NASA/TP–2012-216013

The National Aviation Operational Monitoring Service (NAOMS): A Documentation of the Development of a Survey Methodology

Mary M. Connors
NASA Ames Research Center

Robert Mauro
Decision Research
Eugene, Oregon

Irving C. Statler
Ames Associate
NASA Ames Research Center

March 2012

NASA STI Program...in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NASA Aeronautics and Space Database and its public interface, the NASA Technical Report Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA Programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counter-part of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include organizing and publishing research results, distributing specialized research announcements and feeds, providing help desk and personal search support, and enabling data exchange services.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at http://www.sti.nasa.gov

- E-mail your question via the Internet to help@sti.nasa.gov

- Fax your question to the NASA STI Help Desk at 443-757-5803

- Phone the NASA STI Help Desk at 443-757-5802

- Write to:
  NASA STI Help Desk
  NASA Center for AeroSpace Information
  7115 Standard Drive
  Hanover, MD 21076-1320

NASA/TP–2012-216013

The National Aviation Operational Monitoring Service (NAOMS): A Documentation of the Development of a Survey Methodology

Mary M. Connors
NASA Ames Research Center

Robert Mauro
Decision Research
Eugene, Oregon

Irving C. Statler
Ames Associate
NASA Ames Research Center

National Aeronautics and
Space Administration

NASA Ames Research Center
Moffett Field, California

March 2012

The use of trademarks or names of manufacturers in this report is for accurate reporting and does not constitute an official endorsement, either expressed or implied, of such products or manufacturers by the National Aeronautics and Space Administration.

Available from:

NASA Center for Aerospace Information
7121 Standard Drive
Hanover, MD 21076-1320
301) 621-0390

This report is also available in electronic form at http://www.sti.nasa.gov
or http://ntrs.nasa.gov/

Table of Contents

List of Figures

List of Tables

Glossary and Acronyms

AIDS ....Accident/Incident Data System (FAA)
ALPA ....Air Line Pilots Association
APA ....Allied Pilots Association
APMS......................Aviation Performance Measuring System
ASAP ....The Aviation Safety Action Programs
ASIST ....Aviation Safety Investment Strategy Team
ASRS ....Aviation Safety Reporting System (FAA/NASA)
ASMM ....Aviation System Monitoring and Modeling (component of AvSP)
ATC ....Air Traffic Control
ATP.........................Airline Transport Pilot
AvSP ....Aviation Safety Program (NASA) (succeeded by AvSSP)
AvSSP.....................Aviation Safety and Security Program (NASA)
BTS .........................U.S. Bureau of Transportation Statistics
CAST ....Commercial Aviation Safety Team
CATI ....Computer-Administered Telephone Interview
C-NAS ....Commercial National Airspace System
CPHRE....................Centers for Public Health Research and Evaluation (Battelle)
FAA ....Federal Aviation Administration
FOQA......................Flight Operational Quality Assurance
GA...........................general aviation
GAO........................Government Accountability Office
GTOW.....................Gross Takeoff Weight
ICAC.......................In-Close Approach Changes
JIMDAT..................Joint Implementation Measurement Data Analysis Team
MEL ....Minimum Equipment List
NAIMS....................Natinal Airspace Information Monitoring System
NAOMS ....National Aviation Operational Monitoring Service
NAS ....National Airspace System
NASA......................National Aeronautics and Space Administration
NextGen ....Next Generation Air Transportation System
NMACS ....Near Midair Collision System
NRC ....National Research Council
NTSB ....National Transportation Safety Board
OEDS ....Operational Error and Deviation System
OMB ....Office of Management and Budget
PDARS....................Performance Data Analysis and Reporting System
PDS ....Pilot Deviation System
RIS ....Runway Incursion System
SAQ ....Self-Administered Questionnaire
SDRS ....Service Difficulty Reporting System
SE............................Safety Enhancement
VPDS ....Vehicle/Pedestrian Deviation System

Executive Summary

The National Aviation Operational Monitoring Service (NAOMS) project was formulated to address a need that had been expressed by many aviation stakeholders – for a method to routinely monitor the status of the National Airspace System (NAS) and to identify potential problems as they arise. The goal of NAOMS was to determine if this need could be met through the use of a comprehensive, system-wide, statistically–sound survey approach.

NAOMS was developed as an element of the Aviation System Monitoring and Modeling (ASMM) project of NASA's Aviation Safety Program (AvSP) during the years from 2000 to 2005. ASMM was designed to provide the technologies to facilitate a data-driven basis for proactive safety decisions. The ultimate goal of the ASMM was to identify, and in some cases correct, predisposing conditions that that could lead to accidents and thereby escape from the retroactive, accident-based policy-making syndrome,

Fundamentally, NAOMS was a research project. Its primary purpose was to develop and assess the utility of a survey methodology for detecting relative changes over time in the frequency of safety-related events in the NAS. The NAOMS project sought to assess the value and limitations of this concept, to uncover potential flaws in the approach, and, if demonstrated to be viable, to provide a basis for developing a continuing monitoring service that would serve the aviation system over the long term.

NAOMS was designed to provide a statistically sound basis for detecting changes over time in the occurrence of events that might compromise safety by surveying all of the 'front-line' constituents of the aviation system and fusing the information provided by these different operator groups into a system-level perspective. This approach would monitor the performance and safety of the overall national air transportation system and detect and evaluate the effects of new technologies or procedures as they are inserted into that system by seeking reports from the operators of the system (i.e., from flight crews, air traffic controllers, cabin crews, mechanics, etc.). NAOMS was not designed to measure safety of the air-transportation system in any absolute sense; nor was it designed to identify the precursors or the causal factors of unsafe events. Its primary purpose was to assess the effects introduced by changes in technologies, procedures, or training on system safety by monitoring the experiences of operators. NAOMS was intended to measure event-based experiences of reporters and to provide early indications of changes that deserve further investigation. Such further investigation would rely on other sources of information.

This report is a documentation of the development of the survey methodology and a demonstration of its capabilities. Its focus is on the identification and resolutions of issues; including meeting the requirements to achieve data reliability, establishing the research protocols, resolving selection issues, designing the survey instrument, deciding on data-collection procedures, and developing a plan for analyzing the data. Also reported are the results of a NAOMS survey of air-carrier pilots that was conducted to demonstrate the sort of information NAOMS could provide. Although the demonstration was limited to a survey of the air-carrier pilots, the issues addressed are relevant for the design and implementation of surveys of other constituents of the air-transportation system. However, the resolutions of these issues will likely differ for each constituent group.

Survey Development

One measure of the methodological success of a survey is respondent acceptance. During the forty-five months of this initial experiment, a total of 26,170 survey interviews were conducted. The air-carrier-pilot response to the NAOMS survey was enthusiastic. Respondents' very positive feedback served to increase the NAOMS team's confidence in its approach.

A number of air-carrier pilots were excluded from the sample because they could not be located or because they were ineligible. Of the pilots contacted and found to be eligible, NAOMS achieved a survey response rate of 81%. The exceptional response rate achieved by NAOMS, far exceeding that of typical survey efforts, indicates the quality of the survey design as well as the NAOMS team's effectiveness in enlisting the participation of the aviation community.

During the first 2 years of the project, the NAOMS team concentrated on resolving methodological issues in order to establish a solid survey foundation, to maximize data quality, and to ensure the validity of the statistical results. A number of questions required thoughtful resolutions before the air-carrier-pilot survey could be launched. These questions included:

1. How should respondents be selected, located, and engaged for the survey?
2. What sample size would be needed to achieve desired levels of precision?
3. How far back in time can respondents accurately recall safety events?
4. How do pilots organize their recollection of events?
5. What is the best survey-collection mode to use?
6. Would a purely random or a panel sampling design work best?

These questions and the extensive research accomplished to reach an answer for each are discussed in detail in this report. As a result of these investigations, a source of potential interviewees was identified, the needed sample size was established, and an ordering to facilitate pilot recollection of events was determined. Also, as a result of these investigations, decisions were made to utilize a telephone-based survey mode, a recall period of 60 days, and a random design for selecting interviewees.

Another major activity during the first 2 years centered on the content and design of the survey instrument itself. The NAOMS team needed to determine the types of events to include. Extensive input from the aviation community aided these determinations. The general framework of the survey reflected its purposes. Issues concerning the overall design were resolved early in the process, while many of the internal ordering and structuring questions awaited the results of the experimental investigations. It was determined that the framework of the questionnaire should comprise four sections. Section A measured baseline information for establishing event rates. Section B captured the experiences of the reporters with safety-related events over time. Section C was designed to shine a "spotlight" on a particular area of interest to the aviation community at a particular time. Section D collected feedback information from respondents. Section B was the "heart" of the survey. In this section, the questions about events were expected to remain relevant for many years, thus providing a basis for long-term trend analyses, which was the primary objective of the NAOMS project. The statistical reliability of the collected data was assured by careful design, and was demonstrated in subsequent analyses.

Results

The data-gathering period of the study was sufficient to reliably measure trends for about half (43 of 91 questions with numeric answers) of the questions posed in Section B. The remaining 48 events addressed by questions in Section B occurred too rarely to collect sufficient data within the limited time period of the demonstration for reliable statistical analysis of change. The results of analyses of these 43 questions are presented in Section 7 of this report. They illustrate the capabilities of the NAOMS survey methodology and provide evidence that the NAOMS survey can reliably identify relative changes over time in the rates of occurrence of safety-relevant events.

Although data were gathered over a 45-month period, 9 months were used to conduct tests needed to resolve methodological issues. Most of the analyses in this report are based on the last 36-months of data collection. Changes over time were measured over this period for data grouped into (12) quarters. Of the 43 questions from Section B whose results were analyzed, 27 showed no significant differences in event rates over the duration of the survey. Fourteen questions showed reductions in event rates over this time. Only 2 of the 43 questions (use of reserve fuel and expedite or divert landing due to medical emergency) showed increasing rates of events over the time period of the survey. These results indicate that when changes over time occurred, they were predominantly in the direction of increased safety.

In addition to assessing linear trends over time, year-to-year changes were examined. Seventeen questions revealed significant year-to-year changes. For most questions, these changes mirrored the linear trends by quarters. However, changes across time could also follow a non-linear pattern. For 3 questions in which there were no significant linear trends, there were significant year-to-year non-linear effects.

The time period represented by each quarter approximated a season of the year, providing an opportunity to assess factors related to seasonal weather (e.g., icing, thunderstorms) or other factors (e.g., tourist travel) that vary regularly across the year. Statistically significant seasonal effects were observed for 21 of the 43 questions analyzed. One interesting demonstration of seasonal effects relates to the cyclical pattern shown in the data from the question involving bird strikes. Here, events reported were highest in summer and fall, lower in spring than in summer, and still lower in winter, with no year-to-year variation.

Although the primary purpose of NAOMS was to demonstrate a methodology for assessing changes over time, the NAOMS survey methodology also proved capable of revealing other important findings in the data that were unrelated to time changes. The aircraft flown by pilots in the NAOMS survey were classified into 4 size categories: small, medium, large, and wide-body. Notable differences often distinguished these aircraft groups. Significant differences among aircraft-size categories were found for 36 of the 43 questions analyzed. For many events, it is the small and/or the wide-body aircraft pilots who experienced the highest event rates.

However, the differences among aircraft size categories were complex, reflecting the differing operations performed by these aircraft.

Reports were further classified according to type of operation (cargo or passenger) in which the reporting pilot was engaged. Of the 43 questions analyzed, 4 were specific to events involving passengers, and, therefore resulted in very few, or no, reports of those events from cargo pilots. For 13 events, there were no differences between operation types. For 20 events, rates were higher in cargo operations than in passenger operations. For 6 events, rates were higher in passenger operations than in cargo operations. For both the aircraft-size categories and the passenger/cargo categories, the NAOMS survey revealed differences for events that were stable over time, as well as for events that changed over time.

While the results from Section B was the prime interest to the demonstration of the NAOMS survey, the survey instrument also included a section (Section C) that could supply quick-turnaround information on a particular topic of current industry concern. The topic of this section was envisioned to change as new pressing questions arose. For much of the NAOMS project, the topic-of-interest in Section C was In-Close-Approach-Changes (ICAC), a subject of particular interest at the initiation of the NAOMS study. The results from this survey section revealed the frequency with which ICAC requests were made, the air-carrier pilots' responses to these requests, and the aftermath of making in-close approach changes on the remainder of the flight. A second topic for Section C (that included inquiries on training, protocols and procedures) was initiated later in the NAOMS study period to address needs advanced by the Commercial Aviation Safety Team's (CAST) subgroup, the Joint Implementation Measurement Data Analysis Team (JIMDAT). Representatives of the JIMDAT and representatives of the NAOMS team worked collaboratively to produce and integrate into the main survey a new version of Section C that reflected the JIMDAT's particular information needs. This experiment proved successful in capturing information that had been judged to be unattainable through other approaches.

The CAST/JIMDAT expressed interest in the operation of a survey of Part 121 pilots, but NASA could not commit to continuing the service after the expiration of NASA's Aviation Safety Program. The Air Line Pilots Association (ALPA) offered to perform the service provided that NASA could develop a less costly alternative to telephone interviews. To address this issue, the NAOMS team developed and tested a Web-based version of the NAOMS survey, which was transferred to ALPA to be operated on behalf of the CAST.

The Web-based survey itself was well received by the respondents. However, preliminary data collected in the limited time that could be devoted to this effort prior to the conclusion of the NASA project resulted in a lower response rate, leaving open the question of whether the Web-based approach could produce the same high-quality data as the telephone survey.

Future Applications

The NAOMS research project demonstrated that changes over time, across the system, could be measured reliably using a carefully crafted survey methodology to learn of the experiences of the system's operators. Information from such a survey could be the first indication of a developing new situation or trend deserving of investigation that could be accomplished through the use of other sources of relevant information. For example, the results reported in Section 7 of this report show that the rate of instances in which pilots reported beginning the takeoff roll while another aircraft was on the runway decreased between 2002 and 2003. There is no evidence of interaction of this trend with either aircraft-size category or type of operation. This pattern suggests that a discrete system-wide intervention may have been the cause. What could it have been? The rate of these events appears higher in small and medium aircraft and in passenger operations. Why? Another example is the rate that pilots reported utilizing their reserve fuel supplies. This rate increased linearly across the observation period. This pattern suggests a slowly changing cause. This increase in the use of reserve fuel was observed in medium, large, and wide-body aircraft but only for passenger operations. The increase in the use of reserve fuel could reflect changes in economic conditions or weather patterns or other factors - but why was it not observed in cargo operations? NAOMS was not designed to answer these questions, but it can identify the foci for further investigation.

To move beyond research to an operational implementation of NAOMS, some changes would need to be made in the methodology, but many of these changes would follow naturally. The selection criteria used in the NAOMS experiment for air-carrier pilots was a compromise, dictated by project realities that resulted in some otherwise eligible respondents not being included in the selection pool. The NAOMS team attempted to determine if this limitation affected the results. For the great majority of questions, no effect of these criteria was observed on the temporal pattern of responses. A further limitation in the selection pool resulted from the absence of pilots who chose to remove their names from the public registry, an option that became available after the NAOMS project began. While data were not available to examine temporal effects related to the opt-out option, no evidence was found in the experience level of pilots or in differences in their responses to questions in Section B to indicate that the option to opt out of the public registry created a substantial bias in the NAOMS data results. However, in a full-scale operational implementation of a NAOMS-like survey, neither the NAOMS selection criteria nor the “opt-outs” from the public registry would be an issue, as the participant pool would be selected from the full FAA Airman Registry and based on prevailing relevant requirements for the particular user group (e.g., air-carrier pilot, air traffic controllers, etc.). Nor would the ability to include the analyses of rare events be an issue. For the NAOMS study, the data-gathering period was insufficient to obtain reliable results for rare events. This limitation would not exist in an operational application of NAOMS in which data gathering would be continuous.

Summary

Over the years, a need has been expressed for a methodology to routinely assess the status of the National Airspace System (NAS) with the goal of identifying factors that could affect the safety of its operations. The results reported here demonstrate that a statistically reliable, scientifically sound survey of the various operators of the aviation system could be that methodology. Through this research effort, the NAOMS project has laid a solid foundation for developing a continuing monitoring service to capture the status of the NAS and to identify changes that could compromise safety.

1. Introduction

1.1 Background

In response to a Presidential declaration of a national goal to reduce the fatal aircraft accident rate by 80% within 10 years after the tragedy of TWA 800 in February 1997, the National Aeronautics and Space Administration (NASA) undertook a new program in aviation safety. In collaboration with the Federal Aviation Administration (FAA) and the National Transportation Safety Board (NTSB), NASA formed the Aviation Safety Investment Strategy Team (ASIST), which organized a series of conferences with participants from all sectors of the aviation industry to examine the options and recommend an approach for NASA to develop the enabling technologies that addressed the President's goal. The result of these deliberations was a three-thrust program covering:

1. *Accident Prevention:* Preventive technologies to eliminate accident precursors.
2. *Accident Mitigation:* Stay-alive technologies to decrease fatalities in survivable accidents.
3. *Aviation System-Wide Monitoring and Modeling:* Technologies to identify existing accident precursors in the aviation system, and to forecast and identify potential safety issues to guide the development of future safety technology.

Following the ASIST workshops, formative planning and preliminary studies of approaches were conducted during the years of FY98 and FY99. NASA's 5-year Aviation Safety Program (AvSP) was initiated in FY00 with the goal of developing technologies that, depending upon implementation, could reduce the aircraft accident rate by a factor of five within ten years, and by a factor of ten within twenty years.[1]

One of the projects within the AvSP, the Aviation System Monitoring and Modeling (ASMM) project, addressed the need to provide decision makers with system-wide analysis tools for identifying and correcting the predisposing conditions that could lead to accidents. System-wide analysis is a concept with extraordinary potential benefits and challenges because of the complexity and diversity of objectives and relationships among the manufacturers, users, and regulators of the modern aviation transportation system.

ASMM was primarily concerned with aiding the decision makers in gaining insight into the health and safety of the national air-transportation system. This would be accomplished by providing technologies to facilitate efficient, comprehensive, and accurate analyses of data collected from various sources throughout the aviation system during normal daily operations. ASMM addressed the need of aviation policy makers for reliable measures of events related to aviation safety, particularly as capacity becomes constrained and the skies become more crowded. As portrayed in Figure 1, fatal accidents are only a portion of the data relating to overall aviation safety. The focus of the ASMM project was to ensure that precursors of the next potential accident as indicated by the incidents and problematic trends in routine operations are reliably detected, identified, assessed, and managed.

[1] After 9/11, the Aviation Safety Program (AvSP) became the Aviation Safety and Security Program (AvSSP).

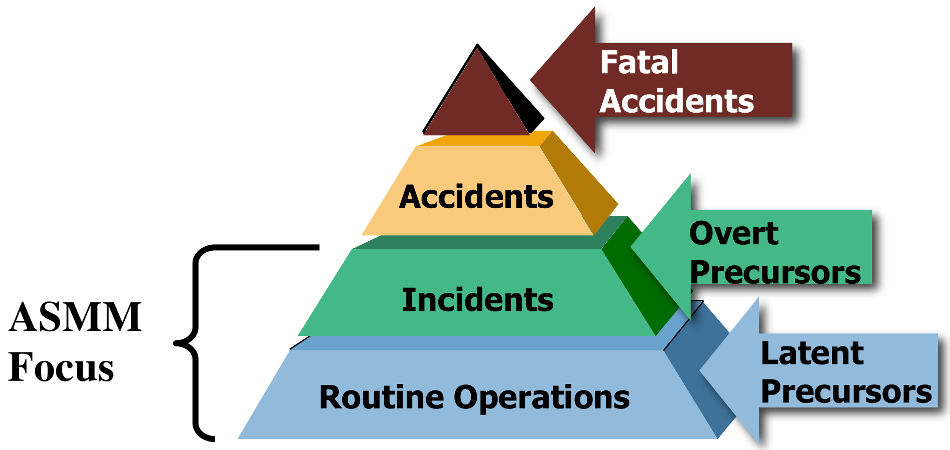


Figure 1. Aviation System Monitoring and Modeling (ASMM) focus on precursors.

The ASMM project developed computational tools that provide decision-makers with regular, accurate, insightful information on the performance and safety of the aviation system. The ability to monitor continuously, convert the collected data into reliable information, and share that information for collaborative decision making is the basis for a revolutionary, proactive approach to managing the aviation-transportation system for the prevention of accidents.

The ASMM objectives were to develop the technologies to:

1. Identify causal factors, accident precursors, and off-nominal conditions in the aviation data.
2. Provide health, performance, and safety information to decision makers.
3. Ensure seamless aviation information services.

The ASMM project consisted of the following four elements:

The Data Analysis Tools Development element developed capabilities to automatically extract information from large databases of textual and numerical data and to present meaningful displays of that information. These data-analysis tools were developed to meet the needs of the Intramural Monitoring and Extramural Monitoring elements described next.

The *Intramural Monitoring* element provided individual air-service operators (i.e., air carriers and air traffic control facilities) with the tools needed to monitor their own performance continuously, effectively, and economically within their own organizations. This was the 'bottom-up' part of a two-phase strategy for monitoring the system. The goal was that eventually information extracted from these various data sources would be shared to gain insight into the performance and safety of the national air-transportation system.[2]

The *Extramural Monitoring* element was the 'top-down' part that complemented Intramural Monitoring. It developed a comprehensive system-wide statistically sound survey as an

[2] That goal was not achieved in the time frame of this project because it was not possible to overcome the concerns of owners of data for potential misuse of proprietary data.

approach for monitoring the performance and safety of the overall national air-transportation system and for detecting and evaluating the effects of new technologies or procedures as they are inserted into the system. The method employed was to seek the perspectives of the front-line operators (i.e., flight crews, air traffic controllers, cabin crews, mechanics, etc.). This concept was called the National Aviation Operational Monitoring Service (NAOMS).

The fourth element of the ASMM project was the *Modeling and Simulations* element that addressed the need to support predictions and safety risk assessments by developing and validating system-wide models and simulations.

Although the products of these four elements of the ASMM project all have stand-alone value and many have, in fact, been deployed to the industry as such, they had been planned and developed in concert to be complementary, interdependent, and interrelated. The information extracted by the tools of these four elements could be merged into a system-wide framework enabling aviation policy makers to collaborate in aviation safety-risk management. This information was to be shared while respecting the proprietary rights to some sources of data and sensitivities to potential misuse should they be released outside the owning (i.e., commercial) organization. While this was not achieved during the life of the ASMM project, it is being addressed in activities initiated by the industry after the Aviation Safety and Security Program (AvSSP) ended.

The present report focuses on the development of the NAOMS survey methodology. For more information on the ASMM Project generally, the reader is referred to ASMM Project Plan (AvSSP, 2004) and Statler. 2007, which documents the achievements of the ASMM project during the life of the AvSP/AvSSP from FY00 to FY05.

The goal of the NAOMS research project was to develop survey methodology to support a continuing, comprehensive and coherent survey of all of the operators of the aviation system (i.e., its pilots, controllers, mechanics, flight attendants, and others). The information provided by these operator groups was to be integrated to provide a multifaceted picture of national aviation system safety that would complement and supplement information extracted from other national aviation safety data sets. The goal of developing a survey methodology that included all operator groups was not achieved by the NAOMS project largely due to the cost of designing and implementing a statistically sound survey for each new constituency and the practical limitations on the project's time and funds. Only pilots could be surveyed within the time and resource limitations of the project as a demonstration of the methodology. Air-carrier pilots were the primary respondent group for a NAOMS survey conducted from April 2001 until December 2004, but, in 2002, there was also a 9-month survey of General Aviation (GA) pilots[3].

Although the data presented relate only to the air-carrier-pilot survey and to the period during which these data were gathered, the report presents sufficient information on the research and

---

[3] This report has been delayed due to decisions by AvSP/AvSSP management that diverted the personnel and resources needed to perform and interpret the analyses to higher priority activities during 2005, 2006, and 2007, and by subsequent management decisions. Recently, the NASA leads of the NAOMS team were authorized to re-engage, to analyze the data, and to produce this report.

development underlying the NAOMS concept and methodology to assess its viability for future applications to all operators of the aviation system.

The following is a general overview of the content of this report. In Section 2, the NAOMS concept, its purposes and objectives, and the rationale for its implementation as a continuing service to the industry are described. The established methodology for acquiring survey data and the measures of a successful survey used by NAOMS are described in Section 3. Section 4 describes the development process of designing and implementing a survey instrument appropriate for air-carrier pilots. The conduct of the air-carrier-pilot survey and preliminary findings about the data are described in Section 5. Section 6 discusses the approach to analyzing the data which evolved, in large part, during the process of collecting and assessing the data. Section 7 presents the results of the analyses of relative changes over time performed on the survey data obtained in interviews with air-carrier pilots during almost three-years from February 2002 through December 2004. Section 8 describes the approach and results from a Spotlight (or Section C) survey conducted for the CAST/JIMDAT (Commercial Aviation Safety Team/Joint Implementation Measurement Data Analysis Team). Section 9 describes the development of a Web-based version of the NAOMS survey of pilots and the transfer of this capability to the Air Line Pilots Association (ALPA). Section 10 discusses the lessons learned during the research and development of the NAOMS survey. A summary, conclusions, and recommendations for NAOMS' future are presented in Section 11.

2. Why NAOMS?

2.1 The Rationale for NAOMS

The value of the information contained in experiential reports on performance of the system from its operators has been well established in the 35 years of the Aviation Safety Reporting System (ASRS). Many U.S. air carriers have recognized the value of voluntary experiential reporting and have implemented Aviation Safety Action Programs (ASAP) internally. However, these voluntary reports only provide information on events deemed important to the reporter, while events that may be important from a system's perspective may not be recognized as such and are not reported. The NAOMS concept was designed to address this limitation. NAOMS was not designed to replace current sources of aviation-safety information, but to supplement these sources.

The need for this kind of information had been documented when the NAOMS research project was initiated. For example, the White House Commission on Aviation Safety and Security (1998) said in their report: "The most effective way to identify incidents and problems in aviation is for the *people who operate the system* (pilots, mechanics, controllers, dispatchers, etc.) to self-disclose the information." The Government Accountability Office's Safer Skies Review (2000) emphasized the need for additional performance measures. The NTSB Report on Transportation Safety Databases (2002) said that there was a "need to address the problem of *under-reporting in current aviation safety data systems*." When the NAOMS project was initiated, there were a number of databases that attempted to capture safety-related information concerning the national air-transportation system. These databases were designed to address particular needs, and, like all databases, had certain limitations. A representative listing of aviation safety data resources that existed at the time of creation of the NAOMS project, their foci and their limitations, is shown in Table 1.

Table 1. Sources of Aviation Safety Data

Data Set	*Value*	*Limitation*
ATC radar tapes	Captures data on aircraft trajectories.	Captures data mainly on aircraft trajectories. Data contain relatively few parameters. Very limited ability to address human performance issues.
Air carrier digital flight (FOQA) data	Captures aircraft performance information continuously from hundreds of sensors.	No information on operative ATC clearances, which makes data difficult to interpret. No insights on human cognition or affect. Considered proprietary data by owners.
FAA Accident/ Incident Data System (AIDS)	Contains data records for incidents gathered from FAA Incident Report Form 8020-5, and teletype preliminary data. Data collected include date, location, pilot qualifications, type aircraft, type of event, pilot experience, limited narrative descriptions and other descriptive data.	The incidents recorded are limited to events investigated by the FAA. AIDS data are not suitable for trend analyses since not all incidents for each category are included in the database. Data are also of limited use for safety analysis since the focus is on investigation of enforcement events.
NASA/FAA Aviation Safety Reporting System (ASRS)	Contains voluntary reports of occurrences that could impact aviation safety. Reports are submitted by pilots, controllers, mechanics, other interested parties, and users of the NAS. Extensive narrative is the norm for each report. All privacy or identifying data are expunged or *sanitized*.	Limited by self-reporting biases, which affect the statistical representativeness of the data. Does not actively collect data on topics of interest. Data cannot be used for trend development since voluntary reports represent a subset of all events.
Bureau of Transportation Statistics (BTS)	Data are typically used in calculation of accident and incident rates. Reports are reliable since data originate with the carriers who are required to report them. BTS Form 41 Reports provide summary activity levels for U.S. certificated air carriers.	Data are collected on levels of flight activities but no information is collected about unwanted aviation events.
General Aviation Avionics and Activity Survey (GAAAS)	Data are derived from a survey that is applied annually to the U.S. GA community. Information is obtained on estimated flying time, number of landings, fuel consumption, lifetime airframe hours, avionics, and engine hours of the active GA aircraft. These data are typically used for the development of incident and accident rates for GA aircraft.	Designed to collect information on GA aircraft activity, the survey is considered unrepresentative due to a low response rate. Some of the estimates have very large estimation errors and as a result any calculated trends or rates using this information in the calculation will have very large margin of error.

continued on next page

Table 1. Sources of Aviation Safety Data (continued)

Data Set	*Value*	*Limitation*
National Airspace Information Monitoring System (NAIMS)	The following systems are the core of FAA's aviation safety monitoring efforts: *PDS (Pilot Deviation System):* Contains pilot deviation reports resulting from violation of a federal aviation regulation or a NORAD Air Defense ID Zone tolerance. *OEDS (Operational Error and Deviation System):* Contains all operational error or deviation reports that occurred in the NAS. *NMACS (Near Midair Collision System):* Contains voluntary pilot-reported near midair collision incidents. *VPDS (Vehicle/Pedestrian Deviation System):* Contains information on incidents involving unauthorized entry or movement on an airport by a vehicle operator or pedestrian. *RIS (Runway Incursion System):* Contains information derived from OEDS, VPDS, and PDS airport surface incidents that created a collision hazard or resulted in loss of separation with an aircraft taking off, intending to take off, landing, or intending to land.	A broad variety of FAA incident and airman enforcement data systems comprise the NAIMS. These systems vary in accuracy and ability to capture events.
National Transportation Safety Board (NTSB) Aviation Accident/ Incident Data System	Contains information collected during investigations of accidents or incidents involving civil aircraft within the U.S., its territories and possessions, and international waters. NTSB is the official source of accident data and their causal factors. Information is from 1983 to present and includes preliminary and final reports, narratives, and findings. The database serves as the gold standard for the counts of accidents in the U.S. Data are used as the source for accident statistic trends and rates.	Addresses commercial part 121 accidents and high profile incidents, but contains very little data on the tens of thousands of aviation safety incidents (accident precursors) that occur annually. The quality of the data varies. While the data are excellent for large aircraft accidents, they are of lesser quality for smaller aircraft accidents.
Service Difficulty Reporting System (SDRS)	Contains GA malfunction and defect reports and air carrier mechanical reliability report subsets. Air carriers, field offices, manufacturers, repair stations, and individuals submit the data.	Focuses on aircraft equipment problems. Only limited information about human operator performance (mechanics) in the context of equipment maintenance and repair. This database suffers from underreporting and poor data quality.

