARS	-0.6332751	0.4316348	0.14	-1.47926	0.212714
ASDE	-0.5749929	0.4047233	0.16	-1.36824	0.21825
STARS & ASDE	0.2015302	0.6916822	0.77	-1.15414	1.557202
ARTS II	-0.3675354	0.4071408	0.37	-1.16552	0.430446
ARTS III	0.0167015	0.3236412	0.96	-0.61762	0.651027
Daily Operations	-0.0013387	0.0024641	0.59	-0.00617	0.003491

N = 871	LR Chi-Squared Stat: 9.48
LL = -300.88941	LR P-value: 0.15
LL_0 = -305.62813	Ordered Test P-value: 0.12

The binary results are similar to the ordered model. Interestingly, nothing is significant at the standard five percent level (STARS is significant at a 7% level and is the only variable significant at a reasonable level).

Table 195 – Binary Logit Results for Radar Variables

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB
STARS	0.4629815	0.1987276	0.07	0.199618	1.073813
ASDE	0.5537793	0.2268481	0.15	0.248115	1.236003
STARS & ASDE	1.342735	0.9269527	0.67	0.347029	5.195347
ARTS II	0.6845135	0.2769926	0.35	0.309698	1.512955
ARTS III	0.9808792	0.3165585	0.95	0.521084	1.846389
Daily Operations	0.9997259	0.0024566	0.91	0.994923	1.004552

N = 970	LR Chi-Squared Stat: 9.47
LL = -259.27121	LR P-value: 0.15
LL_0 = -264.00835	

Table 196 – Multinomial Logit Results for Radar Variables

Variable	Coefficient	Standard Error	P-Value	95% CI LB	95% CI UB
D: STARS	**0.808187**	**0.316595**	**0.01**	**0.187673**	**1.428702**
D: ASDE	**-0.14364**	**0.459665**	**0.76**	**-1.04457**	**0.757286**
D: STARS & ASDE	**-0.32037**	**0.574702**	**0.58**	**-1.44676**	**0.806024**
D: ARTS II	**-0.27576**	**0.373263**	**0.46**	**-1.00734**	**0.455824**
D: ARTS III	**0.157731**	**0.345069**	**0.65**	**-0.51859**	**0.834054**
D: Daily Operations	**-0.01308**	**0.003128**	**0.00**	**-0.01921**	**-0.00695**
B: STARS	**-0.73312**	**0.593334**	**0.21**	**-1.89603**	**0.429793**

Variable	Coefficient	Standard Error	P-Value	95% CI LB	95% CI UB
B: ASDE	**-0.59063**	**0.635622**	**0.35**	**-1.83643**	**0.655165**
B: STARS & ASDE	**-14.0723**	**789.7078**	**0.99**	**-1561.87**	**1533.727**
B: ARTS II	**-1.40895**	**0.780987**	**0.07**	**-2.93965**	**0.12176**
B: ARTS III	**-0.40292**	**0.472102**	**0.39**	**-1.32823**	**0.52238**
B: Daily Operations	**-0.00125**	**0.003835**	**0.75**	**-0.00876**	**0.00627**
A: STARS	**-0.55095**	**0.606281**	**0.36**	**-1.73924**	**0.637336**
A: ASDE	**-0.58461**	**0.516016**	**0.26**	**-1.59598**	**0.42676**
A: STARS & ASDE	**0.871079**	**0.828519**	**0.29**	**-0.75279**	**2.494945**
A: ARTS II	**0.162622**	**0.498048**	**0.74**	**-0.81353**	**1.138779**
A: ARTS III	**0.319417**	**0.434934**	**0.46**	**-0.53304**	**1.171871**
A: Daily Operations	**-0.00116**	**0.003141**	**0.71**	**-0.00731**	**0.004998**

N = 970	LR Chi-Squared Stat: 66.48
LL = -592.08992	LR P-value: 0.00
LL_0 = -625.33029	

Table 197 – IIA Test Results for Radar Variables

Omitted Outcome	Chi-Squared Stat	Degrees of Freedom	P-Value
D	9.7x10^8	14	0.00
B	5.74	14	0.97
A	1.6x10^10	14	0.00
C	5.9x10^10	14	0.00

The multinomial model does little to clarify the results. Additionally, note that this model does not satisfy the IIA assumption when category B incursions are excluded. Although these tests are not particularly powerful, it is important to acknowledge that this model might violate that assumption in some cases. Because this is a series of flags with interactions, the predicted probabilities for each category are depicted in Table 198. The baseline airport has ARTS II.

Table 198 – Predicted Probabilities for Different Radar Combinations

STARS	ASDE	STARS & ASDE	ARTS-II	ARTS_III	Probability of Category D	Probability of Category C	Probability of Category B	Probability of Category A
YES	YES	YES	YES	NO	0.08	0.88	0.00	0.05
NO	YES	NO	YES	NO	0.05	0.91	0.01	0.04
YES	NO	NO	YES	NO	0.12	0.84	0.01	0.04
NO	NO	NO	YES	NO	0.05	0.87	0.02	0.06

Figure 54 depicts the impact of daily operations on severity categories. The only category for which this variable is significant is category D. As seen in other models containing this variable, increased daily operations are associated with increased severity.

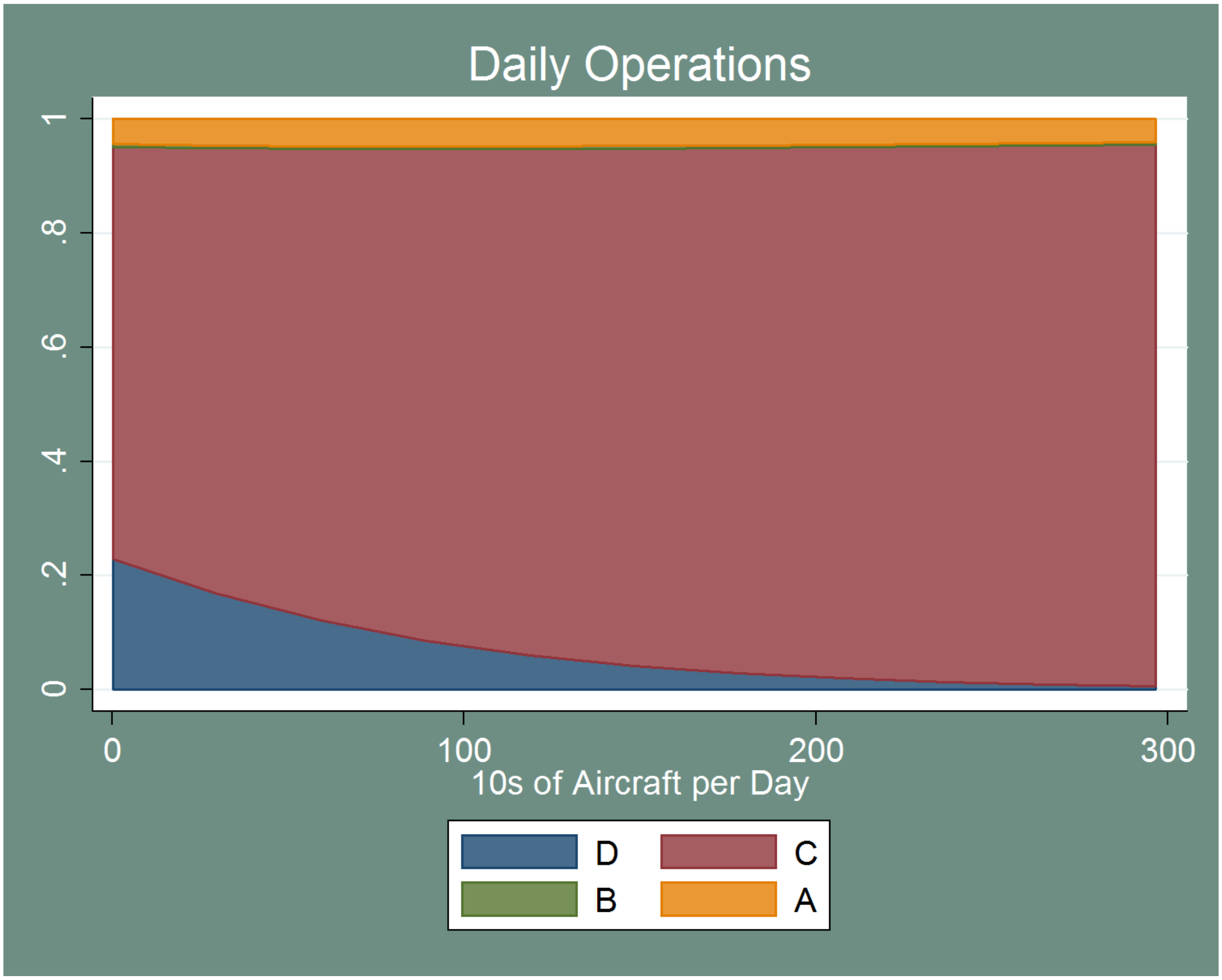

Figure 54 – Impact on Probability of Severity Categories of Daily Operations, Radar

It appears that, as seen in the ordered and binary models, STARS reduces the likelihood of severe incidents; however, this appears to be mostly a reduction in category C. Adding ASDE to STARS actually increases the likelihood of category C compared to only STARS, but ASDE alone also reduces the of likelihood of *conflict* incidents. As mentioned previously, it is possible that these effects are capturing the distribution of radar among airports. That is, ASDE is may be deployed at mostly busier airports that are more likely to have conflict events (due to the higher traffic). Additionally, the model coefficients are not precisely estimated – even when they are statistically different from zero. Thus, this model suggests that STARS may have some benefit in terms of reducing severity, but the results on other radar systems are inconclusive and provide little information beyond that provided by the categorical tests presented in Section 3.3.4.

4.3.4. Controller

These models examine the characteristics of the controller involved in the incident. Recall that the sample is only OE incidents, so in some sense these describe the controller responsible for the incident. The ordered results (including all severity categories) are presented below.

Table 199 – Ordered Logit Results for Controller Variables

Variable	Coefficient	Standard Error	P-Value	95% CI LB	95% CI UB
Age	0.0137436	0.0113773	0.23	-0.00856	0.036043
Time on Shift	0.0002362	0.0004877	0.63	-0.00072	0.001192
Training in Last Year	0.070188	0.2586886	0.79	-0.43683	0.577208
Workload	0.1248622	0.0295402	0.00	0.066965	0.18276
Daily Operations	0.0012988	0.0014647	0.38	-0.00157	0.00417

N = 780	LR Chi-Squared Stat: 25.27
LL = -491.74876	LR P-value: 0.00
LL_0 = - 504.38492	Ordered Test P-Value: 0.00

Note that the ordering assumption for this model is violated. This is consistent with the other ordered models presented in this section that contain all four severity categories. Additionally, very few of the variables seem to explain the variation in incursion severity. The only variable that is significant (for all severity categories or for conflict events only) is controller workload (the number of aircraft the controller is responsible for at the time of the incident). When excluding category D incidents, this variable is only marginally significant at the 10% level. Daily operations are also significant at the 10% level in the conflict only model, but with the opposite sign to that seen in other models. In general, it appears that these ordered models are not particularly informative.

Table 200 – Ordered Logit Results for Controller Variables, Conflict Only

Variable	Coefficient	Standard Error	P-Value	95% CI LB	95% CI UB
Age	0.001266	0.0151437	0.93	-0.02842	0.030947
Time on Shift	0.000753	0.0006008	0.21	-0.00042	0.00193
Training in Last Year	0.07604	0.3602348	0.83	-0.63001	0.782087
Workload	0.061144	0.0337487	0.07	-0.005	0.127291
Daily Operations	-0.00378	0.0021747	0.08	-0.00804	0.000486

N = 712	LR Chi-Squared Stat: 6.15

LL = -270.45674	LR P-value: 0.29
LL_0 = - 273.53366	Ordered Test P-Value: 0.52

The binary logit results are not much more promising. Controller workload is again the only significant variable, and maintains the same effect of increasing severity.

Table 201 – Binary Logit Results for Controller Variables

Variable	Coefficient	Standard Error	P-Value	95% CI LB	95% CI UB
Age	0.8931086	0.089687	0.26	0.733543	1.087384
Time on Shift	1.001394	0.001176	0.24	0.999091	1.003702
Training in Last Year	1.000619	0.000584	0.29	0.999475	1.001764
Workload	1.022499	0.183874	0.90	0.718776	1.454563
Daily Operations	1.076802	0.035294	0.02	1.009803	1.148247

N = 780	LR Chi-Squared Stat: 7.44
LL = -229.47049	LR P-value: 0.28
LL_0 = - 233.19181	

Some additional insights are available from the multinomial model. This model also satisfies the IIA assumption. Controller age and the flag for controller training are still insignificant across all categories. The result for training is not entirely surprising given that most controllers receive runway incursion training frequently enough that 70% controllers are marked as "yes" in the dataset.

Table 202 – Multinomial Logit Results for Controller Variables

Variable	Coefficient	Standard Error	P-Value	95% CI LB	95% CI UB
D: Age	-0.0201249	0.014414	0.16	-0.04838	0.008126
D: Time on Shift	0.0001593	0.000657	0.81	-0.00113	0.001447
D: Training in Last Year	-0.150851	0.364085	0.68	-0.86445	0.562743
D: Workload	-0.3938314	0.078881	0.00	-0.54844	-0.23923
D: Daily Operations	-0.0059526	0.002773	0.03	-0.01139	-0.00052
B: Age	0.0232815	0.023268	0.32	-0.02232	0.068885
B: Time on Shift	-0.0005547	0.001118	0.62	-0.00275	0.001636
B: Training in Last Year	-0.1402045	0.5057	0.78	-1.13136	0.850949

Variable	Coefficient	Standard Error	P-Value	95% CI LB	95% CI UB
B: Workload	0.0437738	0.053595	0.41	-0.06127	0.148819
B: Daily Operations	-0.0038828	0.003307	0.24	-0.01036	0.002599
A: Age	-0.0122581	0.019413	0.53	-0.05031	0.025792
A: Time on Shift	0.0012345	0.000657	0.06	-5.4E-05	0.002523
A: Training in Last Year	0.2191219	0.493681	0.66	-0.74847	1.186718
A: Workload	0.067486	0.039652	0.09	-0.01023	0.145203
A: Daily Operations	-0.0035337	0.002794	0.21	-0.00901	0.001943

N = 780	LR Chi-Squared Stat: 67.22
LL = -470.776	LR P-value: 0.00
LL_0 = - 504.38492	

Table 203 – Result of IIA Test for Controller Variables

Omitted Outcome	Chi-Squared Stat	Degrees of Freedom	P-Value
D	3.059	12	1.00
C	7.297	12	0.84
B	6.789	12	0.87
A	6.123	12	0.91

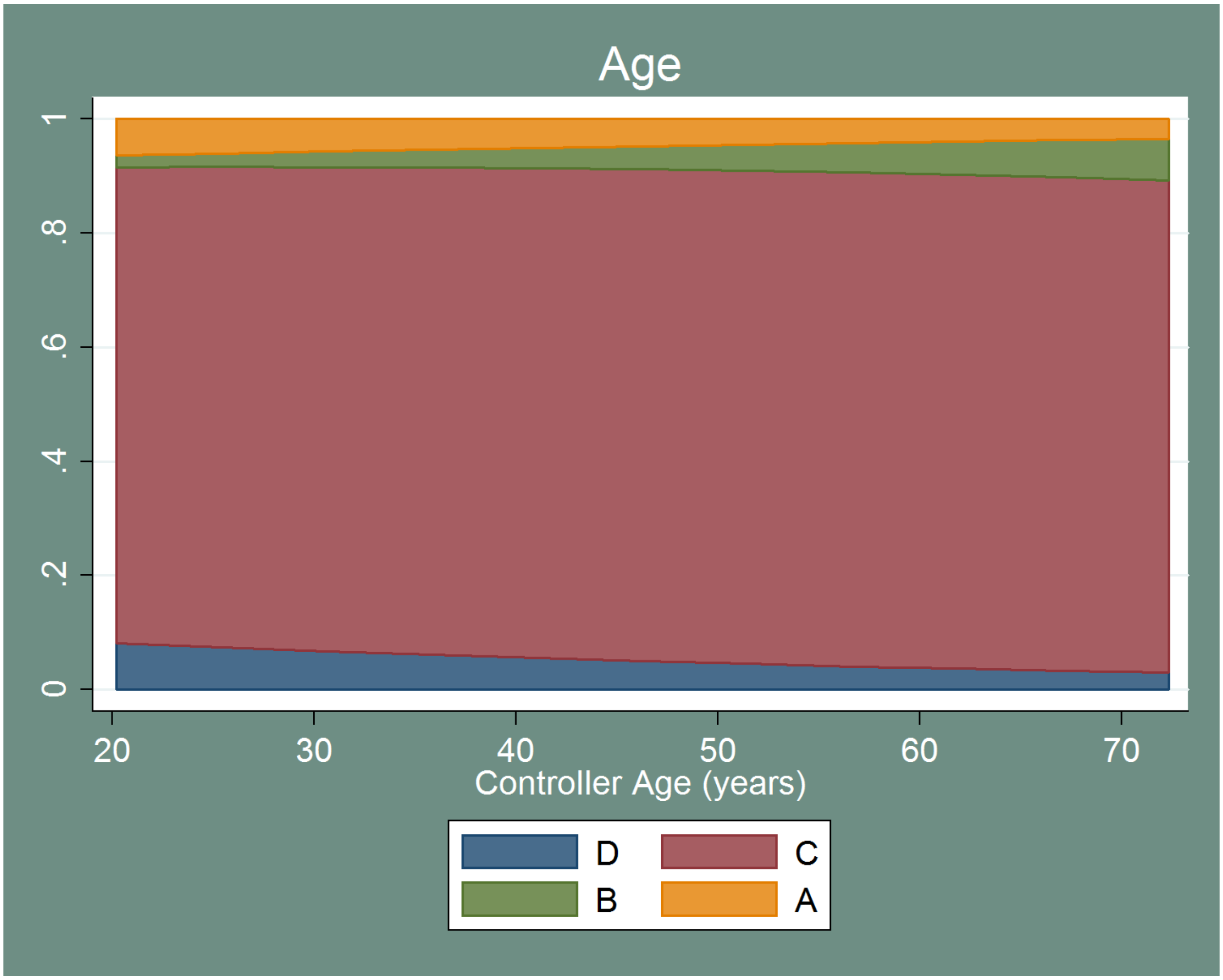

Figure 55 – Impact on Probability of Severity Categories of Controller Age

The result for age is interesting in its non-significance.[98] Figure 55 depicts the impact graphically. While there is some change in probability over the range, the variable is insignificant for any category. Thus, it is indistinguishable in a statistical sense from a graph that showed each category as a straight line over the range of controller age. One might naively expect controller age to contribute to severity – either through lowered reaction times or increased experience. It is impossible to disentangle those two effects without a better measure of these possible causes; those two explanations may both be at play and counteracting each other. Recall that controller age is also capped artificially by forced retirement. All in all, it is possible that current practices already account for the impact of age. Regardless, there is no indication that increased controller age contributes to severity.

Controller workload is highly significant for category D, but not so for other categories. It is significant at a lesser 10% level for category A. This likely explains the dramatic increase in category A and decrease in category D probabilities seen in Figure 56. This is consistent with the effect seen in the ordered and

98 A model was tested with a squared term for age, attempting to account for a nonlinear effect of age as seen in other behavioral contexts. This did not result in any changes to the model and thus was not reported.

binary models, and supports the intuition that controllers can only handle so many planes before safety is compromised.

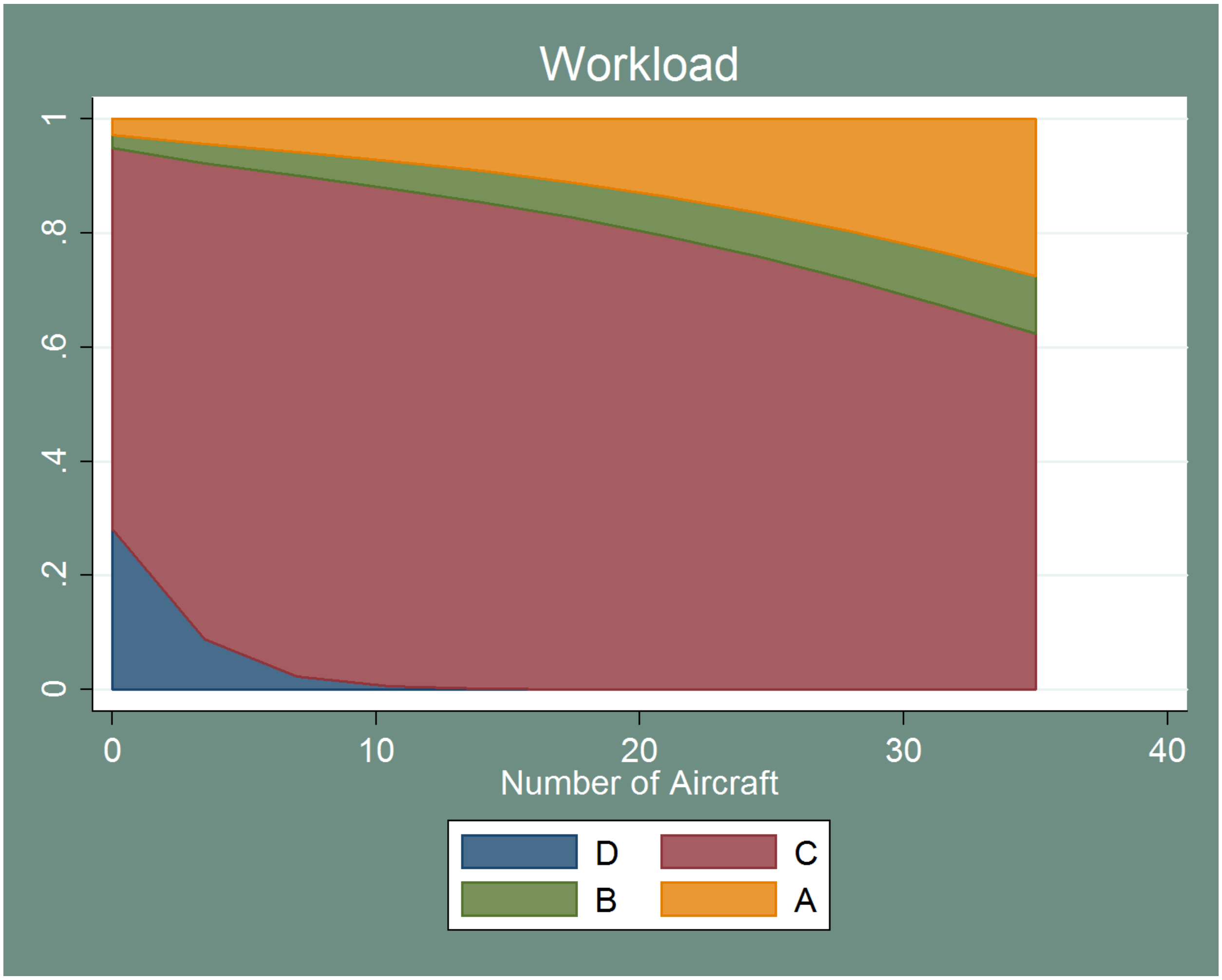

Figure 56 – Impact on Probability of Severity Categories of Controller Workload

Time on shift is significant at the 10% level for category A incursions, and insignificant for all other categories. Figure 57 indicates that the increase in the probability of category A comes mostly at the expense of category C. This hints that time on shift is associated with increased severity, but does not appear to impact category B in a statistically significant manner. Over a reasonable range (an 8-hour shift is approximately 500 minutes) this impact is not large. It is unclear why there are records in the dataset that have a time on shift three times larger than that. It is possible that these extremely long shifts represent a data error in the reported shift start and end times.[99] When estimated excluding shifts longer than eight hours, the impact of time on shift is not statistically different than zero – further contributing to the idea that this is a spurious result.