All of these data resources, while containing valuable information, were limited in their scope or in their quantitative properties. None of these programs could provide decision makers with statistically defensible estimates of the frequencies with which many unwanted events occurred or with reliable information as to whether the frequencies of such occurrences were changing over time. Similarly, the national capacity to measure the effects of aviation safety interventions and their unwanted side effects was limited. There was a need to obtain rapid feedback on the effects of technological and procedural changes to the system and to detect, as soon as possible, changes in the rates of safety-related events that deserve immediate attention. There was a need to escape from a reactive policy-making syndrome to a proactive approach. NAOMS was expressly designed to meet these needs by way of a comprehensive, statistically sound survey on which assessments of relative changes over time in the rates of unwanted events could be reliably based.

2.2 The NAOMS Concept

NAOMS was an attempt to address the performance of the national air-transportation system as an entity in a quantitatively defensible fashion by providing an active, broad-spectrum, safety-data collection capability with sound statistical properties.

The NAOMS concept entailed a scientifically designed survey to obtain the front-line operators' experiences with system performance and to provide a statistically sound basis for evaluating relative changes over time of the rates of incidents that might compromise safety. As indicated in Figure 2, information was to be solicited from the operators of the aviation system about the things they have experienced. The information provided by these operator groups was to be fused to develop a system-level perspective of the performance of the national air-transportation system. NAOMS was not designed to measure safety of the air-transportation system in any absolute sense or to identify the causes of unsafe events, but rather to assess the effects on system safety of changes over time in technologies, procedures, or training. By directly contacting randomly selected individuals, the NAOMS approach complements systems such as the ASRS and the ASAP which solicit from the community more broadly while relying on user-initiated reporting. Information from NAOMS surveys could be the first indication of a developing problem or trend deserving of further investigation using other sources of relevant information.

The key conviction underlying the NAOMS concept was that the operators of the aviation system—its pilots, air traffic controllers, mechanics, flight attendants, and others—collectively have the ability to accurately observe and report the safety events they encounter. The challenge was to collect these data in a systematic and objective manner. NAOMS was conceived as a research project to develop the methodology for a well-designed survey process that could meet this requirement.

NAOMS' goal of a data-driven approach to aviation-system safety analysis required the collection of data that were statistically valid, meaningful, and representative of the system. The NAOMS project team set out to develop a survey methodology to acquire such data. There was nothing that met the needs addressed by NAOMS when the project started over 10 years ago nor is there today.

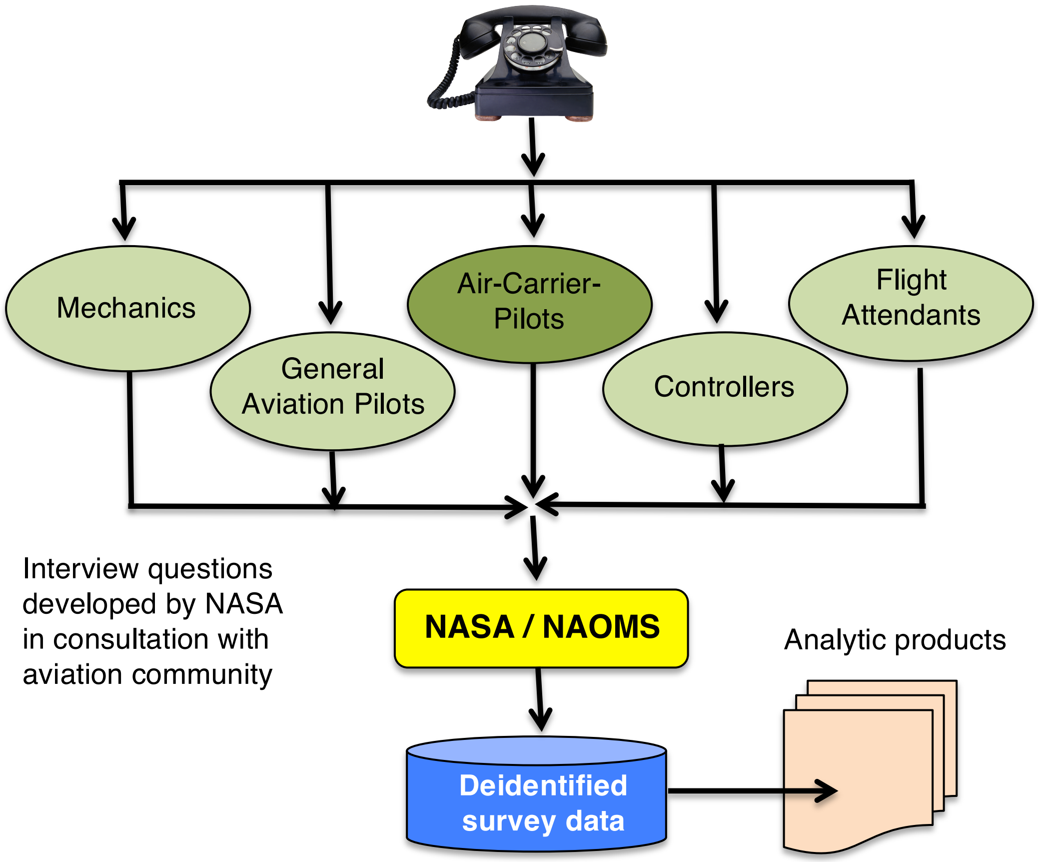


Figure 2. The NAOMS concept.

The nation is preparing for a major transition from the current system to a far more complex national air-transportation system. Yet some of the problems implicated in current accidents and incidents are not themselves adequately understood. Many of these problems entail factors that are not revealed in the data collected by the current automatic recording systems like the Flight Operational Quality Assurance (FOQA) program. The predetermined categories used in FOQA programs are highly airframe oriented, lacking detail on human performance and other important operational factors. The best source of information among the present systems about such aspects of unwanted events is the ASRS and the airline-based programs (e.g., Aviation Safety Action Program). The problems reported to the FAA under mandatory reporting rules are a small extreme subset of the problems that could be precursors of accidents. The NAOMS survey measures reports of event-based experiences This was conceptualized as a viable approach to complement existing systems for obtaining information on relative changes over time of the range of events that could compromise the safety of system operations is the use of surveys.

Surveys could potentially address a broad array of safety issues involving system design, human operator performance, organizational policies, procedures, regulations, publications, charts, aircraft and ATC equipment, airspace structures, and other aspects of the aviation system. The NAOMS survey is a pointer to events that could compromise the safety of the air-transportation system. It would precipitate further investigation using other capabilities and data sources such as those developed under Intramural Monitoring to identify the causal factors and, thereby, suggest an intervention. NAOMS was not designed for investigatory purposes nor for causal analyses and it was not designed to provide sufficient information by itself to enable the design of an appropriate intervention.

The potential benefits of a NAOMS survey system include:

- Providing decision makers in air carriers, air traffic management, and other air- services providers with regular, reliable, and insightful measures of the health, performance, and safety of the national air-transportation system.
- Providing decision-makers with reliable information to determine if changes in technology or procedures introduced into the system are producing expected improvements without inducing unwanted side effects.
- Establishing the baseline of operational performance against which to measure the system-wide impact of changes in the technologies, procedures, or training introduced into the system.
- Contributing to the information extracted from other data sources to support investigations of the causal factors of system-wide safety unwanted events.
- Enabling a quick look or "snapshot" of operational issues of interest to the aviation community at a particular time.

The NAOMS survey measures reports of event-based experiences. While originally intended to be broader, as the project evolved, funding limitations restricted the reporter population for the purpose of demonstrating the concept to pilots. The data presented in this report are those gathered from the air-carrier-pilot population.

NAOMS was a research and development project. It allowed the concept's value to be assessed, flaws to be discovered, and provided a basis for developing a continuing monitoring service. Once this was accomplished, NASA planned to hand off the process to another aviation safety agency with the hope that NAOMS would serve the aviation community by becoming a fully implemented institutional capability based on the lessons learned from this research.

3. Approach

The NAOMS project focused on research and development. It allowed a solid, scientific foundation to be laid for the methodology, the concept's value to be assessed, flaws to be discovered, and a basis for developing a continuing monitoring service to be provided. Once this was accomplished, NASA planned to hand off the process to another aviation safety agency with the hope that NAOMS would serve the aviation community by becoming a long-term institutional survey capability based on the lessons learned from this research. While originally intended to be broader, as the project evolved funding limitations restricted the reporter population for the purpose of demonstrating the concept to pilots. The data presented in this report are those gathered from the air-carrier-pilot population.

Since the purpose of the NAOMS project was to determine if a viable survey method could be developed, the project was heavily "front-loaded" to examine, in as many ways as feasible, those factors that could impact success. An understanding of these factors would also be essential to an operational implementation of NAOMS if the concept proved viable. These factors are examined in this section and throughout the report.

3.1 Literature Review

The science of conducting a survey has evolved over several decades. This is exemplified by the literature on survey methodology included in the bibliography of Appendix A. The body of literature on the science of conducting surveys is very large and has been a subject of research and development for a long time. The fundamental concepts of the design and implementation of a survey instrument have been honed by psychologists, sociologists, political scientists, philosophers, mathematicians, and statisticians.

Correspondingly, the development of NAOMS and the achievement of its objectives required a multidisciplinary team. The core project team included experts in aviation safety, aviation-safety-data systems, survey research, statistics, and system-monitoring design. In addition to the core team members, nationally recognized experts in survey research design and statistics served as project advisors. (See Appendix B for biographical sketches of the NAOMS project team and consultants.) Equally important to the adaptation of this methodology to the domain of aviation safety were the team members and advisors with extensive domain knowledge of, and experience in, aviation operations.

The NAOMS project team conducted an extensive literature review of survey methods and published studies of surveys of pilots and other aviation personnel. The team found very few aviation-specific studies that demonstrated a rigorous survey approach. The use of surveys to acquire data in support of aviation safety has been episodic and not always well done. In particular, there was very little published data relating to pilot recall ability, memory organization, and like matters of great importance to NAOMS.

The concept and the design of the NAOMS survey employs the best practices of surveys used in other policy domains and provides comparable benefits. The information that was available in related literature is summarized in Appendix C.

3.2 Survey Methodology

Survey methods are mature, well understood, and are known to produce reliable data when the survey instrument is properly designed for the objectives, the domain, and the interviewees for which it is intended. For at least five decades, the U.S. federal government has been actively and extensively involved in the collection and analysis of survey data. Surveys have been routinely sponsored by a wide range of federal agencies seeking a wide range of types of information about life in America. Surveys have been used to shape national policy for many decades, especially in areas such as public health, economics, and criminal justice. Following are a few examples of on-going Federal surveys and their sponsoring agencies:

- Survey of Income and Program Participation (Census Bureau) 1984
- Consumer Expenditure Surveys (Census Bureau) 1968
- Annual Housing Surveys (Census Bureau) 1973
- Survey of Consumer Attitudes (National Science Foundation) 1953
- Health and Nutrition Examination Surveys (National Center for Health Statistics) 1959

- National Health Interview Surveys (National Center for Health Statistics) 1970
- American National Election Studies (National Science Foundation) 1948
- Panel Study of Income Dynamics (National Science Foundation) 1968
- National Longitudinal Surveys (Bureau of Labor Statistics) 1964
- Behavioral Risk Factor Surveillance System (Center for Disease Control) 1984
- Monitoring the Future (National Institute on Drug Abuse) 1975

One of the reasons for the popularity of surveys is that they can be used to measure important phenomena that would be very expensive and difficult-to-impossible to measure in other ways. For example, when an entire population or sub-population may have been affected by some event, it would be impractical to try to identify, and learn from, the experience of each individual. But in many cases a representative sample of affected people can be selected and interviewed to learn about their experiences. When the survey is properly conducted, valid conclusions about the entire population may be based on a sample, making the impractical practical. It is often impossible to compare survey results to other benchmarks to assess the accuracy of the surveys because surveys are so often conducted in contexts where no other indicators measure the same phenomena.

There are three key measures of survey success (i.e., is the survey measuring things consistently) that can be used to evaluate whether the survey instrument was properly designed and implemented: (1) response rate; (2) data quality; and (3) validity of statistical results (i.e., is the survey measuring what it says it is measuring). These measures of survey success are also indicators of the confidence one can have in the data. (For example, see Weisberg et al, 1996, and Lyberg et al, 1997.) They are described in the following section.

4. NAOMS Concept Development

As there was little experience in the adaptation of survey methodology to aviation safety, the NAOMS project team approached the development and implementation of NAOMS with methodological thoroughness and probity as their foremost concern. The development of the NAOMS survey entailed the careful adaptation of survey methodologies to the aviation-safety domain to ensure a robust survey process in accordance with the measures of success mentioned above. The NAOMS survey was designed to obtain reliable information on:

- Composition of U.S. commercial and GA flight operations.
- Long-term temporal changes in the frequencies of unwanted events.
- Effectiveness of interventions intended to enhance aviation safety and the occurrence of any unwanted side effects due to such interventions.
- Relative frequencies with which unwanted events occur.
- Other aviation operational and safety topics amenable to quantification and of interest to aviation policymakers and operators.

The design of the survey instrument and the method of conducting the survey that would produce results sufficiently reliable to support aviation-safety policy decisions required careful consideration of many factors. The treatments of many of these factors are likely different for each constituency of operators. The NAOMS project team focused its efforts during the first two years on ensuring methodological rigor in the treatment of these factors, with particular emphasis on their application to surveys of air-carrier pilots. Further, the emphasis of the initial survey was on obtaining reliable information on relative changes over time in the rates of occurrences of unwanted events. Descriptions of the factors considered and how the NAOMS project team addressed them for the survey of air-carrier pilots are presented below. For additional detail on the development of the concept, the reader is directed to the Battelle Reference Report (Battelle, 2007).

4.1 Data Reliability

4.1.1 Response Rates

The NAOMS team followed the general guidance of Dillman (1978) in designing the survey for the most effective ways of maximizing response rate and data reliability. Dillman identified several methods to increase respondent attention to, and understanding of, survey questions in both mail and telephone-interview formats based on psychological research on attention, information processing, and compliance, as well as empirical data from surveys conducted using his recommendations.

The NAOMS project included a field trial, a demonstration test, and the full survey of air-carrier pilots. The response rate (i.e., percentage of those located and eligible who completed the survey) during the NAOMS field trial was encouraging with very low rates of refusal among eligible respondents in any of the modes tested. This indicated that a response rate of at least 70% for eligible respondents could be expected during the full survey. The response rate for the NAOMS survey of air-carrier pilots benefited from the large sample size used. Spanning both the demonstration test and the fully implemented survey, approximately 7,000 air-carrier pilots were interviewed each year from 2001 through 2004. These pilots were randomly selected from a pool of about 40,000 eligible pilots available from the data source used for the survey.

Over the course of the air-carrier survey, about 24% of the pilots drawn from the Airmen's Registry who met the NAOMS selection criteria could not be located. Also, about 23% of those contacted were found ineligible because they had not flown as commercial air-carrier pilots in the recall period. These latter pilots were eligible for possible later random selection. Excluding those who could not be located and those who were ineligible, the air-carrier survey response rate far exceeded the standard set by survey experts. The NAOMS team was gratified to find the enthusiasm among the air-carrier pilot community to participate in an activity related to maintaining and improving the safety of the system.

High response rates increase confidence in the results of a survey because they indicate that the responses obtained are minimally influenced by self-selection bias. Furthermore, a high response rate is an indication of the participants' interest in contributing to the understanding and improvement of the system in which they work and, therefore, in providing accurate responses.

4.1.2 Data Quality

The following indicators are gauges of the quality of the data collected by a survey such as NAOMS:

- *Extreme Outlier Frequency*. Properly obtained and recorded high-quality data contain relatively few extreme outliers resulting from misunderstood questions, intentionally misleading responses, or data-collection errors. Relatively few extreme outliers were found in the data from the air-carrier-pilot interviews, further attesting to the pilots' conscientiousness in responding accurately. The data from the surveys of the air-carrier pilots contained few (much less than 1%) extreme outliers.

- *Survey Completion*. Having agreed to participate, the respondents answered all, or almost all, of the questions presented to them. Fewer than 2% of the respondents to the NAOMS survey of air-carrier pilots broke off their interviews before completion.
- *Respondent Assessment of Survey Quality*. Respondents were provided an opportunity to comment on the survey process in the design of the questionnaire. The responses to the questions regarding the survey spoke very favorably of the NAOMS survey design and process.

4.1.3 Valid Statistical Analysis

Another measure of the success of a survey is that the analyses of the data produce statistically valid results that are viewed as operationally important by the domain experts. This is, of course, the end purpose of conducting the survey. It requires strict adherence to the concerns addressed above by the first two measures of success. However, the achievement of both statistical and operational validity entails other considerations that are more domain and constituency specific. The success of the NAOMS survey in terms of data reliability and data quality, as well as how statistical analysis and operational validity were achieved is discussed in detail later in this report.

4.2 Voluntary and Anonymous

Participation in the NAOMS project was voluntary and the NAOMS project team assured protection of the anonymity of the respondents. The names of potential respondents were drawn from the Airmen Certification Database as published on the Internet. The NAOMS project team sent letters to each of the potential respondents requesting participation. The letter made clear that the recipient was under no obligation to participate but promised anonymity if he or she chose to do so. The very high voluntary response rate to the NAOMS' survey indicates the pilots' confidence in NASA's assurance of anonymity.

All data provided by respondents to the NAOMS survey were held in confidence. The NAOMS team maintained records of who participated in the survey to avoid unnecessary follow-on mailings and to prevent an individual from being asked to participate more than once a year. However, there was no linkage in NAOMS's data repositories between the names of respondents and the data they provided. The confidentiality checks that were established to protect the respondents' identities included the following:

- Although pilots were assigned identification numbers for administrative purposes, the data were stored with complete anonymity.
- The computerized interview form did not contain a pilot-identification number or any other form of identification linking a pilot to a completed interview.
- No link was maintained between the interview form and the pilot data file after the survey interview was completed.

These controls produced the desired results. The responses are functionally anonymous. There were no breaches of anonymity guarantees during the NAOMS survey development and operational periods or thereafter.

Also, because the pilots were being asked to report some information that might be considered sensitive, the questions in Sections B and C of the survey instrument were carefully phrased to avoid asking pilots directly about their performance. (For example, most questions began with the phrase "How many times did an aircraft in which you were a crewmember...".) Considerations such as these resulted in the highly positive responses from the interviewees when asked their opinions of the survey in Section D.

4.3 Emphasis on Experience Rather than Opinion

During the very early phases of formulating NAOMS, the team found that presentations on the survey concept were received with reservations by some elements of the aviation community. This attitude was largely attributed to beliefs that opinion-based surveys are volatile, they can shift rapidly in response to current events, and they can be structured in ways that predispose the responses sought by survey operators. However, the NAOMS survey was not an opinion-based survey. The NAOMS survey focused on collecting data about recalled personal *experiences*, not personal *opinion*. This is a distinction that some have found difficult to grasp. Most of the questions in the NAOMS survey instrument ask respondents whether or not they have personally experienced or observed particular types of events within a specified period of time. This emphasis on recalled experience minimizes the subjectivity and volatility of the data obtained from survey responses.

An exception to this policy of focusing on experience was made for some questions that were specially designed for a study that was performed at the request of the CAST/JIMDAT (discussed in Section 8.3) for which there was interest in pilot views (i.e., opinions) about a variety of operational and training practices.

4.4 Methodological Issues

There were many other issues that required thoughtful resolution before the air-carrier-pilot survey could be initiated. The NAOMS team applied considerable effort to obtain the best answers to the following questions as they applied to the air-carrier pilots:

- What is the appropriate content of the NAOMS questionnaires for the survey of air-carrier pilots?
- How should the questionnaires be structured?
- How far back in time can respondents accurately recall safety events?
- What is the best survey-collection mode to use?
- What should be the source of the respondent names (i.e., the air-carrier-pilot population pool) that would be sampled and what sample size of this pool is needed to achieve desired levels of accuracy?
- How should respondents be selected, located, and engaged for the survey?
- Would a purely random or a panel sampling design work best?

The way that the NAOMS team addressed each of these issues for the air-carrier-pilot survey is described below. The manner in which each of these issues is resolved is not only aviation domain specific, but is likely to be different for air-traffic controllers and for mechanics and for cabin crews. Similar considerations would have to be given to the design of the survey instrument and the manner of its implementation for each of the other constituents. The unanticipated magnitude of the research effort required to prepare the survey instrument for the single constituency of the air-carrier pilots led to the conclusion that there would not be sufficient time or funds available to support surveys of all the other constituents within the ASMM Project.

The research and experiments conducted by the NAOMS team to address each of these issues are described along with the resolution of each issue. Some of these issues were addressed in the first two years of research and questionnaire development. Others were resolved in a field trial of data collection from November 1999 to February 2000. The field trial provided feedback on the interview mode, obtained a first measure of probable response rates, and explored pilot-recall capabilities. The field trial also provided first approximations of the safety-event frequency rates that respondents would be reporting in the NAOMS survey. This information provided a basis for the number of interviews needed for acceptable reliability. The results of the field trial are discussed below as they pertain to each issue.

However, while the field trial resolved most issues, it could not resolve them all. The field trial lasted about 3 months. That was not sufficient time to conclude the optimal recall period or to assess adequately the sampling strategy for data collection; that is, whether it should use cross-sectional (random) selection or panel selection. Also, the field trial revealed some problems with the questionnaire as it was structured at the time. Cognitive experiments that were conducted after the field trial resulted in changes to the order and content of questions in the final survey instrument. The study of these issues had to continue into the first year of the survey using the final survey instrument. Consequently, the first year (actually, the first nine months) of the full survey should be viewed as a second trial or demonstration test in the development phase of the survey. The data from those first nine months were not used in the primary data analysis (Section 7) but were useful in resolving remaining issues.

4.4.1 Questionnaire Content

Issue:

The NAOMS survey needed to collect the following types of content-centered information:

- Measures of respondent risk exposure, such as the numbers of flight hours and legs flown, which act as the denominator of rate estimates, and estimates of the number of safety incidents and unwanted events that respondents experienced during the recall period, which act as the numerator of rate estimates.
- Answers to questions on identified general topics.
- Answers to questions on special focus topics requested by stakeholders.
- Feedback on the quality of the questions asked and of the overall survey process.

Resolution:

The NAOMS team developed a survey instrument composed of a four-part questionnaire for the air-carrier pilots that addressed these informational needs. The final design included a major effort to solicit input from the aviation community on the events to be addressed, question format for each event, order of questions, and structure of survey instrument. This entailed peer reviews by focus group sessions with air-carrier pilots, one-on-one interviews, and two workshops arranged by the NAOMS team. The NAOMS team consulted existing aviation safety data repositories maintained by NASA, the FAA, and the NTSB to identify known safety issues. One source was the ASRS analysts who identified the recurring and emerging safety issues they saw in their reviews of ASRS reports.

This section discusses the design of the resultant survey instrument. Particular emphasis was given to identifying the safety topics that would be the bases for the longitudinal survey (Section B of the survey).

The structure of the NAOMS survey instrument for air-carrier pilots is shown in **Figure 3**. Section A provides information on the interviewee's activity that is used for normalizing the data to establish rates of occurrences. Section B provides information to reveal changes over time of events addressed in the survey questions and constitutes the numerator of the statistical analyses to establish rates of occurrences. Section C addresses a particular subject of special interest to the aviation community at that particular point in time. Section D asks for an evaluation of the survey and any other comments the interviewee cares to make.

Section A. Aviation Activity Data

- Hours and legs by make-model and by crew position during the recall-period and over life-time (total hours only).

Section B. Safety Related Events

- Provides a consistent data set over time.
- Safety-related occurrences such as airborne conflicts, spatial deviation, ground events, weather encounters, equipment problems, pilot-ATC interaction issues, turbulence, passenger issues, and aircraft handling.

Section C. Focus Questions

- Topics driven by government/industry priorities for short-term response.

Section D. Survey Feedback

Figure 3. The structure of the NAOMS survey instrument for air-carrier pilots.

Each of the four sections of the NAOMS survey instrument is discussed below. However, as the focus of the survey and this report is on relative changes over time, the emphasis of this report is on the exposure information from questions in Section A and the safety event questions in Section B.

4.4.1.1 Section A: Flight Activity Levels

The data in Section A of the survey instrument for air-carrier pilots measure the respondent's flight activity during the recall period and over his or her career. The information on flight activity during the recall period is crucial to developing safety rate estimates from data collected by the NAOMS survey. Data from Section A provide the denominators in rate calculations (data from Sections B provide the numerators in rate calculations). Data from Section A can also function as explanatory variables when responses to Section B are statistically modeled.

In the aviation domain, there are two primary flight activity measures: flight hours and flight legs. Flight hours measure the time spent maneuvering aircraft in the air and, depending on the specific measure used, taxiing on the ground. A flight leg is a segment of flight from departure to arrival. Flight hours and legs are measures of flight activity, and by extension, these activity data also measure the risk exposure of a flight crew. These numbers are the key variables used to estimate aviation safety event rates because they are measures of the opportunities of the flight crew to experience a safety event or an accident[4].

The strategy of the NAOMS survey design maximizes the potential for accurate rate estimates by collecting the data on risk exposure (from Section A) simultaneously with the data on safety events experienced from Sections B (and C) so that the events and the flight activities are for the same recall period.

Among other factors, the risk exposure during a flight hour or a flight leg can vary depending on:

- make and model of the aircraft flown
- flight mission (i.e., passenger, cargo, business, utility, recreational, etc.)
- portion of the flight that was international (i.e., outside the U.S.)
- season of the year

The NAOMS survey acquired data on each of these factors using the questions in Section A. Also, the respondent was asked to indicate the flight position he/she held (Captain, First Officer, Relief Pilot, other). The interviewee was also asked for the order-of-magnitude size (i.e., small, medium, or large) of the air carrier for which he/she flew.

The full questionnaire, including questions in section A, can be seen in Appendix D.

[4] The use of flight hours and legs as measures of risk exposure pertains specifically to pilots. Other operators of the aviation system (e.g., air traffic controllers or mechanics) would require some other set of activity measures for Section A of their survey instruments for measures of their risk exposure.

4.4.1.2 Section B: Longitudinal Safety Event Questions

The questions of Section B are longitudinal in nature, that is, they were intended to be collected routinely over a long time period to identify those relative changes with time in reported rates of events experienced by pilots that might be indicative of potential decrease of safety. The same questions were maintained throughout the survey so as to maintain continuity throughout the data-collection process.

Deciding on the topics to be addressed in the questions of Section B was of fundamental importance. A survey of pilots on their perspectives on the performance and safety of the aviation system could entail many aspects such as the aircraft being flown, airports, procedures, policies, training, and air- carrier organizational factors. The NAOMS survey could address a great many topics but it was not practical to consider all. Also, many areas of potential inquiry would entail an opinion, which the NAOMS survey was designed to avoid.

The approach used by the NAOMS team for selecting the highest priority topics and developing questions that would yield valuable insights on them entailed an iterative process of consultation with the community. The NAOMS team consulted existing aviation safety data repositories to identify known safety issues appropriate for the longitudinal study of Section B. One of these sources, the ASRS, is a very large repository of aviation safety reports managed by NASA on behalf of the FAA. The NAOMS team consulted with ASRS analysts to identify the recurring and emerging safety issues the analysts saw in their reviews of ASRS reports.

Input was also obtained from aviation domain subject matter experts (SMEs) who were asked about the safety issues important to them based on their personal experiences. These inputs were obtained during focus group sessions with air-carrier pilots, one-on-one interviews, and the two workshops hosted by the NAOMS project team.

The focus group sessions with air-carrier pilots were held in the Washington, D.C., area in August and September 1998. The focus groups were directed through a range of questions, beginning with open-ended questions to tap the issues that came up naturally and proceeding through more structured questions designed to help recall in specific areas. In addition, the focus group pilots were encouraged to provide feedback on the concept and the possible structure of the questionnaire. The participants included 37 active air-carrier pilots flying both domestic and international routes. Between 2 and 15 pilots participated in each session. Each session lasted 90 minutes and was led by a professional facilitator who encouraged the participants to mention as many different types of safety-related events as possible. This included anything that should not occur during normal air operations. The responses of the participants in these focus groups were de-identified and consolidated.

One-on-one interviews were conducted with active air-carrier pilots flying both domestic and international routes to identify additional events that did not surface in the focus groups. The one-on-one interviews allowed for more intensive discussion between the interviewer and a single pilot than is possible in a group setting. Also, any apprehension that a pilot might feel about mentioning some types of events in front of other pilots is minimized during the one-on-one interviews. Nine such interviews were conducted in the Washington, D.C. area and in Columbus, Ohio, during September 1998. Each interview lasted up to 90 minutes. The responses of the participants in these one-on-one interviews were de-identified and considered in conjunction with the responses from the focus groups.

Decisions about which of the aviation safety topics generated by this process of consultation with the aviation community were driven by a desire to select events serious enough to be good indicators of the safety performance of the aviation system, yet not so serious that they would occur too rarely to be captured reliably in a survey. Questions about a few serious rare events were included in Section B in recognition of the industry's interest in those topics.

The topics selected for Section B needed to be judiciously grouped and ordered, and then translated into one or more carefully crafted questions. The approach to grouping and ordering Section B questions was based on the results of the NAOMS team's research on pilot-memory organization, which is described in Section 4.4.3. The NAOMS team included input from experts in survey methodology and in aviation operations to craft questions that would elicit the desired information on each survey topic.

The questions were arranged to maximize the accuracy of recall within each first-level category of questions and each category was appropriately introduced as a separate set of questions by the interviewer. At the first level, the questions were grouped by general safety-related events, weather-related events, passenger-related events, airborne conflicts, ground operations, aircraft handling-related events, altitude deviations, and interactions with air traffic control.

The full questionnaire, including questions related to section B, is in Appendix D.

4.4.1.3 Section C: Focus Questions

The third section of the NAOMS survey questionnaire was designed to address special focus topics. Whereas Section B questions were expected to persist over a period of many years to provide a basis for trend analyses of long-term aviation safety rate measures, Section C topics were expected to arise and be the subject of data collection generally for a number of months, after which they would be replaced with new topics. Section C was envisioned to focus a "spotlight" on timely concerns of the aviation community.

Three Section C question sets were developed during the NAOMS tests and implementation periods. The first of these, used in the field trial only, dealt with minimum equipment lists (MELs). The second dealt with "in-close" changes to approach and landing clearances (ICACs). The third and final Section C question set involved the development of baseline aviation system performance measures requested by the CAST's JIMDAT.

The full questionnaire, including questions developed for the Section C implementations of the ICAC and the JIMDAT studies, can be found in Appendix D. The results of the ICAC study and the special study for the CAST are presented in Section 8.

4.4.1.4 Section D: Respondent Feedback Questions

Section D of the survey obtained feedback on the interview process and the questionnaire. For example, respondents were asked directly about the quality of the survey process, including the relevance and understandability of questions, and their ability to reply meaningfully and with confidence to those questions. Respondent feedback was very positive and increased the

NAOMS team's confidence in its methodological approach. For example, when asked if any of the questions were confusing, poorly worded, or ambiguous, about 92% of the respondents answered "no". When asked, only a small percentage of the respondents offered comments on the survey and these were, in general, highly positive. In response to the question of the accuracy with which they counted the reported events, over 93% said they were "extremely" or "very" confident of their replies.

4.4.2 Questionnaire Structure

Issue:

The content requirements of the four sections of the NAOMS survey instrument described in Section 4.4.1 drove its structural organization at the high level. The desire of the NAOMS team to minimize respondent burden while maximizing respondents' abilities to recall the information sought by the NAOMS survey influenced other structural aspects, such as how questions in Sections B and C were grouped and the order of questions within each section.

Resolution:

The questions were grouped, ordered, and sequenced to take advantage of the ways people naturally organize memories.

Research on human cognitive and memory organization (e.g., Anderson, 1976; Pryor and Ostrom, 1978; Srull, 1983; Barsalou, 1988; and Sedikides and Ostrom, 1988) shows that memories of similar events are typically stored together in clusters in a person's memory. It is easiest and most efficient to recall all other memories in that cluster, rather than jumping to another cluster. Therefore, the NAOMS team grouped questions about a particular event (or closely related events) as one would expect them to be stored in memory.

The *a priori* expectation about pilots' memory organizational schemes was derived from the organizational scheme used by the ASRS, from discussions with pilots, and from reading aviation literature. The NAOMS team anticipated that pilots might organize safety-related events in terms of their causes (e.g., equipment failures, flight-crew mistakes, weather, etc.), the phase of flight in which they occur (e.g., takeoff, ascent to cruise altitude, etc.), or their seriousness (e.g., minimal, moderate, or severe). These expectations contributed to the design and analysis of the following experiments to explore pilots' organization of safety-related events:

- *Autobiographies:* Pilots were simply asked to recall all of the aircraft safety problems they had witnessed in their careers.
- *Sorting:* The NAOMS team constructed a list of 96 hypothetical safety-related aviation events. Pilots sorted the 96 events into groups, indicating which events seemed most similar or related to one another. The NAOMS team used this experiment to measure pilots' perceptions of linkages among the events (i.e., that events that seem similar or related to one another are clustered together in memory).
- *Recall:* In this experiment, another group of pilots read the 96 events used in the Sorting experiment and later recalled as many of the events as they could. The order in which the events were recalled indicated how those events were stored in memory.

As a result of these experiments, the NAOMS project team concluded that pilots use a hybrid structure combining elements of the flight-phase- and cause-based- organization schemes to organize aviation safety measures. The results of the memory-organization experiments indicated that there were individual differences among pilots and that not all pilots organized memory according to the hybrid scheme. However, the NAOMS team concluded that the hybrid scheme best reflected pilot-memory structure and that using a survey question order based on that structure could best facilitate accurate recall. As a consequence, the safety-event categories used in the final design of Section B of the NAOMS pilot survey were those shown in Table 2.

Table 2. Organization of Safety Events in Pilot Memory

Category	*Scope*
Aircraft equipment	Any aircraft equipment-related problem
Turbulence	Turbulence encounters due to wake or weather
Weather	Weather problems other than turbulence
Passengers	Any passenger-related problems
Airborne conflicts	Any conflicts with other aircraft in the air
Ground events	Runway and taxiway transgressions, ground conflicts, and all other ground-based events
Aircraft handling	Flight crew issues in managing the aircraft
Altitude deviations	Any deviation from assigned altitude
ATC interactions	Events rooted in pilot-ATC interaction difficulties

Having resolved the grouping by survey topic of Section B, the NAOMS team then addressed the issue of how to sequence subtopics/questions within each topic area. Research in cognitive psychology has demonstrated that the recollection of strong memories (i.e., those that are vivid and easy to recall) makes it more difficult afterwards to recall weaker related memories (i.e., more common, everyday occurrences). (See for example McGeoch, 1942; Slamecka, 1968; and Mensink & Raaijmakers, 1988). The NAOMS team followed this guidance in the final version of Section B of the survey instrument and, as much as possible, ordered the questions within each topic group from the least severe event type to the most severe event type.

4.4.3 Recall Period

Issue:

The NAOMS survey needed to have respondents accurately recall not only safety events, but also the time frames in which they occurred. This requirement had implications for Sections A, B, and C of the questionnaire because the questions in all of these sections ask respondents to report on some aspect of experiences that occurred within a specified recall period. The NAOMS team did not know how far back in time pilots could reliably recall the timing of safety events. Asking pilots to reflect over a longer period of time is desirable from the standpoint of survey efficiency because more safety events would be "captured" per interview thereby enabling higher levels of statistical accuracy to be attained for the chosen sample size. However, the accuracy of memories—including memories of when events occurred—fades with time. Hence, the quality of the data declines as the recall period increases. Data quality had to be weighed against cost in choosing a recall period. The goal of the NAOMS team was to find the longest recall period that could be used without unacceptably compromising data quality.

Resolution:

The NAOMS team recognized that accuracy of recall was an important issue to be resolved to meet the objectives so the team members attempted to address it early in the project (see contractor's memorandum dated December 22, 1999, Appendix E). Appropriate recall periods are probably not the same for all aviation constituencies. The NAOMS team focused initially on air-carrier pilots and their recall ability.

The literature on memory and the experience of survey researchers identify three types of misremembering that might impact the accuracy of survey responses. First, respondents might forget events that occurred. Second, they might remember real events but misremember the times in which they occurred. Third, respondents might imagine events that never occurred (perhaps derived from a general sense that certain kinds of events typically happen "roughly once every 'X' weeks"). The NAOMS survey instrument needed to employ a recall period that kept the level of these potential errors acceptably low.

A small experiment was conducted to help identify the most appropriate recall period for air-carrier pilots. Pilots were asked to recall the number of landings they had performed recently. Comparisons of these data with each pilot's flight log showed that pilots could recall these routine events very accurately for one week after their occurrence; thereafter accuracy declined. Therefore, the 7-day recall period was used as the baseline measure of extremely high reliability for evaluating longer recall periods. Longer recall periods were evaluated during the field trial and during the first nine months demonstration test of the survey as discussed below.

During the field trial, the NAOMS team randomly assigned respondents to one of six recall periods: 7 days, 14 days, 30 days, 60 days, 120 days, and 180 days. Comparisons of the number of safety events remembered for each of these recall periods showed that the absolute number of events remembered increased as the recall period increased. However, the reported number of events per flight hour (i.e., the event rate) declined and also the variance of the event rate decreased as the recall period increased. The NAOMS team had hypothesized that more severe events might be "brought forward," i.e. remembered as occurring within the recall period when they had actually occurred earlier. However, it appeared that based on the results of the short field trial, if this occurred at all it was more than offset by the number of events respondents had collectively forgotten. Also, during the field trial the respondents were asked how confident they were in their answers and the results showed that confidence levels dropped as recall periods grew longer.

Based on evidence from experiments on memory recall and the field trial, the NAOMS team concluded that a recall period in excess of 90 days was unlikely to produce an acceptable level of data quality, while a recall period of less than 30 days would require an unacceptably large number of interviews. However, the quantity of data collected during the field trial was not sufficient to settle the issue of what recall period should be used for the full survey. A final decision on this question was deferred to the first 9 months if the full survey. The first 9 months of the full survey (from April 2001 through January 2002) were viewed as a trial period, essentially a second field trial using the final version of the questionnaire. The data from this period were not used in the primary data analysis (Section 7). During this period, half of the respondents used a 90-day recall period and the other half used a 30-day recall period. During the last 2 months a second test was conducted that entailed a *three-way* split: 30, 60, and 90-day recall periods.

The first 9 months of the survey revealed that the total number of all events recalled per flight leg declined for each month added to the recall period as had been observed in the field test. Respondents using a 60-day recall period remembered on average 27% fewer events per flight leg than those using a 30-day recall period; those using a 90-day recall period remembered about 41% fewer events per flight leg than those using a 30-day recall period. On the basis of all the data collected, the NAOMS team decided to use a 60-day recall period for all interviews conducted over the following 3 years of the full survey. This decision was a compromise between the practical requirement to maximize the number of safety events recalled per interview and the desire to reduce memory-related errors.

This study of the effects of the recall period showed an important feature of the NAOMS survey. Regardless of whether a 30-, 60-, or 90-day recall period is used, the resultant event-rate estimates are biased downward compared to a more "instantaneous" recall period such as 7 days. This downward bias results in approximately 63% of events being reported for a 60-day recall period when compared with a 7-day recall period.

4.4.4 Survey Mode

Issue:

Questionnaires can be applied using various techniques or "modes." For example, mailed or computer-based questionnaires are used to obtain self-administered survey responses. Other survey methods rely on the assistance of trained interviewers. Whether self-administered or assisted, each survey mode has advantages and disadvantages and the selection of a survey-administration mode is a compromise among collection cost, respondent satisfaction, response rate, and data quality. The NAOMS team needed to determine the best survey mode for its data-collection system.

Resolution:

The NAOMS team evaluated the three most common survey data-collection modes in use at the time:

1. Self-administered questionnaires (SAQ)
2. Computer-assisted telephone interviews (CATI)
3. In-person interviews

The NAOMS team reviewed the general literature on the relative strengths and weaknesses relative to cost, respondent satisfaction, and response rate of each of these survey modes. (For examples, see Bishop, et al, 1988; DeLeeuw and van der Zouwen, 1988; Dillman, 1978; Groves, 197; Hall, 1995; Jobe, et al, 1997; Jordan, et al, 1980; Krosnick and Green, 1998; Krywan, et al, 1994; Siemiatycki, 1979; Walker and Restuccia, 1984; Weisberg, et al, 1996.) Following are summaries of the literature on the relationship of each of the survey modes to the factors to be considered in selecting the mode:

- *Survey Mode and Cost*. The literature indicates that, on a per-interview basis, in-person interviews typically are the most expensive; telephone interviews often are significantly less expensive; and SAQs generally are the least expensive. However, if efforts are made to achieve the highest response rates possible with SAQs, then the cost of that mode is close to the cost of applying the same questionnaire via the telephone.
- *Survey Mode and Respondent Satisfaction*. In general, respondents tend to favor face-to-face over telephone interviews since the former provide greater opportunity to develop rapport with the interviewer. However, the difference between the modes is not great. Although not necessarily favored, self-administered questionnaires when compared with interviews have the advantage of allowing the respondent more time to think but also are seen as impersonal (Weisberg, et al, 1996).
- *Survey Mode and Response Rate*. In-person interviews were also found superior for response rates. It is widely recognized that in-person surveys can achieve response rates of 70% or greater; telephone surveys can achieve a response rate of 60%; and mail surveys generally achieve response rates of only 10 to 20% unless exceptional efforts are taken (Dillman, 1978).
- *Survey Mode and Data Quality*. Response rate and respondent satisfaction influence data quality. The literature also addresses, although not extensively, two other survey considerations that could negatively impact data quality. These factors are known as satisficing and social desirability bias. Generally, modes that encourage these phenomena also compromise data quality.
  - *Satisficing*. When *satisficing* behavior happens, respondents exert the minimal effort needed to *satisfy* survey requirements. Responses are not always well thought through and are more likely to contain bad data. *Satisficing* behavior can be encountered when surveys require respondents to do a great deal of cognitive work for little or no real reward (Krosnick, 1991). The literature, while not entirely consistent or conclusive, indicates that from the satisficing perspective telephone interviews are least desirable and self-administered questionnaires work as well or better than in-person interviewing.
  - *Social Desirability Bias*. Social desirability bias describes the tendency of respondents to answer questions in a way that present them to interviewers in a respectable light—even if it requires the truth to be "bent" or ignored. The literature suggests that SAQs appear to minimize social desirability bias while telephone interviews maximized this bias.

In summary, the literature shows that SAQs are often the least expensive to run and can provide quality data. However, SAQs also have significant disadvantages. They usually have significantly lower response rates than other survey modes and it is difficult to implement complex skip patterns in the questions in paper-based SAQs. In-person interviews are superior with respect to respondent satisfaction and response rate but they are very expensive to operate. Telephone interviews can achieve a reasonably good response rate at a moderate cost but possibly at the expense of user satisfaction.

The literature review did not provide a clear choice of survey mode to use for NAOMS survey of air-carrier pilots. The NAOMS team decided to use the field trial to conduct additional evaluation of different survey application methods with measures of response rate, data quality, cost, and other key indicators for comparison.

The field trial was designed to evaluate the effectiveness of three different data-collection modes; (1) SAQ; (2) CATI; and (3) in-person interviews. Multiple versions of the questionnaire were developed to test different hypotheses in the field trial. All three of these trials were administered by the Centers for Public Health Research and Evaluation (CPHRE), a division of Battelle that is highly experienced in performing surveys involving sensitive information. For the field trial of each of the three collection modes, letters were mailed in advance to all selected pilots on NASA letterhead and signed by the NASA NAOMS project managers. The letter informed pilots of the study's purpose and the selection process. It committed NASA to maintaining the anonymity of the interviewees and the confidentiality of their survey responses. Respondents were asked to affirm their eligibility and intention to participate on the pre-addressed, postage-paid return postcard. Each postcard contained an indentification number that allowed Battelle's CPHRE staffers to eliminate pilots from the sample pool who were ineligible or who indicated they did not wish to participate in the survey.

Self-Administered Questionnaires. Dillman's principles of design were implemented in several stages during the field trial to maximize response rate and data quality of the SAQ survey mode (Dillman, 1978).

At the outset of the field trial of the mailed SAQ, the NAOMS team conducted a small experiment to determine if the mailing method would influence response rate. Other survey results have indicated that the use of Priority Mail for sending the survey packets increased response rates. Half of the first-mailed self-administered survey packets were sent using U.S. Postal Service (USPS) Priority Mail while the other half was sent by regular USPS First-Class Mail. The response rate was essentially the same for both groups.

The letter sent in advance to pilots selected to be SAQ respondents in the field trial told them to expect a follow-up mailing that would include the questionnaire. A week later each pilot received a packet that included a cover letter, the questionnaire, a pre-addressed postage-paid return postcard to confirm eligibility and to let the NAOMS team know whether or not the addressee agreed to participate, and a pre-addressed postage-paid return envelope for the completed questionnaire. A week after sending the first packet Battelle's CPHRE staffers sent a postcard to participating pilots to remind them to return the questionnaire if they had not done so but to ignore the postcard if they had. One week later a second packet was mailed if the pilot had agreed to participate but the questionnaire had not been returned. Except for a slightly revised cover letter, this packet was the same as the first packet.

Similar correspondence was sent to potential interviewees for the CATI and in-person interviews requesting their participation. Upon receiving confirmation of their willingness to be interviewed, appointments were arranged.

Computer Assisted Telephone Interviews (CATI). Both the CATI and in-person data-collection modes used trained interviewers. Interviewers serve as the interface between survey operators and respondents. Well-trained, poised interviewers are essential to the success of a professional interviewer-based survey effort. The rapport developed between interviewers and interviewees during the NAOMS survey—enhanced by the promise of anonymity—encouraged candor. The NAOMS project team employed CPHRE to conduct its surveys because of CPHRE's experience in performing surveys involving sensitive information.

It might be thought that it would be important to have knowledgeable aviation personnel asking the survey questions. However, the use of professional and highly trained interviewers supported the desire of the NAOMS project team to use the standard survey methodological approach in which each interview is conducted in exactly the same manner. This approach has been verified as the best way to minimize the interviewer's influence on participant responses and to maximize data quality (Fowler and Mangione, 1990).

CPHRE trained nine interviewers for the field trial of the NAOMS survey of air-carrier pilots. They trained as a group for a total of 12 hours. This was followed with supplemental personal training. Interviewers were introduced to the study, its background, and purpose and they were given an overview of aviation and aircraft terminology. The interviewers were briefed on the issues pertaining to different recall periods and to panel or cross-sectional selection; on the letters that had been sent to the respondents; on the need for confidentiality; and on various administrative forms and procedures. Group role-playing helped interviewers to practice administering the questionnaire and to anticipate responses to questions pilots might ask.

Each interviewer was tested by conducting actual interviews with NAOMS team members during which they were challenged with errors, balking, and other faults. Finally, the interview process itself was carefully scripted. Interviewers were very disciplined and did not depart from the script. The NAOMS team developed standard responses for the interviewers for those few questions where pilots might ask for additional clarification.

After being certified, CATI interviewers began making calls from the telephone center at CPHRE using CATI versions of the questionnaire. Battelle's CPHRE validation staff silently monitored the interviewers. The introductory telephone script notified pilots that supervisors could monitor calls for quality assurance purposes.

In-Person Interviews. Certified CPHRE field interviewers called pilots and scheduled appointments for in-person interviews. Pilots were allowed to choose an interview location, including a home or domicile airport. Interviewers conducting in-person interviews were required to wear photo-ID badges identifying them as members of the study workforce. Interviewers also carried a letter signed by the NASA project managers identifying them as legitimate interviewers for the study.

Interviewees were asked to provide mailed feedback on the quality of the interview experience using a form supplied by Battelle CPHRE.

Field Trial Findings. The NAOMS team expected to find positive and negative aspects for each of the collection modes (SAQ, CATI, and in-person interviewing) with respect to the cost, data quality, and response rate. Although in-person interviewing proved effective, this mode was terminated early in the field trial due to the excessive time and cost of implementation. Consequently, findings of the field trials are presented for only CATI and SAQ methods.

Collection Mode Effect on Cost. The field trial of the NAOMS survey provided validation of earlier estimates of the cost of conducting the survey. As expected, the least expensive data collection method proved to be the self-administered questionnaire. Based on field trial numbers, the NAOMS project team estimated that data collection costs for a fully operational program would run $85 per CATI interview and $67 per SAQ. This estimate is in 1999 dollars and is for a survey questionnaire of moderate length.

Collection Mode Effect on Data Quality. A relative measure of data quality is the time spent to complete the questionnaire, the assumption being that the more time a respondent takes to complete a questionnaire the better the quality of the resulting data. During the field trial, respondents who completed a SAQ spent on average about 40% less time on the survey than those interviewed by CATI (see Table 3).

Table 3. Indirect Measures of Data Quality

Mode	*Mean Completion Time (Minutes)*	*Respondents Failing to Complete All Questions (Percent)*
Self-administered (SAQ)	17	4.8%
Telephone (CATI)	29	0.0%

It takes less time to read, understand, and, possibly, skip a question (especially when it is one of many to which the response is “none” or “never”) than it does to listen to each one being read over the telephone and responding. Nevertheless, the presumption is that the significantly shorter amount of time spent in completing the SAQ is indicative of pilots working through the questionnaire quickly, paying less attention to questions, and spending less time trying to accurately recall events. The appropriate response to most of the questions in the NAOMS survey was “0” (i.e., the respondent experienced no occurrence of that safety event during the reference period). This characteristic of the questions in Part B is likely to tempt respondents to skip quickly across questions in the SAQ mode and therefore take much less time to complete than the telephone-interview mode.

Another indirect measure of data quality is the number of missing responses for the questionnaire, i.e., unanswered questions. As shown in Table 3, there were no missing responses for the telephone mode of interview during the field trial. When an interviewer reads a question during CATI, a response is required (including that the pilot chooses not to respond) which explains the fact that all questions were completed, at least during the telephone interviews of the field trial.

Another indicator of the quality of a respondent's data is the relationship between the reported number of events and the total hours (or legs) flown in the recall period. If the questionnaire is capturing accurate responses from pilots about the frequency of events they experience, then pilots with more flight time should experience and report a proportionately greater number of events than those pilots who flew fewer hours.

Statistical analyses of the number of events reported and number of hours flown during the recall period showed positive associations for both CATIs and SAQs (i.e., the number of safety events increased in conjunction with reported flight activity during the recall period). However, the CATI data from the field trial showed a greater number of events being reported per hour flown, suggesting a more deliberate approach to completing the survey than in the SAQ mode.

Collection Mode Effect on Response Rate. The response rates for the SAQ (70%) and the CATI (81%) are both excellent. However, it is important to point out that regardless of mode these rates were achieved as the result of multiple reminders to those pilots who had agreed to participate. With most respondents, more than one request was needed before a successful interview was accomplished. Follow-on contacts were needed because the pilots did not respond to earlier requests, had scheduling conflicts, lost the original mailing, etc.

After due consideration of all of the factors, the NAOMS team decided that the CATI mode was the best compromise for conducting the NAOMS air-carrier-pilot survey

4.4.5 Sample Source and Size

Although it was recognized that some events are so rare that it was unlikely that reliable measures of changes over time could be achieved in this limited experiment, an objective was to attain solid, statistical estimates for many, and hopefully most, of the questions in Section B of the NAOMS air-carrier-pilot survey.

4.4.5.1 Sample Source

Issue:

The NAOMS team had to find a practical, accessible source of information on pilots who would qualify as participants in the air-carrier-pilot survey.

Resolution:

The NAOMS team undertook an investigation to determine the number and distribution of aviation operational personnel working in commercial aviation organizations within the United States. The results of this study are presented in Appendix F. The data for 1998 indicated that there were about 127,000 pilots in the United States with air-transport-pilot (ATP) ratings at the time the NAOMS survey was being designed. It was estimated that there were 50,842 pilots employed by major carriers, 8,314 pilots employed by national (i.e. mid-size carriers), 13,323 employed by regional carriers, 10,468 employed by cargo-only carriers, and 3,713 employed by non-scheduled carriers for a total of 86,660 employed commercial pilots[5]. This was used as the estimate of the full population of active air-carrier pilots at the time of the field trial.

In the process of selecting pilots for the survey, NASA wanted to: (1) maintain the independence of the NAOMS-survey data-collection effort; (2) have a logistically simple process of selecting potential respondents; and (3) avoid any appearance of intruding on the privacy rights of respondents. After careful consideration and investigation of possible sources, the NAOMS team chose to use the publicly available FAA-maintained Airmen Certification Database. This registry identified potential participants who were current in their certification. The Airmen Certification Database, in part because it was publicly releasable, was deemed to be the most practical source of interviewees and would suit the objectives of the experimental demonstration of NAOMS survey.

However, there was no way to be certain that a pilot in the Airmen Certification Database was an air-carrier pilot or was active at the time. Therefore, the NAOMS team introduced criteria to maximize the likelihood that the pool from which the interviewees were selected was composed of pilots who very likely had been actively flying as air-transport pilots during the recall period. The following criteria were established for a pilot to qualify for the sample pool used in the NAOMS survey:

- have an Air Transport Pilot (ATP) certificate
- be U.S. based
- have a current six-month first class medical certificate
- have a multi-engine rating
- have a Flight Engineer certificate

It was recognized that these five criteria (especially the requirement for a Flight Engineer certificate) would eliminate people from the sample pool of the NAOMS survey who in fact flew in the cockpits of air-carrier aircraft. However, they also markedly reduced the number of pilots who otherwise would have been contacted but who were not active air-carrier pilots. These criteria helped minimize the effort to find air-carrier pilots eligible for the NAOMS-survey sample pool.

Another event occurred that might have affected the NAOMS-survey sample pool. A factor that has implications with respect to both response rate and data quality of a survey is self-selection bias which results when prospective respondents decide to "opt out" of participating in the survey. Opting out is generally a reflection of a refusal rate, or unwillingness, to do the survey. In the course of the development of the NAOMS survey experiment, a different kind of opt out occurred—one that was unrelated to the NAOMS survey. Between the time of the field trial and the initiation of the full survey, pilots were granted the option to opt out of the publicly releasable Airmen Certification registry thereby avoiding commercial solicitors and protecting personal privacy. The Air Line Pilots Association (ALPA) and the Allied Pilots Association (APA) provided their members with a convenient way to opt out of being included in this

[5] At this time, there were also 1,772 helicopter pilots. However, these were not included in the pool of air-carrier pilots from which the sample for the NAOMS survey was taken and to which the results of this study would apply.

database. However, members of both the ALPA and the APA did remain in the registry and did participate in the NAOMS survey. Although it is possible that the "opt outs" introduced a bias in the pilot distribution remaining in the registry, it seems unlikely that the desire for privacy would be other than random with reference to safety-related events.

Within a few months of the onset of routine NAOMS-survey data collection, an initial comparison of NAOMS-survey and Bureau of Transportation Statistics (BTS) data showed that the NAOMS-survey data samples were biased towards wide-body flight operations and away from small transport operations. While the BTS data indicate that 35% to 45% of the legs flown by air-carriers in the years 2002 to 2004 were by aircraft in the Small Transport category, only about 10% of the legs reported to the NAOMS survey were flown by Small Transport pilots. Also, about 55% of NAOMS-survey respondents indicated that they flew mainly as air-carrier Captains, whereas a roughly 50/50 split between Captains and First Officers would have been expected. Both of these effects are likely due to the limitation to pilots with both ATP and Flight Engineer certificates as these are probably the older pilots who become Captains and fly wide-body aircraft.

Implications of both the selection criteria used and the opt-out option will be discussed in Section 7.

4.4.5.2 Sample Size

Issue:

The critical factor in selecting the sample size is the number of respondents needed to attain reliable estimates for questions in Section B of the NAOMS air-carrier-pilot survey.

Resolution:

The results of the field trial of the NAOMS air-carrier-pilot survey, coupled with some relatively simple statistical calculations and simulations, provided the basis for selecting the air-carrier-pilot sample size for the NAOMS survey. NASA's principal concern was that the NAOMS survey be able to detect operationally meaningful changes over time in Section B rate data. As will become clear in succeeding paragraphs, the available budget became the constraining factor for the sample size selected.

The field trial demonstrated that the incident types being tracked by the NAOMS survey with the questions of Section B covered a very wide range of event frequencies as shown in Figure 4. Pilots might experience some classes of events frequently (e.g., encountering wind shear), whereas other types of events might never be experienced in an air-carrier-pilot's entire career (e.g., going off the edge of a runway or taxiway while taxiing). Detecting trends or other changes with time, pertaining to relatively high-frequency events, requires a small sample size; whereas detecting rate changes for relatively rare events potentially requires a much larger sample size.

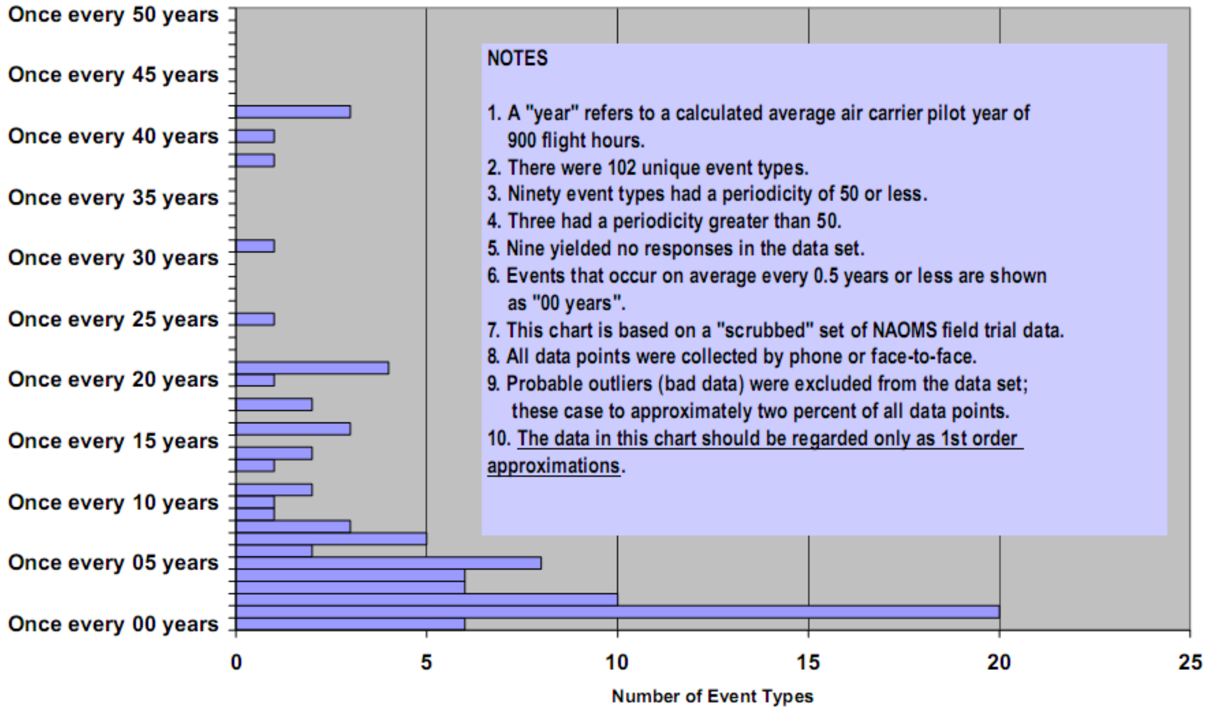


Figure 4. Distribution of the periodicities of NAOMS events.

For an adequate demonstration of the concept, the team determined it needed a sufficient number of events reported in the limited duration of the NAOMS survey to be able identify a 20% change in an event rate with 95% confidence. A 20% change in a safety event rate would be a large change given the maturity of the U.S. aviation system and would generally occur only as a consequence of a safety intervention or a serious system perturbation. If the survey could be continued over a long period of time, trends could be observed with diminishing year-to-year differences.

Simulation studies based on field trial data suggested that as many as 65,000 interviews would need to be conducted each year to achieve rate estimates with tight confidence intervals for every question in Section B of the air-carrier-pilot survey. This number of interviews per year was far in excess of NAOMS's project budget and might have imposed burdens on the respondent community. It was clear that some compromises would need to be made because the NAOMS project had neither the mandate nor the budget for a survey effort of that magnitude.

The dilemma that the NAOMS team faced was that the less-frequently-occurring event types are often those with the greatest safety significance whereas many more-frequently-occurring events could be dismissed as having minimal safety meaning.

The NAOMS team concluded that it was necessary only to show changes in less rare events to demonstrate the capability of the methodology to identify relative changes. Given enough time (were the industry to adopt NAOMS as a permanent measure of system safety) relative changes in more rare events, where they occurred, also would be demonstrated. As a result, the NAOMS project team arrived at a compromise between considerations of the available time and funds and the desired statistical accuracy for the largest number of questions. They established a goal of 8,000 completed air-carrier surveys each year. This was done with the understanding that this sample size would permit the NAOMS survey of air-carrier pilots to achieve event rate estimates with satisfactorily tight confidence intervals for a significant portion of the questions of Section B and rate estimates with broader confidence intervals for most of the rest of the questions. In fact, a simulation study using data on event frequency collected using the records of the first 8,000 interviews of the air-carrier-pilot survey indicated that the NAOMS survey would be able to detect 20% rate shifts for about half of Section B questions with 95% confidence should such changes occur (see Appendix G). Although 8,000 interviews per year was the target of the operational survey of air-carrier pilots, the actual sample attained was about 7,000 interviews annually for each of the 3¾ years of the experiment, bringing the total sample for the April 2001 to December 2004 time period to 26,170 interviews completed. Moreover, as discussed in Section 5, the discovery of the need to consider data by aircraft-size categories further reduced the number of events for which the desired reliability could be expected in this demonstration.

In summary, the NAOMS team determined that the full population of current air-carrier pilots to be addressed by the survey was about 87,000; that although it was not perfect, the most practical source of information on this group was the publicly available FAA-maintained Airmen Certification Database and that 7,000 to 8,000 interviews per year was the best compromise between cost and statistical accuracy.

4.4.6 Selecting Respondents

Issue:

Having determined the approximate size of the full population of active air-carrier pilots and the sample size for acceptable statistical accuracy, the NAOMS team needed to find a way to obtain the names and addresses of the potential survey respondents to construct a “sample pool” from this group. The next challenge was to engage the respondents to achieve an adequate response rate and quality.