99 While it seems likely these records are an error, the research team was unable to find anyone able to certify that these shifts were not possible in extreme/unusual circumstances.

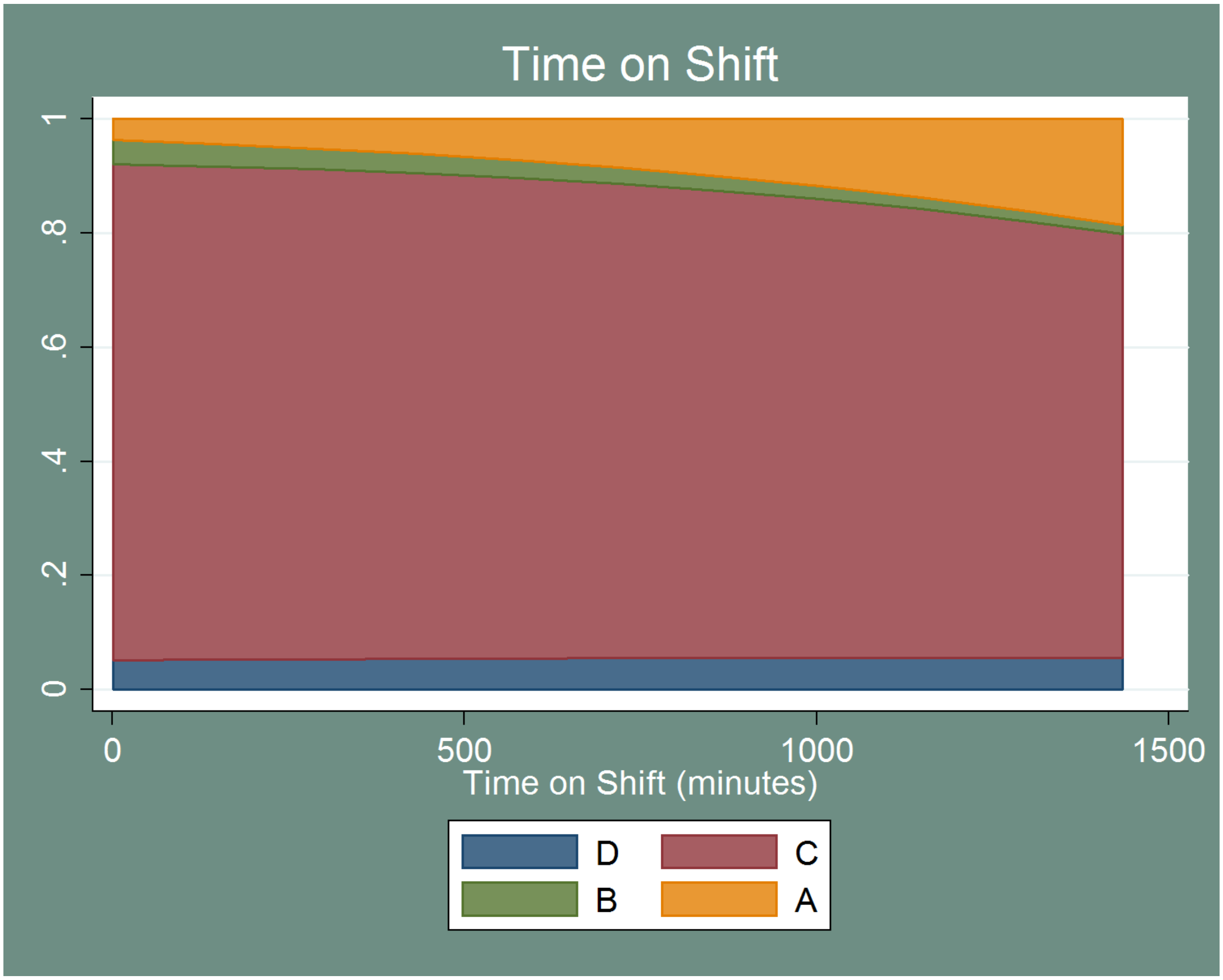

Figure 57 – Impact on Probability of Severity Categories of Controller Time on Shift

Finally, daily operations appear to have a different impact than it does in other models. Increased daily operations appear to increase category C events, but not either of the severe categories. Contrast this effect to that seen in Figure 51.

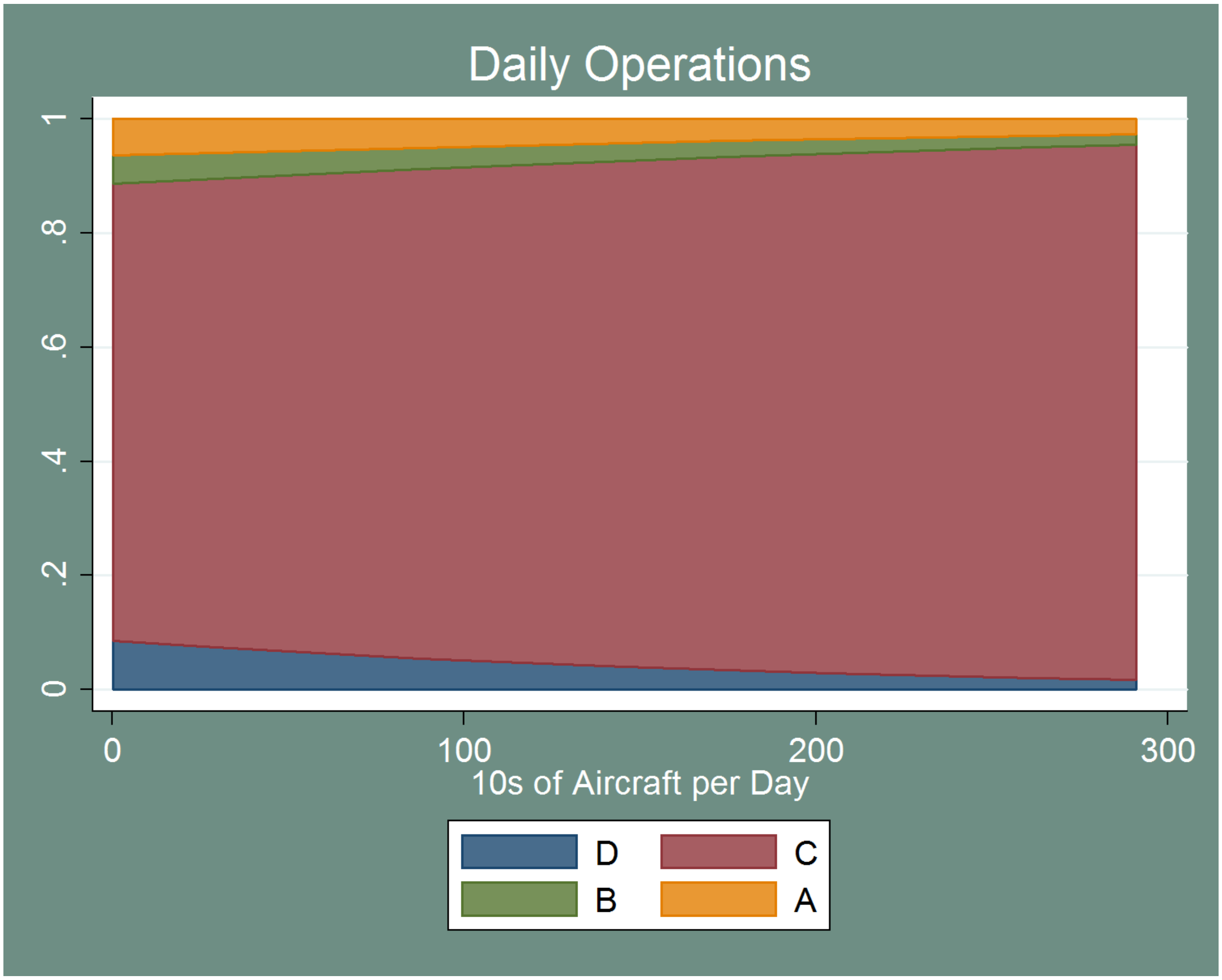

Figure 58 – Impact on Probability of Severity Categories of Daily Operations, Controller

Overall, the controller variables shed little insight into severity. The most useful conclusion is perhaps that controller age does not impact severity. It is also important to note that increased controller workload may contribute to increased severity. Additionally, the impact of time on shift is suspect. Caution when using this model to draw conclusions is warranted. Further research into controllers, possibly including controller information for non-incursions, is highly recommended.

4.3.5. Weather

These models contain many of the weather variables identified in previous sections. However, the advantage of the models is that interactions between variables can be explored. This is especially pertinent for weather variables, as many of them are quite closely related. The results of the ordered model for weather variables are presented below. This model includes all severity categories D through A.

Table 204 – Ordered Logit Results for Weather Variables

Variable	Coefficient	Standard Error	P-Value	95% CI LB	95% CI UB
Cloud Coverage	-0.078913	0.040734	0.05	-0.15875	0.000924
Sea Level Pressure	-0.0644932	0.023233	0.01	-0.11003	-0.01896
Cloud Coverage x Sea Level Pressure	0.0145993	0.004912	0.00	0.004972	0.024227
No Weather Phenomena	-0.4790643	0.382549	0.21	-1.22885	0.270717
Wind Speed	0.0220009	0.023522	0.35	-0.0241	0.068103
Daily Operations	0.0040903	0.001568	0.01	0.001016	0.007164

N = 633	LR Chi-Squared Stat: 19.15
LL = -403.55011	LR P-value: 0.00
LL_0 = -393.97433	Ordered Test P-Value: 0.00

The results of the weather model are a bit surprising. Firstly, cloud coverage appears to decrease severity. This is similar to the result seen in Table 150 where category A incursions had a lower median cloud coverage (i.e., increased cloud coverage is associated with lower severity), although the individual categories were not distinguishable from each other in Table 150. It is possible that this is revealing an overreaction of sorts to increased cloud coverage. That is, operational changes may occur (such as decreased traffic or larger spacing between traffic) that already counteract the increased severity risk due to the lowered visibility. If that were true, these measures appear to overcorrect (in some sense) and end up decreasing the likelihood of category A events during cloudy weather.

A similar pattern is seen for sea level pressure – increased sea level pressure is associated with lowered severity. This is contrary to the results seen in Table 160, which indicated no relationship between severity and sea level pressure. Higher sea level pressure is associated with clearer skies and generally calmer weather, but it is unclear how pressure would directly impact operations on the ground. It is more likely that pressure impacts the pilot population on a given day. Higher pressure, and calmer weather, is more amenable to GA pilots who are much more likely to be involved in category D incursions than their commercial counterparts. It is possible this change in pilot population is also reflected in the severity of OE incidents.

The interaction between cloud coverage and sea level pressure is also significant. Because it is the opposite sign of both cloud coverage and sea level pressure, it has an ameliorating effect on the impact of those variables. That is, if cloud coverage and sea level pressure are both higher, the interaction is a mitigating effect – the impact on severity is less than the variables alone would predict. As noted earlier, a more thorough examination of the impacts of weather on severity is required to better understand these impacts.

Finally, the indicator for no weather phenomena and wind speed are insignificant. One might expect that rain, haze, or fog may impact the severity of an incident, but it does not appear to do so. The exposure variable, as expected, increases the likelihood of a severe incursion.

The results of the test on ordering assumption indicate that this model is invalid. Table 205 presents the same regression, but excludes category D incursions. The indicator for no weather phenomena is now significant, but the interaction between cloud coverage and sea level pressure is not. When examining only conflict events, the assumptions of an ordered model are satisfied. This lends further support to the idea that category D incursions are not ordered in the same way categories C through A are. It also suggests the use of a multinomial model to account for the non-ordered nature of all four categories. Similar to the previous sections, a binary logit is also presented for comparison.

Table 205 – Ordered Logit Results for Weather Variables, Conflict Only

Variable	Coefficient	Standard Error	P-Value	95% CI LB	95% CI UB
Cloud Coverage	-0.1898268	0.071915	0.01	-0.33078	-0.04888
Sea Level Pressure	-0.0785278	0.034337	0.02	-0.14583	-0.01123
Cloud Coverage x Sea Level Pressure	0.0115897	0.008594	0.18	-0.00525	0.028434
No Weather Phenomena	-1.145536	0.55067	0.04	-2.22483	-0.06624
Wind Speed	-0.0433148	0.039672	0.28	-0.12107	0.03444
Daily Operations	-0.0026633	0.002723	0.33	-0.008	0.002674

N = 555	LR Chi-Squared Stat: 15.84
LL = -159.33354	LR P-value: 0.01
LL_0 = -167.25292	Ordered Test P-Value: 1.00

The binary logit results are similar to the ordered results. The variables maintain their signs, but the results in terms of significance are more similar to the conflict only model (Table 205) than the all-inclusive ordered model (Table 204).

Table 206 – Binary Logit Results for Weather Variables

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB
Cloud Coverage	0.8262223	0.059955	0.01	0.716687	0.952499
Sea Level Pressure	0.9189695	0.031839	0.02	0.858638	0.98354
Cloud Coverage x Sea Level Pressure	1.013526	0.008749	0.12	0.996522	1.03082
No Weather Phenomena	0.3159954	0.174439	0.04	0.107101	0.932329
Wind Speed	0.9660382	0.03827	0.38	0.893869	1.044035

Variable	Odds Ratio	Standard Error	P-Value	95% CI LB	95% CI UB
Daily Operations	0.9983222	0.00271	0.54	0.993025	1.003648

N = 633	LR Chi-Squared Stat: 14.82
LL = -141.76193	LR P-value: 0.02
LL_0 = -149.17232	

Overall, the model passes the test for IIA. As noted earlier, though, these tests are not particularly strong but are presented for completeness. The coefficient results from the multinomial model are mixed. As with the other models, it is best to examine the impact of the variables as changes in probability for each severity category. There is only one categorical dependent variable in this model (the flag for no weather phenomena), and its impact is reported in Table 209. The figures following depict the impact of cloud coverage at various levels of sea level pressure.

Table 207 – Multinomial Logit Results for Weather Variables

Variable	Coefficient	Standard Error	P-Value	95% CI LB	95% CI UB
D: Cloud Coverage	-0.00198	0.0498018	0.97	-0.09959	0.095625
D: Sea Level Pressure	0.03155	0.0272032	0.25	-0.02177	0.084867
D: Cloud Coverage x Sea Level Pressure	-0.01379	0.0063144	0.03	-0.02616	-0.00141
D: No Weather Phenomena	-0.01055	0.4756608	0.98	-0.94282	0.921732
D: Wind Speed	-0.05662	0.0304117	0.06	-0.11623	0.002984
D: Daily Operations	-0.01013	0.0027228	0.00	-0.01547	-0.00479
B: Cloud Coverage	-0.07176	0.1249049	0.57	-0.31657	0.17305
B: Sea Level Pressure	-0.03565	0.0676585	0.60	-0.16825	0.096962
B: Cloud Coverage x Sea Level Pressure	0.003223	0.0163779	0.84	-0.02888	0.035323
B: No Weather Phenomena	-0.16004	1.160003	0.89	-2.43361	2.113523
B: Wind Speed	-0.07913	0.0778411	0.31	-0.23169	0.073439
B: Daily Operations	-0.00381	0.0052981	0.47	-0.01419	0.006579
A: Cloud Coverage	-0.24032	0.0883936	0.01	-0.41356	-0.06707
A: Sea Level Pressure	-0.09377	0.0398574	0.02	-0.17189	-0.01565

Variable	Coefficient	Standard Error	P-Value	95% CI LB	95% CI UB
A: Cloud Coverage x Sea Level Pressure	0.014579	0.0101211	0.15	-0.00526	0.034416
A: No Weather Phenomena	-1.51362	0.6315717	0.02	-2.75148	-0.27576
A: Wind Speed	-0.02757	0.0455582	0.55	-0.11686	0.061724
A: Daily Operations	-0.00215	0.0031654	0.50	-0.00835	0.004053

N = 633	LR Chi-Squared Stat: 48.90
LL = -379.10258	LR P-value: 0.00
LL_0 = -403.55011	

Table 208 – Results of IIA Test for Aircraft Variables

Omitted Outcome	Chi-Squared Stat	Degrees of Freedom	P-Value
D	8.34	14	0.87
C	5.15	14	0.98
B	4.39	14	0.99
A	4.05	14	1.00

Table 209 – Change in Probability of Severity Categories for Categorical Variables, Weather

	Category D	Category C	Category B	Category A
No Weather Phenomena	.01	.09	-.00	-.10

The weather phenomena flag has an interesting effect. It decreases the likelihood of the severe categories, while increasing the likelihood of category C and D. This type of impact is not able to be modeled by the ordered model presented previously. It is unclear why good weather would both reduce the probability of the most and least severe events. There is likely an underlying behavioral change in good weather – either in the pilot population or in how controllers manage traffic or elsewhere – that is the source of this impact.

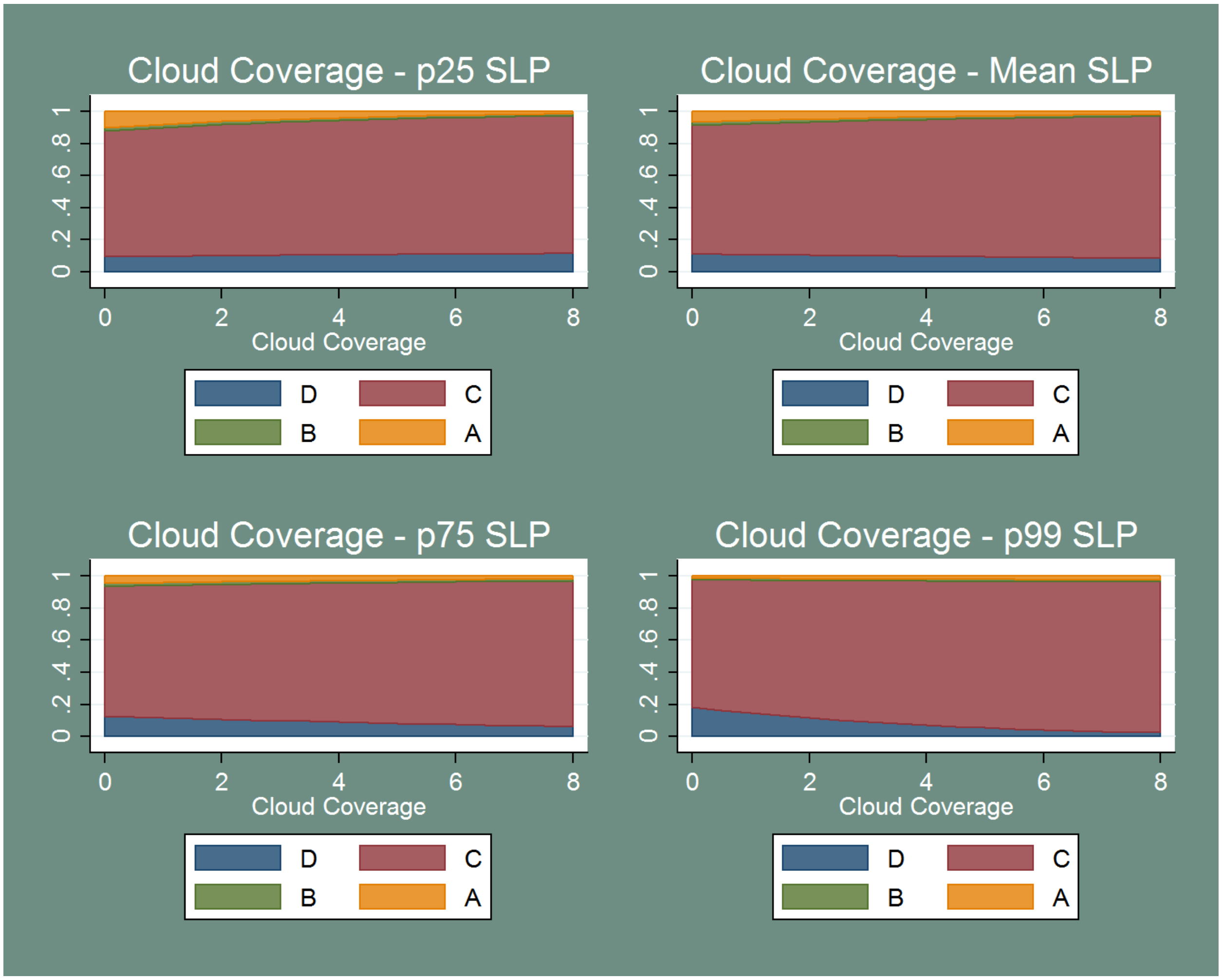

Figure 59 – Cloud Coverage and Sea Level Pressure

The impact of cloud coverage and sea level pressure are also interesting. At relatively low levels of sea level pressure, increased cloud coverage appears to reduce the probability of category A incursions. However, at relatively high levels of pressure increased cloud coverage decreases the likelihood of category D and increases category C. Not only does the impact of cloud coverage on severity change, it appears to decreases severity (at low levels of sea level pressure) and alternatively increases severity (at higher levels of sea level pressure). It is possible that these varying impacts are reflecting operational changes, as well. As with the indicator for weather phenomena, further study is required to truly understand this effect. It is likely that an underlying behavioral factor – such as visibility – is truly at play here, and spurious correlation cannot be ruled out.

Wind speed and exposure both appear to decrease the likelihood of a category D incursion. The mechanism for exposure is clear: more traffic increases the likelihood of a conflict event. Wind speed is another matter. As with the many of the weather variables, the only conclusion that can be drawn is that there is a correlation, and the general direction of that correlation. It is likely that underlying behavior that is impacted by the weather in turn impacts severity, rather than weather leading directly to increased or decreased severity. Thus, weather and related behavioral changes appear to be fertile ground for further research.

4.3.6. "Bouillabaisse"

The models discussed above focus on testing specific sets of variables. The goal of the model presented in this section is to best *predict* severity, given the variables available. This model was developed by picking the most relevant parts of the previous models and combining them. Fit statistics were used to help identify those models that were "better" in the numerical sense. While a limited approach, the goal is to best fit to the data rather than test specific hypotheses. The models presented in this section are prone to overfit, and may not be generalizable to other datasets or time periods. In other words, this represents the best guess at predicting the severity of runway incursions but may not be the best explanatory model.