Resolution:

As noted above, the NAOMS team chose to use the publicly available FAA-maintained Airmen Certification Database to obtain the names of potential respondents who designated themselves as U.S.-based commercial aviation pilots flying multi-engine aircraft. However, the NAOMS team knew that a portion of these potential respondents might be ineligible because the information in the Airmen database is not always current. Most of the Airmen database records were about two years old even though pilots can update their information online.

For all three interviewing modes used in the field trial, only those pilots were selected for whom telephone numbers could be obtained and for whom the addresses provided in the Airmen database were currently correct. As telephone numbers are not listed in the Airmen database, the NAOMS team used *Telematch*©, a service that matches names and addresses with telephone numbers in order to identify pilots that could be in the sample pool of potential respondents. Missing addresses and/or the inability to locate telephone numbers reduced the sample pool by about 25%.

The workshops, briefings, presentations, and focus groups during the formative years of the NAOMS project that are discussed elsewhere in this report not only engaged the aviation community in the development and design of the survey of air-carrier pilots but they also served to advertise and encourage participation from the community when they were asked. By the time the field trial and the full survey were initiated, the NAOMS team found a receptive pool of air-carrier pilots. The NAOMS team developed a very successful routine of systematic mailings followed up with multiple reminders that explained the objectives, requested their participation, enabled them to choose whether to participate or not, and assured anonymity. The response rate during the 3¾ years of the air-carrier pilot survey was consistently well above the "gold standard" for surveys.

4.4.7 Random versus Panel Sampling Approaches

Issue:

Surveys are often conducted by randomly selecting a participant for interview from the sample pool. This is a "random" design. Another approach is to select a participant from the sample pool and ask him or her to periodically complete the survey over an extended time period. This is called a "panel" design. Panel design can allow researchers to measure the experiences of that individual over time. The "Nielsen" survey used to monitor television-viewing habits is a panel survey design with which most people are familiar.

The NAOMS team needed to evaluate the advantages and disadvantages of the random and the panel designs to decide which was most suitable for the objectives of the NAOMS survey.

Resolution:

It was not possible to fully explore the issue of random versus panel designs during the brief duration of the field trial because the panel approach would have required repeated interviews with the same pilot over the course of a year. Therefore, as in the resolution of the recall period discussed above, a decision on the use of random or panel sampling was postponed until the initial operational phase of the program.

The panel design introduced a number of logistical challenges that were not present in the random design. However, the panel design also was thought to lead to a high response rate, while the response rate from the random sample was yet to be determined. The goal was to evaluate random and panel designs to determine which gave better insight into the safety issues confronting air-carrier pilots.

During the first 9 months of implementation, the NAOMS survey of air-carrier pilots used a split-design whereby approximately half of the interviews were conducted with respondents chosen at random from the sample pool and the other half with respondents who agreed to join a NAOMS-survey panel. A comparison of the results showed that the demographics of the two groups (as reflected by responses to questions in Section A) were similar. Furthermore, the random design resulted in a high response rate comparable to that of the panel design.

Several factors led the NAOMS team to its decision to use the random design. First, the panel approach for the NAOMS survey had been suggested because enlisting respondent cooperation in this manner could produce higher response rate and better response quality. However, during the field trials and the first 9 months of implementation, it became clear that the NAOMS survey of air-carrier pilots would achieve a very high response rate and quality from randomly selected respondents and that virtually all of the randomly selected persons who agreed to participate would complete the survey. Consequently, this presumed benefit of the panel design did not exist.

Second, the repeated observations produced by the panel design can be statistically useful if these repeated measurements could be analyzed within participant across time. However, the de-identification policies favored for the NAOMS survey respondents made this very difficult to implement.

Finally, panel designs are logistically more complex to administer.

With no compelling advantage attributable to the panel approach, there was no reason to incur this added burden. Consequently, after the first 9 months of data collection, the decision was made to use the random (cross-sectional) design for all subsequent data collection efforts.

4.5 NAOMS Development Roadmap

Figure 5 shows the sequence of project development steps from project commencement in 1998 through December 2004 when NAOMS-survey data collection ended. During the two subsequent years (2005–2006) the NAOMS team developed a streamlined version of its data-collection system, adapted it to a Web-based survey, and transferred the capability to the ALPA for continued surveys of the Part 121 pilots on behalf of the CAST.

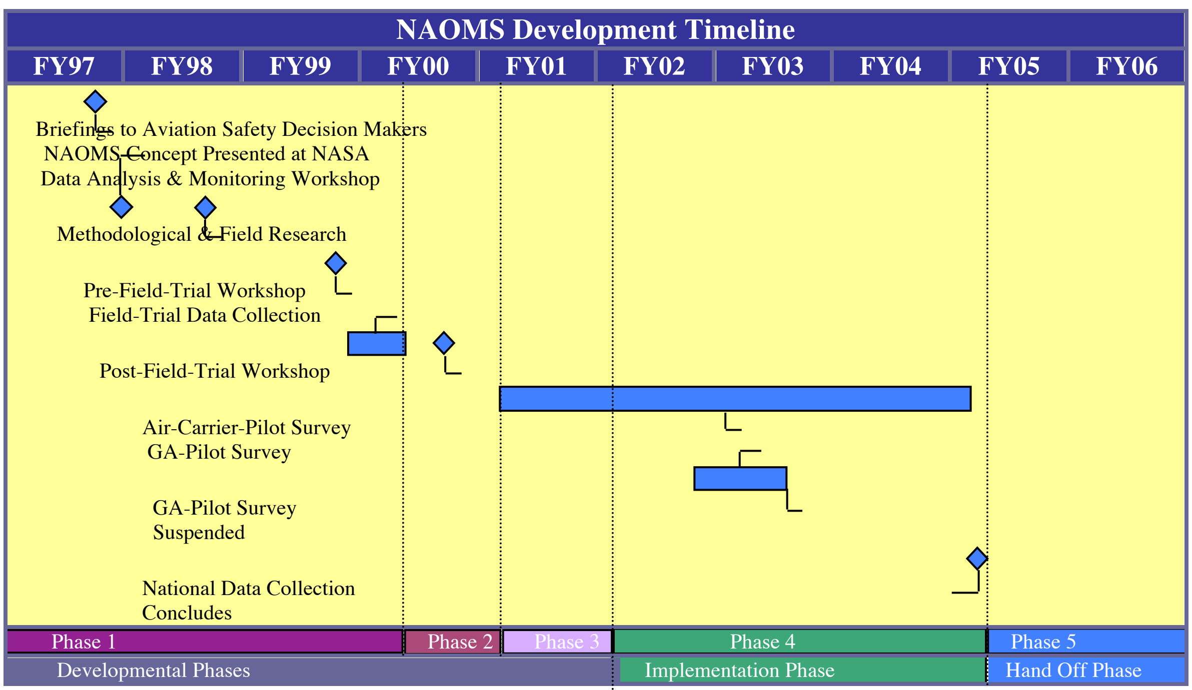


Figure 5. NAOMS development timeline.

The following paragraphs briefly describe the key steps shown in Figure 5. Other sections of this report address these topics in greater detail. Phases 1, 2, and 3 were considered the Development Phases.

Phase 1:

Literature Review: The team examined published studies in order to learn about what worked with previous survey research efforts involving aviation personnel, including pilots. It was determined that little had been published that was specific to aviation. Some information was gleaned from analogous work done in other professional operating domains.

Questionnaire Development: With the help of focus groups, the team developed a first draft survey instrument; a "straw man" for what eventually became the NAOMS air-carrier pilot survey.

Methodological Research: The NAOMS team conducted research focused on topics such as the demographics of pilots and other aviation operating personnel; pilot event recall and memory organization; the identification of high priority aviation safety topics; and related matters.

Pre-Field-Trial Workshop: An industry workshop was held in May 1999, in part to provide an opportunity for industry participants to review and critique the draft questionnaire. That draft was revised in response to input from workshop participants.

Field Trial: Once the research and methodological experiments had been conducted, the team developed a sampling plan. The team then ran a small-scale field trial from November 1999 to February 2000 with several purposes. The NAOMS team was seeking additional feedback on the survey instrument, obtaining its first measure of probable response rates, further exploring pilot-recall capabilities, and—importantly—obtaining first approximations of the likely safety event frequencies that respondents would be reporting once NAOMS survey of air-carrier pilots began.

Phase 2:

Post-Field-Trial Workshop: A second workshop was held in March 2000. Attendees were briefed on field trial tests of the NAOMS-survey questionnaire and interview process. Additional revisions were made to the questionnaire in response to input offered at the workshop.

Phase 3:

First Nine Months of Air-Carrier-Pilot Survey: The NAOMS survey of air-carrier pilots using the final survey questionnaire started in April 2001. However, the first 9 months were used to investigate issues such as recall period and panel vs. random sample that had not been fully resolved by the field trial.

Phase 4:

Air-Carrier-Pilot Survey Implementation: The NAOMS full survey of air-carrier pilots using the final survey questionnaire and final protocols started in February 2002 and continued until December 2004.

GA-Pilot Survey Implementation: The concept of NAOMS was to adapt and implement the survey process to collect data on aviation safety from a variety of aviation operations groups, including pilots, flight attendants, maintenance technicians, air traffic controllers, and others. Unfortunately, time and funding were insufficient to survey every operational group. However, a survey of GA pilots was developed and interviews were conducted from August 2002 until May 2003 when funding for this activity was cut. The present report addresses results only of the survey of air-carrier pilots.

Phase 5:

Hand Off: The NAOMS team developed a streamlined, Web-based version of its air-carrier-pilot survey questionnaire and transferred the operation to ALPA.

4.6 Reviews and Approval Process

4.6.1 Aviation Industry

The adaptation of the classical survey methodologies to each domain of aviation operations and to the specific objectives of the survey of that domain required special considerations and decisions in the design of each survey instrument. The NAOMS team made an extraordinary effort to engage the aviation industry and academia in the development of the survey concept. Validation of these decisions entailed presentations in a variety of venues to representatives of the aviation operations industry, the unions, the FAA, and academia who provided peer reviews of not only the fundamental methodology but also the decisions regarding objectives, content, instrument design, and the specific questions. Content for the survey was developed through a series of focus groups, personal interviews with active commercial pilots, discussions with various professional aviation organizations, and also through workshops and briefings to the FAA, industry, pilots' union, and academia that were conducted during 1998 through 2000. These workshops and briefings served the dual purpose of providing information to the community and gathering the input from the community.

4.6.1.1 Focus Groups and Personal Interviews

Focus groups and personal interviews were used to identify the safety issues of greatest interest to line pilots and to provide feedback on the proposed structure of the questionnaire. A total of 37 active air-carrier pilots (flying both international and domestic) were interviewed during three focus groups convened in the Washington, D.C., area during August and September 1998. During each focus-group session, the pilots were asked to list all of the safety-related events that pilot experienced. During nine one-on-one interviews, pilots were asked to describe all the safety-related events they themselves had witnessed during their careers. The NAOMS team developed a first-draft survey instrument (a "straw man") for what eventually became the air-carrier-pilot survey based on the inputs from these focus groups and interviews.

4.6.1.2 Workshops

The NAOMS project team sponsored two workshops during the development phase. The first of these preceded the field trial; the second followed it.

The first workshop was convened in Alexandria, Virginia, on May 11, 1999, when the concept of NAOMS was still in its early development. This first workshop was to acquaint stakeholders with the objectives and methods of NAOMS and for participants to review and critique the NAOMS concept, the survey methodology, and the draft of the questionnaire for the field trial. There were 76 participants representing the FAA, Department of Transportation, Department of Defense, air-carrier operators, manufacturers, software vendors, unions, and academia in this first workshop. Although the meeting was open, the attendees were largely respondents to invitations that sought participants who were experienced in survey methodologies, air transportation operations, and/or aviation safety to serve as peer reviewers of the NAOMS concept and the adaptation of survey methods to aviation. The NAOMS team also used this forum to enlist the support of this community in implementing the program and especially to elicit their participation in the field trial. Appendix H provides the workshop agenda, the list of attendees, the questions addressed during the sessions, and a summary of the feedback from workshop participants on each of the questions. The draft questionnaire for the field trial was revised in response to these comments and recommendations.

The FAA Office of System Safety conducted a survey of FAA staff members who attended this first NAOMS Workshop. In general, the comments were supportive of the concept, but there were questions about its implementation and the design of the questionnaire, many stemming from the incorrect expectation that a NAOMS survey by itself would provide sufficient information to identify causal factors and appropriate intervention strategies and, thereby, affect accident rates. In response, it was explained that NAOMS was designed to be primarily a pointer to events with a trend towards decreased system safety; other sources of data would be required to identify the causal factors of that trend. The FAA inputs and NASA responses to those inputs are summarized in Appendix I.

The second workshop was held in Washington, D.C., on March 1, 2000. Its purpose was to update stakeholders on progress being made toward NAOMS implementation and, in particular, to report on the findings of the field trial. There were 37 invited participants in this second workshop with representations similar to that of the first. Appendix J presents the workshop agenda, the attendance list, the issues posed by the NAOMS team and addressed during the sessions, and a summary of the feedback from the workshop participants on those issues.

The NAOMS team considered all of the feedback from both workshops in the final design of the survey instrument for air-carrier pilots.

4.6.1.3 Field Trial

After the research and methodological experiments had been completed, the focus groups and the first workshop had met, and all the inputs consolidated, the NAOMS team developed a sampling plan using a random approach and conducted small-scale field trial data collection from November 1999 to February 2000. The field trial had several objectives. The NAOMS team sought additional feedback on the survey instrument, obtained its first measure of probable response rates, further explored pilot-recall capabilities, and most importantly obtained coarse preliminary approximations of the safety event frequencies that respondents would be reporting once NAOMS became fully implemented. In April 2000, after all field trial data had been collected, the NAOMS team produced the report on field-trial results that is provided in Appendix K. The results, most of which were summarized in the discussions of the issues and resolutions in Section 4, pointed the way toward implementation of the full air-carrier-pilot survey. As was previously mentioned, the 3-month field trial was not able to answer all the questions and so certain studies were continued and completed within the first nine months of the full survey using the final questionnaire.

4.6.1.4 Briefings

In addition to the two workshops, the focus groups, and the individual interviews, from 1998 through 2005 the NAOMS team conducted 13 briefings to external stakeholder organizations about the program's status and reported preliminary findings to other federal agencies. These briefings are listed in Table 4.

Table 4. Briefings of NAOMS by Project Team

Date	*Location*	*Audience*	*Subject*	*Preliminary Data Included?*	*Comments*
3/5/98	Washington*	Aviation Safety Investment Team (ASIST)	NAOMS concept	No	
11/13/98	Ames*	NASA ASRS Advisory Subcommittee	Review development approach	No	
1/26/00	Langley*	Aviation Safety Program Executive Committee (AvSPEC)	Program overview; partial field trial results	Yes	Preliminary results of 516 pilot interviews as of January; planned conclusion by February 2000.
8/28/02	Ames	Participants in ICAC study	ICAC results	Yes	Share information on ICAC from various ASMM sources.
12/5/02	Langley	AvSSP leadership	Program overview; preliminary results	Yes	Report on project status and results to date.
4/9/03	Washington*	FAA senior management	Detailed program view; results to date	Yes	Share preliminary air-carrier-pilot data results.
5/7/03	Ames	National Research Council (NRC) Review Committee	NAOMS program review	Yes	Describe organization, approach, methodology, and results to date.
8/5/03	Newport, RI*	FAA and CAST/JIMDAT	NAOMS overview and status	Yes	Present NAOMS concept; discuss applicability to JIMDAT needs.
6/16/04	San Francisco*	CAST/JIMDAT	Construction of Section C for JIMDAT study	No	Discuss questions on the Safety Enhancements selected by JIMDAT for demonstration study.
9/1/04	Washington*	FAA ATO	Program overview; Section C ICAC study results	Yes	Overview of preliminary data; request FAA Performance Analysis Office join in NAOMS.
9/8/04	Ames*	FAA Tech Center	Program overview	Yes	Data briefing.
1/26/05	Washington*	CAST/JIMDAT	JIMDAT study (Section C) results	Yes	Data analysis report.
1/28/05	Washington*	CAST	JIMDAT study (Section C) results	Yes	Data analysis report.

Note: Meetings highlighted in green and noted with an asterisk (*) included participation by representatives of the FAA.

4.6.2 Government Reviews

4.6.2.1 Programmatic Reviews

All elements of the ASMM Project, including NAOMS, were conducted in full compliance with the descriptions, definitions, and requirements in the NASA Procedure Guidelines (NPG) 7120.5a. The documentation of ASMM Project Plan and its annual revisions were written accordingly and reviewed and approved by the AvSSP Office. Further, the progress and plans for NAOMS including assessment of technical status, cost, schedule progress versus plans, and identification of problems, issues, concerns and any major interactions with industry were part of the ASMM's Project Manager's monthly report to the AvSSP Manager.

In addition to these regularly scheduled program reviews, the NAOMS team participated in various other ad hoc reviews that had been established to communicate information on AvSSP to diverse committees and agencies. For example, the Agency sponsored Independent Implementation Reviews and independent AvSSP Executive Council meetings to assess project stability, technical progress, technology impact, and technology relevance.

4.6.2.2 Special Required Reviews

Certain reviews were required because the survey dealt with the public and entailed human subjects.

Office of Management and Budget (OMB). The OMB requires review and approval of all surveys of the public conducted by a Federal agency. Consequently, representatives of the OMB were provided with requested information in the submission for approval of the air-carrier survey and in the subsequent submission for approval of the GA-pilot survey in accordance with OMB #2700-0099 (see Appendix L). The issues addressed in OMB's review process included justification, why the information was necessary and non-duplicative, to whom it would be useful, how the data were to be collected and processed, assurance of confidentiality of respondents, the time burden on respondents, the cost to government, collection procedures and analysis plans, methods of maximizing response rates, and identifications of the survey methodologists and statisticians. The NAOMS survey was announced publicly in the Agency Federal Register of May 23, 2000, as required by OMB.

Institutional Review Board. The NAOMS concept and planned implementation were reviewed and approved by the NASA Ames Research Center's Institutional Review Board (ARC IRB) responsible for review, approval, and monitoring of research protocols involving human subjects. The ARC IRB operates under the guidance of the NASA Institutional Review Board as outlined in Ames Procedural Requirements (APR) 7170.1.

4.6.2.3 Other Special Reviews

FAA Review. In 2003, subsequent to a briefing by the NAOMS team to FAA senior management, the FAA convened an ad hoc group from various elements within the FAA to evaluate the NAOMS survey concept and to make recommendations on the FAA's role to senior management. This committee documented its concerns, questions, and suggested revisions in an internal FAA memorandum. The committee concluded that information on trends of rates of events, the main objective of the study, was very useful and they recommended that the FAA support the NAOMS survey of air-carrier pilots, at least through the processing of the existing data.

National Research Council Review. In 2003, a Panel of the National Research Council (NRC) conducted a review of the entire Aviation Safety and Security Program. During its review of the ASMM Project, the Panel reviewed the NAOMS survey methodology, the survey-instrument design, the findings of the field trial, and a few preliminary findings from the survey of air-carrier pilots. The NRC Panel's primary finding with respect to the NAOMS project was that NAOMS resources might be combined with those of the ASRS to enhance the ASRS capability. NASA's response was to clarify the obvious misunderstanding that the Panel had about NAOMS and its relationship to the ASRS. The ASRS is a voluntary reporting system and therefore is not a sound basis for statistical evaluations of frequencies of occurrences or trends. NAOMS was designed to complement information gained from voluntary reporting systems such as the ASRS by providing statistically reliable information on the relative frequency of events, thereby allowing assessment of temporal changes and of the impact of intervention strategies.

5. Conduct of the Air-Carrier-Pilot Survey

The survey began in April 2001 and interviews continued until December 2004. During the 45 months of this initial experiment, a total if 26,170 telephone interviews of air-carrier-pilots were completed. The interviews were conducted evenly over each 12-month period to ensure that seasonal impacts on some of the events (e.g., weather-related events) would be detected reliably.

5.1 The First Nine Months

As was discussed in Section 4, the three months of the field trial were not long enough to determine the optimal recall period or to determine whether to use random selection or panel selection of interviewees. Therefore, the first nine months of the full survey (from April 2001 through January 2002) were used to collect data with which to resolve these questions. These first nine months were viewed as a trial period and were considered part of the NAOMS Development Phase. It was essentially a second field trial using the final version of the questionnaire.

5.1.1 September 11, 2001

Five months after the start of the full NAOMS survey in April 2001 (about mid-way through the nine-month Development Phase), the nation suffered the attack of September 11, 2001. The repercussions from this attack were felt nationwide and resulted in impacts to the national air transportation system, which included an immediate reduction in the number of aircraft flown, followed by a return to more normal operations.

The NAOMS study relies on rates, i.e. the number of events experienced for a given period of exposure. The NAOMS team believed that, although the occurrences of a few of the events addressed by the questions in Section B of the survey might be affected by reduced volume of traffic in the national airspace, the large majority of questions would be unaffected since they were independent of how many aircraft were flying at a given time. The fact that when 9/11 occurred different recall periods, as well as panel and random modes, were still being investigated made a formal pre- and post-9/11 analysis unfeasible. However, the team did note the rates reported following September 11th and informally compared them with those reported prior to September 11th.

As anticipated, for the vast majority of questions there was little evidence of rate changes between the two periods. There was one clear exception to this general observation. This exception was to Question CP3, which dealt with a pilot leaving the cockpit to attend to a passenger disturbance. After 9/11, the FAA issued strict instructions for structural and procedural changes restricting access to the flight deck and the flight crew leaving the flight deck throughout flight. The results of those restrictions were clearly evident in the NAOMS-survey data, even with the cursory analysis the team was able to conduct.

Both the similarity of rate results for most questions and the clear reduction in reported rates of pilots leaving the cockpit provided some assurance that the NAOMS-survey approach was working as intended. By the end of the nine-month Development Phase, air traffic had returned to normal, the 60-day recall and the random selection mode decisions had been made, and the stage was set for the acquisition of the data that became the basis of the analyses shown in Section 7. The data from the first nine-month period were not used in the primary data analysis of this report.

5.2 The Data

The air-carrier-pilot response to the NAOMS survey was enthusiastic. The NAOMS survey achieved exceptional completion rates—81% of those contacted and found to be eligible completed their interviews.

5.2.1 Questions

In March 2003, midway through the air-carrier-pilot survey, a simulation study was conducted to estimate sample-size requirements for each event addressed in the questions of Section B of the air-carrier-pilot survey using the records of the first 8,000 interviews that included 30-, 60-, and 90-day recall periods. Results were obtained for 88 event questions. According to the analysis of data run at that time, the NAOMS survey would be able to detect 20% rate shifts with 95% confidence for 47 events (53%) with 8,000 interviews per year.

After the completion of the data collection from the air-carrier-pilot survey in December 2004 and prior to the final analyses of the data, a power analysis was performed on the 60-day recall data that were collected over the final three years to determine which questions provided sufficient information to support reliable analyses of relative changes. Questions that displayed a sufficiently large number of events to allow detection of a 20% change in event rate with 95% confidence, should one occur, were retained for further analysis. Of the 91 questions in Section B of the air-carrier-pilot survey with numeric answers, it was determined that respondents had reported enough events on 43 questions to support further analysis (based on approximately 7,000 interviews per year). The remaining questions (those reflecting rarer events) would require a larger sample size to produce sufficient power to justify further analyses.

5.2.2 Re-Classification of Pilots

During the survey of air-carrier pilots, a similar survey of GA pilots was conducted for about nine months. A number of interviewees contacted during the GA survey were found to have been operating primarily as air-carrier pilots during that period. When the interviewer made that determination, he/she shifted from the GA survey questionnaire to the air-carrier survey and collected data based on that survey. By the time the GA survey ended, 407 pilots selected as part of the GA survey process had been identified as having flown as air-carrier pilots during the recall period. As these 407 pilots had not been required to meet the selection criteria that had been imposed for eligibility in the sample pool used for the NAOMS survey, they were not included in the analyses of the survey of air-carrier pilots. However, these 407 pilots were valuable in assessing the impact of the selection criteria, as will be discussed in Section 6.

The raw data file of the NAOMS air-carrier-pilot survey is defined as the original capture of the 25,763 (26,170 minus 407) survey-participants' responses to the series of questions. A small subset of the 25,763 interviews (208 interviews) could not be used for analyses because respondents did not provide key operational risk exposure data (i.e., either the hours or legs flown during the recall period). The balance of 25,555 air-carrier pilots supplied the needed risk-exposure data (see Table 5).

Table 5. Number of Interviews

Recall Period	*Number of Interviews Conducted*	*Information on Exposure Missing or in Error*	*Number of Interviews Available for Analysis*
30-day	3,152	16	3,136
60-day	18,611	144	18,467
90-day	4,000	48	3,952
Total	*25,763*	*208*	*25,555*

The usable data were further reduced because the aircraft make/model could not be identified from the information recorded during 12 interviews. The elimination of these 12 cases left a total of 25,543 interviews, of which 3,134 interviews used 30-day recall, 18,460 interviews used 60-day recall, and 3,949 interviews used 90-day recall.

Furthermore, there were 52 cases of aircraft identified during the interviews that fell below the 5,000 lb. gross takeoff weight (GTOW) lower boundary that had been set for the Small Transport category. Eliminating these resulted in a total of 25,491 valid interviews, of which 3,128 interviews used 30-day recall, 18,422 interviews used 60-day recall, and 3,941 interviews used 90-day recall.

Finally, as discussed in Section 5.1, the data collected during the first nine months of the survey were used to resolve issues of recall period and panel versus random selection and are not included in the analyses described in Section 7. The results presented in Section 7 are based on the data collected over the last three years of the survey for which only the 60-day recall period was used.

5.2.3 Data Cleansing

In any large survey, erroneous data are inevitable due to interviewer typing errors, misunderstanding of the question by the interviewee, misinterpretation by the interviewer, etc. These errors often appear as outliers and are typically omitted from final analyses of actual valid data. A multistage process was used to eliminate erroneous data.

First, interviews reporting impossible or extremely improbable flight times were eliminated. The flight time per leg was calculated for each interview. Interviews that showed less than 20 minutes per leg or more than 20 hours per leg were deleted from all further analyses. There were 45 such situations in the data of Section A collected over the final years using 60-day recall period. Eliminating these reports reduced the number of interviews for the analysis in Section 7 to 18,377 interviews with 60-day recall,

Second, data for a question were deleted from consideration if the interviewee provided no information in response to that particular question. Responses to other questions for that interviewee were not deleted.

Third, for the remaining interviews and events, a case-by-case evaluation of the responses to each survey question was performed to determine whether the participant's responses indicated that the participant had misunderstood or misinterpreted the question or the interviewer had entered an incorrect value. Experienced aviation-safety researchers and an experienced air carrier pilot (retired Boeing 777 Captain) judged whether the number of events entered was reasonably likely given the type of aircraft flown and the number of hours of flight time and legs reported during the recall period. Responses to questions judged to be operationally impossible or extremely improbable were deleted for that question. The number of responses deleted varied from question to question, but in no case was that number in excess of 0.2% of the total number of interviews for a single question[6]. This supports the confidence that the NAOMS team had in the data quality achieved in the survey.

Finally, a statistical filter was applied. The distributions of responses for each question considered here were examined separately and trends across time modeled. The observations that departed greatly from the general trend were then removed and the analyses repeated. These analyses were then repeated again, this time eliminating only those observations that departed by more than eight standard deviations from the mean rate for the question. In no case did the results of the analyses using data cleaned using one procedure differ substantially from the analyses conducted on the data cleaned using the other procedure. Therefore, for each question, responses in which the reported rate differed by more than eight standard deviations from the mean were removed. This procedure always eliminated substantially less than 1% of the responses.

[6] As we were examining the data from this operational perspective, it became evident that interviewees were interpreting certain of the questions differently. This was particularly evident in the responses to questions pertaining to communications with ATC. Some counted the event as having occurred once if they had a problem with communication on a flight segment while others apparently counted the number of times they may have attempted communication on a single leg.

In the results presented in Section 7, the number and proportion of cases removed is reported for each question. For comparison, the results of an analysis based on the data containing the likely erroneous observations are also included for each question. In general, the data-cleaning procedure had little effect on the analyses other than to reduce the variance and improve the fit of the statistical models.

For example, in the case of the question asking about evasive maneuvers to avoid an in-flight collision, a total of 42 observations (0.23% of the observations) were removed. For these observations, the mean reported rate was more than 63 times greater than the mean rate for the remaining responses. Yet there was very little difference in the estimated linear trend across quarters. The slope changed from -.018 when only the cleaned data are used to -.014 when the suspect data are included.

6. Analysis Methodology

6.1 Overview

A primary purpose of NAOMS was to explore the use of survey methods for the detection of relative changes, and primarily in temporal shifts, in potentially hazardous events. Hence, the results presented in this report focus on analyses of changes over time in the rates of events reported in response to the questions contained in Section B of the NAOMS survey. For each event, rates per 1,000 legs were calculated for each quarter. These data were then examined for time trends using nonlinear regression techniques. The effects of two factors—aircraft type and type of operation—on event rates were also examined because the rate at which hazardous events are encountered may depend on the type of activity in which the pilots are engaged.

6.2 Criterion Variables

Most of these questions were phrased to elicit the number of times that an event occurred during the recall period. However, the frequencies of these events are dependent on the opportunities for the events to occur. Therefore, data on the number of hours and the number of legs (i.e., flight segments from departure gate to arrival gate) the pilot flew during the reporting period were collected. Rates of events per 1,000 legs or 1,000 hours were then calculated for each quarter. In general, hazardous events are more likely during the dynamic flight that accompanies the departure and approach and landing phases of flight. Hence, the number of legs flown is usually a better measure of exposure than the number of hours flown. Therefore, the rates of events per 1,000 legs (or the equivalent) for each quarter were the primary data analyzed.

6.3 Quarterly Data

Interviews were conducted continually throughout the years during which the NAOMS survey was administered. During the main portion of the NAOMS survey, interviewees were asked to report on their experiences during the preceding 60-day period. Therefore, the particular days covered by the reporting periods for each interviewee could differ. For ease of presentation and analysis, the data were grouped into quarters according to the mid-point of the recall period. Interviews for which the mid-point of the recall period fell between December 12 and March 11 were categorized into the first (winter) quarter of each year. Interviews for which the mid-point of the recall period fell between March 12 and June 10 were categorized into the second (spring) quarter of each year. Interviews for which the mid-point of the recall period fell between June 11 and September 10 were categorized into the third (summer) quarter of each year. Interviews for which the mid-point of the recall period fell between September 11 and December 10 were categorized into the fourth (fall) quarter of each year. This system tends to somewhat obscure differences between adjacent quarters because some of the experiences reported by respondents contacted in one quarter could have occurred simultaneously with experiences reported by interviewees contacted during the previous quarter.