The model presented in this section represents *only* the best prediction given this single data set and the models run above; no result from this model should be taken as proof of any causal relationship or a directive to change any particular policies, practices, or operations.

No ordered or binary logit results are presented for this set of variables. The previous models all point to a multinomial framework as being the most useful in explaining all four severity categories. The multinomial results are presented below. Note that no weather variables were included in this model. While potentially interesting, due to limited weather data availability, inclusion of the weather variables reduced the sample size of the model dramatically. Given the indeterminate conclusions that could be drawn from the weather variables, they were excluded in favor of a larger sample size.

Table 210 – Multinomial Logit Results for Best Prediction Model

Variable	Coefficient	Standard Error	P-Value	95% CI LB	95% CI UB
D: Workload	-0.3463247	0.06921	0.00	-0.48197	-0.21068
D: Commercial Carrier	-0.6666587	0.317992	0.04	-1.28991	-0.04341
D: Takeoff	0.0884049	0.264781	0.74	-0.43056	0.607366
D: Daily Operations	-0.0073125	0.003083	0.02	-0.01336	-0.00127
D: # of Hotspots	0.0107307	0.055405	0.85	-0.09786	0.119323
D: AC/AT % of Operations	0.8213881	0.454675	0.07	-0.06976	1.712535
D: # of Runway Intersections	0.0400874	0.079078	0.61	-0.1149	0.195077
B: Workload	0.0579266	0.059663	0.33	-0.05901	0.174863
B: Commercial Carrier	-1.299852	0.537806	0.02	-2.35393	-0.24577
B: Takeoff	0.0624177	0.44852	0.89	-0.81666	0.9415
B: Daily Operations	0.0031241	0.00443	0.48	-0.00556	0.011806

Variable	Coefficient	Standard Error	P-Value	95% CI LB	95% CI UB
B: # of Hotspots	-0.3429516	0.132618	0.01	-0.60288	-0.08303
B: AC/AT % of Operations	0.4746274	0.664255	0.48	-0.82729	1.776543
B: # of Runway Intersections	0.2052403	0.135995	0.13	-0.06131	0.471786
A: Workload	0.0641569	0.044642	0.15	-0.02334	0.151654
A: Commercial Carrier	-0.9109452	0.452898	0.04	-1.79861	-0.02328
A: Takeoff	0.9690437	0.316484	0.00	0.348747	1.589341
A: Daily Operations	0.0024284	0.003431	0.48	-0.0043	0.009153
A: # of Hotspots	-0.0124248	0.073299	0.87	-0.15609	0.131239
A: AC/AT % of Operations	-0.5265661	0.574409	0.36	-1.65239	0.599254
A: # of Runway Intersections	0.1109985	0.097971	0.26	-0.08102	0.303019

N = 947	LR Chi-Squared Stat: 100.00
LL = -537.93165	LR P-value: 0.00
LL_0 = -587.933	

Table 211 - Results of IIA Test for Best Prediction Model

Omitted Outcome	Chi-Squared Stat	Degrees of Freedom	P-Value
D	7.08	16	0.97
C	7.04	16	0.97
B	12.00	16	0.74
A	12.71	16	0.69

Table 212 - Change in Probability of Severity Categories for Categorical Variables, Best Prediction Model

	Category D	Category C	Category B	Category A
Commercial Carrier	-.03	.09	-.03	-.03
Takeoff	.00	-.05	.00	.05

The precise impacts are depicted in Table 212 and subsequent figures. Many of the relationships expressed in this model are consistent with those described in the individual models above. Commercial carrier status reduces the probability of categories A and B and increases the probability of category C, as seen in Table 186. Additionally, commercial carrier status appears to reduce the probability of category D incursions. Although not seen in the Aircraft Model (as category D incursions were excluded), this is likely explained by the tendency for commercial pilots to operate at busier airports. Takeoff continues to be a dangerous time for aircraft and increases the likelihood of a category A incursion. Takeoff also has a marginal increase in the likelihood of category D; however, the coefficient on takeoff for category D is not precisely estimated, making this effect statistical noise rather than a true effect.

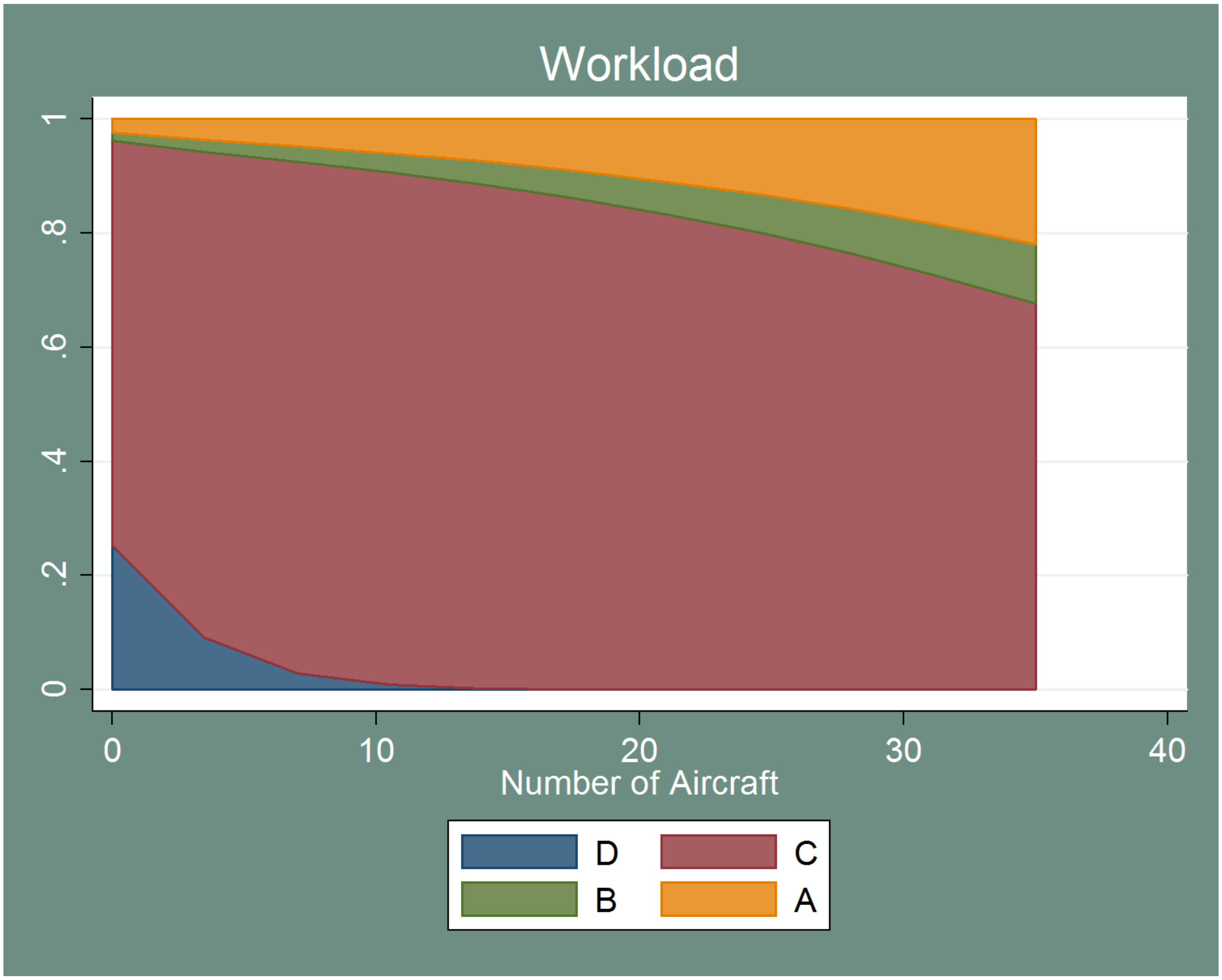

Figure 60 - Impact on Probability of Severity Categories of Controller Workload, Best Prediction Model

Controller workload, although not significant for all severity categories, has a fairly dramatic effect. As controller workload increases, the probability of higher severity incursions also increases. This evidence clearly supports the hypotheses that increased complexity, of which controller workload is but one part, increases the likelihood of a severe event. A related hypothesis is that increased complexity also leads to more incursions overall (higher frequency instead of severity). While this model does not directly answer that hypothesis, it does indicate that complexity increases severity. A model focusing on the frequency of runway incursions may find that increased complexity leads to more incursions in addition to higher severity incursions.

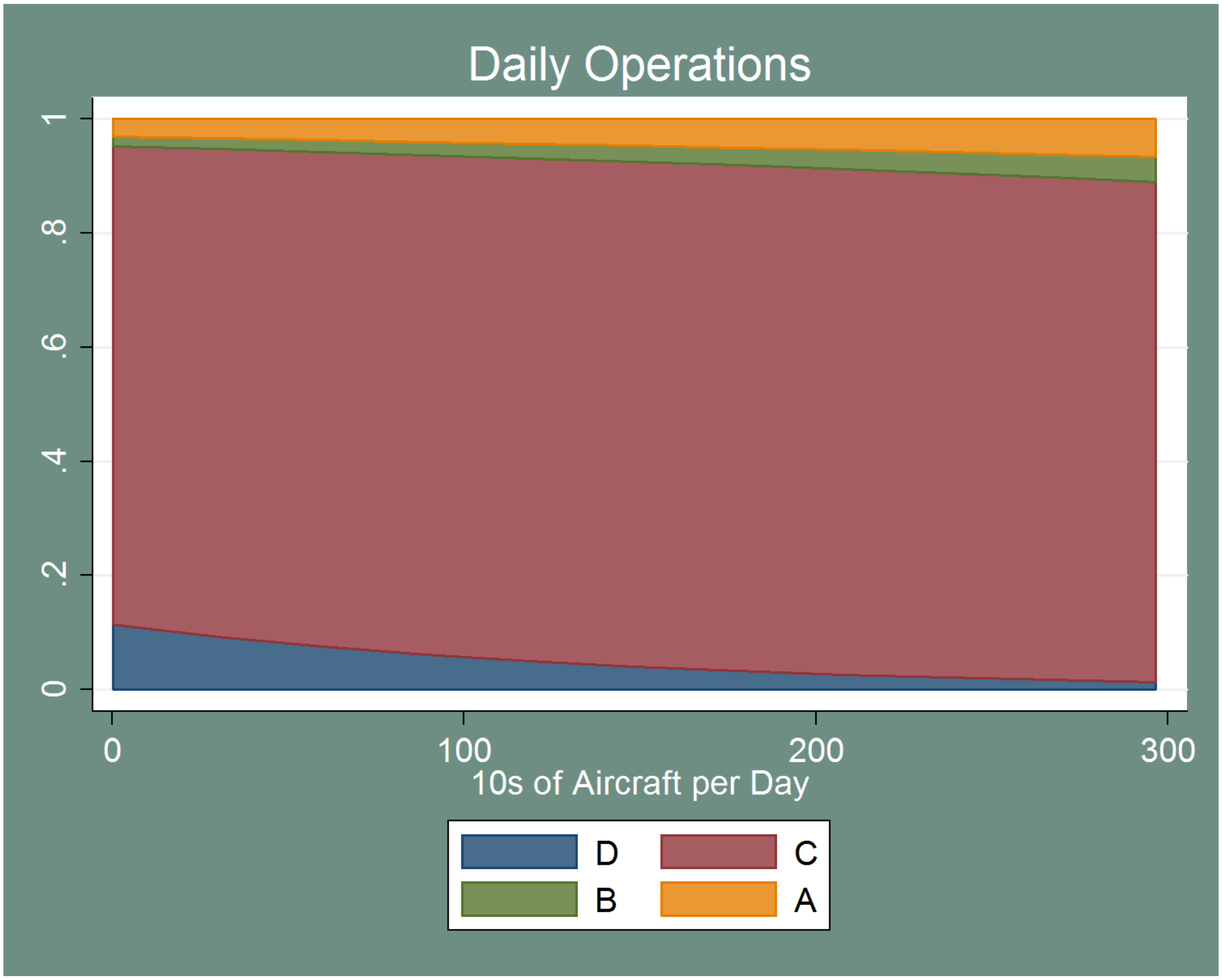

Figure 61 - Impact on Probability of Severity Categories of Daily Operations, Best Prediction Model

Increased daily operations have a similar impact to that seen in other models. This is encouraging and lends additional support to the idea that increased operations contribute to increased severity. The mechanism for this may be as simple as increasing the probability that two planes will be at the same runway at the same time. On the other hand increased operations may put additional strain on controllers and result in more severe errors that way. The truth is likely a mix of both, but this result indicates that busier airports are more likely to have more severe events than less busy airports.

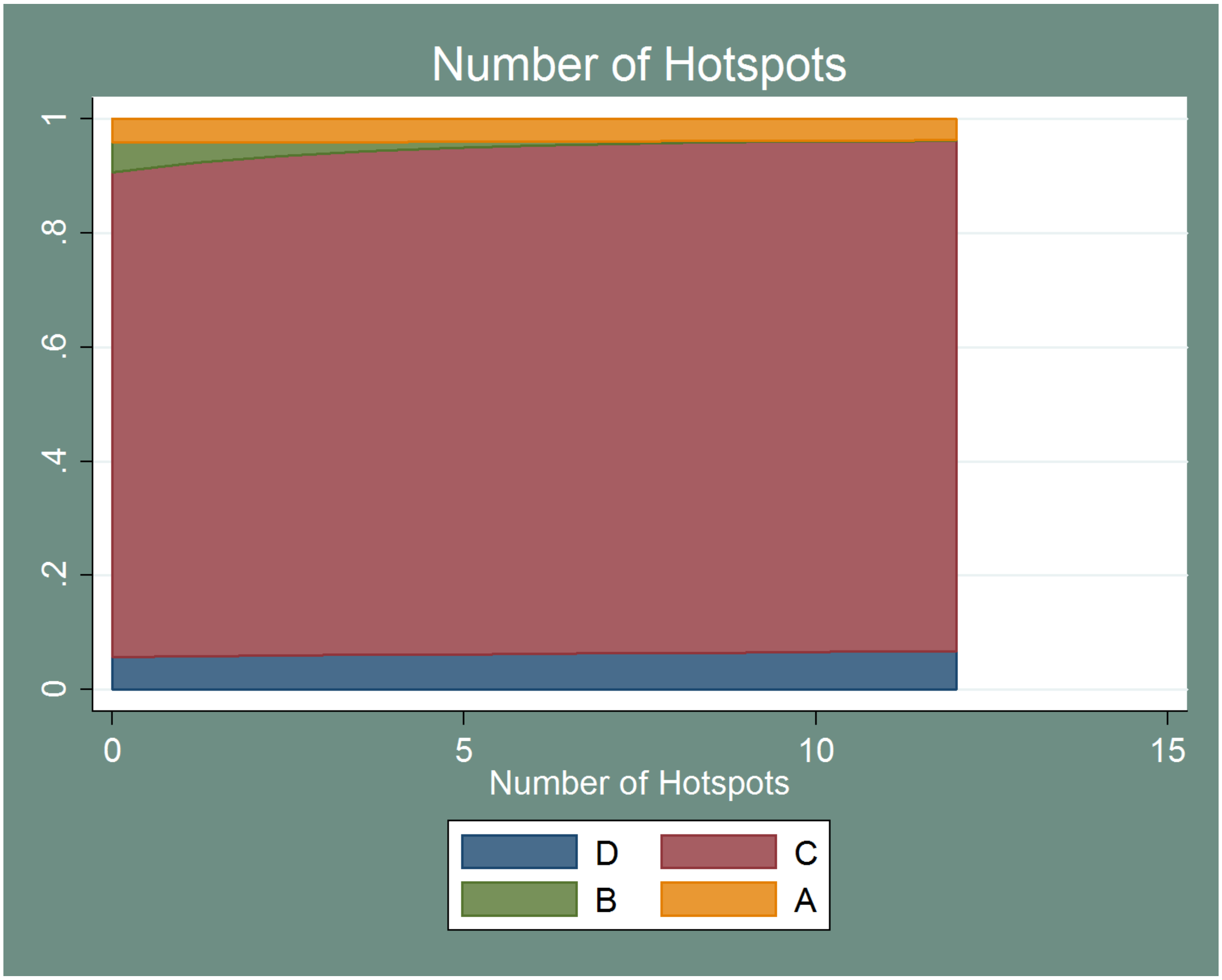

Figure 62 - Impact on Probability of Severity Categories of Number of Hotspots, Best Prediction Model

The number of hotspots at an airport has an almost identical effect to that seen in the airport model. This suggests that there is an effect here, rather than being an artifact of the data. The reduction in the probability of category B incursions is still surprising. The mechanism for this is unclear. In some sense, this is reducing severity, as the probability category A remains unchanged. A more focused look at the hotspot program and its impact on incursion severity could better understand the effect depicted above.

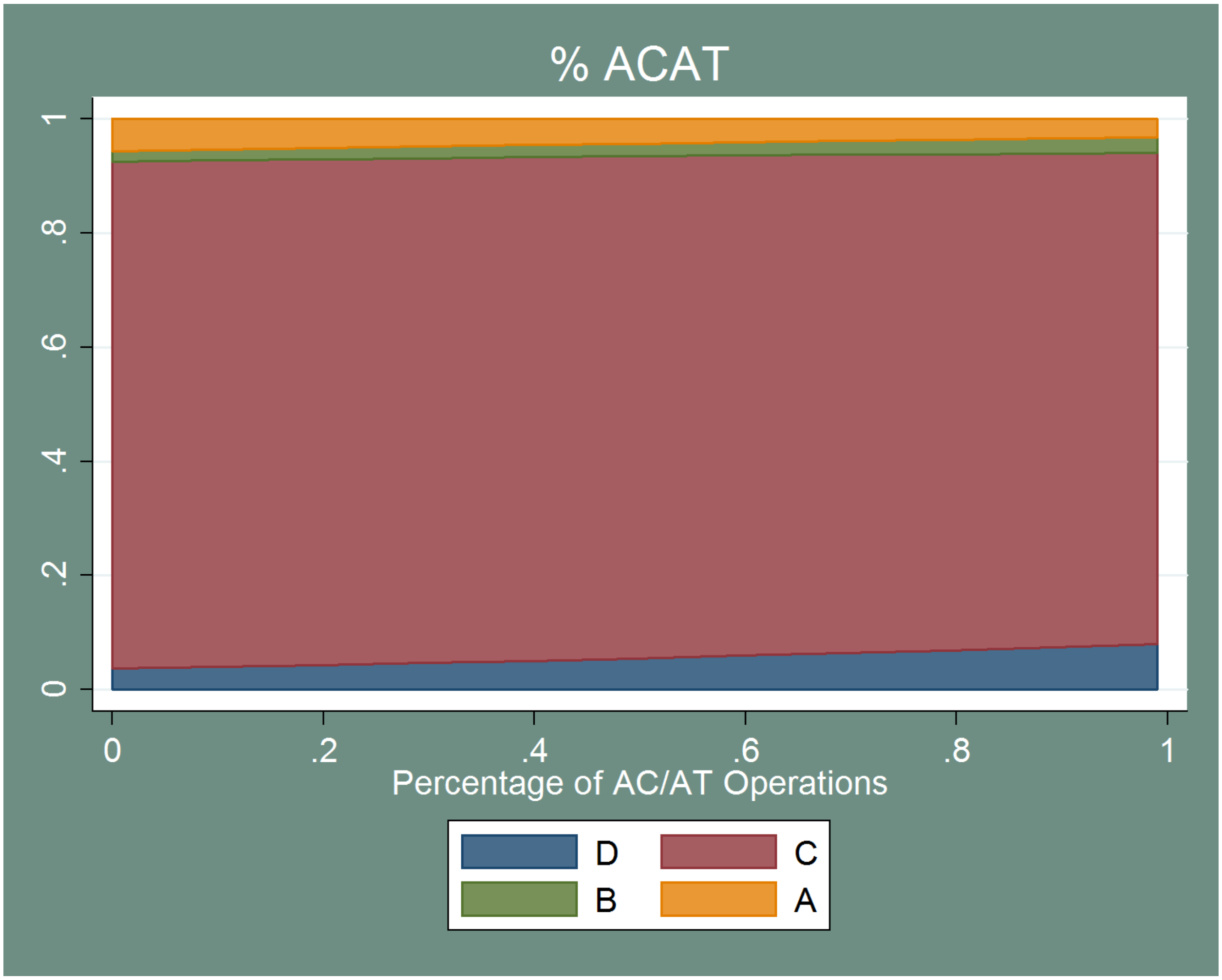

Figure 63 - Impact on Probability of Severity Categories of Percent AC/AT Traffic, Best Prediction Model

Percent of total traffic that is air carrier has a fairly weak effect. This is similar to the impact seen in the airport-specific model. The overall impact appears to be to reduce severity slightly. However, the variable is not significant at the 5% level for any of the severity categories. That the impact is approximately the same is nonetheless encouraging. This likely represents the airport-wide impacts of commercial carrier status. Commercial carrier status reduces the probability of severe categories for individual flights; it is not a stretch to assume that predominately-commercial airports might experience some larger reduction in severity.