6.4 Temporal Patterns

Temporal patterns can take many forms. Rates may increase or decrease linearly. They may also demonstrate curvilinear patterns, e.g., rising and then holding steady or decreasing and then rising. Cyclical patterns may also occur. For example, many potentially hazardous events (e.g. in flight icing, weather diversions) are related to seasonal fluctuations. Without a priori hypotheses concerning the form of the temporal patterns that might be observed, an analysis strategy was adopted that provided reasonably powerful tests for distinguishing potentially meaningful patterns from random fluctuations. Two types of analyses were performed. First, for each question analyzed, a test of linear trend across quarters was performed. These tests could detect generally increasing or decreasing trends. However, these trends could be obscured by seasonal fluctuations. So a second analysis was performed on the data from each question which took into account possible seasonal patterns while testing for temporal patterns across years. These tests can reveal both linear and curvilinear patterns across years as well as seasonal patterns in the data.

6.5 Types of Aircraft and Operations

In addition to the temporal analyses described above, two other sets of analyses were conducted designed to determine whether observed patterns were consistent across the sample or were limited to particular types of aircraft or operations.

Many factors can affect the likelihood of experiencing a potentially hazardous event. These include whether the aircraft is engaged in cargo or passenger transport, whether the flight is conducted largely at low altitudes “in the weather” or at higher altitudes, whether small airports or large airports are used, etc. The NAOMS survey did not include questions soliciting information on all of these factors. However, aircraft type is identified in the survey and it can serve as rough proxy for many of them. Aircraft tend to be used in specific types of operations because they are designed with particular missions in mind. For example, turboprops are used to connect small airports to large airports and fly at lower altitudes. Wide-body aircraft are used for long-haul flights connecting large airports and fly at high altitudes.

There were too many makes and models of aircraft and too few observations of each to use aircraft make/model directly in the analyses. Therefore the aircraft were grouped into categories by standard weight classes based on GTOW:

Small Transports: GTOW greater than or equal to 5,000 lbs. but less than 100,000 lbs.

Medium-Size Transports: GTOW greater than or equal to 100,000 lbs. but less than 200,000 lbs.

Large Transports: GTOW greater than or equal to 200,000 lbs. with a single aisle.

Wide-Body Transports: GTOW greater than or equal to 200,000 lbs. with two aisles.

In the NAOMS survey questionnaire, respondents were asked to estimate the percentage of time that they spent flying each aircraft they flew within the recall period. Based on these data, each interviewee was assigned to one of the four aircraft-size categories above determined by the category in which he/she flew the most. In a few cases, pilots reported flying exactly the same amount in aircraft that were in two different categories. On these occasions, the pilot's report was assigned to one of the tied categories at random. The pilots were also asked to report the relative amount of time they spent flying passengers or only cargo during the recall period. Pilots were categorized as flying passenger or cargo operations depending on whether they reported flying passenger or non-passenger cargo aircraft during the majority of the time flown in the reporting period.

To determine the impact of aircraft category and whether this affected the observed temporal patterns, predictors were added representing aircraft category to the statistical model used to test for year and seasonal effects. Then predictors were added representing the interactions between aircraft categories and the year and seasonal effects. Interaction terms measure the extent to which the effect of one factor (e.g. year or season) depends on another factor (e.g. aircraft category). The results were then tested both for aircraft category and for whether the interaction terms as a group improved the ability to predict the event rates. To determine whether aircraft category affected the observed temporal patterns, the interaction of aircraft category with year, aircraft category with season, and aircraft category with season by year as a group were taken into account and tested to determine whether it improved the ability to predict the event rates (identified as interactions with year/season in the results below). In most cases, the interaction terms did not improve the ability to predict the event rates—indicating that the observed temporal patterns did not differ significantly across aircraft categories. For those questions for which there was a significant interaction, the nature of that interaction was explored in more detail. So for some questions, although the rates of events differed by aircraft category, there was no significant interaction between a temporal pattern and aircraft category. This indicates that the temporal pattern (if any) was not significantly different across aircraft categories but that the reported rates of events for some types of aircraft were significantly higher or lower than the rates reported for other types of aircraft. Corresponding analyses in which the type of operation—passenger or cargo—was substituted for the type of aircraft are also reported in Section 7.

6.6 Statistical Methods

For most of the events, the probability that any pilot would experience an event on any single flight is very low. Indeed for most questions, most pilots did not report an event during the previous 60 days. Given this distribution of responses, statistical analyses based on the normal distribution are inappropriate. Analyses based on a Poisson distribution are frequently used to model this sort of data. The Negative Binomial describes a family of distributions of which the Poisson may be regarded as a special case. For most NAOMS survey questions, a Poisson distribution fits the data quite well (see Figure 6).

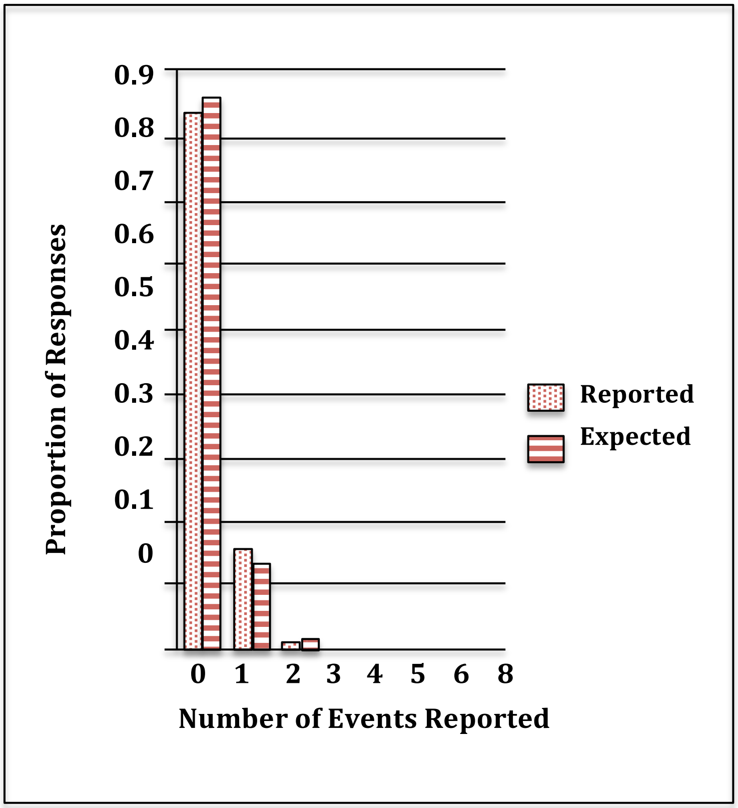


Figure 6. Comparison of actual data to Poisson estimates: Bird strike reports.

In general, when a criterion (dependent) variable is a count of instances of rare independent events and there is no natural upper limit on the counts, a Poisson distribution will provide a reasonably good fit to the data. However, in many cases data will demonstrate greater variance than would be expected if the underlying distribution were Poisson. In these cases, a different Negative Binomial distribution may provide a better fit. Therefore, maximum likelihood regression techniques based on the Negative Binomial distribution were used in analyzing these data[7]. The 95% confidence level for these events that are shown in Section 7 reflect the Negative Binomial Distribution analysis used. These techniques, both the Poisson and the Negative Binomial Distributions, generally produce results that are analogous to those obtained from the commonly used ordinary least squares techniques used to analyze normally distributed data[8].

[7] More information on Negative Binomial regression may be obtained from Hilbe, J.M. (2008). Negative Binomial Regression. Cambridge University Press, Cambridge, UK, and Cameron, A.C. & Trivedi, P.K. (1998). Regression Analysis of Count Data. Cambridge University Press, Cambridge, UK.

[8] In Poisson and Negative Binomial regression, the natural logarithm of the count of the number of events is modeled as a linear function of the predictors. Measures of exposure (such as the number of legs flown) are entered as unparameterized predictors or “offset variables” in the model. Hence, the effect of the exposure is taken into account in estimating the effects of the other predictors.

7. Results

7.1 Introduction

In the analysis of the data, priority was given to assessing temporal changes because the main objective of this experiment was to demonstrate that the methodology could reliably measure relative changes over time of events occurring in the national airspace system. Such methodology would be key to assessing the impact across the system of technologies and procedures implemented in the interests of efficiency, modernization, or to address the many challenges suggested by the concept of operations of the Next Generation Air Transportation System (NextGen). However, the data are also useful for assessing other relative differences, some of which are described below.

Many of the events that were the subject of the NAOMS survey occurred at a rate sufficient to allow for analysis over the limited period of time of the present study. However, for some questions more time would be needed to attain a sufficient response to support reliable analysis. Of the 91 questions of Section B of the air-carrier-pilot survey with numerical answers, 43 questions resulted in sufficiently high response rates to support reliable trending analysis. The remaining questions, those reflecting rarer events, would require a longer period of time for data collection before a reliable analytic threshold could be reached to assess changes over time. When interpreting results for the 43 questions presented here, appropriate caution should be exercised. These data were gathered between 2002 and 2004; the results obtained for an event may not reflect the patterns that would be observed today.

Section 6 described the statistical approach applied to each of the 43 questions analyzed. For each question, linear trends across the survey period (12 quarters) were assessed, as were year-to-year changes and seasonal effects.

Once data were acquired, it became apparent that aircraft-size category (small, medium, large, or wide body) had an important influence on event encounters as did the type of operation (passenger or cargo). The effects of these additional variables, as well as interaction effects, are reported for each question. A significant effect for aircraft-size category indicates that the rate of the event differed depending on the size of the aircraft. Similarly, a significant effect for operation indicates that the rate of the event differed depending on whether the reporting pilot primarily flew passenger- or cargo-carrying aircraft during the recall period. A significant interaction effect between year or season and aircraft category indicates that the pattern observed across years or seasons differed depending on the category of aircraft. Similarly, a significant interaction between year or season and type of operation indicates that the pattern observed across years or seasons differed depending on the operation flown.

7.2 Statistical Summary of the Data

This section presents a summary of the results for the 43 questions analyzed for each of the variables investigated (i.e. linear trends [downward, upward, no change]; year-to-year changes; seasonal changes; effect of aircraft category; and effect of operations). The results are summarized in Table 6. See Appendix M for full analyses of each of these questions analyzed.

Table 6. Results Summary

Question		*Linear Trend1*	*Trend Direction*	*Year*	*Season*	*Interaction Year*Season*	*Aircraft Category*	*Interaction Category with Year/Season*	*Operation*	*Interaction Operation with Year/Season*
AC1:	Bird strikes				***		***		***	
AC2:	Evasive action to avoid In-flight collision	*	↓	**	*		***		***	
AC3:	Loss of separation	*	↓	**	*		***			
AD1:	Deviation from assigned altitude	*	↓				***		***	
AH1:	Use of reserve fuel	***	↑	***			***	***	**	***
AH2:	Accept ATC clearance unable to comply						***		***	*
AH3:	Lost sight of aircraft	*	↓	**	**		***			
AH3A:	Lost sight of aircraft – Less than 3 miles	*	↓		***		*			
AH6:	Inadvertent deviation from vector									
AH9:	Hard landing			**	*		**		***	
AH12:	Takeoff improper configuration						***			
AH13:	Unusual attitude					**	**			
AH14:	Stick shaker/stall warning	**	↓	*			***	*	***	
AT1:	Unable to contact ATC				**		*	*		
AT1A:	Unable to contact ATC on ground						***	***	***	***
AT1B:	Unable to contact ATC – terminal area						**	*		
AT1C:	Unable to contact ATC – en route			*	***	**	***	***	*	
AT2:	High/fast approach due to ATC				**		***		*	
CP1:	Expedite/divert medical emergency	**	↑	*			***		INS	INS
CP3:	Left cockpit to handle disturbance						***		INS	INS
ER1:	Divert due to equipment problem						***		***	
ER3:	Cargo shift						***	**	***	**
ER5D:	Smoke, fire, fumes in the galley	*	↓				2		INS	INS
ER5E:	Smoke, fire, fumes in passenger compartment	**	↓	*				*	INS	INS

continued on next page

Table 6. Results Summary (continued)

Question	*Linear Trend1*	*Trend Direction*	*Year*	*Season*	*Interaction Year*Season*	*Aircraft Category*	*Interaction Category with Year/Season*	*Operation*	*Interaction Operation with Year/Season*
ER5F: Smoke, fire, fumes – Other	**	↓	*						
GE2: Collision/near collision with ground vehicle	*	↓	**	**					
GE3: Hydroplane increased stopping distance				***		*		***	
GE4: Rejected takeoff						***		***	
GE8: Begin takeoff – another aircraft on runway	***	↓	***			***		**	
GE9: Landing – another aircraft on Runway	*	↓	**	***					
TU1: Encounter severe turbulence				***	*	***	*	***	
TU1A: Encounter severe turbulence – IMC						***		***	
TU1B Encounter severe turbulence – clear air				***		***	**	***	**
TU2 Encounter wake turbulence				**		***	*	***	
WE1 Lacked Wx info while airborne				***	**	***	***	***	***
WE1A Lacked Wx info while airborne – non U.S.			***	**	**	***	***	***	**
WE1B: Lacked Wx Info while airborne – ATIS	*	↓	**		**	*	**	***	**
WE2 Fail to receive ATC app. to avoid Wx				***	**	***	***		
WE3 Diversion to alternate due to Wx						***	**	***	***
WE4 Icing encounter causing anomalies	*	↓	*			***		**	
WE5 Encountered wind shear/ microburst				***	***	***		***	*
WE6: Wind shear/microburst avoidance man.				*		***		**	

****=p<.001; **=p<.01; *=p<.05; INS=insufficient data*

1. *In this summary analysis, for a few of the events the results presented exclude the effect of an anomalous (in most cases the first) quarter. See Appendix M for details.*
2. *Significant effect due to the inclusion of small aircraft, few of which have galleys.*

In the following sections, results for each of the variables analyzed are summarized, followed by example questions demonstrating significant findings for each of these variables. The first example presented in Section 7.1.1.1 (GE8: Began Takeoff While Another Aircraft was on the Runway) illustrates a case of a downward linear trend over time. For this one example, the results of the full analyses for this question, not only of the downward trend, are shown. The complete presentation of results for this question is provided to demonstrate the analyses performed for each of the 43 questions. Similar full analyses for all questions analyzed are presented in Appendix M[9]. For upward linear trend, no trend change, year-to-year change, seasonal change, aircraft category effect, and operations effect only a single analysis is shown that illustrates the particular pattern being illustrated[10,11] .

7.3 Changes over Time

To determine if newly implemented technologies, procedures, training protocols, etc., have unanticipated impact on the national airspace system, it is necessary to assess if exposure to safety-related events remains stable or changes over time. Developing a methodology for assessing such changes was the primary objective of the NAOMS study.

7.3.1 Trending Data

Of the 43 questions in the dataset used for analysis, 16 questions resulted in significant linear changes over time. For 14 of these 16 questions the reported rates decreased over time (indicating less exposure to a potential safety-related event over time). For two questions the reported rates increased over time (greater exposure to a potential safety-related event over time). The remaining 27 questions showed no significant linear changes over the period of the study. Three examples from the NAOMS survey discussed below illustrate the three linear effects: a positive change (reduced event rate), a negative change (increased event rate), and no significant change over time.

7.3.1.1 Reduced Rates over Time

An example of the 14 questions that showed reductions of event rate over time is question GE8: Began Takeoff while Another Aircraft was on the Runway.

The rate of instances in which pilots reported beginning takeoffs while another aircraft was on the runway decreased linearly across the observation period (see Figure 7). This trend remained statistically significant when the data from the first quarter of 2002 is omitted from the analysis. Across the study period, a mean rate of 0.33 events per 1,000 legs was reported. The lowest rate

---

[9] While results of significance testing are presented for the questions that follow, the interested reader should refer to Appendix M for a full description of statistical findings.

[10] As stated earlier, each survey question is best viewed individually; i.e., as a separate study. Although results are given here as examples of significant findings, the reader should refer to Appendix M for the full analysis of these events/questions as well as for all other questions analyzed.

[11] The number of reports gathered in the first quarter (and to a lesser extent, the last quarter) was fewer than in other quarters. When a particular question resulted in anomalous-quarter data that could affect the interpretation of findings, results are reported (here and in Appendix M) both with and without the anomalous-quarter data in the analysis.

was observed in the first quarter of 2004 (0.20) and the highest rate was observed in the first quarter of 2002 (0.93).

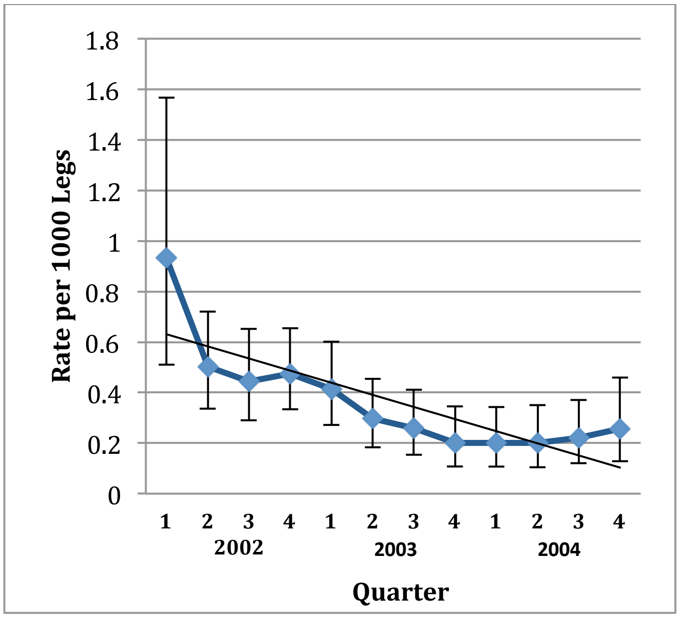


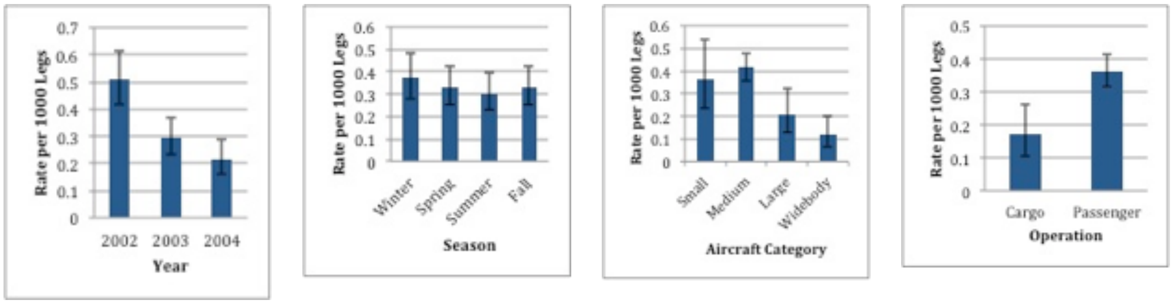
Figure 7. GE8: Began takeoff while another aircraft was on runway; By quarter.

Significant differences between years were observed for the event GE8: Began Takeoff while Another Aircraft was on the Runway (see Figure 8A). The reported rates of these events decreased linearly across the years in the sample.

No differences between seasons were observed (see Figure 8B). No year by season interaction was observed.

There were significant differences in the reported rates of this event (GE8: Began Takeoff while Another Aircraft was on the Runway) among aircraft-size categories (see Figure 8C). Pilots of medium aircraft reported the highest rates, significantly higher than those reported by pilots of large or wide body aircraft. The difference in reported rates between small and wide body aircraft was also statistically significant. No other differences between aircraft categories were significant. No interactions between aircraft category and year/season were observed for this event.

Pilots of cargo aircraft reported significantly lower rates of beginning takeoffs with another aircraft on the runway than did pilots of passenger aircraft (see Figure 8D). There were no significant interactions between operation and years/seasons.



A. By Year B. By Season C. By Aircraft Category D. By Operation

Figure 8. Downward linear trend example: Began takeoff while another aircraft was on runway.

7.3.1.2 Increased Rates over Time

An example of the two questions that showed increasing rates of event over time is question AH1: Use of Reserved Fuel.

Across the study period, a mean rate of 4.13 events in which reserve fuel was used per 1,000 legs was reported. The lowest rate was observed in the fourth quarter of 2002 (2.94) and the highest rate was observed in the fourth quarter of 2004 (5.62). The reported rate of use of reserve fuel increased linearly across the observation period (see Figure 9).

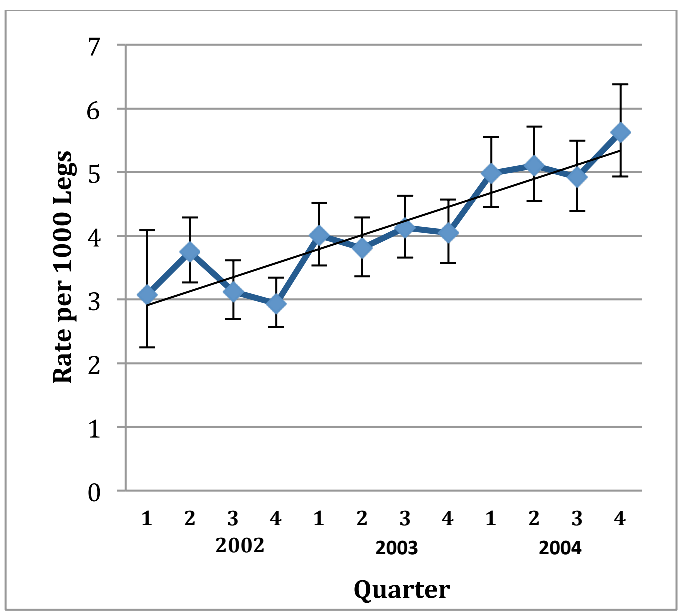


Figure 9. AH1: Use of reserve fuel; By quarter.

7.3.1.3 No Significant Changes over Time

An example of the 27 questions that showed no significant differences in event rates over time is Question AH6: Inadvertently Deviated from Assigned Vector.

Across the study period, a mean rate of 1.60 events of inadvertent deviations from assigned vectors per 1,000 legs was reported. The lowest rate was observed in the first quarter of 2003 (1.27) and the highest rate was observed in the fourth quarter of 2002 (1.83). The rate of reported inadvertent deviations from assigned vectors did not change across the observation period (see Figure 10).

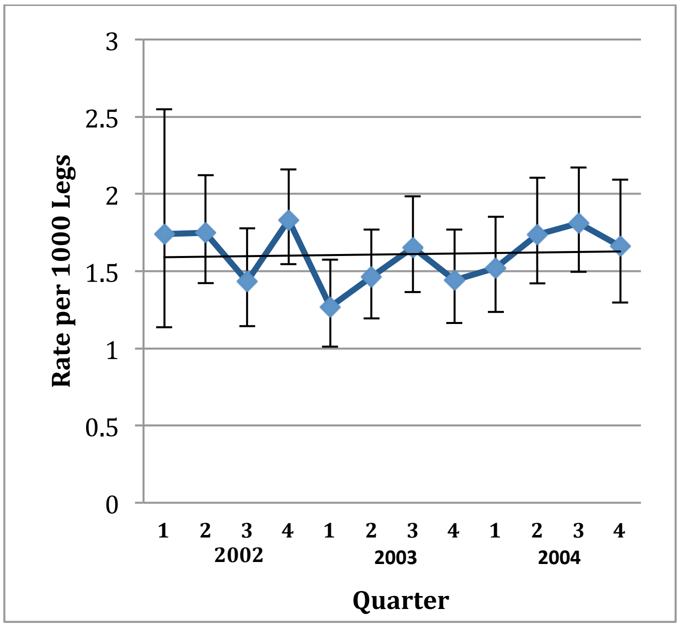


Figure 10. AH6: Inadvertently deviated from assigned vector; By quarter.

7.3.2 Year-to-Year Changes: Non-Linear Effects

As shown in Table 7, seventeen questions revealed significant year-to-year changes. For most questions the year-to-year changes mirrored the linear trends. However, changes across time could also follow a non-linear pattern. For example, a successful change in air traffic control procedures could result in an immediate drop in the rate of events followed by a steady lower rate. In this case, the overall pattern might be approximated by a decreasing linear trend if the decrease was large enough and if the period-to-period variation was small. However, the actual temporal pattern would be non-linear and would be best tested using an analysis strategy that could detect non-linear patterns. To model non-linear temporal patterns, more complex curves could be fit to the data (e.g., a quadratic trend superimposed on a linear trend). Alternatively, one could re-conceptualize the data as yearly shifts potentially superimposed on seasonal cycles. We took the latter tack in the analyses reported here. For four questions, year-to-year changes were found where there was no significant linear trend. For three questions, in which there were significant linear trends, there were also significant year-to-year non-linear effects. Question AH9: Experience Hard Landing is an example of an event demonstrating significant year-to-year changes but without a significant linear trend.

Significant differences between years were observed (see Figure 11). The rate of hard landings in 2004 was significantly lower than in the previous years. The mean rate of hard landings in 2003 was higher than in 2002, but this difference was only marginally statistically significant. However, these differences are a product of a more complex pattern. Only the first quarter of 2002 demonstrates a substantially lower rate than the corresponding quarter in 2003. In addition, after a drop in the rate of hard landings in the first quarter of 2004 the rate climbs steadily back towards the mean for 2003.

To examine the year-to-year changes for each question, see Appendix M.

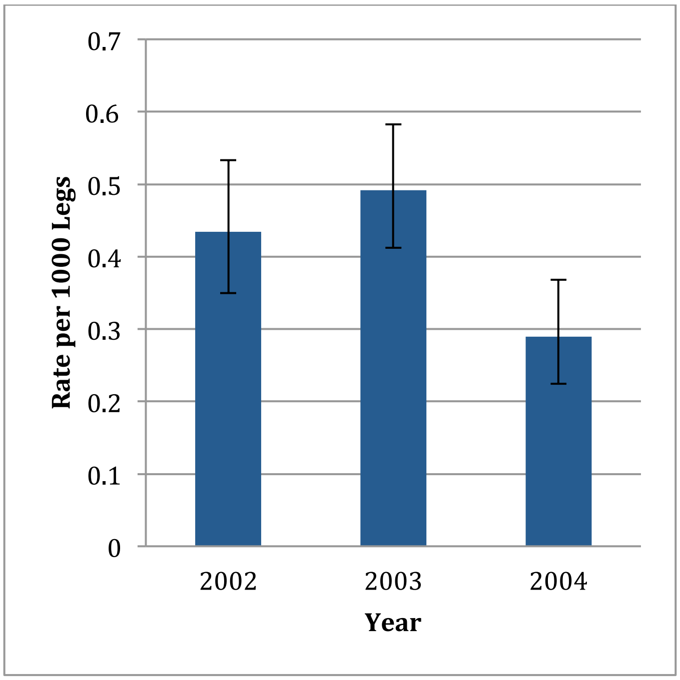


Figure 11. AH9: Experience a hard landing; By year.

7.3.3 Seasonal Effects

Regular patterns across quarters in the rate of reported events could reflect the action of factors related to seasonal weather (e.g., icing, thunderstorms) or to other factors (e.g., tourist travel) that vary regularly across the year. If seasonal patterns are not taken into account, the variation due to seasons could obscure year-to-year differences in the reported rates of events. In the analyses reported here, the effects of year and season were estimated simultaneously so the effects attributed to either factor take into account the effects of the other.

For 21 of the 43 questions analyzed, statistically significant seasonal effects were observed (see Table 6). For many of these events, seasonal effects appear to be directly relevant to identifiable weather conditions. In other cases, the seasonal effects may be unrelated or only indirectly related to weather conditions. One particularly interesting example of seasonal effects relates to the cyclical pattern demonstrated in the data from Question AC1: Experience Bird Strikes.

There is a significant seasonal pattern and no year by season interaction (see Figure 12). The rate of reported bird strikes is highest in the summer and fall (which do not differ), lower in the spring than in the summer, and still lower in the winter than in the spring.

To examine the findings for various seasons for each question, see Appendix M.

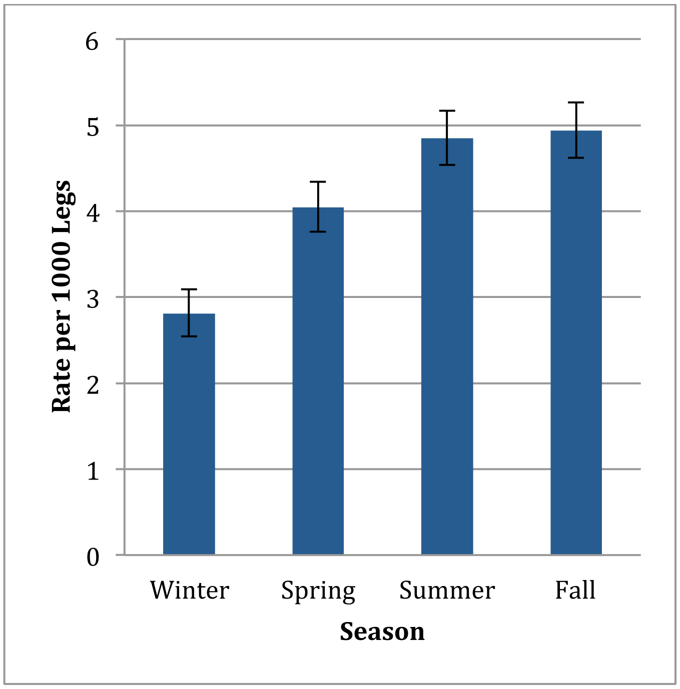


Figure 12. AC1: Experienced bird strike; By season.

7.4 Aircraft-Size Categories

Although the primary purpose of NAOMS was to demonstrate a methodology for assessing changes over time the NAOMS-survey method also proved capable of revealing other important relations in the data beyond changes over time. In the analyses reported here, aircraft were grouped into four size categories. The effects of these categories on rates of reported events were then examined. Differences in the reported rate of an event between aircraft of different categories may reflect factors directly related to aircraft size. For example, the probability of reports of an action taken in response to a medical event is greater on larger aircraft because more passengers are carried on larger aircraft. However, an aircraft's mission often dictates its size. Therefore, differences between aircraft of different sizes in the reported rates of events may reflect differences in the operations—such as the airports used or routes flown—in which these aircraft are engaged.

In the NAOMS-survey data, significant differences among aircraft-size categories were found for 36 of the 43 questions analyzed (see Table 7). For many events, it is the small and/or the wide body aircraft pilots who experience the highest event rates.

The following event showed no trending or year-to-year differences but did exhibit clear differences in the rates of events experienced by the different categories of aircraft. An example of an event for which significantly different rates of occurrences were reported as a function of aircraft-size category was question ER1: Divert Due to Equipment Problem.

There were significant differences between aircraft categories in the rates of these events (see Figure 13). Pilots of small aircraft reported the highest rates of these events, significantly higher than all other aircraft. Pilots of wide body aircraft reported the second highest rates of these events, significantly higher than medium and large aircraft, which did not differ.

There were no interactions between aircraft category and year/season.

To examine the findings for various aircraft categories for each question, see Appendix M.