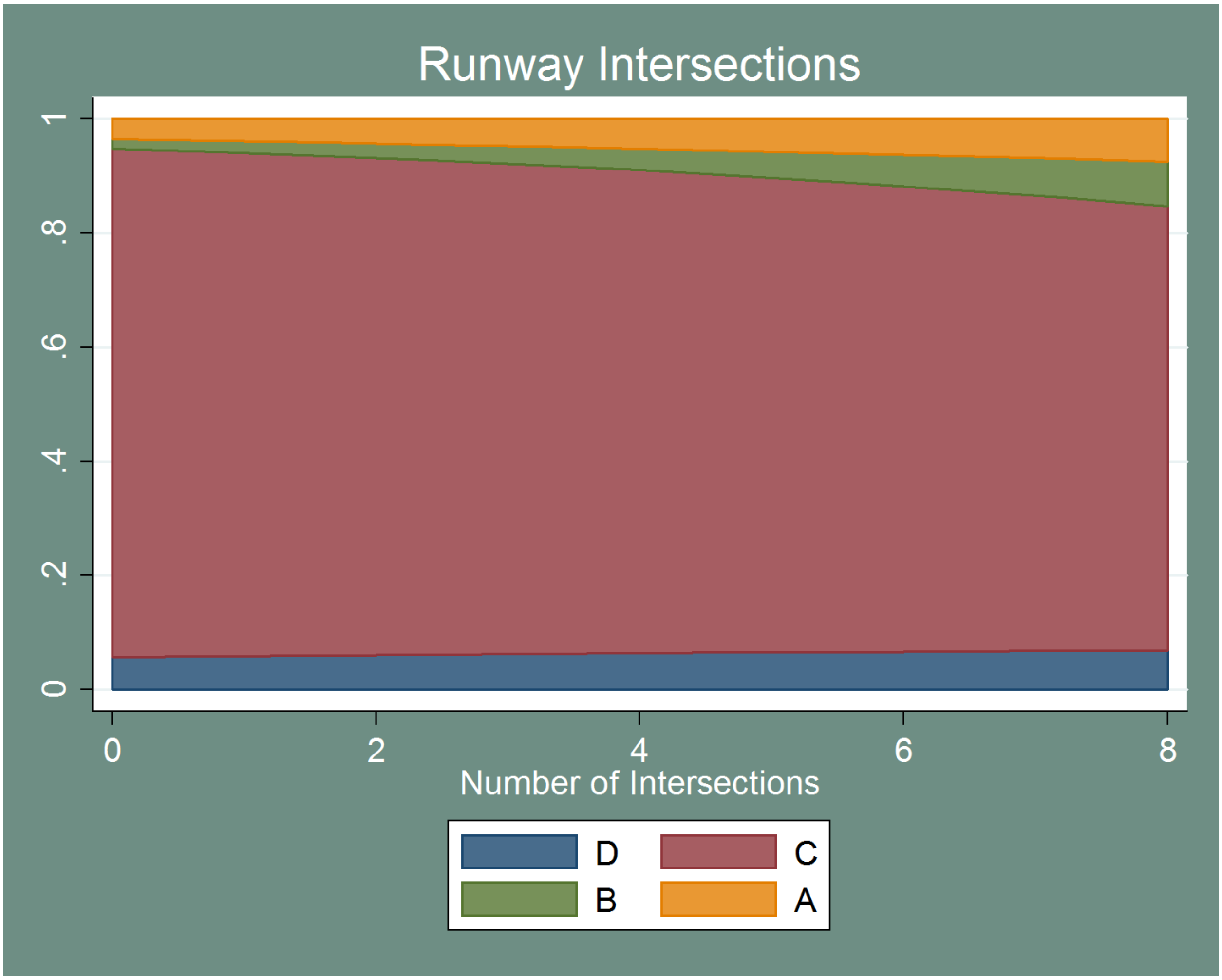

Figure 64 - Impact on Probability of Severity Categories of Number of Intersections, Best Prediction Model

Number of intersections has a similar impact to that seen in the airport model – increased intersections indicates a higher probability of a severe incursion. That many of the airport variables maintain their effect and significance hints that airport characteristics may play a role in incursion severity.

Overall, the best prediction model maintains many of the relationships seen in the constituent models. Commercial carrier status contributes to a higher probability of conflict events, but a lower probability of severe events. Takeoff is associated with more severe events. Increased runway intersections are also associated with higher probabilities of severe events. Hotspots present an interesting case and likely require additional research to understand the nature of their impact. Lastly, a final warning against overfit is warranted. The goal of this model was to generate a model with the best fit for the data rather than a true understanding of the causes of severity. It is promising that the conclusions of the separate models hold through into this combined model. However, caution should be used when using this model to make generalized statements.

5. CONCLUSIONS AND NEXT STEPS

5.1. What have we learned?

The research described in this report covered many different aspects of runway incursions. While some of the results were inconclusive, many provide specific steps for further research. Some of the insights were relevant to incident type distribution as well as severity. While not a central objective of this research, these conclusions regarding incident type can provide valuable insights. A summary of the results is contained in Summary of Modeling Results.

One major conclusion is that **OE incidents tend to be more severe than other incident types**. The reasons for this are unclear at the moment. It is potentially a product of the nature of controller errors. Alternatively, current training practices may already be effective in preventing category D OE incursions. This disparity in incident severity has policy implications: policy directed at a particular incident type will *not* impact severity uniformly.

Another strong conclusion is related to the regional distribution of incident type and severity. Both **incident type and severity vary systematically between regions**. Differences in pilot populations as well as traffic levels may impact this. This also indicates that any policy action will have disparate effects between regions, and that must be taken into account when crafting policy goals and responses.

Commercial carrier status also has a clear impact on severity. **Commercial carriers tend to be involved in less severe incidents**. Despite this lower severity overall, commercial carriers are more likely to be involved in conflict events, potentially due to operating at busier airports. However, once the conflict versus non-conflict dynamic has been controlled for, commercial carriers are less likely to be involved in severe incidents. This relationship holds true even for OE incidents – suggesting that pilot skill and experience may play a role even when they are not responsible for the error.

The phase of flight during which the incursion occurs appears to impact severity as well. **Incursions during takeoff appear to be more likely to be severe than those when the aircraft at fault is taxiing or landing**, once other controlling variables are included.

The preliminary models described in this report also give **no indication that controller age impacts incident severity**. It may be that there is no effect of age or it may be that the increased experience associated with increased age counteracts any impacts. While still preliminary, it is encouraging that there is no effect, and the results do not suggest a change to current policies surrounding controller age.

Controller workload – the number of aircraft a controller is responsible for – plays a significant role in severity. **Increased workload is associated with higher probabilities of severe events**. The positive relationship between severity and workload conforms to current expectations. Nonetheless, it is helpful to quantify that relationship – five additional aircraft increase the probability of a category A incursion by approximately .03 – and to provide statistical evidence to support intuition.

Airport layout also appears to influence severity; this is intuitive, but with the analysis offers tangible statistical support. Evidence indicates that more runway intersections are associated with higher probabilities for severe events. There is also evidence that more runways (for any fixed number of intersections) help reduce the probability of severe events. These two results combined indicate that **more parallel (or at least non-intersecting) runways may be a way to reduce the likelihood of severe events**.

In general, **it does not appear that radar systems play a role in severity**. There is some marginal evidence that STARS may help reduce severity, but it is tentative, at best. It is possible that radar still helps reduce the *rate* of runway incursions; however, such frequency models could not be run with the data provided for this study.

Increased daily operations appear to increase the likelihood of conflict events, but do not affect severity. It is unclear from this study why this is the case, but a likely hypothesis is that there is an increased chance for an interaction between two aircraft as operations increase, increasing the likelihood of any given incursion being a conflict event. Again, this variable likely has some role in the frequency of runway incursions, and further study is required to understand the total impact of this variable.

Finally, while not based on statistical tests, there are a series of observations about the general distribution of variables and severity that are informative. These insights may not be as useful from a policy perspective, but provide a richer context for understanding incursion severity:

- Pilot incursions are the most common type of incursion – occurring more than four times as often as controller errors and approximately twice as often as V/PDs.
- Incursions during LAHSO are very rare.
- No severe incidents (category A or B) have occurred during a LAHSO.
- No pilot having more than 5,000 hours in a make and model has committed a severe incursion.

5.2. Further Research Ideas

In addition to the conclusions outlined in the paper and mentioned above, several questions arose during the research process. These questions fall into two major categories. The first category is additional research on how variables impact severity. The second category is extensions of this research into other areas.

Throughout the paper, specific variables that would benefit from particular follow-up were identified. These can serve as springboards for more focused research into how severity may be impacted, likely with a combination of additional data, statistical analysis, and support from human factors and other safety experts. A full list of these variables and topics can be found in Appendix D: Future Research. Beyond those specific variables, there are three major classes of variables that require in-depth study.

First, and perhaps the largest group, are pilot variables. Some of the pilot variables were addressed through cross tabulations. However, a modeling effort focused on pilot variables and PD incidents would be beneficial. The results in this paper pertain mostly to OE incidents, which represent the smallest absolute number of incidents. A better understanding of pilot incidents would help in minimizing the impact of that category. A similar suggestion holds true for V/PD incidents, although information surrounding them is less available. This suggestion cross-cuts all the below suggestions as they too may vary by type of incident.

Secondly, the weather variables warrant further examination. It is clear that there is some relationship between severity and weather conditions, but it is unclear what the specific causes are. Likely, that relationship is being driven by underlying behavioral responses to weather rather than the weather itself. This preliminary research identified a need to understand these variables and a more focused examination might better explain their impact.

Thirdly, the controller variables require a more thorough examination. The controller variables in particular were plagued by data problems and small sample sizes.[100] It is surprising that *no* controller attributes contributed to the severity of controller incursions. A more focused examination – perhaps using more accurate controller data – might reveal some trends.

Aside from specific variables to follow up on, another fruitful area for research would be frequency modeling. As noted throughout the research, these insights pertain only to severity, *given that an incursion has already occurred*. It is possible that many of these variables contribute to the underlying rate of incursions, but not their severity. Additionally, some variables may impact *both* severity and frequency. Frequency and severity are two sides of the same safety issue. To gain a complete understanding of the problem, frequency must also be modeled. **This is the most beneficial next step,** even if that frequency research were focused only on OE incursions.

5.3. Clarification of the Rating of Runway Incursions

A final word on the ranking system is warranted. In addition to information about factors influencing severity, the deep scrutiny of the ranking system provided insights how incursion severity is ranked.

Throughout the results, there were often disparate impacts between conflict and non-conflict events. Factors such as commercial carrier status showed no impact for all severity categories, but when excluding category D events, a relationship with severity emerged. Moreover, only one of the ordered models satisfied the assumptions underlying the ordered logit model – *and that model excluded category D events*.[101] Additionally, the multinomial models reveal that some variables explain category D incursions but *none* of the conflict categories.

All this evidence combines to suggest that category D incursions are a distinct group from the remaining three categories. Furthermore, there is evidence that category D incursions do not follow a smooth ordering with the other three categories. This has implications for any modeling effort that chooses to focus on all severity categories. Any model will need to account for a "two-stage" process, distinguishing between conflict and non-conflict and then attempting to identify severity.

---

100 See Appendix B: Data Issues for a full list of problems identified in the data.

101 Strictly speaking, the ordered model assumes that the impact is the same across categories (rather than testing that the particular order of categories is important).

This also has implications for understanding the danger posed by any given event. If category D events are not part of a smooth ranking, the current system may not be properly capturing the risk inherent in some category D events. A simplistic example is when an aircraft lands on a runway without clearance or communication with the tower. If another aircraft is present on the runway, the incursion would likely be an A or B; however, if the airport is otherwise empty the same pilot error would be rated a D. This bears serious consideration as, in this example, the behavior in question is inherently quite risky and *likely would be a serious event* in the presence of other aircraft, someone the pilot could not possibly control.

While not informed by rigorous research, the results of this effort would imply that incursion severity is truly (at least) a two-stage process. The first stages relates to the riskiness of a behavior (landing on a closed runway or forgetting an aircraft on the airfield versus stopping one foot past a hold-short line or giving a clearance to cross a runway to a non-existent flight number). The second stage relates to the likelihood that another aircraft will be nearby when the incident occurs. That is, the axes would be the riskiness of the behavior and the possibility of a conflict.

Changes to the ranking system would require significant involvement by many players at the FAA and ICAO, but such coordination may offer a considerable benefit in an effort to respond to safety risks. Those interested in using data on incursions to reduce the likelihood of future collisions need to take a serious look at how to best classify incursions along however many axes of risk are most appropriate to model.

APPENDIX A: RUNWAY INCURSION DEFINITION

This is an excerpt from the Manual on the Prevention of Runway Incursions, First Edition.[102] It is reproduced unedited from that document.

102 International Civil Aviation Organization (2007).

Chapter 6

CLASSIFICATION OF THE SEVERITY OF RUNWAY INCURSIONS

6.1 SEVERITY CLASSIFICATION

6.1.1 The objective of runway incursion severity classification is to produce and record an assessment of each runway incursion. This is a critical component of risk measurement, where risk is a function of the severity of the outcome and the probability of recurrence. Whatever the severity of the occurrence, however, all runway incursions should be adequately investigated to determine the causal and contributory factors and to ensure risk mitigation measures are implemented to prevent any recurrence.

6.1.2 Severity classification of runway incursions should be assessed as soon as possible after the incident notification with due regard for the information required in 6.2. A reassessment of the final outcome may be applied at the end of the investigation process.

6.1.3 For the purpose of global harmonization and effective data sharing, when classifying the severity of runway incursions, the severity classification scheme in Table 6-1 should be applied. See Figure 6-1 for examples of severity classification.

Table 6-1. Severity classification scheme

Severity classification	Description*
A	A serious incident in which a collision is narrowly avoided.
B	An incident in which separation decreases and there is significant potential for collision, which may result in a time-critical corrective/evasive response to avoid a collision.
C	An incident characterized by ample time and/or distance to avoid a collision.
D	An incident that meets the definition of runway incursion such as the incorrect presence of a single vehicle, person or aircraft on the protected area of a surface designated for the landing and take-off of aircraft but with no immediate safety consequences.
E	Insufficient information or inconclusive or conflicting evidence precludes a severity assessment.

* Refer to Annex 13 for the definition of "incident"

6.2 FACTORS THAT INFLUENCE SEVERITY

To properly classify the severity of a runway incursion the following information is required:

a) *Proximity of the aircraft and/or vehicle*. This distance is usually approximated by the controller or from the aerodrome diagram. When an aircraft flies directly over another aircraft or vehicle, then the closest vertical proximity should be used. When both aircraft are on the ground, the proximity that is used to classify the severity of the runway incursion is the closest horizontal proximity. When aircraft are separated in both horizontal and vertical planes, the proximity that best represents the probability of collision should be used. In incidents in which the aircraft are on intersecting runways, the distance from each aircraft to the intersection is used.

b) *Geometry of the encounter*. Certain encounters are inherently more severe than others. For example, encounters with two aircraft on the same runway are more severe than incidents with one aircraft on the runway and one aircraft approaching the runway. Similarly, head-on encounters are more severe than aircraft moving in the same direction.

c) *Evasive or corrective action.* When the pilot of an aircraft takes evasive action to avoid a collision, the magnitude of the manoeuvre is an important consideration in classifying the severity. This includes, but is not limited to, hard braking action, swerving, rejected take-off, early rotation on take-off, and go-around. The more severe the manoeuvre, the higher its contribution to the severity rating. For example, encounters involving a rejected take-off in which the distance rolled is 300 metres are more severe than those in which the distance rolled is less than 30 metres.

d) *Available reaction time.* Encounters that allow the pilot little time to react to avoid a collision are more severe than encounters in which the pilot has ample time to respond. For example, in incidents involving a go-around, the approach speed of the aircraft and the distance to the runway at which the go-around was initiated needs to be considered in the severity classification. This means that an incident involving a heavy aircraft aborting the landing and initiating a go-around at the runway threshold is more severe than one that involves a light aircraft initiating a go-around on a one-mile final.

e) *Environmental conditions, weather, visibility and surface conditions.* Conditions that degrade the quality of the visual information available to the pilot and controller, such as poor visibility, increase the variability of the pilot and controller response and, as such, may increase the severity of the incursion. Similarly, conditions that degrade the stopping performance of the aircraft or vehicle, such as wet or icy runways, should also be considered.

f) *Factors that affect system performance*. Factors that affect system performance, such as communication failures (e.g. “open mike”) and communication errors (e.g. the controller’s failure to correct an error in the pilot’s readback), also contribute to the severity of the incident.

6.3 RUNWAY INCURSION SEVERITY CLASSIFICATION CALCULATOR

A runway incursion severity classification (RISC) calculator is available on CD (see Appendix H for a description). The calculator was developed to assist States in assessing the severity of runway incursion events. Use of the RISC calculator should also enable a consistent assessment to be made. Alternatively, the severity of runway incursions can be classified manually using the guidance contained in 6.1 and 6.2.

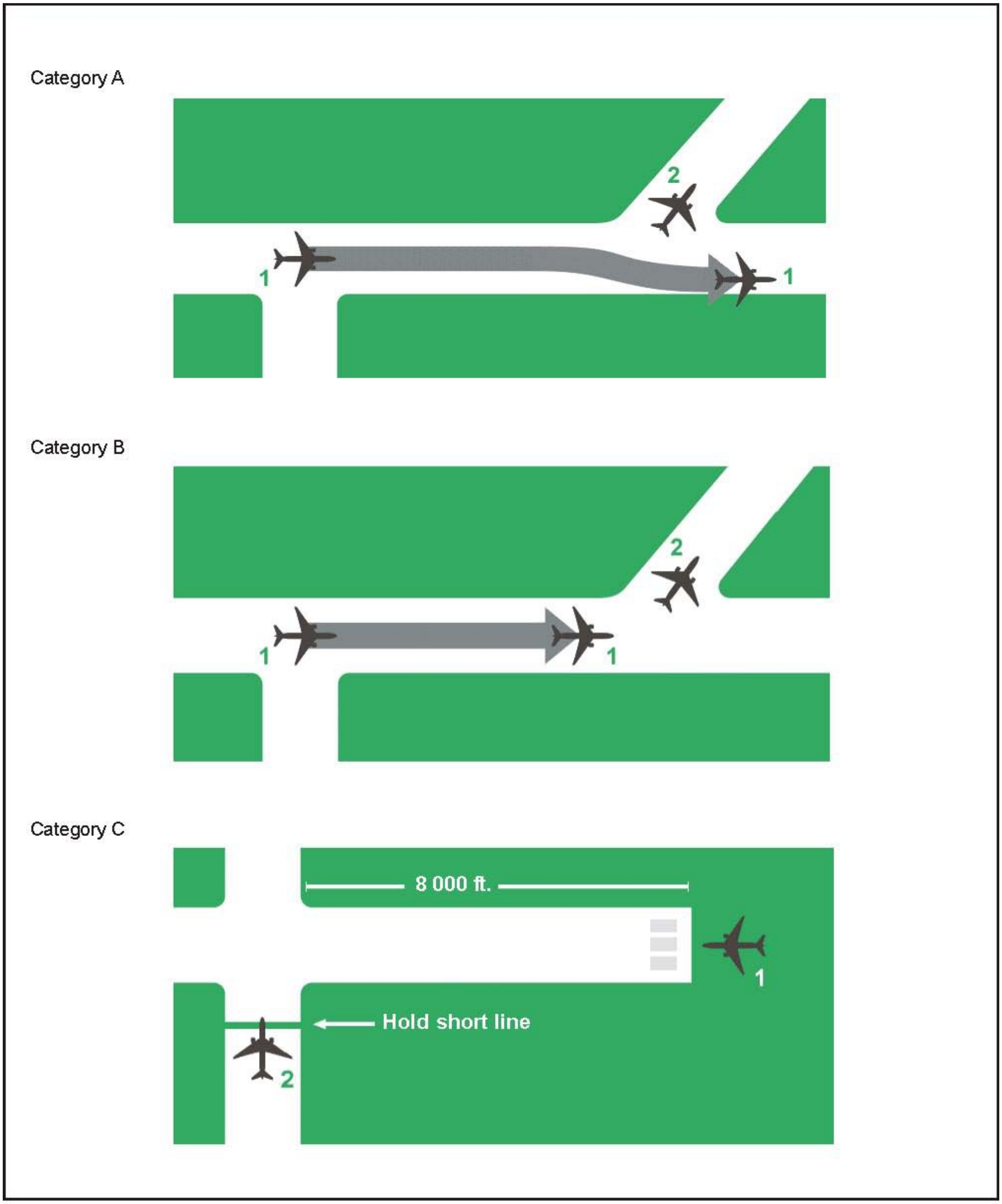

Figure 6-1. Severity classification examples

APPENDIX B: DATA ISSUES

This appendix contains the output from the "Issue Tracker" maintained as part of the data cleaning process.

ID	Variable Name	Issue	Comments	
1	all	Missing Airport: CGI		
2	all	Missing Airport: ESN		
3	all	Missing Airport: FTG		
4	all	Missing Airport: GKY		
5	all	Missing Airport: GLS		
6	all	Missing Airport: GTR		
7	all	Missing Airport: GYH		
8	all	Missing Airport: HXD	only incident has no time associated	o
9	all	Missing Airport: JST		
10	all	Missing Airport: MER		
11	all	Missing Airport: MYV		
12	all	Missing Airport: OMN	note, airport closed	
13	all	Missing Airport: SPG		
14	all	Missing Airport: UCA		
15	all	Missing Airport: VBG		
16	all	Missing Airport: VCV		
17	ALL	definitions of the variables	Many of the variables, while having a plain-text definition, do not indicate how they were gathered (e.g., What is a "short taxi"), over how long (e.g., average rainy days)	Requ
18	geo_bullseye_flag	definitions of the variables		int
19	geo_mult_rwy_crossing_flag	definitions of the variables		Does p

ID	Variable Name	Issue	Comments	
				rwy
20	geo_num_hotspot	definitions of the variables		This ir diag
21	geo_num_rsa_isect	definitions of the variables		Use d inters
22	geo_num_taxi_x_runway	definitions of the variables		In this taxiw
23	geo_rwy_close_flag	definitions of the variables		Aga crite applic are NC a pos I'm a but I en H criteria les bot
24	geo_rwy_crossing_flag	definitions of the variables		Does get to the onl m
25	geo_rwy_num_isect	definitions of the variables		
26	geo_rwy_num_t_isect	definitions of the variables		
27	geo_rwy_parallel_flag	definitions of the variables		are the airpor rw
28	geo_taxi_short_flag	definitions of the variables		Cou Sug air flagge the
29	lahso_flag_ap	definitions of the variables		

ID	Variable Name	Issue	Comments	
30	locid	definitions of the variables		
31	runway information	KWA	need to manually pull	
32	runway information	PFN	need to manually pull	
33	runway information	OMN	need to manually pull (NOTE: 5010 contains information for an airport that has taken over the code)	
34	traffic variables	One airport sums to 111%	operations percenages at Kalaeloa Airport (JRF) sum to 111%	
35	weather variables	combined weather stations	Many airports in a region seem to share weather data, even if they are not correct (e.g., Hyannis, MA and Lawrance, MA)	
36		15 airports not included in dataset that are in RI		
37	acft_evas_actn_code	27 missing values	Replaced missing with unknown	
38	acft_gnd_spd_kt_qty	1133 missing values	codebook (but not in data) contains ACFT_GND_SPD_UNKN_FLAG, is this "Y" for these 1133?	w
39	acft_model_desc	odd naming conventions	used in forming aircraft groups. No longer needed	

ID	Variable Name	Issue	Comments	
			in database. Keeping issue open to grouping is completed	
40	acft_obstn_code	Missing values and "0" coded values	What does missing mean? Also, the codebook has no definition of "0" but it appears 174 times. Note: of the zeroes, 18 have descriptions of obstructions	th "c
41	acft_phase_code	17 missing values	12 missing do not have descriptions. The variable in general agrees with phaseofflight. Letting the 12 missing stay missing.	
42	acft_tcas_equip_code	21 missing values		
43				
44	ctlr_actn_contem_code	Incorrect coding	one value coded as "X", presumed to be a "Y"	
45	ctlr_actn_contem_code	missing	440 missing	
46	ctlr_actn_taken_code	missing	10 missing, same as ctlr_alert_cod e	
47	ctlr_alert_code	missing	6 missing do not have description. Demographic only variable, letting the 6 be missing	

ID	Variable Name	Issue	Comments	
48	ctlr_alert_otr_desc	missing	1336 missing, equal to those in ctlr_alert_code that are not "other"	
49	ctlr_area_spl_code	unclear meanings/codings	It's unclear what this variable is capturing in the first place, as sometimes this appears to list facilities (tower), locations (southwest), positions (LC1), areas (Area 7), or just a single number (6). There are 144 unique values in this field (though some are clearly the same with different abbreviations)	Da
50	ctlr_asst_req_flag	Y,N,Missing	27 Y, 12 missing, 1465 N	
51	ctlr_aware_dvlp_flag	Y,N,Missing	258 Y, 10 Missing, 1236 N	
52	ctlr_birth_date	Missing	479 Missing	199 m may h samp value ar
53	ctlr_certif_date	Missing	457 missing, a few wrong entires (7 years before birth, 2 years after birth). No entires	185 m s

ID	Variable Name	Issue	Comments	
			prior to 1980, many certified 40+ years after birth	
54	ctlr_certif_type_ncode_oe	unclear meanings/codings	initial vs recerficiation?	"Contr In pr oper posi v event c pr
55	ctlr_contrib_code	result of one-to-many merge	appears to have been a one-to-many merge from OE events to contributing factor codes, as a result, events appear in the db multiple times (i.e., once for each contributing code). We are not using the code in our model, so have dropped the field and removed duplicate lines	
56	ctlr_contrib_prev_30mo_qty	Missing	This is mutually exclusive with ctlr_prim_prev_30_mo_qty_oe. Treat missing as missing	
57	ctlr_curr_shft_end_time	Missing	456 missing	184 h

ID	Variable Name	Issue	Comments	
58	ctlr_curr_shft_start_time	Missing	456 missing, same records missing as end time	184 h
59	ctlr_dstrctn_flag	Y,N,Missing	200 Y, 15 Missing, 1289 N	
60	ctlr_fctr_med_certif_flag	Y,N,Missing	3 Y, 351 missing, 1150 N. may not be useful with so few Y	
61	ctlr_fpl_date	Missing	627 missing. The 4digit dates presented make no sense.	
62	ctlr_perl_code	Missing	449 missing	181 m
63	ctlr_prev_shft_end_time	Missing	598 missing. 454 missing current end time. 2 missing current but have previous. 144 miss previous but have current	297 m will h
64	ctlr_prev_shft_start_time	Missing	598 missing, same missing as prev end time. 454 missing current start time. 2 missing current but have previous. 144 miss previous but have current	297 m will h
65	ctlr_prim_prev_30mo_qty	missing	528 missing	259 m need to

ID	Variable Name	Issue	Comments	
66	ctlr_psn_comb_desc	Missing	27 Missing. May require additional parsing to collapse in a usable categorical variable. Currently 642 unqiue values	will us
67	ctlr_psn_fnctn_otr_desc	Missing	1464 Missing. Will need to parse better. 27 unique values	loo positi
68	ctlr_psn_min_qty	Missing	440 Missing. Some large values (555 minutes)	177 m will l
69	ctlr_psn_otr_desc	Missing	80 missing	
70	ctlr_sctr_psn_code_oe	Missing. Additional Parsing	12 missing. Needs to be parsed further to be collapsed into small set'	
71	ctlr_trng_reltd_1yr_flag	Y,N,Missing	859 Y, 448 Missing, 197 N	180
72	ctlr_trng_reltd_desc	Missing. Additional Parsing	654 Missing. Lots of different answers. May need to parse to collapse into aseries of flags	unli m
73	ctlr_wrkld_acft_qty	Believability	Range from 0 to 35. are these reasonable? 90th% is 8	
74	ctlr_wrksked_desc	Missing. Additional Parsing	456 missing additional parsing required to turn into usable	This i

ID	Variable Name	Issue	Comments	
			variable	
75	event_alt_ft_qty_oe	795 missing values	Some rounded to nearest 500 feet particularly at values 5000 and under. Since these should all be ocurring at ground level, what does it mean to have altitude of >0 ?	b relev
76	event_asp_code_oe	92 missing values		In relev surfac
77	event_asp_otr_desc_oe	1488 coded "missing" values	Missing values, but all accounted for. Missing for event_asp_co de_oe equal to "Other".	
78	event_cat flags	Missing, non mutually exclusive	replaced missing to no	ne releva have a is a
79	event_cat_atcs_flag_oe	79 missing values	replaced missings with no	che check cont the repla
80	event_cat_human_flag	variable name	replaced missings to no	che check cont the repla
81	event_class_code	missing	71 missing values	D/E/m
82	event_class_code	No codebook for values	the codebook does not have any explanation of event classes,	D/E/m

ID	Variable Name	Issue	Comments	
			are these incursion ratings?	
83	EVENT_RI_FLAG	Reliability	How reliably does this field indicate Ris? 994/1086 of the "Y"s match to Ris. 418/1504 are missing	(FROI prelim fiel source
84	fac_atc_ctl_code_oe	2 missing values		It is r "Rac
85	fac_class_code	missing, codebook misreported	108 missing values. 52 missing on relevant set. the codebook appears to have values for fac_type_code rather than fac_class_code, as a result, it is unclear wht the codes (1-14, plus ATC-7) mean. Appears to be for pay levels (based on traffic/complexity)	Need FAA F belie over ti the co com busy, r
86	fac_cnflct_alert_code_oe	43 missing values	22 missing on relevant set	22 r hav
87	fac_equip_layout_flag_oe	4 missing values	1 missing on relevant set	c
88	fac_equip_unsatfy_flag_oe	5 missing values	1 missing on relevant set. This observation also missing on fac_equip_lay	c

ID	Variable Name	Issue	Comments	
			out_flag	
89	fac_id_code	missing	2 missing on relevant set. Should not use over RI data.	
90	fac_primary_code	No codebook for values	the coebook does not have any explanation of this code, all 1504 obesrvatrions (no missing) are "P"	
91	FAC_RADAR_ACDS_FLAG	Variable in codebook but not dataset		need
92	FAC_RADAR_AMASS_FLAG	Variable in codebook but not dataset		nee
93	FAC_RADAR_ARTS_FLAG	Variable in codebook but not dataset		need
94	fac_radar_artsii_flag_oe	Value of "Y" and missing	Assume that missings are "N"s, but it is possible that some records may be missing for other reasons.	
95	fac_radar_artsiia_flag_oe	1,202 missing. Value of "Y", "N", and missing	Unclear on difference between N and missing	
96	fac_radar_artsiiia_flag_oe	1,156 missing. Value of "Y", "N", and missing	Unclear on difference between N and missing	
97	fac_radar_artsiiie_flag_oe	867 missing. Value of "Y", "N", and missing	Unclear on difference between N and missing	
98	fac_radar_asdeii_flag_oe	1192 missing. Value of "Y", "N", and missing	Unclear on difference between N and missing	
99	fac_radar_asdeiii_flag_oe	1007 missing. Value of "Y", "N", and missing	Unclear on difference	

ID	Variable Name	Issue	Comments	
			between N and missing	
100	FAC_RADAR_ASDEX_FLAG	Variable in codebook but not dataset		
101	FAC_RADAR_ASR11_FLAG	Variable in codebook but not dataset		
102	fac_radar_asr9_flag_oe	639 missing. Value of "Y", "N", and missing	Unclear on difference between N and missing	rec
103	fac_radar_brd_band_flag_oe	Variable in dataset but not codebook, 1194 missing, others are Y, N		
104	fac_radar_britiv_flag_oe	1189 missing. Value of "Y", "N", and missing	Unclear on difference between N and missing	verify
105	fac_radar_cenrap_flag_oe	1,202 missing. Value of "Y", "N", and missing. 1 Y.	Unclear on difference between N and missing	
106	fac_radar_darc_flag_oe	Variable in dataset but not codebook. 1,119 missing. Value of "Y", "N", and missing.	Unclear on difference between N and missing	nev
107	fac_radar_dbrite_flag_oe	898 missing. Value of "Y", "N", and missing	Unclear on difference between N and missing	verify
108	fac_radar_earts_flag_oe	1,188 missing. Value of "Y", "N", and missing	Unclear on difference between N and missing	
109	FAC_RADAR_EBUS_HOST_FLAG	Variable in codebook but not dataset		
110	FAC_RADAR_HOST_FLAG	Variable in codebook but not dataset		
111	FAC_RADAR_MODEL1_FLAG	Variable in codebook but not dataset		
112	fac_radar_modes_flag_oe	964 missing. Value of "Y", "N", and missing	Unclear on difference between N and missing	what (m
113	fac_radar_nrw_band_flag_oe	Variable in dataset but not codebook. 1,200 missing. Value of "Y", "N", and missing.		
114	FAC_RADAR_OASIS_FLAG	Variable in codebook but not dataset		

ID	Variable Name	Issue	Comments	
115	fac_radar_otr_desc_oe	1,155 missing values. Inconsistent entiries.	Missing values, but the 349 yes for fac_radar_otr_flag_oe is equivilent to the number of desc entries	v
116	fac_radar_otr_flag_oe	956 missing. Value of "Y", "N", and missing	Unclear on difference between N and missing	
117	FAC_RADAR_STARS_FLAG	Variable in codebook but not dataset		
118	fac_radar_trsn_desc_oe	1503 missing.	Missing values, but consistent with 1 "Y" in fac_radar_trsn_flag_oe	
119	fac_radar_trsn_flag_oe	18 missing values.	1 Y	
120	FAC_RADAR_URET_FLAG	Variable in codebook but not dataset		
121	fac_rgn_code	No codebook for values	Found an alternate site with codings	
122	fac_rgn_code	missing values	variable agrees with reg_n and duplicates values. Use reg_n for region information	
123	fac_type_code	missing	1202 missing values	all R
124	fctr_cmplx_asp_flag_oe	1,160 missing. Value of "Y", "N", and missing	Unclear on difference between N and missing	
125	fctr_cmplx_emerg_flag_oe	1,184 missing. Value of "Y", "N", and missing	Unclear on difference between N and missing	
126	fctr_cmplx_expr_flag_oe	1,011 missing. Value of "Y", "N", and missing	Unclear on difference between N	

ID	Variable Name	Issue	Comments	
			and missing	
127	fctr_cmplx_flow_ctl_flag_oe	1,166 missing. Value of "Y", "N", and missing	Unclear on difference between N and missing	
128	fctr_cmplx_na_flag_oe	Variable in dataset but not codebook. 1,202 missing. Value of "Y", "N", and missing	Unclear on difference between N and missing	
129	fctr_cmplx_nbr_acft_flag_oe	632 missing. Value of "Y", "N", and missing	Unclear on difference between N and missing	
130	fctr_cmplx_otr_desc_oe	1,097 missing. All values unique descriptions.	Missing values, but more than for 614 coded as "other" in fctr_cmplx_otr_flag_oe equal to "Other".	
131	fctr_cmplx_otr_flag_oe	867 missing. Value of "Y", "N", and missing	Unclear on difference between N and missing	
132	fctr_cmplx_rwy_cond_flag_oe	1,164 missing. Value of "Y", "N", and missing	Unclear on difference between N and missing	
133	fctr_cmplx_rwy_config_flag_oe	875 missing. Value of "Y", "N", and missing	Unclear on difference between N and missing	
134	fctr_cmplx_spl_event_flag_oe	1,181 missing. Value of "Y", "N", and missing	Unclear on difference between N and missing	
135	fctr_cmplx_trrn_flag_oe	1,196 missing. Value of "Y", "N", and missing	Unclear on difference between N and missing	
136	fctr_cmplx_wx_flag_oe	1,089 missing. Value of "Y", "N", and missing	Unclear on difference between N and missing	

ID	Variable Name	Issue	Comments	
137	fctr_com_err_flag_oe	missing	116 Y, 569 N, 363 missing on relevant set	Fact missi aggre ou
138	fctr_com_otr_desc	inconsistent with flag	Some non-missing values are Y, some are N on fctr_com_err_flag. 63 non-missing values	Fact missi aggre ou
139	fctr_comptr_entry_flag	Missing	31 Y, 756 Missing, 261 N on relevant set	Fact missi aggre ou
140	fctr_comptr_otr_desc	inconsistent with flag	1039 missing, 9 non-missing. All non-missing are missing on fctr_comptr_entry_flag	Fact missi aggre ou
141	fctr_coord_flag_oe	Missing	88 Y, 707 N, 253 Missing on relevant set	Fact missi aggre ou
142	fctr_coord_gnd_lcl_flag_oe	Missing	73 Y, 245 N, 730 Missing	Fact missi aggre ou
143	fctr_coord_gnd_lcl_otr_desc	inconsistent with flag	45 non-missing values. 35 are missing on flag, 10 are Y	Fact missi aggre ou
144	fctr_data_post_flag	missing	39 Y, 85Missing, 924 N on	Fact missi

ID	Variable Name	Issue	Comments	
			relevant set	aggr ou
145	fctr_flight_strip_flag	Missing	38 Y, 762 missing, 248 N	Fact missi aggr ou
146	fctr_flight_strip_otr_desc	inconsistent with flag	16 non-missing values. 13 missing on flag, rest are Y	Fact missi aggr ou
147	fctr_gnd_opn_flag	Missing	144 Y, 673 Missing, 231 N	Fact missi aggr ou
148	fctr_inapp_disp_flag	missing	40 Y, 762 Missing, 246 N	Fact missi aggr ou
149	fctr_inapp_disp_otr_desc	inconsistent with flag	9 non-missing values. 2 Y, 7 missing on flag	Fact missi aggr ou
150	fctr_incdnt_area_flag	missing	48 Y, 729 Missing, 271 N	Fact missi aggr ou
151	fctr_info_exchg_flag_oe	missing	48 Y, 684 missing, 316 N	Fact missi aggr ou

ID	Variable Name	Issue	Comments	
152	fctr_info_exchg_otr_desc	missing	39 non-missing. 25 missing flag, 1 N, 13 Y	Fact missi aggr ou
153	fctr_misid_flag	missing	31 Y, 279 N, 738 Missing	Fact missi aggr ou
154	fctr_misid_otr_desc	inconsistent with flag	9 non-missing. 1 Y on flag, 8 missing	Fact missi aggr ou
155	fctr_psn_relf_brfg_flag_oe	missing	40 Y, 905 N, 103 Missing	Fact missi aggr ou
156	fctr_psn_relf_otr_desc	inconsistent with flag	13 non-missing. 2 Y on flag, rset missing	Fact missi aggr ou
157	fctr_radar_disp_flag	missing	42 Y, 123 Missing, 883 N	Fact missi aggr ou
158	fctr_rdbk_flag	Missing	98 Y, 713 Missing, 237 N	Fact missi aggr ou
159	fctr_rdbk_otr_desc	inconsistent with flag	110 non-missing values. 2 N on flag, 23 Y, rest missing	Fact missi aggr ou

ID	Variable Name	Issue	Comments	
160	fctr_trng_flag	missing	95 Y, 3 missing, 950 N. Likely useable	Fact missi aggre ou
161	fctr_visl_data_flag	missing	111 Y, 231 N, 706 Missing	Fact missi aggre ou
162	fctr_visl_data_otr_desc	inconsistent with flag	171 non-missing values. 23 Y on flag, 17 N on flag, rest missing	Fact missi aggre ou
163	fctr_wx_flag_oe	missing	41 Y, 4 Missing, 1003 N	Fact missi aggre ou
164	loc_drctn_deg_qty	Missing	4 unique vlaues. 27 0s, 1 6, 1 140, 2 180s. Cluster on 0 makes little sense. 1017 missing	
165	loc_dstc_nm_qty	Missing	769 missing, 275 0s. 2 1s. 2 2s. Seems like that makes sense for ground incidents	
166	loc_fix_code	missing	14 non missing vlaues. Mostly unique (2 instances of KFXE)	

ID	Variable Name	Issue	Comments	
167	loc_intxn_id_code	Inconsistent entires	Field needs to be parsed to be useable. Lots of unique entires. Not filled out in conjuntion with rwy and twy code.	These t quite muc with co
168	loc_rwy_code	Inconsistent entires	Field needs to be parsed to be useable. Lots of unique entires. See loc_intxn_id_code	These t quite muc with co
169	loc_twy_code	Inconsistent entires	Field needs to be parsed to be useable. Lots of unique entires. See loc_intxn_id_code	These t quite muc with co
170	opn_proc_defic_flag	missing	3 missing, 984 N, 61 Y. Desc field only filled out for Y values	varia
171	opn_proc_spl_desc	inconsistent with flag	103 non-missing values. 1 on N, rest on Y. 5 missing values have Y for flag	
172	opn_proc_spl_flag	Missing	107 Y, 3 Missing, 938 N	
173	opn_psn_comb_code	missing	1 missing value. Unclear on "N". Interpreted to mean combined, not approved	
174	opn_sctr_comb_code	missing. Unclear coding	Coding unclear. No A listed in ATQA dictionary.	will us

ID	Variable Name	Issue	Comments	
			174 missing	
175	opn_sctr_comb_code_oe	174 missing values		will us
176	opn_sctr_comb_code_oe and ctlr_psn_comb_desc_oe	inconsistent information	These two variables do not always agree	will us
177	oprtr_flt_id_nbr_oe	Lots of unique entires	Will require parsing. 2 missing values	drop wh
178	oprtr_flt_id_nbr_oe	3 missing values and inconsistent formatting of string variable	Airport callsign- Most begin with 3-character airline code, but not all follow that practice (e.g., "AIRPORT", "BICEP", "AP 525")	drop wh
179	plt_rprt_nmac_code_oe	missing	16 missing. 164 unknown	va
180	sepn_hrzntl_ft_qty	missing	289 missing values. Extends up to 13050 which seems high to be in a RI	su
181	sepn_hrzntl_min_qty	missing	all missing (1048)	
182	sepn_ver_ft_qty	missing	588 missing.	su
183		final/prelim status	Are the fields taken from the ATQA database final or preliminary? Is it possible that some of the OEs reported are classified as OE preliminarily	FROM fina based

ID	Variable Name	Issue	Comments	
			then later reclassified as PD? How would you tell which those were?	
184		Need to figure out how to get all 3000 days of operations data		
185	acft_alt_ft_qty	Missing	Dropped variable	
186	acft_alt_unk_flag	Missing, Yes, No values	6179 missing	
187	acft_ctl_arpt_no_twr_flag	Missing, Yes, No values	6252 missing	Treat r
188	acft_ctl_class_a_flag	Missing, No values	Dropped variable	
189	acft_ctl_class_b_flag	Missing, Yes, No values	Dropped variable	
190	acft_ctl_class_c_flag	Missing, Yes, No values	dropped variable	
191				
192	acft_ctl_class_d_flag	Missing, Yes, No values	Dropped variable	
193	acft_ctl_class_e_flag	Missing, Yes, No values	Dropped variable	
194	acft_ctl_class_g_flag	Missing, Yes, No values	Dropped variable	
195	acft_ctl_otr_desc	Missing Values	6418 missing. All filled in values correspond to a value of Y for the acft_ctl_otr_fl ag	
196	acft_ctl_otr_flag	Missing, Yes, No values	6245 missing.	
197	acft_ctl_trsa_flag	Missing, Yes, No values	6222 misisng	
198	acft_ctl_twr_flag	Missing, Yes, No values	5325 missing	
199	acft_ctl_unk_flag	Missing, Yes, No values	6258 missing.	
200	ACFT_FIRST_FLT_CODE	In ATQA, not in dataset	Tracks "first flight of day for pilot". Not in current	

ID	Variable Name	Issue	Comments	
			dataset	
201	acft_mkmd_make_desc	Inconsistant Entries	Used in forming aircraft groupings. Keeping issue open til groups are completed	
202	acft_mkmd_model_desc	Inconsistant Entries	Used in forming aircraft groupings. Keeping issue open til groups are completed	
203	acft_phase_apch_flag	Y,N, Missing	dropped variable	
204	acft_phase_cmb_flag	Y,N, Missing	dropped variable	
205	acft_phase_crz_flag	Y,N, Missing	dropped variable	
206	acft_phase_dscnt_flag	Y,N, Missing	Dropped Variable	
207	acft_phase_lndg_flag	Y,N, Missing	Dropped Variable	
208	acft_phase_otr_desc	Missing values	193 non-missing values. Appear to correspond to Y on the flag variable. However, numbers differ indicating some Y have no description	
209	acft_phase_otr_flag	Y,N, Missing	194 Y, 4174 missing, 2066 N	
210	acft_phase_unkn_flag	Y,N, Missing	Dropped variable. use phase of flight	

ID	Variable Name	Issue	Comments	
			instead	
211	acft_rule_flt_code	UNK and missing	Recoded DVFR and SVFR into VFR and converted variable to numeric	
212	acft_sua_desc	Missing Values	6432 missing. All non-missing values correspond to a Y value for acft_ctl_sua_flag	
213	acft_sua_flag	Missing, Yes, No values	6259 missing	
214	acft_tcas_code_	UNK and missing	271 TCUNKN, 6088 missing	
215	acft_tcas_invlvd_desc_	no entries	All missing	
216	acft_transpndr_code	UNK and missing	177 UNK, 6085 missing	
217	acft_type_code	missing	751 missing. Unclear how to treat missings	
218	acft_type_otr_desc	Missing	Non missing entires appear to correspond to OTR in the type code	
219	clnc_hrzntl_ft_qty	Missing	Only 5 non-missing entries. Likely not relevant varaible	
220	clnc_hrzntl_unkn_flag	Y,N,Missing	26 Y, 6304 Missing, 104 N. Not relevant, but seems inconsistant with missings in clnc_hrzntl_ft_qty	

ID	Variable Name	Issue	Comments	
221	clnc_no_flag	Y,N,Missing	192 Y, 6218 Missing, 24 N. Likely not relevant	
222	clnc_slant_ft_qty	Missing. Not in ATQA data dictionary	1 non-missing value. No description in data dictionary	
223	clnc_ver_ft_qty	Missing	9 non-missing values. Likely not relevant	
224	clnc_ver_unkn_flag_pd	Y,N,Missing	32 Y, 6301 Missing, 101 N	
225	dev_air_airspd_flag_pd	Many Missings	what do missing values indicate different from "Y" or "N"	
226	dev_air_asp_flag	Many Missings	what do missing values indicate different from "Y" or "N"	
227	dev_air_atc_alt_clnc_flag_pd	Many Missings	what do missing values indicate different from "Y" or "N"	
228	dev_air_atc_crs_clnc_flag_pd	Many Missings	what do missing values indicate different from "Y" or "N"	
229	dev_air_carlss_flag	Many Missings	what do missing values indicate different from "Y" or "N"	
230	dev_air_far_otr_flag	Many Missings	what do missing values indicate different from "Y" or "N". 6291 missing	

ID	Variable Name	Issue	Comments	
231	dev_air_miss_rprt_flag_pd	Many Missings	what do missing values indicate different from "Y" or "N". 6299 missing	
232	dev_air_plt_unqlfy_flag_pd	Many Missings	what do missing values indicate different from "Y" or "N". 6298 missing	
233	dev_air_too_low_flag_pd	Many Missings	what do missing values indicate different from "Y" or "N". 6301 missing	
234	dev_air_vfr_ifr_rqrd_flag_pd	Many Missings	what do missing values indicate different from "Y" or "N". 6297 missing	
235	dev_asp_viol_arsa_name	only one value, rest missing	Not in ATQA data dictionary either	
236	dev_asp_viol_code	Many Missings	6,091 missings. 50 values of "NONE", 35 "UNK"	
237	dev_asp_viol_sua_desc	inconsistent	4 entires are non-missing. Only 2 of dev_asp_viol_ code have an entry of SUA	
238	dev_asp_viol_tca_name	all missing	All entries are missing	
239	dev_sfc_carlss_flag	Many Missings	4,306 missing. 2,052 N, 76 Y	
240	dev_sfc_enter_flag	Many Missings	1,543 missing; Y and N values	