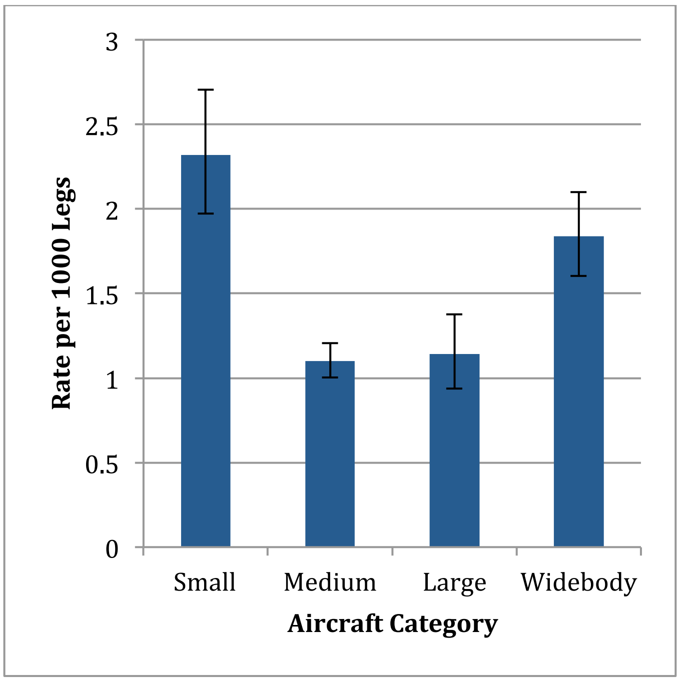


Figure 13. ER1: Divert due to equipment problem; By aircraft-size category.

7.5 Operations

In the analyses reported here, reports were also classified according to the type of operation—cargo or passenger—in which the reporting pilot was engaged. There are a number of reasons why passenger-carrying pilots and cargo-carrying pilots may experience different events. For example, cargo aircraft often operate from different airports at different times of day with fewer other aircraft in the air space than do passenger aircraft. If the rate of an event were affected by characteristics of the airport or the time of day then this could result in a difference in the rate of occurrence of the event between types of operations.

Of the 43 questions analyzed, four events were passenger-oriented and resulted in very few, if any, reports of those events from cargo pilots. About 70% of the 39 remaining questions showed differences between cargo and passenger aircraft. For 20 events, the rates were higher in cargo operations than in passenger operations. For 6 events, the rates were higher in passenger operations than in cargo operations. For 13 events, there were no differences between operation types. A summary of the impact of operations on rates of reported events is given in Table 7.

As with the aircraft size categories discussed above, the NAOMS-survey methodology revealed relative differences between passenger flights and cargo flights for events that were stable over time as well as for those that changed over time. The following is an example of an event, GE4: Rejected Takeoffs that showed significantly greater rates for cargo than for passenger flights while remaining stable over time.

Pilots of cargo aircraft reported a higher rate of rejected takeoffs than did pilots of passenger aircraft (see Figure 14).

There were no significant interactions between operation and year/season.

To examine the findings for cargo and for passenger operations for each question, see Appendix M.

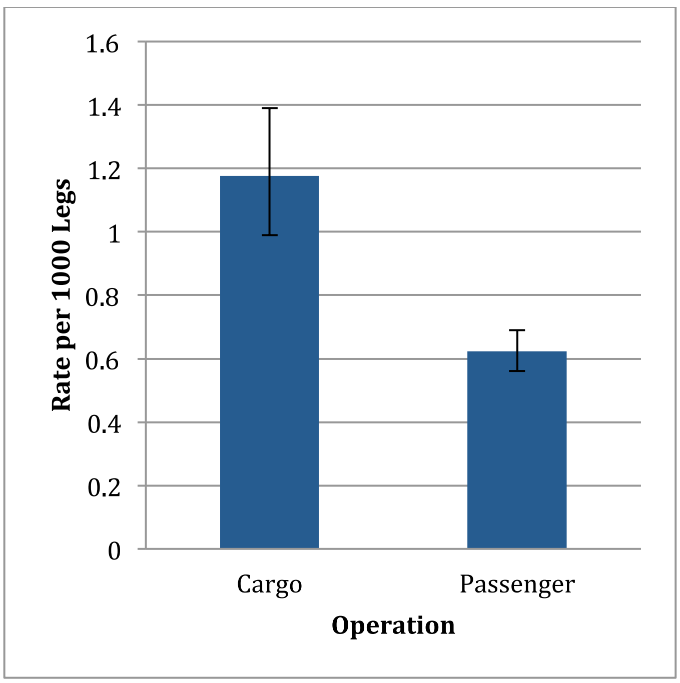


Figure 14. GE4: Rejected takeoffs; By operation.

7.6 Interaction Effects: Aircraft Categories, Operations, and Changes Over Time

In some cases, the factors that impact a temporal pattern affect all aircraft in the system similarly (e.g., global economic shifts). However, in other cases the factors that affect a temporal pattern affect particular segments of the system differently. For example, changes in icing forecast capabilities would be expected to influence predominately aircraft engaged largely in low to moderate altitude operations. In these cases, the temporal pattern will be different for some segments of the system. In the analyses reported here, this condition would result in a statistically significant interaction between variables representing the temporal pattern (year, season) and variables representing the aircraft categories and/or types of operations. Statistically significant interactions between either aircraft category or operation or both and year/season were observed for 18 of the 43 questions analyzed (see Table 6). Some of these interactions reflect small but statistically reliable differences between aircraft categories or operations in the observed temporal patterns. In other cases, these interactions reveal substantially different temporal patterns in the different groups. These patterns are discussed in more detail in Appendix M for each event analyzed.

7.7 Effects of Non-Random Bias

In any survey, there are always possibilities of biases that can affect the validity of generalizing the survey sample to a fuller population. When these biases are random, there are statistical techniques to correct for them. However, corrections for non-random biases require complex analyses entailing both statistical and operational knowledge.

7.7.1 The Effect of the Selection Criteria

Many pilots obtain commercial pilot certificates but do not work as commercial pilots. To include as many air-carrier pilots as possible in the sample pool of the NAOMS survey while excluding those not engaged in air-carrier operations, only pilots who were based in the United States and had: (1) an Airline Transport Pilot certificate, (2) a current first-class medical certificate, (3) a multi-engine rating[12,], and(4) a Flight Engineer's certificate were included in the sample pool. Many, although not all, pilots engaged in air-carrier operations have these certificates and ratings. Hence, the trends described previously in this Section are based on data from a large part of the air-carrier pilot population, but not the entire air-carrier pilot population. It is possible that the air-carrier pilots who were not in the population from which the NAOMS-survey sample was drawn could have been involved in different operations and had different experiences. This issue is of limited interest because it would never be expected to arise in any full-scale implementation of a NAOMS-like survey, which would sample from the full FAA Airman Registry. However, for completeness, the results reported previously were examined for evidence of differences from the remainder of the air-carrier pilot population that were not included in the NAOMS-survey sample as a consequence of the selection criteria.

To explore possible effects of the selection criteria used for the NAOMS survey sample pool, the data obtained from the pilots who matched the selection criteria were compared with data obtained during the same period from commercial pilots who did not match the selection criteria. For about 9 of the 45 months during which the survey of air-carrier pilots was conducted, a survey of GA pilots was conducted as well. The pool from which the GA pilots were selected was composed of pilots who had not met the selection criteria for the air-carrier pilot pool used in the NAOMS survey. However, some of interviewees contacted for the GA survey were found to have been operating primarily as air-carrier pilots during that survey period. Whenever interviewers made that determination, they shifted from the GA-pilot-survey questionnaire to the air-carrier-pilot survey and collected data based on that survey. The GA-pilot survey was conducted from August 21, 2002, to May 15, 2003. During the 9 months of the survey of GA pilots, 407 pilots had been identified as having flown as air-carrier pilots during the survey

[12] The requirement for a multi-engine rating was not a factor in this study because the experiment considered only aircraft of greater than 5,000 pounds GTOW and there were no single-engine aircraft that fit that condition that had been identified in the data set used for analyses.

period. After scrubbing the data, there were 402 of these interviews with viable responses. As these 402 pilots had not been required to meet the selection criteria for eligibility in the NAOMS-survey sample pool, they were not included in the analyses of the survey of air-carrier pilots reported previously. The results from the 402 pilots who did not match the NAOMS-survey selection criteria for the air-carrier-pilot survey were compared with the results from 5,944 interviews of pilots that matched the NAOMS-survey criteria collected during the four quarters that spanned the 9 months of the GA-pilot survey.

The commercial pilots matching the NAOMS-survey selection criteria (referred to as 'Match' in the following tables and discussion) differed from the commercial pilots in the public FAA airmen registry who did not meet these criteria (referred to henceforth as 'No Match') on a number of measures included in Section A of the survey instrument (see Table 7). Compared to the commercial pilots in the Match group who met the selection criteria, the commercial pilots in the No Match group who did not meet the selection criteria were less experienced ($t(6344)=9.38$, $p<.001$), flew more legs during the survey period ($t(419.62)=10.25$, $p<.001$), flew more hours during the survey period ($t(446.90)=5.42$, $p<.001$), and flew shorter legs ($t(495.54)=11.11$, $p<.001$). In some cases the mean differences between these groups were quite large. However, for all of these measures there was substantial variation within the groups. On all of these measures, the standard deviations were larger than the differences between the groups—indicating that on all measures there was considerable overlap between the groups.

Table 7. Mean Experience of Pilots by Match to Selection Criteria

Measure	*Selection Criteria*	
	Match	*No Match*
Hours – 60 days	97.57 (47.54)	112.12 (52.34)
Legs – 60 days	38.26 (34.62)	68.97 (59.41)
Hours/leg	3.75 (2.75)	2.50 (2.14)
Total hours	10,349.72 (6,822.43)	7,068.58 (6,271.77)

Note: Standard deviations are in parentheses.

The distribution of aircraft by size categories flown was different between the pilot samples ($X^2(3)=798.4$, $p<.001$). Nearly half (49.8%) of the pilots in the No Match group flew small aircraft compared to 7.0% of the pilots in the Match group. The proportion of pilots who did not match the criteria who flew large (6.0%) or wide-body (9.0%) aircraft was much smaller than the respective proportions of pilots who did match the criteria (large: 15.0%; wide-body: 30.3%). The pilots in the Match group were also more likely to fly cargo operations (22.0%) than were the pilots in the No Match group (6.0%; $X^2(1)=58.2$, $p<.001$). The groups also differed in the positions flown ($X^2(2)=27.55$, $p<.001$). With a completely unbiased sampling of the full air-carrier-pilot population, one would expect an equal proportion (i.e., 50/50) of Captains and First Officers in both groups. However, results revealed that a larger proportion of pilots who matched

NAOMS-survey criteria and a smaller proportion of pilots who did not match the criteria flew as Captains (Match: 54.0%; No Match: 47.5%). These comparisons suggest that the NAOMS-survey selection criteria resulted in sampling more experienced pilots who were Captains, flew larger aircraft, and flew longer hours per leg. In contrast, the pool of air-carrier pilots that were identified during the GA-pilot survey selection process were those who had not been included in the NAOMS-survey pool and tended to be the less experienced pilots who flew the smaller aircraft for fewer hours per leg.

In summary, the two samples of air-carrier pilots differ in a variety of ways. Some of these differences affect the rates of reported events. These groups differ in the types of aircraft used and the type of operations flown. As noted, these factors affect the rates of some reported events. In some cases, aircraft category or operation affects the overall rate of an event. In other cases, these factors may affect the temporal pattern such that over time the rates for one aircraft category or operation may change differently from those for another category or operation. This added complexity does not affect the validity of the reported analyses. It simply reflects the reality of a multifaceted airspace system. Whenever aircraft category or operation affects the rates of a reported event, these factors must be taken into account when examining temporal patterns, as was done in the analyses reported in Appendix M .

Negative binomial regression analyses were conducted to determine whether there were any differences between the two groups in the temporal patterns of rates of events that cannot be explained by differences in aircraft or type of operation flown. The trend analyses performed previously could not be duplicated because data for the two groups were available for only 4 quarters (Summer 2002, Fall 2002, Winter 2003, and Spring 2003). It is not possible to distinguish between year and season effects with only these data. Only differences between the groups across these 4 quarters could be analyzed.

Significant interactions between the selection criteria and quarter were observed on five questions. For these five events, there were differences between the two pilot groups in the temporal patterns that cannot be explained by the effects of aircraft category and type of operation (see Table 8)[13]. For one of these five events, the time trends for the pilot groups differed but no significant general linear trend was observed for either group. For the other four questions, linear trends were observed.

For AT1A (Unable To Communicate With ATC While On The Ground), both the pilots in the Match group and those in the No Match reported generally increasing rates over the observed period. However, the increasing trend was significant only for pilots who matched the criteria. For AT1C (Unable To Communicate With ATC While En Route), pilots who matched the selection criteria demonstrated generally increasing rates of these events while pilots who did not match the criteria reported generally decreasing rates. For AT2 (Fly An Undesirably High/Fast Approach Due To An ATC Clearance), both the pilots in the Match group and those in the No Match group reported generally decreasing rates of these events over the observed period, but the rate of decrease was greater for the pilots that did not match the criteria. For WE5 (Encountered Wind Shear Or Microburst), both the pilots in the Match group and those in the No Match group

[13] For seven questions there were statistically significant differences in the overall event rates but not in the temporal patterns between the pilot groups that cannot be explained by the effects of the category of aircraft or types of operations flown.

reported generally increasing rates over the observed period, but the rate of increase was greater for the pilots who did not match the selection criteria.

Table 8. Time Trends Across Four Quarters by Selection Criteria

Item	*Selection Criteria*	
	Match	*No Match*
AT1A: Unable to contact ATC on ground	↑***	↑
AT1C: Unable to contact ATC en route	↑*	↓***
AT2: High/fast approach due to ATC	↓*	↓**
WE5: Encountered wind shear/microburst	↑**	↑**

*Note: Arrow indicates direction of trend. Statistical significance of the trend is indicated by the symbols: * p<.o5 ** p<.01 *** p<.001*

In summary, for the great majority of events, no effect of selection criteria on the temporal patterns of the rates of reported events was observed. For five of the 43 events examined, a significant interaction between selection criteria and quarter of observation was observed, with four of these events showing a significant linear pattern. However, for only one of these five events was the observed trend across quarters significant and different in direction between the sample of pilots that matched the selection criteria and the sample of pilots that did not match the criteria. For the other three events, the effect of the selection criteria was merely to alter the observed degree of increase or decrease.

The comparisons presented above must be viewed cautiously given the small sample of 402 pilots that did not match the selection criteria. However, little evidence was found that the specificity of the NAOMS survey-selection criteria substantially influenced the temporal patterns reported earlier in this section.

7.7.2 The Effect of the Opt-Out Option

In 2000, between the time of the field trial and the initiation of the full survey of air-carrier pilots, Congress instructed the FAA to permit pilots to have their names removed from the public registry, which the NAOMS survey was using as its sampling pool. Shortly thereafter, major pilot unions provided pilots with a convenient way to use this option. It is at least possible that the pilots who chose to opt out of the public registry differ from the pilots who did not opt out of the public registry. If there are differences, then the temporal patterns in rates of events described previously might not reflect those found throughout the NAS; they may be peculiar to those pilots who chose not to opt out of the public registry. It is unlikely that a pilot's exposure to the sorts of events addressed in NAOMS-survey questionnaire is related to the decision to opt out of the public registry, but it is possible. Hence, in this section, we compare the rates of events reported by pilots who opted out of the public registry with the rates of events reported by the pilots who did not opt out of the public registry. As in the case of the effect of the selection criteria, this issue is of limited interest because it would never be expected to arise in any full-scale implementation of a NAOMS-like system that would rely upon random sampling from the full FAA Airmen Registry.

For the decision to opt-out to substantially influence the previously reported temporal patterns it would need to be: (1) highly correlated with the criterion (event rate) measures; (2) highly correlated with the temporal pattern; and (3) not highly correlated with the other predictors (aircraft category, type of operation) included in the analysis. It would seem unlikely for all three of these relations to hold simultaneously. The authors didn't have a way to examine the latter two conditions. Nevertheless, they examined the available data for evidence of significant differences between the two groups of pilots.

Data from the field trial (see Section 4.6.1.3), which was conducted before pilots were allowed to opt-out of the public Airmen Registry, provided some bases of comparison with data from the full survey. The selection criteria for pilots used in the sample pools for both the field trial and the full survey were identical. These data cannot be used to assess whether there were differences in the temporal patterns of the events experienced between groups because the field trial occurred over a different period than the final NAOMS survey. However, a comparison of demographic information from Section A of the survey of the pilots during the field trial with comparable information from the pilots in the full survey could provide some evidence about how the pilots who chose to opt-out differ from those who did not opt-out. Also, responses to questions from Section B can be compared for evidence of differences in the rates of events reported by the pre-opt-out and the post-opt-out groups.

The questions used in Section A for the field trial were not all identical to those used in the final full survey. However, the question dealing directly with experience ("Approximately how many hours of total time have you flown a commercial aircraft during your career?") was identical for both groups. Of the 626 interviews from the field trial, 7 provided no information in answer to this question and the remaining 619 responses were compared with the replies to this question from the 25,555 pilots interviewed during the full NAOMS survey. The pilots who completed the NAOMS survey and thus had not opted out of the public Airmen Registry reported an average of 10,426 hours of total commercial flight time while the pilots who completed the field trial prior to having the option to opt out reported an average of 10,235 hours of total commercial flight time. The difference in reported total flight time was not statistically significant.

The responses to questions in Section B from the field trial and from the survey were also compared. The field trial was conducted about a year before the final NAOMS survey began. Therefore, differences in the event rates between the field trial and the final NAOMS survey could result from either temporal shifts or differences in the populations sampled. There were 25 first-level questions from Section B that were presented in basically the same way in both the field trial and in the final survey for which comparisons of responses could be made. After deleting interviews for lack of information or improbable recorded values for hours or legs flown, 601 valid interviews remained from the field trial for all of the recall periods used in that experiment. Given the results of the memory studies indicating the greater accuracy obtained from the shorter recall periods, the analyses were limited to the data obtained from pilots who were asked to recall information from the 30 days or 60 days prior to the interview. These data were compared to the data from the corresponding recall periods obtained from the full survey. During the field trial, there were 99 interviews over 81 days that used 30-day recall and 101 interviews over 73 days that used 60-day recall. For comparison, the first 81 days of 30-day recall (690 valid interviews) and the first 73 days of 60-day recall (602 valid interviews) were selected from the full survey. The number of valid reports varied slightly with each question of Section B because sometimes no information was provided or the recorded data were improbable.

The proportions of participants in each group (i.e., the post-opt-out NAOMS full survey and pre-opt-out field trial) reporting one or more events were compared using an exact probability test.

Significant differences between the pre- and post-opt-out groups for both recall periods were observed for three of the 25 questions examined. For three additional questions significant differences were observed only for the 60-day recall period. The results of these analyses are presented in Table 9. In every case in which there was a significant difference between the groups, the proportion of pilots reporting an event was higher in the pre op-out group than in the post opt-out group.

Table 9. Significant Differences in Event Rates between Pre- and Post-Opt-Out Groups

	Post-Opt-Out		*Pre-Opt-Out*	
	Mean Event Rate	*% of Pilots Reporting Event*	*Mean Event Rate*	*% of Pilots reporting Event*
Accept ATC clearance unable to comply	2.45	2.9 (20)	24.04	24.2 (24)
Inadvertent deviation from vector	6.26	3.9 (27)	21.63	13.3 (13)
Unable to contact ATC	45.2	30.9 (212)	123.20	60.2 (59)
Accept ATC clearance unable to comply	2.67	4.5 (27)	16.08	25.7 (27)
Inadvertent deviation from vector	3.48	6.0 (36)	22.21	18.4 (18)
Unable to contact ATC	26.67	26.2 (158)	153.16	61.2 (60)
Use of reserve fuel	4.02	8.3 (50)	11.58	16.5 (16)
Hard landing	0.56	0.7 (4)	2.52	6.0 (6)
Left cockpit to handle disturbance	1.10	2.1 (10)	2.80	9.0 (9)

Note: Mean event rates are per 1,000 legs. The number of pilots reporting one or more events is in parentheses.

All three of the questions demonstrating differences between pre- and post-opt out for both recall periods reflected responses to questions involving contact with the ATC. The questions pertaining to interactions with ATC were not posed in precisely the same way in the field trial and in the full survey; these differences, and/or the differences in the time of response, could have been responsible for the observed differences between groups.

In summary, there is little evidence to indicate that the option to opt out of the public registry created a substantial bias in the NAOMS survey data.

7.8 Discussion of Results

The results summarized above and presented in more detail in Appendix M illustrate the capabilities of the NAOMS-survey methodology. Although limited in time, this project demonstrated that statistically reliable significant linear trends could be observed. The analyses demonstrated that both temporal shifts that followed a linear pattern and temporal shifts that did not follow a linear pattern could be reliably measured and distinguished from seasonal patterns. The analyses including aircraft-size categories and type of operation categories demonstrated that non-temporal information can be gleaned from the NAOMS survey and that it can be used to determine where in the airspace system particular temporal patterns occur. For this study, the 45-months of data gathering was sufficient to obtain reliable results for about half the questions posed. This limitation would be relaxed in an operational investigation in which data would be gathered continuously.

The effects of two possible limitations of these data—potential biases introduced by the sample selection criteria used for the NAOMS survey and potential biases introduced by changes in the FAA Airmen Registry—were also investigated. The effects seen were limited. If these factors did introduce biases into the collected data, they still would not affect the conclusions about the utility and viability of the methodology. The methodology is shown to work within the sample used for the demonstration of the concept.

A comparison of the reports of pilots who matched the selection criteria with those of pilots who did not match these criteria indicated that there were few consistent differences in reported event rates due to the selection criteria. This conclusion by itself does not justify generalization of the rates of events based on the NAOMS survey to the full air-carrier population. However, it does provide evidence that this bias in the sample pool used in the NAOMS survey had little effect on the reported temporal patterns.

Prior to the start of the NAOMS survey, the FAA allowed pilots to opt out of the public Airmen's Registry. Comparisons between the reports of pilots who participated in a survey prior to the change in the FAA registry policy and reports of pilots who participated in the NAOMS survey revealed some differences. Although it was not possible to fully ascertain the effect of the potential bias introduced by the change in registry policy on the observed temporal patterns in event rates, the available evidence suggests that there were few differences in the experiences of pilots who opted out of the public registry and those who did not. The differences that were observed were in the direction of possibly higher event rates in the experiences of pilots not in the registry. Despite the lack of convincing evidence of a consistent bias caused by pilots deciding to opt out of the public Airmen Registry, the full Airmen Registry should be used in any future efforts similar to NAOMS.

The results of the analyses of potential biases support the idea that the background of the pilot-reporter does not affect the likelihood of encountering an event or reporting it. Most of the questions asked in the NAOMS survey simply ask how many times an event was experienced while flying an aircraft during a specified recall period. These questions demand little if any judgment on the part of the reporter. An event is an event regardless of the reporter.

Although conclusions can be drawn about the methodology, one should not rely upon the findings on specific events to draw conclusions about today's national airspace system. In order to do so, a study would need to be conducted in a relevant time frame. The methods used in any such study should follow the recommendations mentioned above and detailed later in this report. But, with that caveat, the results of the analyses reported here could be helpful in formulating hypotheses for future study.

For example, the rate of instances in which pilots reported beginning the takeoff roll while another aircraft was on the runway decreased between 2002 and 2003. This pattern suggests that a discrete intervention may have been the cause. The lack of interaction between aircraft category and type of operation suggests that the cause of the decrease had an affect across the system. What could it have been? The rate of these events appears higher in small and medium aircraft and in passenger operations. Why? The rate of events in which pilots reported needing to utilize their reserve fuel supplies increased linearly across the observation period. This pattern suggests a slowly changing cause. This increase was observed in medium, large, and wide-body aircraft but only for passenger operations. The highest rates were observed in wide-body aircraft and cargo operations. The increase in the use of reserve fuel could reflect changes in economic conditions, weather patterns, or other factors—but why was it not observed in cargo operations? NAOMS was not designed to answer these questions but it can provide focus for future research.

In summary, NAOMS successfully developed and tested a survey methodology that can provide a reliable basis for determining temporal patterns in potentially hazardous events in the national airspace system.

8. Analyses of Special Topics

8.1 Overview

The primary objective of NAOMS was to assess changes in event rates over time for the events addressed by the questions in Section B. Another objective of NAOMS was to provide the capability for quick research on topics of special interest to the aviation community. Section C of the NAOMS survey questionnaire addressed this goal. In contrast to Section B, Section C was designed to cast a "spotlight" on a particular issue at a particular time, not to measure changes over time. During the nearly 4 years of the full survey of air-carrier pilots, two topics were studied to assess the value of including a "spotlight" look at a selected topic. The first of these was the safety implications of In-Close Approach Changes (ICAC), a subject of interest to the aviation community at the time the NAOMS development began. An ICAC is defined as an unrequested clearance amendment when the aircraft is within 10 miles of its destination airport. This subject was the topic of Section C for the most of the survey. The second topic, implementation of selected safety interventions, was addressed later in the survey in response to interest expressed by the Civil Aviation Study Team. These two studies are discussed in this section.

8.2 The Study of In-Close Approach Changes

In-Close Approach Changes, the primary topic of Section C of NAOMS, was also being investigated by other elements of the ASMM, providing a means of evaluating how NAOMS could complement other kinds of safety data. The description of ICAC used in this study was that an aircraft is given an unrequested clearance amendment when it is within 10 miles of its destination airport involving change in runway assignment, approach speed, or approach altitude.

Sixteen questions related to ICACs were developed for Section C of the air-carrier-pilot survey. A total of 600,410 approaches were reported. Pilots reported experiencing ICAC requests from ATC on about 6% (34,291) of their flights. They almost always (93.5%; 32,054 approaches) accepted these requests. On 6.5% (2,076) of these flights, pilots reported experiencing a problem. These problems might have occurred in the absence of an ICAC and other problems might have developed if the ICAC had not been used. However, a reasonable conclusion is that ICACs contribute to many of these unwanted events. The problems most often mentioned following an ICAC were unstabilized approaches and long landings (see Table 10).

Table 10. Problems Associated with ICAC

Type of ICAC Problem	*Number Reported**	*% of Itemized Problems among Accepted ICACs*
Unstabilized approach	1,205	3.76
Long or fast landing	1,128	3.52
Wake turbulence	407	1.27
Missed approach	404	1.26
Ground conflict	99	0.31
Airborne conflict	96	0.30
Out-of-limit winds	64	0.20
Landing without clearance	13	0.04
Other	914	2.85

** The number of problems reported adds up to more than the number of ICACs because a single ICAC could result in multiple events.*

The data displayed in Figure 15 demonstrate that the number of operations had a major impact on the likelihood of ICACs. However, the variability in ICAC rates at every level of traffic suggests that other factors, such as the configuration of the airport, the distribution of landing times, etc., also may affect the likelihood of an ICAC.

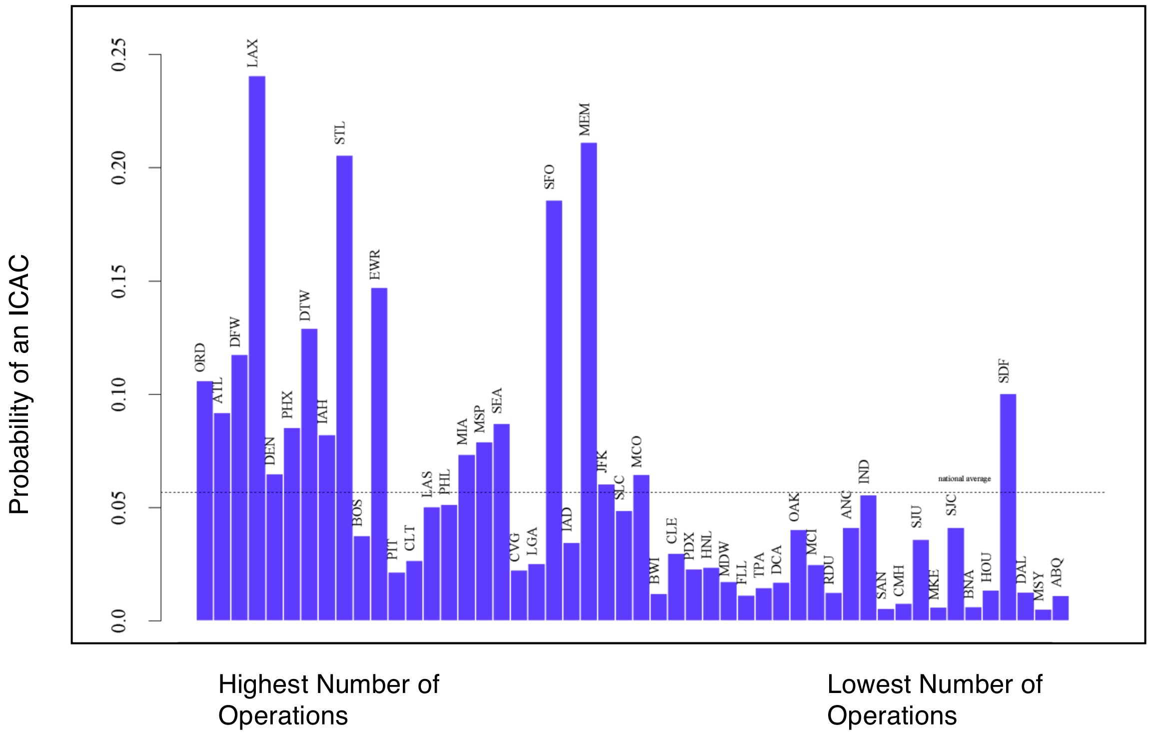


Figure 15. Probability of ICAC for 50 busiest U.S. airports.

The ICAC study demonstrates the ability of NAOMS to support the needs of the aviation community for information on targeted special topics. The ICAC study also provided an opportunity to assess how NAOMS-survey data could be used in conjunction with other safety-related data to present a fuller picture of what is happening in the national airspace. A study was performed under the ASMM project to demonstrate whether the fusion of information generated by ASMM tools (APMS, PDARS, and NAOMS) and other sources (ASRS and APMS) could provide more insight into safety events than each could do alone. The contribution of NAOMS to this larger effort was an understanding of the relative frequency of events and the context in which they occurred. Although limited by time and data availability, the ASMM study, involving multiple sources, provided valuable information on how the various sources might best be used. A summary of the results of this study is found in Statler, I.C., Morrison, R., and Rosenthal, L.J. (2003).

8.3 The Study with the CAST/JIMDAT

As the AvSSP was coming to its close in 2004, there was a strong emphasis on transitioning all of the technologies that had been developed under that program to the aviation industry. By that time, NAOMS was the sole remaining major development under the ASMM project without an identified permanent home. During the two years following the completion of the survey of air-carrier pilots, the NAOMS team focused on transferring the capability to an organization that could continue the service after NASA ended the project. Consequently, efforts turned to that objective by pursuing contacts and briefings with several organizations during 2004. Among these many meetings were discussions with the JIMDAT, a subgroup of an international industry group, the Civil Aviation Safety Team (CAST). For several years, the CAST had been examining accidents in depth, identifying common problem areas, and proposing interventions to mitigate those problems. This activity resulted in many recommended Safety Enhancements (SEs). The CAST organized the JIMDAT to monitor and measure the effectiveness of these SEs as they became implemented by the industry. Meetings with and briefings to the JIMDAT and the CAST by the NAOMS project team produced an interest in the possibility of using the NAOMS survey for their purpose. Consequently, NASA undertook a special study in collaboration with the JIMDAT to evaluate the use of the NAOMS survey to assess the efficacies and the extent of implementation of SEs.