ID	Variable Name	Issue	Comments	
241	dev_sfc_fltpln_flag	Many Missings	4,340 missing; Y and N values. Only N and missing on relevant set	
242	dev_sfc_lndg_clnc_flag	Many Missings	4,031 missing; Y and N values	
243	dev_sfc_lndg_flag	Many Missings	3,967 missing, does not seem to match up with Clearance flag. Presumably you can't land on the wrong runway but have clearance for landing on that runway	
244	dev_sfc_otr_desc	Many Missings	Parsed variable for other common answers and created flags for : crossed hold short line and landed on closed runway	
245	dev_sfc_otr_flag	Many Missings	4018 missings. See dev_sfc_otr_desc	
246	dev_sfc_tkof_clnc_flag	Many missings	3,851 missing; Y and N values	
247	dev_sfc_tkof_rwy_twy_flag	Many missings	4,209 missing; Y and N values	
248	dev_sfc_wx_minm_flag	Many missings	4,332; Y and N values	
249	dev_type_air_flag	Many missings	Does not appear to be mutually exclusive with dev_type_sfc. Spot checking	

ID	Variable Name	Issue	Comments	
			a couple of events indicates they are runway incursions.	
250	dev_type_sfc_flag	Overlap with dev_type_air_flag	no missings, 367 are yes on this and dev_type_air_flag. All observations in dataset are Y on this variable	
251	ev_air_far_occ1_nbr	Many Missings	6,413 missing, only 21 filled out	
252	ev_air_far_occ2_nbr	Many Missings	6,424 missing, only 10 filled out	
253	event_detect_dev_code	Missings	321 missing	
254	event_detect_dev_otr_desc	Other without description	1 entry for event_detect_dev_code of other with missing value for description	
255	event_lcl_time	Missings	92 missing values. All other appear in 0-2359 range	
256	event_ri_flag	Missings	1,816 missing. How reliable is this flag?	
257	event_utc_time	Missings	92 missing values. All other appear in 0-2359 range. 91 coincide with missing local time. 1 entry has missing lcl, but filled in UTC. Another 1 entry has	

ID	Variable Name	Issue	Comments	
			UTC filled in, but lcl missing	
258	fac_atc_none_flag	Y,N,Missing	42 Y, 6241 missing, 151 N	
259	fac_atc_otr_flag	Y,N,Missing	6 Y, 6258 Missing, 170 N. Yes corresponds to relevant entry in description field	
260	fac_atc_unkn_flag	Y,N,Missing	40 Y, 6258 Missing, 136 N	
261	fac_rprt_fsdo_id_nbr	Missing	1349 Missing. Likely not useful	
262	fac_rprt_loc_id_code	Missing	221 Missing	
263	fact_equip_com_flag	Y,N,Missing	309 Y, 4268 Missing, 1857 N	
264	fact_equip_nav_flag	Y,N,Missing	17 Y, 4442 Missing, 1975 N, likely not relevant	
265	fact_equip_none_flag	Y,N,Missing	4892 Y, 350 N, 1192 Missing	
266	fact_equip_otr_flag	Y,N,Missing	259 Y, 4291 Missing, 1884 N	
267	fact_equip_trnspndr_flag	Y,N,Missing	16 Y, 4444 Missing, 1974 N	
268	fact_equip_unkn_flag	Y,N,Missing	375 Y, 4223 Missing, 1836 N. What is the interpretation? "Unknown equipment malfunctioned" or "unkown if equipment malfunctioned"	

ID	Variable Name	Issue	Comments	
269	fact_wx_avoid_flag	Y,N,Missing	40 Y, 4433 Missing, 1961 N, likely not relevant	
270	fact_wx_inacc_flag	Y,N,Missing	12 Y, 4454 Missing, 1968 N, not relevant	
271	fact_wx_none_flag	Y,N,Missing	5132 Y, 1063 Missing, 293 N, not relevant	
272	fact_wx_otr_flag	Y,N,Missing	236 Y, 4312 missing, 1886 N	
273	fact_wx_unkn_flag	Y,N,Missing	304 Y, 4264 missing, 1866 N	
274	fact_wx_vfr_imc_flag	Y,N,Missing	47 Y, 4435 missing, 1952 N	
275	fctr_equip_altm_flag	Y,N,Missing	1 Y, 4453 Missing, 1980 N. likely not relevant	
276	fctr_equip_autoplt_flag	Y,N,Missing	4 Y, 4451 Missing, 1979 N. likely not relevant	
277	form_i_version_code	Missing	4340 Missing. Few entires of one version	
278	form_pr_version_code	Missing	4267 Missing. Few entires of 8020-17 6/91 and TELEX	
279	lat_long_srce_code	Missing / Not in ATQA Data Dictionary	Not in ATQA data dictionary. All values missing	
280	loc_arpt_code	Missing / Not in ATQA Data Dictionary	6183 missing. See loc_arpt_id_code	

ID	Variable Name	Issue	Comments	
281	loc_arpt_id_code	Missing	160 missing. Appears to be the legitimate version of loc_arpt_code . This field name is not in the data dictionary	
282	loc_city_name	missing	99 missing	
283	loc_drctn_degm_qty	Missing / Unclear Definition	6242 missing. Unclear what this is measuring	
284	loc_dstc_nm_qty	Missing / Unclear Definition	6186 missing. Unclear what this is measuring	
285	loc_intxn_id_code	missing	All observations missing. This appears to relate to enroute space, so may be legitimately empty	
286	loc_lat_deg_qty_pd	Missing	1 entry, 6433 missing	
287	loc_lat_min_qty_pd	Missing	1 entry, 6433 missing	
288	loc_lat_ns_code	Missing	all entires missing	
289	loc_long_deg_qty	Missing	1 entry, 6433 missing	
290	loc_long_ew_code	Missing	all entires missing	
291	loc_long_min_qty	Missing	1 entry, 6433 missing	
292	loc_nav_fac_code	missing	6371 missing. If the airport code is filled out, would this be filled	

ID	Variable Name	Issue	Comments	
			out as well?	
293	loc_oceanic_flag	missing	Only N values in dataset. 6297 missing	
294	loc_state_code	missing	101 missing	
295	loc_tfc_ptrn_code	UNK and missing	Signfiicant missings. Others are unable to be parsed into existing categories. Use variable as is for demographic purposes only	
296	oprtr_flt_id_nbr	Missing	Use RI flight number information as it appears to be filled out more.	
297	oprtr_ga_flag	Y,N,Missing	4175 M, 1758 Missing, 501 N	
298	oprtr_type_code	UNK and missing	Coded missing values as unknown. Suggest deriving this information elsewhere for ALL observations instead of just PDs	
299	plt_asp_viol_flag	Y,N,Missing	502 Y, 4140 Missing, 1792 N. likely not useful	
300	plt_atc_instrn_desc	Missings	Number of non-missing observations doesn't line up with number of Y flags. Likely not	

ID	Variable Name	Issue	Comments	
			useful	
301	plt_atc_instrn_flag	Missings	2572 Y, 2752 Missing, 1110 N. Likely not useful	
302	plt_birth_date	Illogical dates	Some birthdates indicate pilots were < 10 years old. Recoded those to missing. One indicates pilot was approx. 98 years old. Leaving it for now, but does seem odd	
303	plt_cert_nbr	Missing	536 missing. Appear to vary in length (7-9 numbers. Need to check what valid codes are from external source)	
304	plt_cert_otr_desc_pd	needs to be parsed for other major certification categories		
305	plt_certif_atp_flag	Y,N,Missing	1927 Y, 3066 missing, 1441 N	
306	plt_certif_cfi_flag	Y,N,Missing	882 Y, 3798 Missing, 1754 N	
307	plt_certif_coml_flag	Y,N,Missing	1615 Y, 3324 missing, 1495 N	
308	plt_certif_frgn_flag	Y,N,Missing	104 Y, 1989 N, 4341 Miss. No indication other than foreign. Maybe want to look at	

ID	Variable Name	Issue	Comments	
			other pilot info to determine country of origin?	
309	plt_certif_mil_flag	Y,N,Missing	61 Y, 4386 missing, 1987 N	
310	plt_certif_none_flag	Y,N,Missing	20 Y, 4403 missing, 2011 N. What is the interpretation here? Can you fly without a certification?	
311	plt_certif_otr_flag	Y,N,Missing	187 Y, 4300 missing, 1947 N	
312	plt_certif_pvt_flag	Y,N,Missing	2090 Y, 3081 Missing, 1263 N	
313	plt_certif_rcrntl_flag	Y,N,Missing	5 Y, 4414 Missing, 2015 N. very few Ys, likely not helpful. Could we lump this into something else?	
314	plt_certif_stdnt_flag	Y,N,Missing	410 Y, 4146 Missing, 1878 N	
315	plt_certif_unkn_flag	Y,N,Missing	178 Y, 4310 Missing, 1946 N	
316	plt_city_name	Missing	510 Missing	
317	plt_ck_2yr_flt_rvw_date	Missing	3466 missing. Need to check how pilot checks relate to one another	
318	plt_ck_atp_flt_test_date	Missing	6226 missing	

ID	Variable Name	Issue	Comments	
319	plt_ck_cmptncy_flt_date	Missing	5723 Missing	
320	plt_ck_flt_test_date	Missing	5796 Missing	
321	plt_ck_inst_test_date	Missing	5369 Missing	
322	plt_ck_otr_date	Missing	6052 Missing	
323	plt_ck_otr_desc	Parse of useful categories	Parsed for other common answers and created a flag for solo endorsement	
324	plt_ck_profic_ck_date	Missing	4901 Missing	
325	plt_ck_rte_ck_date	Missing	5571 missing	
326	plt_ck_simltr_date	Missing	5640 Missing	
327	plt_dstrctn_flag_	Y,N,Missing	1218 Y, 3598 Missing, 1618 N. Likely not useful	
328	plt_duty_l24hr_hr_qty_pd	missing	Can't simply recode this as GA pilots will not have an "on duty" number but will have a flight time number. May have to just take variable as is or simply recode missings to 0. Re-opened on 10/24/2011	
329	plt_enfrc_code	Missing, UNK	coded unknown on missing and made variable numeric. Also recoded "1MORE" to "YES" and "NONE" to "NO"	
330	plt_fatig_flag_	Y,N,Missing	156 Y, 4343	

ID	Variable Name	Issue	Comments	
			missing, 1935 N.	
331	plt_flt_l24hr_hr_qty	Missing	2390 missing	
332	plt_flt_leg_hr_qty	Missing	2248 Missing. 3 entires > 24 (values of 30, 40, 70.4)	
333	plt_inadqt_acft_flag	Y,N,Missing	33 Y, 4429 Missing, 1972 N	
334	plt_inadqt_aip_flag	Y,N,Missing	165 Y, 4344 Missing, 1925 N	
335	plt_inadqt_arpt_flag	Y,N,Missing	918 Y, 3892 Missing, 1624	
336	plt_inadqt_atc_flag	Y,N,Missing	634 Y, 4110 Missing, 1690 N	
337	plt_inadqt_avion_flag	Y,N,Missing	132 Y, 4372 Missing, 1930 N	
338	plt_inadqt_crew_flag	Y,N,Missing	132 Y, 4382 Missing, 1920 N	
339	plt_inadqt_english_flag	Y,N,Missing	65 Y, 4406 Missing, 1963 N	
340	plt_inadqt_otr_flag	Y,N,Missing	447 Y, 4157 Missing, 1830 N	
341	plt_inadqt_preflt_flag	Y,N,Missing	215 Y, 4318 Missing, 1901 N	
342	plt_inadqt_trmnlgy_flag	Y,N,Missing	483 Y, 4141 Missing, 1810 N	
343	plt_inadqt_unkn_flag	Y,N,Missing	408 Y, 4199 Missing, 1827 N	
344	plt_inadqt_wx_flag	Y,N,Missing	34 Y, 4434 Missing, 1966 N	

ID	Variable Name	Issue	Comments	
345	plt_inst_code	UNK and missing	Recoded missing to unknown and converted to numeric.	
346	plt_locat_tfc_flag_	Y,N,Missing	12 Y, 4445 Missing, 1977 N	
347	plt_lost_flag__	Y,N,Missing	445 Y, 4175 Missing, 1814 N	
348	plt_med_first_flag	Y,N,Missing	1959 Y, 3067 Missing, 1408 N. Medical flags likely not useful	
349	plt_med_last_date	Missing	1119 Missing	
350	plt_med_none_rqrd_flag	Y,N,Missing	54 Y, 4401 Missing, 1979 N	
351	plt_med_outdt_flag	Y,N,Missing	60 Y, 4407 Missing, 1967 N	
352	plt_med_scnd_flag	Y,N,Missing	1222 Y, 3656 Missing, 1556 N	
353	plt_med_self_certif_flag	Y,N,Missing	4 Y, 4432 Missing, 1998 N	
354	plt_med_spl_flag	Y,N,Missing	36 Y, 4413 Missing, 1985 N	
355	plt_med_thrd_flag	Y,N,Missing	2250 Y, 2994 Missing, 1190 N	
356	plt_med_unkn_flag	Y,N,Missing	289 Y, 4267 Missing, 1878 N	
357	plt_mkmd_hr_qty	Missing	Created variable plt_mkmd_hr_round that is roudned to nearest 10	

ID	Variable Name	Issue	Comments	
			hours. That variable may require additional rounding on the upper end, or a flag indicating over soem threshold	
358	plt_mkmd_l90d_hr_qty_	Missing	Created plt_mkmd_l90d_hr_round that is rounded to nearest 10 hours. May require additional rounding at highest levels or variable indicating above som threshold.	
359	plt_none_flag_	Y,N,Missing	921 Y, 3847 Missing, 1666 N. Unclear what this flag indicates?	
360	plt_not_scan_flag_	Y,N,Missing	81 Y, 4395 Missing, 1958 N	
361	plt_otr_flag_	Y,N,Missing	669 Y, 4035 Missing, 1730 N. Not all Ys have descriptions	
362	plt_ovrwrkd_flag	Y,N,Missing	66 Y, 4406 Missing, 1962 N	
363	plt_pr_unkn_flag	Y,N,Missing	574 Y, 4190 Missing, 1670 N. Not sure if useful	
364	plt_resp_tcas_adzy_flag_	Y,N,Missing	0 Y, 4452 Missing, 1982	

ID	Variable Name	Issue	Comments	
			N. Makes sense as we want ground incidents. Possible some of the missings are possible Ys	
365	plt_rtng_gldr_flag	Y,N,Missing	136 Y, 4332 Missing, 1966 N	
366	plt_rtng_lta_flag	Y,N,Missing	21 Y, 4413 Missing, 2000 N	
367	plt_rtng_mel_flag	Y,N,Missing	3109 Y, 2336 Missing, 989 N	
368	plt_rtng_mes_flag	Y,N,Missing	60 Y, 4392 Missing, 1982 N	
369	plt_rtng_none_flag	Y,N,Missing	220 Y, 4279 Missing, 1935 N	
370	plt_rtng_otr_flag	Y,N,Missing	380 Y, 4187 Missing, 1935 N	
371	plt_rtng_rotor_flag	Y,N,Missing	338 Y, 4197 Missing, 1899 N	
372	plt_rtng_sel_flag	Y,N,Missing	4710 Y, 1338 Missing, 386 N	
373	plt_rtng_ses_flag	Y,N,Missing	488 Y, 4109 Missing, 1837 N	
374	plt_rtng_unkn_flag	Y,N,Missing	255 Y, 4278 Missing, 1901 N	
375	plt_scan_flag	Y,N,Missing	154 Y, 4346 Missing, 1934 N	
376	plt_sick_flag	Y,N,Missing	13 Y, 4444 Missing, 1977 N	
377	plt_total_hr_qty	Missing	generated plt_total_hr_r	

ID	Variable Name	Issue	Comments	
			ound which is rounded to nearest 10	
378	plt_total_l90d_hr_qty	Missing	Created plt_total_l90d_hr_round which is rounded to nearest 10	
379	plt_trnspndr_off_flag_	Y,N,Missing	11 Y, 4446 Missing, 1977 N	
380	plt_unkn_flag_	Y,N,Missing	380 Y, 4218 Missing, 1836 N	
381	remarks field	Not in dataset	No remarks field was included in dataset	
382	rprt_atchmnt_flag	Y,N,Missing	5168 Y, 1251 Missing, 15 N	
383	sepn_ft_code	missing	365 misisng	
384	sepn_hrzntl_ft_qty	missing	use the RI variable for demographic purposes. Not worth the effort to combine the two.	
385	sepn_hrzntl_unkn_flag	missing, inconsistent	6261 missing - all missing on sepn_hrzntl_ft_qty as well. 18 Y values have missing sepn_hrzntl_ft_qty. 137 have N values, but missing sepn_hrzntl_ft_qty.	
386	sepn_long_min_qty	missing	6433 missing	
387	sepn_long_unkn_flag	missing	6274 missing	

ID	Variable Name	Issue	Comments	
388	sepn_loss_acft_air_flag	Missing, Yes, No, values	4,146 missing.	
389	sepn_loss_acft_gnd_flag	Missing, Yes, No, values	3819 missing. 22 Yes on both this and sepn_loss_acft_air_flag. Narrative not included	
390	sepn_loss_na_flag	Inconsistent. Missing values	1135 missing. How does this relate to sepn_loss_unk_flag?	
391	sepn_loss_obstn_flag	Missing, Yes, No values	4350 missing	
392	sepn_loss_persnl_flag	Missing, Yes, No values	4334 missing	
393	sepn_loss_unkn_flag	Missing, Yes, No values	4282 missing	
394	sepn_loss_veh_flag	Missing, Yes, No values	4314 missing	
395	sepn_no_flag	Missing, Yes, No values	Variable used for demographic only. Missings only appear with missing and no appears only with no. safe to assume missing = no	
396	sepn_no_flag	drop observations that are Y on this and one of the other sepn_ flags	Only 9 observations with this property. Can be dropped later (flag generated as suspect_sepn_flag) if desired.	
397	sepn_slant_ft_qty	Not in data dictionary	Variable in dataset, but not in data dictionary	S
398	sepn_ver_ft_qty	Missing values	Use RI information for demographic	

ID	Variable Name	Issue	Comments	
			purposes. Not worth the effor to combine the two.	
399	sepn_ver_unkn_flag	Missing, Yes, No values	6272 missing	
400		Hand Matching	When hand matching records, a range of +/- 30 min was used to determine if two records were candidates for matches.	
401	acft_invlvd_code	Missing values, "UNK", caps out at "FOUR+"	What do missings mean? Presumably you can't have a PD without a pilot/aircraft	FROM In mar comple
402	acft_invlvd_qty	conflicts with acft_invlvd_code	Some with acft_invlvd_code == "TWO" have qunaitities above 2. Some with missing codes have filled out quantities	
403	arpt_ctl_code	175 missing.	Assuem that non-controlled airports will have no runway incursions. Possible to drop these from the analysis?	FRO airp there
404	dev_air_acft_equip_flag	many missing	what do missing values indicate	

ID	Variable Name	Issue	Comments	
			different from "Y" or "N"	
405	event_utc_date	The only date provided for the incident is the UTC date	The UTC date can vary from the local date (e.g. when the local time is late and the UTC offset is large). Is other date information available in the dataset?	
406	RPRT_OTR_PR_OED_NBR		This appears to be indicate that a PD report had a preliminary OE/D number? Volpe does not have this field in the data provided	
407	ac1cat	more possible entries than needed	Dropping variable and creating our own categories from flight number	
408	ac1id	values of "N/A"	"N/A" values make sense for V/PD. There are 29 incidents with ac1id = "N/A" but are not V/PDs. The 29 incidents are OE and OD. Quick scan of record narratives show no aircraft involved. Entires seem	FRO even ATC is cleara OE, but

ID	Variable Name	Issue	Comments	
			valid. Ac2id exhibits same patterns	
409	ac1id	contains more information than needed		
410	ac1type	values of "N/A"	Similar set up to ac1id. 29 OE/OD N/As look legitimate. Ac2type exhibits same patterns	
411	ac1type	more possible entries than needed		Varia to th
412	ac2cat	more possible entries than needed	Creating our own categories from flight number	
413	ac2id	contains more information than needed		
414	ac2type	more possible entries than needed		Varia to th
415	acmainttaxi	Value of "Y" and missing	recoded missing as N	
416	adjustedrank	1595 missings, 1 "E"	Dropping all cases with missing adjusted rank as those are surface incidents	
417	amassinservice	ASDE-X option shows up before reasonably possible	Is it possible these are ASDE-3 aswell?	r
418	arptempveh	value of "N", "Y", and missing	Unclear on difference between N and missing	miss
419	arptid	one instance were not equal to locid	Lake Hood airport. Locid == LHD, arptid == ANC. Likely	FROM are bas is adjacen

ID	Variable Name	Issue	Comments	
			just a typo	
420	assessmentremarkscomments onriscm	not needed		
421	catrank	will use adjrank intsead		
422	city	not needed		
423	collision	Entry of "UNK", N/A, and 179	179 is most confusing. "YES" entry is not given a rank. All "Y"s are rank "A". No missing values for this field. UNK is rank C along with the 179 entry. All other's are N/A	will re 179 an
424	constrnpersl	Value of "Y" and missing	Assume that missings are "N"s, but it is possible that some records may be missing for other reasons	tre
425	crsdholdshortlineonly	inconsistent entries	Missing, Y, N, and specific taxi ways are listed. THe missings appear to be mostly incidents where the hold short line was not in play. However, some narratives suggest that a hold short line was crossed (such as entering	In gene some a on the are m

ID	Variable Name	Issue	Comments	
			runway without clearance)	
426	crsdrwyortwy	inconsistent entries	Missing, N, Y, UNK, specific runways all listed	In gene some a on th are m
427	ctlrtrng	value of "N", "Y", and missing. Also value of "N/A"	Unclear between N, N/A, and missing. How does this compare to controller fields in OE database?	
428	dateevent	not needed		
429	dateincdtclsfdrisi	not needed		
430	daterptrcvd	not needed		
431	daylightsavings	not needed		
432	dayofthewk	not needed		
433	dupreport	100 non-missing entires	How are the non missing entires to be interpreted? 2 cases were dupreport entry is equivalent to report number	
434	dupreport	not needed		
435	enteredrwy	inconsistent entries	Missing, N, Y, UNK, specific runways all listed	In gene some a on the is YES,
436	faafemp	Value of "Y" and missing	Assume that missings are "N"s, but it is possible that some records may be missing for	