A working group was formed composed of members of the JIMDAT and the NAOMS team. This group decided that they should focus their experiment on the SEs that were concerned specifically with training and procedures and that these could be appropriately addressed in Section C of the pilot-survey questionnaire, the section that addresses special topics. The CAST and the JIMDAT specified the topics they wanted addressed by the questions in the new Section C of the survey instrument. These topics were:

- Basic flying activity (during the recall period-60 days)
- Ground proximity warning systems
- Approach related issues
- SOP related issues
- Recurrent training
- Safety reporting
- Corporate safety

The joint working group designed a new Section C that incorporated a total of 34 first level questions on these topics. The full air-carrier-pilot questionnaire, including the specific questions used in the survey for the JIMDAT, is provided in Appendix D. To meet the needs of the JIMDAT, this section included some judgments on the part of the pilots in addition to reports of their experiences.

This new section became part of the air-carrier pilot questionnaire and replaced the Section C concerned with ICAC that had previously been part of the main survey. Selection of respondents continued as before from the pilot pool of the NAOMS survey described in Section 4.

Telephone interviews were completed with 1,148 additional air-carrier pilots over a three month period from October 1, 2004, to December 17, 2004, to collect the data for this special study. The responses to the questions in Section A were used for normalization of the responses to the questions in Section C. The responses of this group of pilots to the questions in Section B (which remained unchanged) were included in the data set used for the analyses reported in Section 7.

A complete set of the results for all of the questions in Section C used for the JIMDAT study is provided in Appendix N.

The pilots were cooperative and responsive to the set of questions. The results indicated that pilots were aware and supportive of their companies' procedures and policies and they presented a favorable view of their corporate culture and practices. For example, in response to questions related to their airline's SOPs, 97% reported that their airline's written SOPs describe how to perform recovery from unusual attitudes and departure from controlled flight; 96% said the SOPs describe how to avoid approach and landing accidents; and 99% said they addressed how to fly non-precision approaches. On the topic of Corporate Safety, 93% said their airline had a Director or a VP of Safety and 85% said they had observed a strong commitment to safety among senior management. 94% said that, if they had a safety concern, there was a mechanism available to them for bringing that concern to the attention of senior management and 89% felt it was effective.

Their responses regarding the training they received also were generally positive. For example, 97% said their most recent recurrent training had addressed the criteria for initiating and the procedures for executing a go-around and missed approach and 95% rated the quality of their most recent C-FIT prevention training as Excellent or Good.

The pilots interviewed were receptive to the safety reporting programs. For example, 86% reported that their airline participated in the Aviation Safety Action Program; 71% favored the establishment of an ASAP; 65% felt that there had been positive changes resulting from such pilot-reporting programs; and 98% said that, if the situation arose in the future, they would submit an ASAP report.

The survey identified several specific areas where there was room for improvement. For example, while 100% said they had received CRM training, only 62% felt that this training changed how they managed the flight deck. Also, 73% said their airlines had a Flight Operations Quality Assurance Program, but for pilots whose airlines did not have such a program only 39% said they favored its establishment.

The caution that was expressed with regard to the results of the ICAC study in the previous section pertains as well to this study. This was a limited study for the purposes of demonstrating a concept and the results should be viewed as preliminary.

The analyses of the responses to the questions of the revised Section C were briefed to the JIMDAT and then to the CAST. Both groups were very pleased with the results of this small study. The CAST/JIMDAT expressed interest in the operation of a survey of Part 121 pilots but NASA could not commit to continuing the service. The Air Line Pilots Association (ALPA) offered to perform the service on behalf of the CAST provided that NASA would develop a less costly alternative to telephone interviews. The hand-off of the NAOMS air-carrier-pilot survey from NASA to ALPA is discussed in the following Section.

9. The Hand-Off Phase

9.1 Background

Two separate but related issues influenced the decision to stop collecting air-carrier-pilot data in December 2004. First, the limited term of the NASA Aviation Safety and Security Program (AvSSP), of which NAOMS was a part, expired in FY05. Second, NASA felt that sufficient data had been gathered to meet the core objective of assessing the feasibility of the methodology. As a research and development organization, NASA's goal was to develop NAOMS, demonstrate its merits and limitations, identify its flaws and recommended improvements, and provide others with the knowledge needed to develop, expand, and operate such a survey over the long-term, as had been planned. Consequently, as the AvSSP was coming to an end, NASA management made the decision to focus the NAOMS team on transferring the capability to the aviation industry.

After the collaborative study described in Section 8.3, the CAST/JIMDAT became interested in using NAOMS and, in view of its pending termination by NASA, the ALPA offered to operate a survey based on NAOMS of Part 121 pilots on behalf of the CAST. However, ALPA stipulated that the CATI (telephone-interview) mode NASA had used in the pilot surveys would be too costly for them and requested that a less expensive way to collect data be developed. Consequently, NASA decided that the best path forward was to find a means of greatly reducing NAOMS operating costs—even if some reduction in data quality resulted. NASA turned its attention to determining if a more streamlined, yet still acceptable, process could be identified to meet the ALPA requirements and move the process forward to support aviation industry safety efforts.

9.2 A Web-Based Survey

Since the inception of the NAOMS project, Web-based surveys have become increasingly common. The NAOMS team hypothesized that the most cost-effective method of administering the survey would be Web-based data collection. The NAOMS team undertook to:

- develop a Web-based implementation of the NAOMS survey for routine operation of a commercial-pilot survey
- provide a comparison of statistical results of the Web-based survey to a NAOMS telephone survey
- transfer an operational Web-based survey to the ALPA

The NAOMS team also recommended that additional savings could be achieved by soliciting respondent participation, except for the initial contact, by e-mail rather than the U.S. Postal Service, which would eliminate much of NAOMS' paper-and-postage-driven expenses. Taken together, these actions appeared to have the potential for substantially reducing NAOMS-survey data-collection costs.

Web-based data collection is a recent variation of the self-administered questionnaire (SAQ) survey mode that the NAOMS team had considered in the formative years. There are considerations regarding the use of a paper-based questionnaire that pertain to an electronic version as well. For example, the quality of data collected via the SAQ mode is generally poorer than that of data acquired by CATI.

The NAOMS team had not examined Web-based data collection as a modal option during the formative years of NAOMS because the approach was too new. There was limited information on the validity of survey results obtained through this means. The cost advantages of the Web-based data-collection approach were clear but there were important unanswered questions about how these changes might impact survey quality. However, much has been published on this topic in recent years. The developing consensus in the literature is that Web-based data collection can produce consistent, quality results at significantly reduced costs (especially relative to CATI or face-to-face interviews) for some survey data applications. Nevertheless, the NAOMS team was not at all certain that a Web-based survey system for pilots would achieve the same response rates and level of quality achieved by the CATI survey.

Trust in the interviewing organization is critical to ensuring a good survey. Throughout the NAOMS project, the team investigated possible ways that the survey could be continued if the concept proved successful. The question of trust was one of the reasons that NASA had difficulty engaging an appropriate agency to take over NAOMS. The airline operators and the unions have shown they are reluctant to entrust their identifiable data to their regulatory agency, the FAA, although the FAA was in many ways the most logical agency to whom to transfer custodianship. The purposes and objectives of the CAST motivated the team to attempt to find an agency or organization to continue the survey after the NAOMS project ended. However, even the transfer of NAOMS' methods and procedures to the ALPA was accompanied by certain concerns. While ALPA members may trust their union with this information, pilots who are members of other unions may not. Also, it is not possible to obtain a full picture of the performance of the air-transportation system solely from the perspective of the pilots, which is the extent of ALPA's interest in NAOMS.

In addition to cost savings, Web-based data collection makes it practical for respondents to stop and ponder a response, consult logbooks, or do other types of data lookup that are impractical with a phone-based interview. However, the NAOMS team had several concerns:

- *Interface*. A NAOMS Web-survey application would need to have a pleasing interface that would present questions to the respondents with the same clarity as phone interviewers. It would need to incorporate the capability built into the CATI system to skip over irrelevant questions and drill-down into pertinent follow-on questions and it must be able to satisfy respondent requests for clarifying information. All of this has been demonstrated with Web-based survey technology, but the NAOMS survey instrument is far more complicated and lengthier than typical Web-based surveys.
- *Response Rate*. As noted earlier, the NAOMS survey achieved exceptionally high response levels during the 2001–2004 evaluation period using the CATI mode. The NAOMS team recognized that two of the key reasons for achieving this response rate were the trust that NASA brought to the data-collection effort and the Dillman design method that had been used for the survey instrument. These advantages would be lost under a cost-based Web-based system administered by another organization.

  Also, NAOMS-CPHRE professional interviewers were very effective at establishing rapport with respondents. Evidence of this is that interviewees ended fewer than 2% of the telephone interviews prematurely even though they often ran more than one-half hour. In the absence of a human interviewer, it would be relatively easy for a respondent to disengage from a Web-based survey session. This concern is supported by the research cited earlier that shows SAQs tend to receive less complete responses than CATI surveys.
- *Quality of Responses*. As noted earlier, it is possible for respondents to be more deliberate while completing a Web-based survey than during a CATI phone interview; it is also possible to skim rapidly through a Web-based survey and respond to questions with minimal thought. It is not possible to do this in a CATI setting because the interviewer paces the survey and presents questions one-by-one in the order structured by the survey instrument.
- *Potential Data Discontinuity*. A major concern is that the change from the CATI to a Web-based collection mode would affect the information provided by respondents and would, thereby, create a discontinuity between the several years of data collected via the CATI

mode and the data collected under a Web-based mode. Therefore, it might be difficult to interpret differences between the data collected by the CATI and Web-based approaches. One might not be able to differentiate between differences in the data that represent real measured changes in aviation safety and differences that are consequences of the change in data-collection mode.

Consequently, the NAOMS team undertook a study with the following objectives:

1. Determine whether it is practical to accomplish the collection of NAOMS survey data of commercial pilots through Web-surveys rather than by CATI.
2. Assess the relative merits of NAOMS Web-based vs. CATI-based data-collection approaches for commercial pilots against the following metrics:
   - Survey response rate: The percentage of eligible respondents who complete the interview.
   - Consistency of results: The differences between rates of events derived from Web-based and CATI data.
   - Subjective quality assessments: The percentage of respondents who have high or very high confidence in the accuracy of their answers.

The NAOMS team created a Web-based version of the NAOMS CATI instrument that had been used for the air-carrier-pilot survey. After trials of several commercial off-the-shelf Web survey products, the NAOMS team found that *Illume™,* vended by DatStat, Inc., offered the required programming flexibility with which to construct a Web-based NAOMS survey instrument.

The team developed an interface and ways to deal with the concerns for security associated with Web-based data collection and conducted a brief field test to uncover any deficiencies. This was followed by a small trial of Web-based interviews of 1,000 air-carrier pilots randomly selected from the same subject pool as the NAOMS-survey study. The survey invitations to the selected group were issued on NASA letterhead and were the same as had been used for the CATI survey, except that there were no follow-up mailings as there had been for the CATI survey to those who had agreed to participate but had not responded.

The trial ran for two months. Of the 1,000 pilots selected for the experiment, 128 completed the survey. Whereas the large majority of participants who were located and received invitations to the CATI survey responded to the NAOMS team, only a small portion of those whose participation was sought for the Web-based survey responded. Furthermore, of those who participated in the Web-based survey, about 10% began but failed to complete the survey questionnaire whereas less than 2% of the participants in the earlier CATI survey broke off before completing the interview. Consequently, the response rate realized in this small test of the Web-based survey was substantially lower that the 81% response rate for the CATI survey mode. This low response rate was disappointing but not wholly unexpected for the reasons described previously. It is possible that ALPA, with its access to better communication channels to air-carrier pilots, could achieve greater participation than was possible during the brief test conducted by the NAOMS team.

Other aspects of the Web-based survey experiment had more favorable outcomes. The invitees who agreed to participate were asked to rate their ability to efficiently navigate the survey document as Excellent, Very Good, Good, Fair, or Poor. Ninety-three percent of the respondents reported their experiences as being Excellent or Very Good. The median time to complete the

Web survey was 31 minutes. This is close to the average time required to complete the equivalent CATI survey. The respondents were asked to rate the time to complete the survey as Excellent, Very Good, Good, Fair, or Poor. Seventy-nine percent of the respondents viewed the time required to complete the survey as Excellent or Very Good.

The flight activity data reported by Web-based respondents to the questions in Section A were very similar to the equivalent data provided by CATI respondents. In particular, the average number of reported hours and legs flown for the two respondent groups were virtually identical. Compared to respondents to the CATI survey, proportionally more respondents to the Web-based survey worked for small carriers and proportionally fewer worked for cargo carriers. Although there were some other differences of note, the Web-based data set was small and incomplete making it impossible to compare the event rates reported in the Web-based data to the rates obtained using the CATI survey.

The trial of the Web-based version of the NAOMS-survey data-collection system produced mixed results. It successfully developed a Web-based version and attendant procedures. The application itself worked well and most respondents liked its navigational approach. However, the response rate was substantially lower than that of CATI, resulting in a dataset too small for reliable analysis. The feasibility of obtaining adequate data quality through a Web-based survey would require a larger data collection effort than was available to NASA in this initial development and testing effort.

9.3 Transfer of Web-Based Data Collection System

Although the NAOMS project had officially ended, NASA collaborated with ALPA throughout FY06 to ensure an effective and graceful transfer of the operation of the Web-based NAOMS survey. The NAOMS team conducted training sessions on the Web application for ALPA staff members in early FY07. In January 2007, the operational data-collection system and associated software licenses were conveyed to ALPA for continuing operation of a survey of Part 121 pilots on behalf of the aviation industry.

10. Lessons Learned: Opportunities for Technical Improvement

In addition to its primary objectives, the NAOMS project had the objective of identifying what worked well and what didn't work well, and of making recommendations for future work. The NAOMS team placed great emphasis on developing and designing a survey process that would not only test the concept but also provide researchers and implementers with guidelines on how to proceed. Investigations into sample size; memory and recall period; questionnaire content, structure, and organization; preferred mode; number of events anticipated; number of interviews required; etc. served both the demonstration of the concept and also informational needs of those engaging in follow-on activities. But this experiment also revealed steps needed to improve the NAOMS survey process. These steps are described here.

10.1 Improving the Sampling Process

The most important improvement needed is to identify a means for selecting active air-carrier pilots from a fully representative pool of respondents. The NAOMS survey sampled the air-carrier-pilot population by drawing from the public Airmen's Registry published by the FAA. This registry describes the rating held by a pilot, but does not indicate whether he or she is flying currently as an air-carrier pilot. The NAOMS team used the ratings to identify likely active air-carrier pilots. To minimize false positives, the pilots were required to hold both ATP and Flight Engineers certificates to be eligible for inclusion in the sample pool of air-carrier pilots used for the NAOMS survey. This selection approach, and notably the requirement of the Flight Engineer Certificate, had the undesirable side effect of biasing the NAOMS-survey sample set of pilots toward Captains and pilots who fly the larger, wide-body transports. By contrast, First Officers and short-haul flights by regional carriers were under-represented in the NAOMS air-carrier-survey data.

Approval from the FAA to use the full Airmen Certification Database from which more representative samples could be selected could eliminate this potential bias. Fortunately, this bias in the NAOMS survey represented by over sampling of captains and pilots of large aircraft does not impact analyses of relative changes over time. Also, the size of the bias can be estimated, making it possible to compensate for this bias if required for other applications.

10.2 Achieving Statistical Significance for Rare Events

Many of the most serious aviation safety events occur very infrequently. This is the result of many decades of effort to make the aviation system as safe as possible. Part 121 commercial aviation accidents are exceedingly rare. It takes a great deal of data gathered over a long period of time to develop statistically reliable estimates of extremely rare event rates. Given the available time and financial resources only about 7,000 air-carrier-pilot interviews per year could be collected. This sampling rate was not sufficient to obtain statistically reliable estimates of the rates of some rare event but it was sufficient to identify relative changes over time of a number of aviation safety events. This served the purpose for the demonstration of the concept.

After stratification of the data by four air-carrier-size categories, there were sufficient reports to enable statistically reliable analyses for nearly half of the 95 events addressed by the questions in Section B. This limitation could be eased by increasing the number of interviews per year and by extending data collection over a longer period of time and, preferably, continuously.

10.3 Addition of Clarifying Drilldown Questions

When pilots are asked whether they experienced the unwanted events covered by the questions in Section B during the preceding 60 days, the overwhelming response is “no.” In fact, the typical NAOMS interviewee experienced no more than a very few safety events during the preceding 60 days. When a respondent says he/she experienced a safety event, it may be desirable, in some circumstances, to ask additional clarifying questions. For example, about 30% of all NAOMS-survey respondents said they flew more than one make and model of aircraft in commercial service. The particular make and model that the interviewee said he/she flew the majority of the time during the recall period was used in the NAOMS analyses. For future projects, an interesting clarifying question that could be asked of pilots flying multiple types of aircraft is the make/model he/she was flying at the specific time that the reported event was experienced.

10.4 Non-Response Bias

Survey response rates are typically calculated as the proportion of individuals who are contacted and complete some substantial portion of the survey. Individuals who are in the target pool but who cannot be contacted are not considered. However, if this segment of the population differs in important ways from the individuals who are contacted, then the calculated response rate can be misleading. Assessing the effects of this type of "non-response" to a survey is extremely difficult and rarely done. Further efforts to locate and survey individuals who cannot be found are often ineffective. The NAOMS team judged that, in this case, expending the resources needed to attempt to perform an assessment of this type of non-response bias was impractical. For future research, pilot unions and other industry representatives could be helpful in locating pilots who are selected for interview.

Also, pilots who are contacted but who are ineligible because they had not flown during the recall period could be questioned further, though this may be of limited benefit since the only useful question from the NAOMS questionnaire is the single question about lifetime experience. All the other questions pertain to flying experience during the recall period.

It is interesting to note that the FAA has recently taken substantial steps to update the Airmen Registry and ensure that it is up-to-date. If these efforts are successful, the likelihood of potential non-response bias would not be an issue for any future implementation of a NAOMS-like system.

10.5 Opportunities for Reduced Cost of Operations

NASA made a substantial investment in the research and development of the concept of the NAOMS survey. Much of that investment relates to the development of methodology, the engagement of the aviation stakeholders in the program, and the creation of prototypic NAOMS-survey analytic outputs. However, the majority of costs relate to data collection. If the NAOMS approach is to go beyond the concept demonstration phase, the costs associated with survey data collection must be reduced.

One factor in the cost of NAOMS-survey data collection was the investment to maximize response rate. To this end, the NAOMS team took its guidance from the "Total Dillman Method" that advocates various advance and follow-on mailings. This method led to high participation but increased the cost. If a continuing (i.e., long-term) survey were to be undertaken, the community would become sufficiently informed so the multiple mailings used for the NAOMS research project might not be necessary.

Also, the NAOMS team selected the CATI data-collection method which, while not exorbitantly expensive, required training and professional interviewers, which were a significant fraction of NAOMS-survey data-collection costs. The fully allocated cost of data collection for the experimental NAOMS survey of air-carrier pilots from 2001 to 2004 ranged from $85 to $100 per interview, depending on the length of the Section C topical section. This was deemed to be an acceptable cost during the proof-of-concept and early operational periods of the NAOMS survey but is not sustainable over the long run. High unit data-collection costs also create a barrier to expanding the survey to other constituent groups such as GA pilots, air-traffic controllers, mechanics, cabin attendants, and others in keeping with the original NAOMS vision. The NAOMS team actively explored ways to reduce costs and identified two main avenues of interest: (1) changing to a Web-based data-collection mode; and (2) soliciting respondent participation by e-mail. Taken together these actions have the potential for reducing NAOMS survey unit data-collection costs, possibly by as much as 80%. However, there are important unanswered questions about the impact these changes would have on the ability to achieve the exceptional high response levels realized by the NAOMS survey during the 2001–2004 evaluation period. The transition from a NAOMS CATI instrument to a Web-based instrument has been accomplished and is available for follow-on research. A properly designed research study would be needed to assess the effects of implementing these changes on the completeness and accuracy of the information collected by a NAOMS-like survey.

11. Summary, Conclusions, and Recommendations

The NAOMS research project succeeded in demonstrating that it is feasible to use a survey of the operators of the air-transportation system to obtain statistically reliable information on the occurrences of safety-related events as a means for assessing safety-related changes in the National Airspace System. The telephone-based NAOMS survey produced reliable estimates for a three-year period of the relative changes over time in the rates with which many unwanted events occurred. This project demonstrated the viability of using the NAOMS survey to provide a basis for trend analyses of long-term aviation safety rate measures.[14]

The successful achievement of the objectives of the NAOMS project to gain information that had not previously been attainable was due in large part to the care and attention with which the NAOMS team resolved a number of key methodological issues related to content and structure of survey instruments, the selection of potential respondents, the length of the recall period used in the survey, and like matters. The NAOMS team overcame each one of these methodological hurdles with carefully designed experiments, consultations with domain experts, and knowledgeable pragmatic decisions. The NAOMS survey achieved scientific integrity by using meticulously designed survey instruments and a carefully crafted statistical sampling design customized to the constituency. The value of viewing the aviation system through the eyes of its operators was demonstrated in a survey of 26,170 survey interviews with air-carrier pilots. The exceptional survey response rates achieved by the NAOMS survey far exceed those typical of survey efforts. This serves to affirm the quality of the survey design as well as the effectiveness of the NAOMS system for enlisting the participation of the aviation community. The statistical validity of the collected data was assured by this careful design and was demonstrated by the subsequent analyses.

Other studies conducted as part of NAOMS dealt with the use of specifically designed questions in Section C of the NAOMS survey. This section demonstrated the potential value of including a "spotlight" look at a selected topic (see Section 8). Two topics that were of high-interest to the aviation community at the time were studied to demonstrate that the NAOMS survey could acquire reliable, stable data for rapid feedback on the effects of technological or procedural changes.

A crucial, practical indication of the success of NAOMS was its ability to support the aviation community in the assessment of safety risks and the efficacy of government/industry interventions. The NAOMS team cultivated a close working association with the aviation industry and organized labor including the CAST. One of the studies reported in Section 8 was performed in collaboration with the CAST group. That study demonstrated the potential value of the survey concept in quickly providing reliable information regarding special topics of interest,

[14] Telephone use (hard wired or cellular) has changed dramatically since the NAOMS project was conducted. Future use of telephones for survey purposes will need to take these changed conditions into account, although the NASA methodology (and others that give advance notice and provide for scheduled interviews) should be less affected by these changes than those that attempt to conduct a survey on first contact.

such as the impacts on the system of new procedures, technologies, or training that the CAST had recommended. Section C was used to demonstrate the possibility of establishing performance baselines against which the effects of ongoing and future safety interventions recommended by the CAST could be measured. As a consequence of this demonstration, the CAST concluded that NAOMS afforded the best means currently available to assess the impacts of particular interventions.

It is important to recognize that NAOMS was a research project whose purpose was to provide proof of a concept and to assess the feasibility of the methodology being proposed. NAOMS sought to test the methodology, identify what procedures worked well and what did not work well, and offer recommendations for future work. These recommendations could not be made until the analyses of the survey data had been accomplished. However, analyses of the data, as was done to demonstrate that they produce statistically valid results, does not translate into an assessment of the operational performance and safety of the air-transportation system nor is it appropriate to attempt to make that leap.

Criticism of NAOMS and confusion as to its objectives followed from speculating on system safety based on the premature release of raw, unscrubbed, and unanalyzed data from a partially executed project.[15] It was only recently that some of the team members have been able to re-engage to complete the research project and produce this report documenting the achievement of its primary objective: providing statistically valid estimates of trends in the rates of problematic events experienced by flight crews. There is much more analysis that could be done to glean what else might be learned from viewing the aviation system through the eyes of these 26,170 air-carrier pilots. However, the analyses performed so far provide justification only for the results reported here, with no implications regarding other aspects of system performance.

Although conclusions can be drawn related to the methodology, no conclusions can be drawn from these data about the safety of today's National Airspace System. For that, a study would need to be conducted in a relevant time frame and performed in accordance with the recommendations outlined in the previous section of this report. However, with that caveat, there are findings in these data that can be helpful in formulating hypotheses for future research.

In accordance with the appropriate objectives of a research project, the NAOMS approach was successful in identifying flaws and needed improvements. Although NAOMS solicited responses only from air-carrier pilots, it is reasonable to assume that these lessons learned could also apply to other respondent groups (GA, ATC, maintenance personnel, cabin attendants, etc.). The following modifications should be made before the NAOMS survey concept is translated into a fully operational national aviation safety-monitoring service.

First, although NAOMS demonstrated the ability of the methodology to measure changes over time and other relative changes for a large segment of the air-carrier community, an

---

[15] The data collected over nearly 4 years of the NAOMS survey were released to the public in response to a FOIA request, despite the fact that the analyses had not been completed and the project was still a work-in- progress. While the NAOMS team had followed strict procedures to protect the confidentiality and anonymity of reporters, NASA engaged in a further large redaction effort to assure anonymity before release of these raw data.

operationally implemented survey should select from a pool of the entire air-carrier population (as well as from the entire pool of other participants as these groups are added to the survey). Second, the sample size and duration of the survey must be adequate to yield statistically reliable estimates of event rates, including those of extremely rare events that may constitute high-level threats to aviation safety. Third, the costs of conducting a fully operational survey, and especially one that taps all participant communities, must be reduced. The majority of NAOMS costs related to identifying respondents and conducting the data collection. The NAOMS team identified several ways to potentially reduce these costs: (1) shifting to a Web-based data-collection mode; and (2) soliciting respondent information via e-mail. Taken together these changes have the potential for substantially reducing survey costs. However, both Web-based survey and e-mail contact approaches need to be examined further for their impact on the quality of the data and the response rates.

The goal of NAOMS was to conduct the research that would enable the creation of a permanent service of a continuing, comprehensive, and coherent survey of all of the operators of the aviation system by an appropriate agency. NAOMS demonstrated the viability of the concept. Even today, no other feasible approach for obtaining statistically valid data about a wide range of safety events from the various system operators has been identified. Therefore, the NAOMS research project team recommends that steps be taken to begin development of a fully operational survey of the operators of the air-transportation system.[16] Such a system is needed to:

- routinely measure the safety of the national air-transportation system in a quantitatively precise way
- assess trends in the performance of the national air-transportation system that may have implications of compromised safety and identify the factors associated with those trends
- identify safety and efficiency effects of new technologies and/or procedures inserted into the operating environment

NextGen will involve numerous changes to the NAS. A NAOMS-like survey provides the means to identify the effects of such policy changes across the National Airspace System.

[16] The primary step needed involves cost reduction. In particular, Web-based surveys require further investigation to supplement the research done under the NAOMS project.

12. References

Anderson, J.R., 1976. *Language, Memory, and Thought*. Hillsdale, NJ: Erlbaum .

AvSSP, 2004. *Aviation System Monitoring and Modeling 2.1 Project Plan*. February 2004.

Barsalou, L.W., 1988. The Content and Organization of Autobiographical Memories. In U. Neisser, and E. Winograd (Eds.), *Remembering reconsidered: Ecological and Traditional Approaches to the Study of Memory* (pp. 193-243). New York: Cambridge University.

Battelle, 2007. NAOMS *Reference Report: Concepts, Methods, and Development Roadmap*. NASA Contract NNA05AC07C. November 30, 2007.

Bishop, G.F.; Hippler, H.J.; Schwarz, N.; and Strack, F., 1988. A comparison of response effects in self-administered and telephone surveys. In R.M. Groves, P.P. Biemer, L E. Lyberg, J.T. Massey, W.L. Nichols II, and J. Waksberg (Eds.), *Telephone survey methodology* (pp. 321-340). New York: Wiley.

De Leeuw, E.D., and van der Zouwen, J., 1988. Data quality in telephone and face-to-face surveys: A comparative meta-analysis. In R.M. Groves, P.P. Biemer, L.E. Lyberg, J.T. Massey, W.L. Nicholls II, and J. Waksberg (Eds.), *Telephone survey methodology* (pp. 283-299). New York: John Wiley.

Dillman, D.A., 1978. *Mail and Telephone surveys: The total design method*. New York: Wiley.

Fowler, F.J. and Mangione, T.W. 1990. Standardized Survey Interviewing. Newbury Park, CA: Sage.

GAO, 2000. Aviation Safety: Safer Skies Initiative Has Taken Initial Steps to Reduce Accident Rates by 2007. GAO/RCED-00-111. June 2000.

Groves, R.M., 1979. Actors and questions in telephone and personal interview surveys. *Public Opinion Quarterly, 43*, 191-205.

Hall, M.F., 1995. Patient satisfaction or acquiescence? Comparing mail and telephone survey results. *Journal of Health Care Marketing, 15*, 54-61.

Jobe, J.B.; Pratt, W.F.; Tourganeau, R.; Baldwin, A.K.; and Rasinski, K.A., 1997. Effects of interview mode on sensitive questions in a fertility survey. In L.E. Lyberg, P.P. Biemer, M. Collins, E.D. de Leeuw, C. Dippo, N. Schwarz, and D. Trewin (Eds.), *Survey measurement and process quality* (pp. 311-329). New York: John Wiley & Sons.

Jordan, L.A.; Marcus, A.C.; and Reeder, L.G., 1980. Response styles in telephone and household interviewing: A field experiment. *Public Opinion Quarterly, 44*, 210-222.

Krosnick, J.A., 1991. Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5, 213-236.

Krosnick, J.A., and Green, M.C., 1998. The impact of interview mode on data quality in the National Election Studies. Unpublished memo to the Board of Overseers of the National Election Studies. Columbus, OH: The Ohio State University.

Krysan, M.; Schuman, H.; Scott, L J.; and Beatty, P., 1994. Response rates and response content in mail versus face-to-face surveys. *Public Opinion Quarterly, 58*, 381-399.

Lyberg, L.E., Biemer, P.P., Collins, M., de Leeuw, E.D., Dippo, C., Schwarz, N. and Trewin, D.

(Eds.) 1997. *Survey measurement and process quality*. New York: John Wiley & Sons.

McGeoch, J.A., 1942. *The Psychology of Human Learning*. New York: Longmans, Green.

Mensink, G.J.M., and Raaijmakers, J.G.W., 1988. A Model of Interference and Forgetting. *Psychological Review*, 95, 434-455.

NTSB, 2000. *Transportation Safety Databases*. NTSB Report Number: SR--02-02. 9/11/2002.

Pryor, J.B., and Ostrom, T.M., 1981. The Cognitive Organization of Social Information: A converging-operations approach. *Journal of Personality and Social Psychology*, 41, 628-641.

Sedikides, C., and Ostrom, T.M., 1988. Are Person Categories Used when Organizing Information about Unfamiliar Sets of Persons? *Social Cognition*, 6, 252-267.

Siemiatycki, J., 1979. A comparison of mail, telephone, and home interview strategies for household health surveys. *American Journal of Public Health, 69*, 238-245.

Srull, T.K., 1983. Organizational and Retrieval Processes in Person Memory. *Journal of Personality and Social Psychology*, 44, 1157-1170.