ID	Variable Name	Issue	Comments	
			other reasons. Possibly not of interest	
437	farpart91121135mil	Data dictionary says it should pertain to PDs only. Apepars as if some OD/OEs have non-missing/"N/A" entries.	Does this apply only to PDs? 3 missings, 1 unkown. Most OD/OE are N/A, but not all. Suspect that some of the N/A OE/OD could be coded as something else	The tr in
438	finalprelimrptstatus	not needed		
439	frngacorpilot	Value of "Y" and missing	Assume that missings are "N"s, but it is possible that some records may be missing for other reasons. Appears as if some missing have mention of snowplows in narrative. We may need to parse the narrative	
440	holdshortinstrissued	inconsistant entries	Missing, N, Y, Unk all listed. One entry of TWY C	
441	holdshortrdbk	inconsistant entries	Missing, N, Y, Unk all listed. Missings seem to coincide with missing holdshortinstrissued	treat r check
442	horizontaldistanceormileage	inconsistant entries	Transformed to numeric. Ranges	

ID	Variable Name	Issue	Comments	
			coverted to means and miles converted to feet.	
443	hour	not needed		
444	inatqa	2 missings	How reliable is this field? Matching on report number gives 508 that are inatqa="N" but match exactly to records in ATQA OE data	
445	inatqa	not needed		
446	intersectingrwydeptorarr	Missing, N/A, UNK, Y		agre
447	lahso	B,N,N/A,Y, and missing	2 B's, 327 missings, 2 N/As	missing year, l
448	landedtaxiingin	Missing, N/A, UNK, Y	What is the relationship between Missing and N/A? What are UNK?	in g sense missing FAA w
449	lawenforcement	Value of "Y" and missing	Assume that missings are "N"s, but it is possible that some records may be missing for other reasons	replace
450	lndgdeptdtwyorclsdrwytwy	Missing, N/A, and Y values	What is the relationship between Missing and N/A	
451	lndgordeptdwoclrccomm	Missing, N/A, and Y values	What is the relationship between Missing and N/A	

ID	Variable Name	Issue	Comments	
452	localtime	not needed		
453	meetsicaostandardsri	Missing, Yes, No, UNK values	missing, No, and UNK all coincide with missing adjusted rank. All Yes values have an adjusted rank. Further indicates we should ignore anything without an adjusted rank	
454	meetsicaostandardsri	not needed		
455	min	not needed		
456	month	not needed		
457	narrative	not needed		
458	notes	not needed		
459	oecombpstns	Y,N, UNK, MISSING, as well as which specific positions were combined	Converted specific listings to Y and all others to N	
460	originalnarrative	not needed		
461	pdfassmt	Missing, N/A, Yes	only 2 N/A. What is this?	
462	pdfassmt	not needed		
463	phaseofflight	inconsistent entries	Collapsed to 1 variable for each aircraft and normalized codings.	
464	report_part	not needed		
465	preicaocatrank	not needed		
466	ri	not needed		
467	riscranking	not needed		
468	riscscenario	not needed		

ID	Variable Name	Issue	Comments	
469	prvtcitzveh	value of "N", "Y", and missing. Also value of "N/A"	Unclear on difference between N and missing and N/A	
470	ri	Some observations coded as "N". Should these be exlcuded from analysis?	two options for the field: "Y" or "N", no missing values. When compared against the "adjustedrank" field, some coded as "N" are given a rank of D. Some "N" are given no rank. All "Y" are given a rank. Presume that subset of non-missing ranks are the relevant dataset.	
471	state	not needed		
472	rwytwyconstrn	Value of "Y" and missing	Assume that missings are "N"s, but it is possible that some records may be missing for other reasons. Appears as if some missing have mention of snowplows in narrative. We may need to parse the narrative	
473	servicearea	not needed		
474	time	not needed		
475	timezone	not needed		

ID	Variable Name	Issue	Comments	
476	snowrmvlveh	Value of "Y" and missing	hand checked records containing the word "snow." Hand coded those that indicated snow removal was happening at the airport.	
477	sutdentpilot	Value of "Y" and missing	Hand checked records containing the word "student" (case insensitive) th at did not have a Y flag on this variable. Found some that were student pilots, hand coded those as Ys.	
478	taxiingoutfordept	Missing, N/A, UNK, Y, N	Recoded missing and N/A to N.	
479	toorlndgrwy	not needed		
480	tiph	Value of "Y" and missing	Data dictionary says this should contain location. No location is given.	
481	tiphdptdwoclrnc	Value of "Y", "N", and missing		
482	typeerrorcode	not needed		
483	utcoffset	not needed		
484	v78	not needed		
485	toorlndgrrwy	inconsistent entries	N, N/A, UNK, YES, as well as specifics listed. 2	

ID	Variable Name	Issue	Comments	
			missing	
486	toorlndgrrwy	Missing, UNK	Lots of missings. 10 UNK. Unclear how they relate	91 = a CF 121 = a CF 125 = a CFR Pa AIRP PAS PA 129 = a CFR P 135 = a MA
487	tugs	missing values		
488	typeerrorcode	No codebook for values	No way to decode numerical values	FROM scenar don't t m u
489	verticaldistance	inconsistant entries and many missings	Recoded to numerica. took means for ranges.	
490	vmc	Y, N, UNK, and missing		rep
491	vpdsauthonarptormvntarea	Lots of missings. Yes, No, Missing	Coded missings to "No Vehicle". Note that some "Y"s are non-V/PDs	

ID	Variable Name	Issue	Comments	
492		Need to decide about how to handle LHD	LHD is a seaplane base	
493		Sample Size	249 OE/OD records in the RI database did not match to any ATQA data	
494		Sample Size	1048 OE/OD records matched between the RI and ATQA data	
495		Sample Size	456 records in the ATQA data did not match to RI data. Need to explore why these don't match	
496		Sample Size	1668 PD records in the RI database did not match to ATQA data	
497		Sample Size	4361 records matched between RI data and ATQA	
498		Sample Size	2074 records in ATQA data did not match any PDs in the RI data	
499	fac_cnflct_alert_ncode_oe	Don't understand categories	Categories are unclear. Can't seem to find any information on the internet. Need Greg's input on the meaning of	

ID	Variable Name	Issue	Comments	
			the categories.	
500	fac_atc_artcc_code_pd	Should be missing for all our observations. 21 observations.	Spot checked records appear to be runway incursions.	
501	fac_atc_fss_code_pd	Should be missing. 12 observations	Overlaps heavily (9/12) with fac_atc_artcc_ code. These records appear to be runway incursions.	
502	fac_atc_tracon_code_pd	should be missing. 67 non-missing	spot checked records appear to be runway incursions.	
503	acft_lcl_flt_code_pd	Recode Unknown to Missing	Recoded unkown to missing and converted variable to numeric.	
504	acft_phase_taxi_flag_pd	Dropped Variable	use phaseofflight instead.	
505	acft_tkof_flag_pd	Dropped variable	use phase of flight instead	
506	acft_phase_turn_flag_pd	Dropped variable	use phaseofflight insterad	
507	trfcmix	Contains info for two AC	Split variable into one for each aircraft and encoded into human readable parts	
508	bullseye	18/30 Core 30 airports had incorrect data	Variable is of insufficient quality and cannot be used for	

ID	Variable Name	Issue	Comments	
			analysis at this time.	
509	taxiways crossing runways	15/30 Core 30 were incorrect	Variable is of insufficient quality to be used for analysis at this time.	
510	plt_dstrctn_desc	Parse for other common responses	Created flags for: instruction (i.e. giving instruction), student (i.e. instructing student OR student was asking questions - th is may be combined with instruction), check list, traffic, passenegers, radio related, and weather	
511	plt_inadqt_otr_desc	Parse for common answers	Parsed for common responses and created a flag for inadquate knowledge of signs and markings.	
512	plt_rtng_otr_dsec	Look for common responses that need another flag	Variable appears to contain mostly duplicate information. Thos cases where respondents listed specific aircraft (i.e. DC9) appear to be covered by the other	

ID	Variable Name	Issue	Comments	
			flags present for that record.	

APPENDIX C: STATISTICAL CONCEPTS

C.1. Two-way Chi-Squared Tests

The difference between relative frequency and overall frequency raises the need to test for differences in the two. This is where a (two-way) Chi-Squared test[103] can be useful.

The Chi-Squared test compares the observed values of an n by k table to their expected values. The observed values are the observed frequencies of the intersection of two categories (represented by the row and column labels). In this case, the expected value for a cell of the table is the marginal percentage for the column applied to the row total.[104] For example, in Table 1, the marginal percentage for OE column is approximately 14.4% (1,268/8,812). The row total for category A incursions is 132. Thus, the expected value for category A OE incursions is approximately 19 (.144 x 132). A generalized way to calculate the expected value is:

$$E_{i,j} = \frac{\sum_{i=1}^{n} O_{i,j} \cdot \sum_{j=1}^{k} O_{i,j}}{N}$$

where:

Ei,j = Expected value for cell i, j

Oi,j = Observed value for cell i, j

N = Total observations

n = number of rows

k = number of columns

Constructing the expected values in this way is a test of independence between the rows and columns. That is, this tests for an association between the rows and columns. The test statistic is calculated by finding the difference between the observed and expected values for each cell, and then totaling them, shown formulaically as:

$$X^2 = \sum_{i=1}^{n} \sum_{j=1}^{k} \frac{\left(O_{i,j} - E_{i,j}\right)^2}{E_{i,j}}$$

103 While there exist other Chi-squared tests, the two-way Chi-Squared test is the most commonly used and the only one appearing in this report; all further references will drop the “two-way” term.

104 Note that the opposite formulation of marginal percentage for the row applied to the column total is equivalent.

This test statistic is distributed Chi-Squared with degrees of freedom (n – 1)*(k – 1). In Table 1, this results in 6 degrees of freedom. Similar tests will be applied in the following sections regarding other combinations of variables.

C.2. Box and Whisker Plots

The box and whisker plot concisely presents the percentiles of the distribution and outliers. The core of this plot type is the box. The box represents the middle 50% of the distribution. The lower bound of the box represents the 25th percentile, the middle line represents the 50th percentile (or median), and the top of the box represents the 75th percentile. The second component of the plot type is the whiskers. These whiskers attempt to represent a “reasonable” range of the data. Specifically, the whiskers encompass the data that is within 1.5 times the interquartile range of the 25th and 75th percentiles. Data outside the whiskers are represented by dots, and are considered outliers. An annotated example follows.

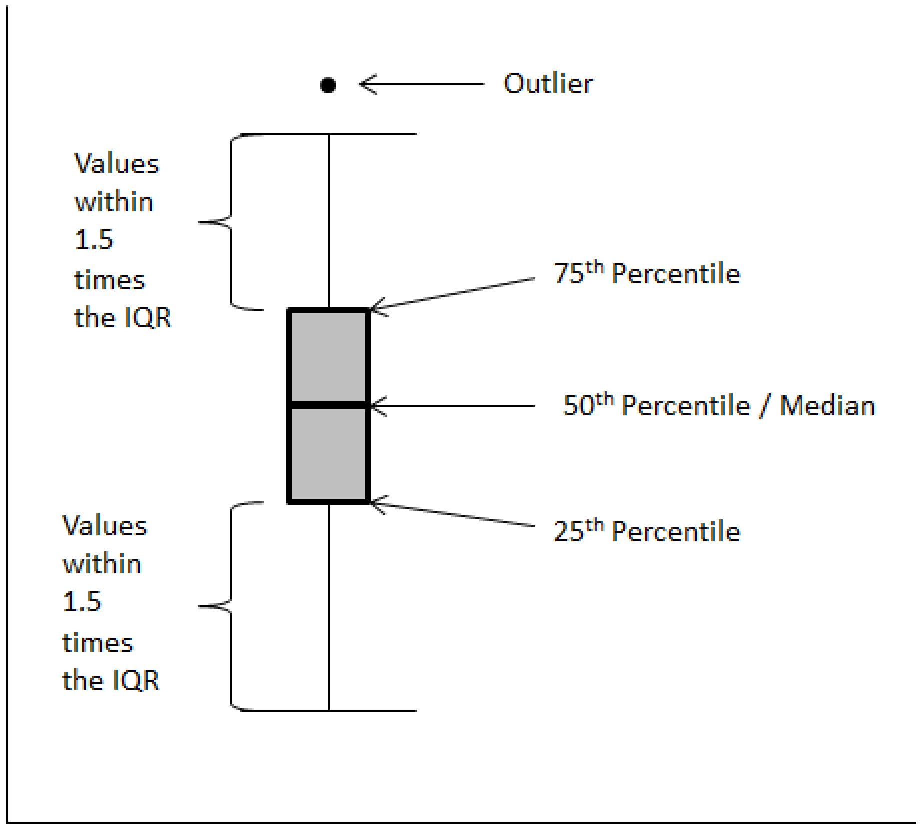

Figure 65 - Annotated Box and Whisker Plot

C.3. Kruskal-Wallis Tests

The Kruskal-Wallis test is an extension of the Mann-Whitney (or Wilcoxon) rank-sum test to two or more categories. The procedure for this test replaces each observation with its rank in the overall dataset and then calculates the mean rank for each category. This procedure jointly tests if the categories have statistically different mean ranks (i.e., if the ranks are distributed randomly among the categories). In other words, a significant test statistic indicates that the categories have different distributions of the continuous variable. This test is particularly useful for small samples, as it requires no asymptotic distributional assumptions. Because the test examines ranks rather than observed values, the exact distribution of the test statistic can be calculated. However, for data with several groups and a moderate number of observations in each group, the distribution is well approximated by the Chi-Squared distribution.[105] More information on the calculations underlying the Kruskal-Wallis rank test can be found in Siegel & Castellan (1988).

Given that the Kruskal-Wallis test indicates that the groups are jointly significant, it may be interesting to determine *which* groups are in fact different. The mean ranks can be compared in a pairwise fashion to determine this. However, this introduces a significant statistical problem, multiple comparisons.

105 Siegel & Castellan (1988).

For example, if there are four groups to compare, there are 6 total pairwise comparisons. Suppose further that that standard significance level of 5% is assumed (i.e. the null hypothesis is rejected incorrectly 5% of the time). Lastly, for this example, suppose that none of the groups *actually* differ (i.e., the null hypothesis is true for all comparisons). Thus:

$$P(\text{reject at least one } H_0 \mid H_0 \text{ true}) = 1 - P(\text{reject no } H_0 \mid H_0 \text{ true}) = 1 - (0.95)^6 \approx 0.2649$$

Thus, for six comparisons the likelihood of rejecting at least one null hypothesis when all are known to be true is greater than 25%. *Put simply, even of all 4 groups are the same, there is a 25% probability of falsely identifying one difference as statistically significant.* Therefore, a correction to the statistical significance criteria is required to compare the groups pairwise and avoid falsely identifying groups as significant.

A simple correction is to compare each test at a smaller significance level. The one employed in this analysis (referred to as the Bonferroni method) uses a pairwise significance rate of α/k, where α is desired significance level for the overall set of tests and k is the number of tests. This ensures that the overall false rejection rate among all the tests combined is no greater than the desired overall false rejection rate. Thus, in the above example, a pairwise significance level of .0083 (0.05 / 6) ensures that the overall false rejection rate is less than or equal to .05.[106]

C.4. Interpreting Regression Output

The two main outputs of the regression models presented in this report are the coefficients and standard errors. These two values are then used to compute the remaining output presented in the tables (the p-value and the confidence intervals). In general, each piece of output has the same interpretation across models, but where there are differences they will be noted.

The piece of output that receives the most attention is the estimated coefficient. The coefficient represents the impact of the independent variable on the dependent variable. For example, as in Table 183, the estimated coefficient for “# of Aircraft Involved” represents how the dependent variable (the probability of a category A incursion) changes with respect to the value of “# of Aircraft Involved.” In this particular example, the coefficient is positive, indicating that the dependent variable increases as the independent variable increases.

For ordered models, the sign of the coefficient indicates the direction of the effect. That is, positive values indicate that the probability of a category A incident (for ordered models) or a severe incident (for binary models). Negative values indicate a complementary decrease in probability. This convention

106 Note that in some sense the multiple comparison problem applies to the analysis as a whole, as well. While the criteria for statistical significance were adjusted for the Kruskal-Wallis tests, they were not done so on a report-wide basis. In other words, this paper examines a large number of variables and presents the associated test statistics. In all likelihood, there is a high probability that at least one of the tests falsely identified a significant relationship when there is none. However, it is impossible to determine which particular test might be reporting erroneously. More focused research can further corroborate the findings in this analysis.

is not true for the multinomial models. In those instances, it is not the absolute size or sign of a coefficient that is important; rather, it is the size and sign of that coefficient relative to the other coefficients presented in the model that are important.

Coefficients for the binary models are presented as odds ratios. These are direct transformations of the coefficients, but work multiplicatively with respect to the odds of a severe incursion. Thus, if the odds ratio is less than one, the odds of a severe incursion decrease as the independent variable increases. If the odds ratio is greater than one, then the odds of a severe event increase as the independent variable increases.[107]

Finally, it is important to note that the coefficients do not directly translate to changes in probability. For all models presented in this report, the coefficients must be combined and then transformed to understand the direct impact on probabilities. In many cases, this transformation is mathematically complex. Thus, for the multinomial models the relevant graphs and tables indicating the change in probability are provided. As the ordered and binary models were not of primary interest, no such calculations were done for those models. Such a calculation could be performed using the coefficients provided in the model.

The second major category of output presented is the standard errors. The standard error measures how precisely the coefficient was estimated. Smaller standard errors indicate that the coefficient was precisely estimated.

The p-value is calculated with the coefficient and the standard error. The p-value measures how likely it is that the estimated coefficient is different from zero (or different from one in the case of an odds ratio). Coefficients of zero indicate that there is no relationship between the given variable and the dependent variable. The P-value approximates how likely it would be to observe the estimated coefficient if the *actual* value of the coefficient was zero. In other words, the p-value represents how likely it is that the estimated coefficient was a product of a random association between the dependent variable and the independent variable. In general, it is standard practice to accept that a random process did not generate the estimated coefficient if the p-value is less than .05.

The last piece of information presented is the 95% confidence interval (CI). The confidence interval represents an alternative description of the uncertainty surrounding a parameter estimate. It consists of two values, the lower bound (LB) and upper bound (UB). These values represent the endpoints of an interval representing the “most likely” values for the estimated coefficient. The estimated coefficient is the midpoint of this interval and the width of the interval is determined by the standard error. The confidence interval provides two pieces of information. First, the interval represents plausible values of the estimated coefficient, given the data on hand.[108] Second, if the confidence interval contains zero, this is equivalent to a p-value greater than or equal to .05. Thus, the p-value and confidence interval both capture the uncertainty surrounding the coefficient estimate.

107 As a side note, some independent variables represent the “status” of an aircraft, such as Commercial Carrier status. These variables are “flags” and are measured as binary (0 or 1) variables. Thus, an increase in one of these variables is going from 0 to 1 (i.e., from not Commercial Carrier status to Commercial Carrier Status).

C.5. A Question of Interpretation: Bayesian versus Frequentist Models

Regardless of the model implemented, there is an overarching concern about the interpretation of results, which cascades backwards into how the models themselves are run. There are two major schools of thought regarding the interpretation of estimation results: Bayesian and Frequentist. Discrete choice models can be implemented in either context. The difference lies in how the results are obtained and interpreted.

C.5.1. Frequentist Econometrics

Most people who have some statistics or econometrics training have been taught frequentist methods. There are a variety of statistical packages that implement a wide array of frequentist methods for any number of models. By and large, frequentist econometrics is the most common type of econometric study. Frequentist techniques in general are outlined in Section 4.1.2.

Treating β as fixed constants is a direct contrast to Bayesian econometrics, as discussed in the following section.

C.5.2. Bayesian Econometrics

The basis of Bayesian econometrics is the use of Bayes' Rule.[109] Bayes' Rule can be written as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

For example, if event A is having a disease and event B is a positive result from a test for that disease, P(A|B) is the probability of having the disease given a positive test result and can be calculated as above. The essence of this

formula is that it combines information about the data – in this case the outcome of the test (the factor P(B|A)/P(B)) – and information about the unconditional probability of the outcome – being sick (the factor P(A)). In this example, Bayes' Rule would be:

$$P(sick|positive\ test) = \frac{P(positive\ test|sick)P(sick)}{P(positive\ test)}$$

The formula above can be extended to a regression context and used to describe a wide variety of models. Suppose the regression model has data y and parameter set θ.[110] The above formula can be rewritten as:

---

108 Specifically, the confidence interval is an interval that contains the "true" value of the coefficient with some probability, in this case 95% probability. However, no one confidence interval can be said to contain the true value of the parameter. It is important to remember that the confidence interval is estimated from the data available, and thus would change as the data changes.

109 Koop (2003).

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

This relationship can be simplified, removing extraneous information about y. It reduces exactness of the expression but maintains the most important part of the relationship defined in Bayes' Rule (i.e., the proportional relationship between θ and y). When simplified, the relationship is expressed as:

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

p(θ) is referred to as the "prior distribution" and represents the information available about θ before looking at the data. This information can come from previous research or the researcher's informed beliefs. p(y|θ) is called the "likelihood" and represents the probability distribution of the data given a parameter set. Finally, p(θ|y) is called the "posterior distribution" and captures all available information on θ – information available from the data and from the prior distribution.[111] This framework can be used to estimate parameters for a variety of models based on differing likelihood functions.

As Koop notes, the probability distribution p(θ|y) is "of fundamental interest for an econometrician interested in using data to learn about parameters in a model."[112] Bayesian methods focus on the interpretation and analysis of p(θ|y) to understand the relationship between θ and y.

C.5.3. Making the Decision: Comparing and Contrasting

The equations above outline the two big departures between the frequentist and Bayesian schools of thought. First, the two methods generate different results. The result of Bayesian estimation is the posterior distribution p(θ|y) and is a probability distribution for θ. There is no single value for θ, rather each value has a probability of being observed. The probability of observation is informed by the likelihood function (i.e., the data) and the prior distribution. A result with higher variances indicates increased uncertainty about the probability of any single value. This distribution can be summarized through statistics such as the mean, median, or variance, but the fundamental result is a probability distribution.

This is subtly different than the frequentist result, which is a point estimate of the "true" value of β. That is, in frequentist statistics, β has a value that can be determined to some precision given the data (an estimate of β), and there is variance around that point that can be characterized as a function of the data. A wider variance around β implies less certainty about the estimate, much as increased variance in a Bayesian posterior implies increased uncertainty about each possible value. For frequentists, the

110 Bayesian econometrics often uses θ instead of β to reduce confusion when comparing the two methods.

111 Koop (2003).

112 Ibid.

fundamental result is this point estimate; this is contrasted with the Bayesian fundamental result, which is a probability distribution.

The second difference that these equations illuminate is the inclusion of prior information. The inclusion of prior information in the estimation of parameters is unique to Bayesian analysis. The inclusion of the prior is a way to introduce additional information not contained in the data into the estimation. These prior beliefs about the distribution of the parameters can be highly specific or only loosely defined. In the extreme case, the researcher can choose an uninformative prior, essentially saying that there are no prior beliefs. This is akin to specifying a distribution with infinite variance for the prior and forces the estimation to rely completely on the data. When an uninformative prior is specified, the estimation results are similar to frequentist estimations in the sense that they rely solely on the data (i.e., the likelihood function).

A final point worth making is a similarity between Bayesian and frequentist methods. Both discussions above invoke the term “likelihood.” In fact, both methods employ the same likelihood function. The likelihood in this case merely characterizes the probability of observing the data, given a set of parameters. The difference lies in how this likelihood is treated. For Bayesians, it forms one part of the posterior distribution. Frequentists seek to find the β that maximizes this function.

Comparative Characteristics of Bayesian Methods

The previous sections outlined the basic structures of the Bayesian and Frequentist frameworks and how they compare to one another. Each paradigm has practical advantages and disadvantages when compared with the other. Bayesian methods can be more informative on small samples. Bayesian analysis can also provide more theoretically pleasing estimation results.

Bayesian analysis can have some advantages where the data do not provide much information with which to estimate parameters (namely due to the lack of prior information being included in frequentist analysis). One instance of this is when examining data with small sample sizes. Xie et al. address the small sample size question and compare the results from the Bayesian analysis to a frequentist analysis.[113]

The authors performed their comparison in the context of an ordered probit model.[114] The authors find that when using the full sample of 76,994 observations, a Bayesian model with uninformative priors (i.e., $p(\theta)$ contains very little information and the data is relied upon to provide almost all of the information about θ) is almost identical to the frequentist model. A variety of other priors were fit to the entire sample and all models provided similar results to the frequentist model. The authors then examined the models on a subsample of 100 records. A frequentist model and a Bayesian model with an informative prior were fit to this small sample and compared to the full sample results. The Bayesian model with the informative prior provided results that were significantly closer to those observed on the entire sample.

---

113 Xie*, et al.* (2009).

114 More information on ordered probit models is contained in Section 4.1.

This study reveals two important points regarding the use of Bayesian in an applied sense. First, Bayesian methods can provide real gains when examining small samples. While this may not be a relevant advantage given the current objective of modeling incursion severity across the many thousands of incursions-to-date, further rounds of research may wish to analyze small subsamples. Secondly, the advantages of Bayesian hinge upon the definition of the priors. Given an uninformative prior, the Bayesian results mimicked the frequentist results. Thus, when examining runway incursion severity, a relatively unexplored field with few prior beliefs about the impacts of variables, Bayesian methods may not provide a substantial advantage.

In addition to the beneficial small sample properties, Bayesian analysis is more theoretically pleasing. As an example, consider Griffiths et al; the authors compare Bayesian estimation with a variety of priors to the standard frequentist estimation results in the context of a probit model of mortgage types.[115] In this case, the researchers used a truncated uniform prior distribution. That is, the authors had the prior belief that a coefficient is positive, and all positive values are equally likely. The mean and variance of the posterior distribution were similar to the results from the frequentist estimation. However, the Bayesian results were truncated at zero, whereas the frequentist results imply a distribution that normally distributed around the estimate, regardless of where it falls. For a variable that must be positive, this frequentist result may be incorrect. This may be especially true for variables with small effects, that is, for variables with estimated effects that are not very different from zero. The Bayesian estimates, by virtue of being truncated at zero, have a slightly different distribution – the mean and variance may be similar, but impossible values will have zero probability. Figure 67 demonstrates this graphically.

Figure 66 - Bayesian versus Frequentist Parameter Estimates

The red bar (top) displays a hypothetical Bayesian estimate. The width of the bar represents the distribution for the parameter estimated. Note that the bar is truncated at zero, indicating that the distribution of θ does not extend past zero in that direction. The blue bar (bottom) represents the variance around a frequentist point estimate, β. The variance can extend outside of the reasonable range for the parameter, in this case extending to negative values. Finally, note that the point estimate β is equal to the mean of the distribution of θ (represented by a vertical line in the bar). This need not be the case in general.

This discrepancy – truncated versus unconstrained – extends to predicted probabilities, as well. The use of a probit model confines the frequentist point estimate of the probability to be between zero and one. However, there is some variance about that point which may include illegitimate values (probabilities outside the zero to one range).[116] The predicted probabilities obtained from the Bayesian estimation were truncated at zero and one respectively, constraining results to be within the valid interval. Figure 68 provides a simplified graphical explanation of this phenomenon. Griffiths et al. note that this is not a

115 Griffiths*, et al.* (2006).

result of using a truncated prior but rather to the differences in how estimations are generated for Bayesian and frequentist methods.[117]

Figure 67 - Bayesian versus Frequentist Probability Estimates

The red bar (top) represents the probability estimate from a Bayesian estimation while the blue bar (bottom) represents that from a frequentist. The frequentist point estimate of the probability, P_f, is confined to be in the valid range of zero to one. However, the variance around this point (representing uncertainty of the estimate) can extend into unreasonable ranges. This does not invalidate the frequentist estimate, and is merely an undesirable side effect of the frequentist interpretation. The Bayesian probability estimate, P_b, is again a distribution. This distribution is truncated to remain in the valid range of zero to one.

The ability to confine predicted probabilities to the appropriate bounded interval is advantageous. Additionally, if priors about the sign but not magnitude of a coefficient exist, Bayesian methods offer superior estimation results. However, as noted earlier, few if any priors exist in the runway incursion context.[118] It is also unclear how useful predicted probabilities may be in this context. Regardless, Bayesian methods will likely provide results that are theoretically superior compared to the frequentist methods. The degree of superiority will however vary, and in some situations, can be quite small.

However, Bayesian methods are more difficult to implement than frequentist methods. First, inference about the effects of individual components of θ is difficult using the posterior distribution, leading to less clear policy direction. Further complicating matters is that $p(\theta|y)$ may not be written as a simple formula (i.e., there is no closed form for $p(\theta|y)$). In these cases, simulation is required to deduce $p(\theta|y)$, requiring additional programming, computing resources, and time.

Comparative Characteristics of Frequentist Models

Frequentist methods often are at a disadvantage where Bayesian methods are advantageous, and vice versa. Frequentist estimation, by relying solely on the data to produce results, is subject to the weakness

116 As an aside, it is important to compare this to the problems with OLS as mentioned above. OLS results are unconstrained; when predicting a probability, OLS point estimates may be outside the range of zero to one. Here, the point estimates produced by a probit model are constrained to the appropriate interval, but the uncertainty surrounding that estimate may include unreasonable values. In some sense, constraining point estimates is an improvement over the unbounded OLS estimates, even if the uncertainty may result in unwanted values for part of the interval.

117 Griffiths*, et al.* (2006), p. 8.

118 The distinction is made here between hypotheses and priors. Hypotheses are statements that are to be tested. There may be a multitude of hypotheses in the runway incursion context. Priors are beliefs that have influence over the model estimation process, and are not testable in the same way that hypotheses are. Another take on this distinction is that priors are assumed to be true in the absence of any data while hypotheses are intended to be tested with data and proven true or false.

in that data. However, frequentist methods do not require prior distributions on the parameters. This has the advantage of not requiring the researcher to specify a prior distribution when no reasonable prior expectations exist. Additionally, Bayesian estimation with an uninformative prior essentially collapses to the frequentist estimate. That is, for a Bayesian without any information from a prior distribution, only information in the data can be used to estimate a result, which is exactly the frequentist technique.

Frequentist methods also have advantages in terms of implementation. Many common statistical packages implement frequentist methods for the models under consideration. Though they may require significant computing power, the requirements are substantially less than those required by Bayesian methods with simulation. The availability of "canned" implementations of frequentist methods also allows different model specifications to be tested quickly. Conversely, a significant portion of resources would be dedicated to implementing Bayesian methods, restricting the focus to a single model with one or two sets of explanatory variables that, given the lack of informative priors for runway incursions, would likely return the same results as frequentist methods.

C.5.4. Conclusion

Both Bayesian and frequentist schools of thought have their merits. Frequentist methods result in point estimates of parameters and are easily implemented. However, frequentist methods do not allow the researcher to include any information that is not in the data, and may suffer from poor performance on small samples. Bayesian methods result in a distribution of a parameter, have improved small sample properties, and allow for the inclusion of additional information. Bayesian methods offer theoretical improvements; however, without strong priors, these theoretical improvements are mitigated. Implementation is a major concern for Bayesian methods, likely requiring a significant resource investment to get the estimation working properly.

C.6. Extensions to the Multinomial Logit Model

The multinomial probit has been suggested as an alternative to the multinomial logit specification, primarily to avoid the unfavorable IIA property. However, due to the computational concerns regarding the multinomial probit, other alternatives have been developed. The two major developments have been nested logit models and random parameter models.

Nested models are one way to address the IIA property of multinomial logits. Nested models achieve this by grouping the choices into several subsets with similar unobserved differences.[119] Thus, unobserved differences are similar across groups but not between groups. This avoids the unwanted IIA property introduced into the multinomial logit framework. It is important to note that the nesting is merely a statistical artifact and does not imply any sort of decision tree.[120] The interpretation of a decision tree is a behavioral one imposed by the researcher, but is not reflected in the underlying

119 Washington*, et al.* (2011), p. 335.

120 Ibid., p. 338.

statistics. Nested logit models are useful in studying mode choice as it can account for similar unobserved effects between choices (such as between busses and trains).[121] A nesting approach may be useful for modeling runway incursion severity if the different categories have significantly different unobserved effects. For example, a nesting structure with two branches, one with C and D while the other contained A and B, may be applicable.

Another alternative that attempts to relax the IIA assumption of the multinomial logit is the random parameters model.[122] Essentially, this model allows the parameters of the top-level model to vary in a systematic way. That is, a model of mode choice could estimate a coefficient for mode price. That coefficient could then be allowed to vary in a systematic way with education, income, and other variables. This allows the model to be extremely flexible in terms of the correlation structure of the random disturbance terms. However, this model can be difficult to implement. For a more complete discussion, see Greene.[123] Baht and Gossen provides an example of an implementation of this model.[124]

121 Forinash and Koppelman (1993).

122 Greene (2003), p. 728.

123 Ibid.

124 Bhat and Gossen (2004).

APPENDIX D: FUTURE RESEARCH

- Understand the relationship between incident type (OE/PD/VPD) and severity
- Why departures/arrivals on intersecting runways are associated with more serious incursions
- Why departures/arrivals on intersecting runways are more likely to be OEs than PDs
- Use data on number of operations per controller or pilot to understand error rate
- Why LAHSO operations appear to have fewer than expected incursions despite being a riskier operation
- Policy/training implications: why incidents during takeoff are more likely to be OEs than during landing
- Why commercial carriers are involved in less severe incursions despite operating in more complex conditions and locations
- How the impact of commercial carrier status varies with OE and PD incursions
- Cause for the lack of incursions among experienced pilots
- Policy implications: changes to training for experienced pilots or identification of poor quality pilots early
- Investigate the nature of the ordering (if any) of severity between C and D events.
- Models of incursion frequency (rather than severity) may shed light on how other variables impact safety
- Refine and clarify traffic complexity measures
- Better understand differences in controllers between OEP 35 and Non-OEP 35 airports
- Better understand differences relationship between LAHSO capability and incident type
- Understand the relationship between severity and LAHSO capability
- Cause or nature of the relationship between who identifies an incident and severity
- Relationship between time on shift and frequency of incursions
- How changes to operations in adverse weather interact with changes in risk due to the weather
- Cause of increase in V/PDs in cold weather
- Relationship between higher dew points and OE events
- Potential relationship between dew point and conflict events
- Disentangle effects of various visibility-related measurements (i.e., visibility, ceiling, cloud coverage)

- Determine if snow removal vehicles are in more severe incidents that other V/PDs due to runway access alone
- Describe the relationship between nighttime operations, controller actions, and incident severity
- Understand the relationship between “good” weather, controller behavior, and severity
- Understand the relationship between high pressure, controller behavior, and severity
- Further research into pilot instrument ratings should account for the three rating groups (current, past, and never rated) and further investigate whether current and past ratings have the same impact on severity

APPENDIX C: SUMMARY OF MODELING RESULTS

The following table summarizes the results presented in the body of the paper. Rows represent different variables while the different columns represent the variety of tests and models detailed in the report. The general direction of the effect is given. If the estimated relationship had a p-value less than 0.10, it is reported as an X in the table, indicating it was included in the model or a test was performed, but it is deemed insignificant. Empty cells represent that no test was run or that the variable was not included in the model. Positive indicates that increasing values of the variable (or in the case of binary variables, being coded as a yes) increase the severity of an incursion. Negative indicates that opposite – increasing values decrease the severity of the incursion.

Variable	Chi2/Exact or Kruskal-Wallis by Severity	Simple Logit: Odds Ratio	Ordered Logit: Coefficient	Binary Logit: Odds Ratio	Multinomial Logit
ARTS II	X		X	X	
ARTS III	X		X	X	
ASDE	Related	Negative	X	X	
Cloud Ceiling	Related				
Cloud Coverage	Related		Negative	Negative	
Cloud Coverage X Sea Level			Positive	X	
Commercial Carrier		X	Negative	Negative	
Commercial Carrier , Conflict Only		Negative			
Controller Age	X		X	X	Unchanged
Controller Time on Shift	X		X	X	A increases B unchanged C decreases D unchanged
Controller Workload	Related		Positive	X	A increases B & C unchanged D decreases
Daily Operations	Related				
Daily Operations (Aircraft Model)			X	X	Unchanged
Daily Operations (Airport Model)			Positive	Positive	
Daily Operations (Controller Model)			X	Positive	A & B unchanged C increases D decreases
Daily Operations (Radar Model)			Positive	X	A & B unchanged C increases

Variable	Chi2/Exact or Kruskal-Wallis by Severity	Simple Logit: Odds Ratio	Ordered Logit: Coefficient	Binary Logit: Odds Ratio	Multinomial Logit
					D decreases
Daily Operations (Weather Model)			Positive	X	
Dew Point	X				
Differences of AC/AT and GA Percents			X	X	A increases B unchanged C increases D decreases
Employee Alerted to Incident By	Related				
Employee Alerted to Incident By Pilot, Conflict Only		Positive			
Entered Runway Without Clearance	Related				
Evasive Action Taken	Related				
Evasive Action Taken , A & B Only	X				
Foreign Aircraft or Pilot	Related				
Intersecting Runway Departure or Arrival	Related	Positive			
Land and Hold Short Capability at Airport	X				
Landed/Departed on Closed Runway or Taxiway	Related				
Landed/Departed without Clearance Communication	Related	X			
Landed/Departed without Clearance Communication , Conflict Only	Related	Positive			
Night	Related	Positive			
No Weather Phenomena Indicator	Related		X	Negative	
Number of Aircraft Involved	X		Positive	Positive	A & B increase C decreases
Number of Hotspots	Related		Negative	Negative	A unchanged B decreases C unchanged D increases
Number of Runway Intersections	Related		Positive	Positive	A & B increase C decreases D unchanged

Variable	Chi2/Exact or Kruskal-Wallis by Severity	Simple Logit: Odds Ratio	Ordered Logit: Coefficient	Binary Logit: Odds Ratio	Multinomial Logit
Number of Runways	Related		Negative	Negative	A decreases B & C unchanged D increases
OEP 35 Airport Status		Positive			
OEP 35 Airport Status, Conflict Only		X			
Part 139 Airport Status	Related				
Part 139 Airport Status, Conflict Only	X				
Percent of Operations that are Air Carrier/Air Transport	Related		X	X	Unchanged
Phase of Flight: Landing		Positive	X	Positive	
Phase of Flight: Takeoff		Positive	Positive	Positive	
Pilot Instrument Rating	Related				
Pilot Instrument Rating , Conflict Only	Related				
Pilot Instrument Rating: Rated, but not Current		Negativ e			
Pilot Instrument Rating: Current Rating		Negativ e			
Pilot Lost	X				
Pilot Ratings	Related				
Sea Level Pressure Deviation	X		Negative	Negative	
Snow Removal Vehicle Involved	X				
Snow Removal Vehicle Involved, V/PD Only	Related				
Special Procedures	X				
STARS	Related	Negativ e	Negative	Negative	
STARS & ASDE		X	Positive	X	
Taxiing Out for Departure	Related				
Temperature	Related				
Temperature-Dew Point Difference	Related				
Traffic Complexity Code	Related				
Training in Last Year	X		X	X	

Variable	Chi2/Exact or Kruskal-Wallis by Severity	Simple Logit: Odds Ratio	Ordered Logit: Coefficient	Binary Logit: Odds Ratio	Multinomial Logit
Visibility	Related				
Visual Meteorological Conditions		Negative			
Weather					
Wind Speed	Related		X	X	

BIBLIOGRAPHY

Abdel-Aty, Mohamed, "Analysis of Driver Injury Severity Levels at Multiple Locations Using Ordered Probit Models," Journal of Safety Research, 34 (2003), pp. 597-603.

Bhat, Chandra R., and Rachel Gossen, "A Mixed Multinomial Logit Model Analysis of Weekend Recreational Episode Type Choice," Transportation Research Part B: Methodological, 38:9 (2004), pp. 767-787.

Cardosi, Kim, and Alan Yost, Controller and Pilot Error in Airport Operations: A Review of Previous Research and Analysis of Safety Data, Springfield, VA, National Technical Information Service (2001).

DiFiore, Amanda, and Kim Cardosi, Ph.D, Volpe National Transportation Systems Center, Human Factors in Airport Surface Incidents: An Analysis of Pilot Reports Submitted to the Aviation Safety Reporting System (ASRS), prepared for: Federal Aviation Administration, Office of Runway Safety & Operational Services, DOT-VNTSC-FAA-06-14 (2006).

Dow, Jay K., and James W. Endersby, "Multinomial Probit and Multinomial Logit: A Comparison of Choice Models for Voting Research," Electoral Studies, 23:1 (2004), pp. 107-122.

Federal Aviation Administration, "Runway Safety Office Runway Incursion Database," (12/6/2010, 2010).

Federal Aviation Administration, Annual Runway Safety Report 2008, (2008).

Federal Aviation Administration. Runway Safety - Hot Spots List. http://www.faa.gov/airports/runway_safety/hotspots/hotspots_list/. Last Accessed: 9/11/2012.

Federal Aviation Administration. Runway Status Lights Questions and Answers. http://www.faa.gov/air_traffic/technology/rwsl/faqs/. Last accessed: 8/1/2012.

Forinash, Christopher V, and Frank S Koppelman, "Application and Interpretation of Nested Logit Models of Intercity Mode Choice," Transportation Research Record, Transportation Research Board, Innovations in Travel Behaviour Analysis, Demand Forecasting, and Modeling Netoworks:1413 (1993), pp. 98-106.

Government Accountability Office, Aviation Safety: FAA Has Increased Efforts to Address Runway Incursions, Testimony Before the Subcommittee on Aviation, Committee on Transportation, and Infrastructure, House of Representatives, GAO-08-1169T (2008).

Greene, William H, Econometric Analysis: Fifth Edition, Upper Saddle River, New Jersey: Pearson Education, Inc. (2003).

Griffiths, W. E., R. Carter Hill, and Christopher J. O'Donnell, "A Comparison of Bayesian and Sampling Theory Inferences in a Probit Model," (2006).

Horowitz, Joel L., "Reconsidering the Multinomial Probit Model," Transportation Research Part B: Methodological, 25:6 (1991), pp. 433-438.

Horowitz, Joel, "The Accuracy of the Multinomial Logit Model as an Approximation to the Multinomial Probit Model of Travel Demand," Transportation Research Part B: Methodological, 14:4 (1980), pp. 331-341.

International Civil Aviation Organization, Manual on the Prevention of Runway Incursions, International Civil Aviation Organization. (2007).

Islam, Samantha, and Fred Mannering, "Driver Aging and Its Effect on Male and Female Single-Vehicle Accident Injuries: Some Additional Evidence," Journal of Safety Research, 37:3 (2006), pp. 267-276.

Islam, Samantha, and Fred Mannering, "Driver Aging and Its Effect on Male and Female Single-Vehicle Accident Injuries: Some Additional Evidence," Journal of Safety Research, 37:3 (2006), pp. 267-276.

Kockelman, Kara Maria, and Young-Jun Kweon, "Driver Injury Severity: An Application of Ordered Probit Models," Accident Analysis & Prevention, 34:3 (2002), pp. 313-321.

Koop, Gary, Bayesian Econometrics, West Sussex, England: John Wiley & sons Ltd (2003).

Lam, Lawrence T., "Factors Associated with Young Drivers' Car Crash Injury: Comparisons among Learner, Provisional, and Full Licensees," Accident Analysis and Prevention, 35 (2003), pp. 913-920.

Lauer, Charlotte, "Family Background, Cohort and Education: A French-German Comparison Based on a Multivariate Ordered Probit Model of Educational Attainment," Labour Economics, 10:2 (2003), pp. 231-251.

National Weather Service Weather Forecast Office. Glossary. http://www.crh.noaa.gov/ddc/?n=glossary. Accessed: 8/1/2012

Nolan, Michael. Fundamentals of Air Traffic Control, 5th Edition. Clifton Park, NY. Delmar. (2011)

O'Donnell, C. J., and D. H. Connor, "Predicting the Severity of Motor Vehicle Accident Injuries Using Models of Ordered Multiple Choice," Accident Analysis & Prevention, 28:6 (1996), pp. 739-753.

Perera, Loshaka, and Sunanda Dissanayake, "Contributing Factors to Older-Driver Injury Severity in Rural and Urban Areas," Journal of the Transportation Research Forum, 49:1 (2010), pp. 5-22.

Quilty, Stephen M., Preventing Vehicle-Aircraft Incidents During Winter Operations and Periods of Low Visibility, prepared for: Federal Aviation Administration, Airport Cooperative Research Program, Transportation Research Board, (2008).

Rice, John A. Mathematical Statistics and Data Analysis, Third Edition. Belmont, CA. Thomson Higher Education. (2007).

Scarborough, Alfretia, et al., Analyzing Vehicle Operator Deviations, Washington, DC, Federal Aviation Administration (2008).

Schneider IV, William H, et al., "Driver Injury Severity Resulting from Single-Vehicle Crashes Along Horizontal Curves on Rural Two-Lane Highways," Transportation Research Record, (2009), pp. 85-92.

Schneider IV, William H, et al., "Driver Injury Severity Resulting from Single-Vehicle Crashes Along Horizontal Curves on Rural Two-Lane Highways," Transportation Research Record, (2009), pp. 85-92.

Siegel S. & Castellan J. Nonparametric Statistics for the Behavioral Sciences, Chapter 8. McGraw-Hill. (1988).

Washington, Simon P., Matthew G. Karlaftis, and Fred L. Mannering, Statistical and Econometric Methods for Transportation Data Analysis: Second Edition, Boca Raton, FL: Chapman & Hall/CRC (2011).

Xie, Yuanchang, Yunlong Zhang, and Faming Liang, "Crash Injury Severity Analysis Using Bayesian Ordered Probit Models," Journal of Transportation Engineering, (2009), p 7.