Slamecka, N.J.,1968. An Examination of Trace Storage in Free Recall. *Journal of Experimental Psychology*, 7 & Modeling Project Plan 6, 504-513.

Statler, I.C., Morrison, R., and Rosenthal, L.J., 2003. Beyond Error Reporting Toward Risk Assessment. In *Proceedings of the 12th Biennial International Symposium on Aviation Psychology*, Dayton, Ohio, April 14–17, 2003.

Statler, Irving C. (ed.), 2007. *The Aviation System Monitoring and Modeling (ASMM) Project: A Documentation of its history and accomplishments*; 1999-2005. NASA/TP-2007-214556. June 2007.

Walker, A.H., and Restuccia, J.D., 1984. Obtaining information on patient satisfaction with hospital care: Mail versus telephone. *Health Studies Research, 19*, 291-306.

Weisberg, H.F.; Krosnick, J.A.; and Bowen, B.D., 1996. *An Introduction to Survey Research, Polling, and Data Analysis*. Thousand Oaks, CA: Sage., 121-127.

White House Commission on Aviation Safety and Security, 1998.

Appendix A. Sample of Bibliography of Survey Methodology

Aguinis, H.; Pierce, C. A.; and Quigley, B. M. (1995). Enhancing the validity of self-reported alcohol and marijuana consumption using a bogus pipeline procedure: A meta-analytic review. *Basic and Applied Social Psychology, 16*, 515-527.

Allen, B. P. (1975). Social distance and admiration reactions of "unprejudiced" whites. *Journal of Personality, 43*, 709-726.

Anderson, B. A.; Silver, B. D.; and Abramson, P. R. (1988). The effects of the race of the interviewer on race-related attitudes of black respondents in SRC/CPS National Election Studies. *Public Opinion Quarterly, 52*, 289-324.

Anderson, J. R. (1976). *Language, memory, and thought*. Hillsdale, NJ: Erlbaum.

Aquilino, W. S. (1994). Interview mode effects in surveys of drug and alcohol use: A field experiment. *Public Opinion Quarterly, 58*, 210-240.

Aquilino, W. S., and Lo Sciuto, L. A. (1990). Effects of interview mode on self-reported drug use. *Public Opinion Quarterly, 54*, 363-395.

Arabie, P. & Carroll, J. D. (1980). MAPCLUS: A mathematical programming approach to fitting the ADCLUS model. *Psychometrika, 45*, 211-235.

Atkin, C. K., and Chaffee, S. H. (1972-73). Instrumental response strategies in opinion interview. *Public Opinion Quarterly, 36*, 69-79.

Barsalou, L. W. (1988). The content and organization of autobiographical memories. In U. Neisser, and E. Winograd (Eds.), *Remembering reconsidered: Ecological and traditional approaches to the study of memory* (pp. 193-243). New York: Cambridge University.

Baker, R. P.; Bradburn, N. M.; and Johnson, R. A. (1995). Computer-assisted personal interviewing: An experimental evaluation of data quality and cost. *Journal of Official Statistics, 11*, 413-431.

Becker, W. M. (1976). Biasing effect of respondents' identification on responses to a social desirability scale: A warning to researchers. *Psychological Reports, 39*, 756-758.

Begin, G.; Boivin, M.; and Bellerose, J. (1979). Sensitive data collection through the randomized response technique: Some improvements. *Journal of Psychology, 101*, 53-65.

Beldt, S. F.; Daniel, W. W.; and Garcha, B. S. (1982). The Takahasi-Sakasegawa randomized response technique: A field test. *Sociological Methods and Research, 11*, 101-111.

Benson, L. E. (1941). Studies in secret-ballot technique. *Public Opinion Quarterly, 25*, 79-82.

Bickart, B., and Felcher, E. M. (1996). Expanding and enhancing the use of verbal protocols in survey research. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research* (pp. 115-142). San Francisco: Jossey-Bass.

Bishop, G. F., and Fisher, B. S. (1995). "Secret ballots" and self-reports in an exit-poll experiment. *Public Opinion Quarterly, 59*, 568-588.

Bishop, G. F.; Hippler, H. J.; Schwarz, N.; and Strack, F. (1988). A comparison of response effects in self-administered and telephone surveys. In R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nichols II, and J. Waksberg (Eds.), *Telephone survey methodology* (pp. 321-340). New York: Wiley.

Booth-Kewley, A.; Edwards, J. E.; and Rosenfeld, P. (1992). Impression management, social desirability, and computer administration of attitude questionnaires: Does the computer make a difference? *Journal of Applied Psychology, 77*, 562-566.

Brewer, K. R. W. (1981). Estimating marihuana usage using randomized response – some paradoxical findings. *Australian Journal of Statistics, 23*, 139-148.

Brewer, M. B.; Dull, V. T.; and Jobe, J. B. (1989). Social cognition approach to reporting chronic conditions in health surveys. Vital and health statistics. Series 6, No. 3 (DHHS publication no. PHS 89-1078). Washington, D.C.: U.S. Government Printing Office.

Brewer, M. B., and Lui, L. N. (1996). Use of sorting tasks to assess cognitive structures. In N. Schwarz and S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research* (pp. 373-385). San Francisco: Jossey-Bass.

Buchman, T. A., and Tracy, J. A. (1982). Obtaining responses to sensitive questions: Conventional questionnaire versus randomized response technique. *Journal of Accounting Research, 20*, 263-271.

Calsyn, R. J.; Roades, L. A.; and Calsyn, D. S. (1992). Acquiescence in needs assessment studies of the elderly. *The Gerontologist, 32*, 246-252.

Campbell, B. A. (1981). Race-of-interviewer effects among southern adolescents. *Public Opinion Quarterly, 45*, 231-244.

Cantril, H. (1944). *Gauging public opinion*. New Jersey: Princeton University.

Carroll, J. D., & Arabie, P. (1983). INDCLUS: An individual differences generalization of the ADCLUS model and the MAPCLUS algorithm. *Psychometrika, 48*, 157-169.

Cohen, J., & Cohen, P. (1983). Applied multiple regression/correlation analysis for the behavioral sciences. Hillsdale, NJ: Erlbaum.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*, 407-428.

Colombotos, J. (1965). The effects of personal vs. telephone interviews on socially acceptable responses. *Public Opinion Quarterly, 29*, 457-458.

Cotter, P.; Cohen, J.; and Coulter, P. B. (1982). Race of interviewer effects in telephone interviews. *Public Opinion Quarterly, 46*, 278-294.

Davis, C., and Cowles, M. (1989). Automated psychological testing: Method of administration, need for approval, and measures of anxiety. *Educational and Psychological Measurement, 49*, 311-320.

DeLamater, J., and MacCorquodale, P. (1975). The effects of interview schedule variations on reported sexual behavior. *Sociological Methods and Research, 4*, 215-236.

De Leeuw, E. D., and van der Zouwen, J. (1988). Data quality in telephone and face to face surveys: A comparative meta-analysis. In R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls II, and J. Waksberg (Eds.), *Telephone survey methodology* (pp. 283-299). New York: John Wiley.

DeNisi, A. S., and Peters, L. H. (1996). Organization of information in memory and the performance appraisal process: Evidence from the field. *Journal of Applied Psychology, 81*, 717-737.

DePaulo, B. M.; Kashy, D. A.; Kirkendol, S. E.; Wyer, M. M.; and Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology, 70*, 979-995.

Dillman, D. A. (1978). *Mail and telephone surveys: The total design method*. New York: Wiley.

Esaiasson, P., and Granberg, D. (1993). Hidden negativism: Evaluation of Swedish parties and their leaders under different survey methods. *International Journal of Public Opinion Research, 5*, 265-277.

Evan, W. M., and Miller, J. R., III. (1969). Differential effects on response bias of computer vs. conventional administration of a social science questionnaire: An exploratory methodological experiment. *Behavioral Science*, 14, 216-227.

Evans, R. I.; Hansen, W. B.; and Mittlemark, M. B. (1977). Increasing the validity of self reports of smoking behavior in children. *Journal of Applied Psychology*, 62, 521-523.

Finkel, S. E.; Guterbock, T. M.; and Borg, M. J. (1991). Race-of-interviewer effects in a pre-election poll: Virginia 1989. *Public Opinion Quarterly, 55*, 313-330.

Fischer, R. P. (1946). Signed versus unsigned personal questionnaires. *Journal of Applied Psychology, 30*, 220-225.

Fidler, D. S., and Kleinknecht, R. E. (1977). Randomized response versus direct questioning: Two data-collection methods for sensitive information. *Psychological Bulletin, 84*, 1045-1049.

Fowler, F. J., Jr.; Roman, A. M.; and Di, Z. X. (1998). Mode effects in a survey of Medicare prostate surgery patients. *Public Opinion Quarterly, 62*, 29-46.

Franklin, L. A. (1989). Randomized response sampling from dichotomous populations with continuous randomization. *Survey Methodology, 15*, 225-235.

Gano-Phillips, S., and Fincham, F. D. (1992). Assessing marriage via telephone interviews and written questionnaires: A methodological note. *Journal of Marriage and the Family, 54*, 630-635.

Goffman, E. (1959). *The presentation of self in everyday life*. Garden City, NY: Doubleday/Anchor Books.

Goodstadt, M. S., and Gruson, V. (1975). The randomized response technique: A test on drug use. *Journal of the American Statistical Association, 70*, 814-818.

Gordon, R. A. (1987). Social desirability bias: A demonstration and technique for its reduction. *Teaching of Psychology, 14*, 40-42.

Groves, R. M. (1979). Actors and questions in telephone and personal interview surveys. *Public Opinion Quarterly, 43*, 191-205.

Hall, M. F. (1995). Patient satisfaction or acquiescence? Comparing mail and telephone survey results. *Journal of Health Care Marketing, 15*, 54-61.

Herzog, A. R., and Rodgers, W. L. (1988). Interviewing older adults: Mode comparison using data from a face-to-face survey and a telephone resurvey. *Public Opinion Quarterly, 52*, 84-99.

Himmelfarb, S., and Lickteig, C. (1982). Social desirability and the randomized response technique. *Journal of Personality and Social Psychology, 43*, 710-717.

Hinrichs, J. R., and Gatewood, R. D. (1967). Differences in opinion-survey response patterns as a function of different methods of survey administration. *Journal of Applied Psychology, 51*, 497-502.

Hochstim, J. R. (1967). A critical comparison of three strategies of collecting data from households. *Journal of the American Statistical Association, 62*, 976-989.

Horvitz, D. G.; Greenberg, B. G.; and Abernathy, J. R. (1976). Randomized response: A data-gathering device for sensitive questions. *International Statistical Review, 44*, 818-196.

Hough, K. S., and Allen, B. P. (1975). Is the "women's movement" erasing the mark of oppression from the female psyche? *Journal of Personality, 89*¸249-258.

Hutchison, J.; Tollefson, N.; and Wigington, H. (1987). Response bias in college freshmen's responses to mail surveys. *Research in Higher Education, 26*, 99-106.

Jobe, J. B.; Pratt, W. F.; Tourganeau, R.; Baldwin, A. K.; and Rasinski, K. A. (1997). Effects of interview mode on sensitive questions in a fertility survey. In L. E. Lyberg, P. P. Biemer, M. Collins, E. D. de Leeuw, C. Dippo, N. Schwarz, and D. Trewin (Eds.), *Survey measurement and process quality* (pp. 311-329). New York: John Wiley & Sons.

Jordan, L. A.; Marcus, A. C.; and Reeder, L. G. (1980). Response styles in telephone and household interviewing: A field experiment. *Public Opinion Quarterly, 44*, 210-222.

Kiesler, S., and Sproull, L. S. (1986). Response effects in the electronic survey. *Public Opinion Quarterly, 50*, 402-411.

Koson, D.; Kitchen, C.; Kochen, M.; and Stodolosky, D. (1970). Psychological testing by computer: Effect on response bias. *Educational and Psychological Measurement, 30*, 803-810.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*, 213-236.

Krosnick, J. A., and Green, M. C. (1998). The impact of interview mode on data quality in the National Election Studies. Unpublished memo to the Board of Overseers of the National Election Studies. Columbus, OH: The Ohio State University.

Krysan, M.; Schuman, H.; Scott, L. J.; and Beatty, P. (1994). Response rates and response content in mail versus face-to-face surveys. *Public Opinion Quarterly, 58*, 381-399.

Lautenschlager, G. J., and Flaherty, V. L. (1990). Computer administration of questions: More desirable or more social desirability? *Journal of Applied Psychology, 75*, 310-314.

Levin, J., and Montag, I. (1987). The effect of testing instructions for handling social desirability on the Eysenck personality questionnaire. *Personality and Individual Differences, 8*, 163-167.

Locander, W.; Sudman, S.; and Bradburn, N. (1976). An investigation of interview method, threat and response distortion. *Journal of the American Statistical Association, 71*, 269-275.

Martin, C. L., and Nagao, D. H. (1989). Some effects of computerized interviewing on job applicant responses. *Journal of Applied Psychology, 74*, 72-80.

McGeoch, J. A. (1942). *The psychology of human learning*. New York: Longmans, Green.

Mensink, G. J. M., & Raaijmakers, J. G. W. (1988). A model of interference and forgetting. *Psychological Review, 95*, 434-455.

Murray, D. M.; O'Connell, C. M.; Schmid, L. A.; and Perry, C. L. (1987). *Addictive Behaviors, 12*, 7-15.

Newton, R. R.; Prensky, D.; and Schuessler, K. (1982). Form effect in the measurement of feeling states. *Social Science Research, 11*, 301-317.

Nicholls, W. L., II; Baker, R. P.; and Martin, J. (1997). The effect of new data collection technologies on survey data quality. In L. E. Lyberg, P. P. Biemer, M. Collins, E. D. de Leeuw, C. Dippo, N. Schwarz, and D. Trewin (Eds.), *Survey measurement and process quality* (pp. 221-248). New York: John Wiley & Sons.

O'Reilly, J. M.; Hubbard, M. L.; Lessler, J. T.; Biemer, P. P.; and Turner, C. F. (1994). Audio and video computer assisted self interviewing: Preliminary test of new technologies for data collection. *Journal of Official Statistics, 10*, 197-214.

Ostrom, T. M.; Carpenter, S. L.; Sedikides, C.; and Li, F. (1993). Differential processing of in-group and out-group information. *Journal of Personality and Social Psychology, 64*, 21-34.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46*, 598-609.

Pavlos, A. J. (1972). Racial attitude and stereotype change with bogus pipeline paradigm. *Proceedings of the 80th Annual Convention of the American Psychological Association, 7*, 292.

Pavlos, A. J. (1973). Acute self-esteem effects on racial attitudes measures by rating scale and bogus pipeline. *Proceedings of the 81st Annual Convention of the American Psychological Association, 8,* 165-166.

Potosky, D., and Bobko, P. (1997). Computer versus paper-and-pencil administration mode and response distortion in non-cognitive selection tests. *Journal of Applied Psychology, 82*, 293-299.

Pryor, J. B., and Ostrom, T. M. (1981). The cognitive organization of social information: A converging-operations approach. *Journal of Personality and Social Psychology, 41*, 628-641.

Quigley-Fernandez, B., and Tedeschi, J. T. (1978). The bogus pipeline as lie detector: Two validity studies. *Journal of Personality and Social Psychology, 36*, 247-256.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88*, 93-134.

Reiser, B. J.; Black, J. B.; and Abelson, R. P. (1985). Knowledge structures in the organization and retrieval of autobiographical memories. *Cognitive Psychology, 17*, 89-137.

Roenker, D. L., Thompson, C. P., & Brown, S. C. (1971). Comparison of measures for the estimation of clustering in free recall. *Psychological Bulletin, 76*, 45-48.

Roese, N. J., and Jamieson, D. W. (1993). Twenty years of bogus pipeline research: A critical review and meta-analysis. *Psychological Bulletin, 114*, 363-375.

Rogers, T. F. (1976). Interviews by telephone and in person: Quality of responses and field performance. *Public Opinion Quarterly, 40*, 51-65.

Rosenstone, S. J.; Petrella, M.; and Kinder, D. R. (1993, June 21). Excessive reliance on telephone interviews and short-form questionnaires in the 1992 national election study: Assessing the consequences for data quality (NES Technical Report No. 43). Rundus, D. (1973). Negative effects of using list items as recall cues. *Journal of Verbal Learning and Verbal Behavior, 12*, 43-50.

Schlenker, B. R., and Weigold, M. F. (1989). Goals and the self-identification process: Constructing desired identities. In L. A. Pervin (Ed.), *Goal concepts in personality and social psychology* (pp. 243-290). Hillsdale, NJ: Erlbaum.

Schober, M. F., and Conrad, F. G. (1997). Does conversational interviewing reduce survey measurement error? *Public Opinion Quarterly, 61*, 576-602.

Schuman, H., and Converse, J. M. (1971). The effect of black and white interviewers on black responses. *Public Opinion Quarterly, 35*, 44-68.

Sedikides, C., and Ostrom, T. M. (1988). Are person categories used when organizing information about unfamiliar sets of persons? *Social Cognition, 6*, 252-267.

Shimizu, I. M., and Bonham, G. S. (1978). Randomized response technique in a national survey. *Journal of the American Statistical Association, 73*, 35-39.

Siemiatycki, J. (1979). A comparison of mail, telephone, and home interview strategies for household health surveys. *American Journal of Public Health, 69*, 238-245.

Slamecka, N. J. (1968). An examination of trace storage in free recall. *Journal of Experimental Psychology, 76*, 504-513.

SPSS for Windows 8.0.0 [Computer software]. (1997). Chicago: SPSS, Inc.

Srull, T. K. (1983). Organizational and retrieval processes in person memory. *Journal of Personality and Social Psychology, 44*, 1157-1170.

StatSoft, Inc. (1997). STATISTICA for Windows [Computer program]. Tulsa, OK: StatSoft, Inc.

Sudman, S.; Bradburn, N. M.; and Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco: Jossey-Bass.

Tourangau, R.; Smith, T. W.; and Rasinski, K. A. (1997). Motivation to report sensitive behaviors on surveys: Evidence from a bogus pipeline experiment. *Journal of Applied Social Psychology, 20*, 209-222.

Tracy, P. E., and Fox, J. A. (1981). The validity of randomized response for sensitive measurements. *American Sociological Review, 46*, 187-200.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving, and W. Donaldson (Eds.), *Organization of memory* (pp. 381-403). New York: Academic Press.

Tversky, A., and Hutchinson, J. W. (1986). Nearest neighbor analysis of psychological spaces. *Psychological Review, 93*, 3-22.

Walker, A. H., and Restuccia, J. D. (1984). Obtaining information on patient satisfaction with hospital care: Mail versus telephone. *Health Studies Research, 19*, 291-306.

Walker, M. E. (1998) *Standal: A program to format proximities and cluster memberships for performing stand-alone regression* (Version 2.0) [Computer program]. Columbus, OH: The Ohio State University.

Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association, 60*, 63-69.

Weisberg, H. F.; Krosnick, J. A.; and Bowen, B. D. (1996). *An introduction to survey research, polling, and data analysis*. Thousand Oaks, CA: Sage.

Wiseman, F. (1972). Methodological bias in public opinion surveys. *Public Opinion Quarterly, 36*, 105-108.

Zdep, S. M., and Rhodes, I. N. (1976). Making the randomized response technique work. *Public Opinion Quarterly, 40*, 531-537.

Zdep, S. M.; Rhodes, I. N.; Schwarz, R. M.; and Kilkenny, M. J. (1979). The validity of randomized response technique. *Public Opinion Quarterly, 43*, 544-549.

Appendix B. NAOMS Project Team and Consultants

NASA Management

The following NASA Ames managers provided direct oversight for the NAOMS program:

Dr. Mary M. Connors, NAOMS project co-lead. Mary Connors is Asst. Chief for Aviation Programs in the Human-System Integration Division at NASA's Ames Research Center. She received her Ph.D. in Communications from Stanford University and her M.A. in Experimental Psychology from Fordham University. Much of her work with NASA has involved the impact of advanced technologies, including the interaction between humans and information / communications systems. She has authored numerous scientific articles on a range of behavioral science topics including visual perception, communications and space, and aeronautical human factors. She is the senior author of a book ("Living Aloft: Human Requirements for Extended Spaceflight") that is considered a primary source in the field of space-human factors. Most recently, her NASA work has been directed toward the development and implementation of projects to advance aviation safety.

Ms. Linda Connell, NAOMS project co-lead. Ms. Connell is director of the NASA Aviation Safety Reporting System, director of the NASA/VA Patient Safety Reporting System (PSRS), and a research psychologist in the Human-System Integration Division at NASA's Ames Research Center. Ms. Connell has been working at NASA Ames Research Center since 1981 and has participated in numerous studies with domestic and international research teams exploring human factor issues in aviation environments. During her graduate work at San Jose State University, Ms. Connell completed her master's degree in experimental psychology. Ms. Connell also is a pilot and a registered nurse.

Dr. Irving C. Statler, ASMM project manager. Dr. Statler is retired from NASA in 2008 and is currently a volunteer as an Ames Associate. His last official assignment while employed by NASA was as the manager of the project that designed, established, and demonstrated national distributed archives of air-carrier flight-recorded data and of aviation incident reports. Prior to that assignment, Dr. Statler was the Lead for the Aviation System Monitoring and Modeling (ASMM) Project of the NASA Aviation Safety Program from 1999 to its conclusion in 2005. He was the NASA Project Manager for the Aviation Performance Measuring System (APMS) research project since its inception in 1994 and the Performance Data Analysis and Reporting System (PDARS) research project since 1997, both of which became, along with NAOMS, key elements of the ASMM Project. From 1991 to 1994, he was (Acting) Chief of the Aerospace Human Factors Research Division at NASA Ames Research Center. From 1988 to 1991, Dr. Statler was a researcher in that division at Ames. Prior to that, he was Director of the Advisory Group for Aerospace Research and Development (AGARD) to the NATO Military Committee from 1985 to 1988. From 1970 to 1985, he was Director of the US Army Aviation Systems Command's Aeromechanics Laboratory. Dr. Statler received his Ph.D. from California Institute of Technology in 1956.

Contractor Research Team

The following key personnel were instrumental in planning, conducting, and analyzing research for the NAOMS program:

Mr. Loren Rosenthal, contractor's project manager. Mr. Rosenthal is an aviation safety expert and manager of Battelle's aviation safety information technology programs in Mountain View, CA. Mr. Rosenthal's technical expertise includes systems and database design; systems safety; applied statistics; and transportation economics. He received an MBA in Transportation and Public Utilities from the University of Wisconsin in 1977.
Dr. Robert Dodd, contractor's principal investigator. Dr. Dodd is currently a member of the National Transportation Safety Board. He has over 20 years as an aviation safety analyst and investigator. He was a safety engineer for the Air Line Pilots Association, a safety investigator for the National Transportation Safety Board, and worked as a research scientist for Battelle's ASRS program. He obtained his doctorate from Johns Hopkins University in Public Health and is an ATP rated pilot. Dr. Dodd served as the contractor's principal investigator for the NAOMS project and oversaw the day-to-day operations of the research project.

Dr. Jon Krosnick, survey methodologist. Dr. Krosnick is currently a professor in the Department of Psychology at Stanford University. Prior to that, and during the time he was a consultant on the NAOMS project providing council on the design NAOMS survey instrument, he was a professor of psychology at the Ohio State University. He completed his undergraduate degree at Harvard University and his Ph.D. in Social Psychology at the University of Michigan. Dr. Krosnick is an expert in survey design and application, and sits on the Board of Directors for the Gallop organization. He has authored numerous books on survey research and on designing questionnaires.

Dr. Joan Cwi, survey application manager. Dr. Cwi is director of survey operations for Battelle's Center for Health Policy Research and Evaluation (CHPRE). Dr. Cwi obtained her Ph.D. from Johns Hopkins University and has over 20 years experience in survey research. She has managed many large survey efforts relating to public health issues. For NAOMS, Dr. Cwi managed Battelle's application of the surveys during the field trial and the full survey phase of air-carrier and GA pilots.

Dr. Thomas Ferryman, senior statistician. Dr. Ferryman served as the senior statistician for the NAOMS project. Dr. Ferryman is a Battelle research scientist who works as technical group leader in the Statistical and Quantitative Sciences Group of the Pacific Northwest National Laboratory (PNNL) in Richland, Wash. Dr. Ferryman obtained his Ph.D. from the University of California, Riverside and has more than 30 years of experience in statistical research. He has performed work in a wide variety of projects, including environment, health/medical, national security, and aviation safety.

Mr. Daniel Haber, research scientist. Mr. Daniel Haber has nearly a decade of multidisciplinary experience with Battelle. His versatility has enabled him to provide critical systems development and research support to quality, safety, and advanced analytics programs for the NASA Ames Research Center as well as serve as a staff consultant to CompeteColumbus, an executive-level business consortium focused on regional technology-based economic development. In healthcare, Mr. Haber has worked with Central Ohio hospital systems to assess potential investments in health information technologies and a shared-use proton therapy facility. Mr. Haber is currently leading Battelle's strategic business development efforts surrounding healthcare research and quality. Mr. Haber obtained his MBA from the Ohio State University and his BS in Business Administration and Computer Science from Oregon State University.

Dr. Robert Mauro, experimental psychologist and statistical analyst. Robert Mauro is a Senior Research Scientist at Decision Research and an Associate Professor at the University of Oregon where he is a member of the Psychology Department and the Institute of Cognitive and Decision Sciences. Dr. Mauro received an AB in Sociology from Stanford University in 1979, a MS in Psychology from Yale University in 1981 and his PhD in Psychology from Stanford in 1984. Dr. Mauro conducts research and teaches in the areas of decision-making, human emotion, and statistical methods with an emphasis on applied problems in law and aviation. He has worked with a variety of organizations including the National Aeronautics and Space Administration, the Federal Aviation Administration, the Central Intelligence Agency, and private concerns. In aviation, Dr. Mauro has focused on issues related to pilot judgment and decision-making. He has worked on projects related to the design and evaluation of airline procedures, the development and evaluation of pilot training materials, risk assessment, and the statistical analysis of rare events.

Dr. Mike Silver, survey methodologist. Dr. Silver was a doctoral candidate in social psychology at the Ohio State University at the time he was working with Dr. Krosnick as a research associate for the NAOMS project. He has conducted a broad variety of background research for the project and has been instrumental in the analysis of much of the information gathered during background interviews and focus groups.

Mr. J.M. Jobanek, aviation safety analyst. Mr. Jobanek served as an aviation expert senior consultant to Battelle for the NAOMS project. Mr. Jobanek received his undergraduate degree from the United States Naval Academy and holds two masters degrees, one in transportation safety and the other environmental engineering. Mr. Jobanek has over 8,000 flight hours in a variety of civilian and military transport category aircraft. He has flown as a civilian air-carrier pilot and was an air safety investigator for the Air Line Pilots Association. He retired from the U.S. Air Force Reserves as a colonel.

Ms. Andrea Renholds, research statistician. Ms. Andrea Renholds, senior research scientist at Pacific Northwest National Laboratory, has worked for many years in the area of radiation detection via gamma ray spectroscopy. At the time she participated in the analyses of the NAOMS data, she was co-principle investigator for projects involving hyperspectral image analysis. She has experience as a member of multidisciplinary teams working in areas as varied as infield radiation detection data collection and protein research involving micro arrays. Ms. Renholds received her master of science in Mathematics from Brigham Young University in December 2001, specializing in numerical analysis.

Appendix C. Literature Review

Evaluation of the literature at the time of the commencement of the NAOMS project indicated that little scientific survey research on aviation safety had been conducted. Of those that were conducted, none were found to be ongoing systematic surveys.

MITRE conducted a study for the Federal Aviation Administration (FAA) on runway surface operations that relied on information from pilots derived from surveys. The study, titled "Reports by Airline Pilots on Airport Surface Operations," was published in March 1996 (FAA contract DTFA01-93-C-00001). The survey was a lengthy instrument (over 200 questions) that included both formatted questions (questions that required specific responses) and questions that were open-ended. The sampling frame for the survey was not based on a random selection process. Rather, selections were limited to a sample of roughly 10,000 pilots of two airlines that fly throughout the United States. Approximately 2,000 pilots responded for a response rate of about 20%. The survey was a self-administered questionnaire that was distributed to pilots through their airline mail system. Each survey contained only a portion of all 200 questions in an effort to reduce the burden on each individual pilot. A field trial of the survey instrument was conducted and revisions made based on comments from the survey respondents, airline union officials, and airline management.

Other surveys that were discovered were primarily conducted by special interest groups such as pilot unions or trade associations. The results of these surveys were not available, but interviews with the sponsoring organizations indicated that the surveys tended to be limited, episodic and specific to particular subgroups.

The NAOMS team turned to the general survey literature for insights that might be applied to aviation because so little aviation-specific survey literature was available. Two generic issues this literature could address was the selection of data collection mode (described later) and the optimal approach to wording questions to maximize recall accuracy?

C.1 Data Collection Mode Literature

NAOMS evaluated the three most common survey data collection modes: self-administered questionnaires (SAQ), computer-assisted telephone interviews (CATI), and in-person interviews. When choosing a survey administration mode, a number of considerations are relevant. The four criteria influencing this choice are: collection cost, respondent satisfaction, response rate, and data quality.

C.1.1. Mode Literature on Cost

The literature review indicated that in-person interviews are usually the most expensive, telephone interviews are often significantly less expensive, and self-administered questionnaires are typically the least expensive. However, if efforts are made to achieve the highest response rates possible with self-administered questionnaires, then the cost of that mode is close to the cost of applying the same questionnaire via the telephone.

C.1.2. Mode Literature on Respondent Satisfaction

Respondent satisfaction metrics favored in-person interviewing. In a study comparing respondents interviewed in-person and by telephone, Groves (1979) found that a large majority of the respondents to in-person interviews (78%) were satisfied with the experience, whereas only 38% of the respondents to the telephone interviews said they were satisfied.

C.1.3. Mode Literature on Response Rate

When comparing for response rate, in-person interviewing once again proves superior to the other modes. It is widely accepted that in-person surveys can achieve up to 70% response rate; telephone surveys can achieve a 60%r esponse, and mail surveys typically achieve 10-20% response rates unless heroic efforts are implemented (see Dillman, 1978).

C.1.4. Mode Literature on Data Quality

Two other important considerations for NAOMS were not widely recognized in the literature on modes: satisficing and social desirability bias. In general, modes that encourage these phenomena also compromise data quality.

Surveys can require respondents to do a great deal of cognitive work for little or no real reward (see Krosnick, 1991). This can lead to satisficing behavior. When this happens, respondents exert the minimal effort needed to satisfy survey requirements. Responses are not always well thought through and are more likely to contain bad data.

Social desirability bias describes the tendency of respondents to answers questions in a way that present them to interviewers in a respectable light---even if it requires the truth to be "bent" or ignored.

From the satisficing perspective, the literature indicates that telephone interviews are least desirable, and self-administered questionnaires work as well or better than in-person interviewing. Likewise, self-administered questionnaires appeared to minimize social desirability bias, while telephone interviews maximized this bias. The results of this review are described in more detail in the report.

C.1.5. Summary of Data Collection Mode Literature

Self-administered questionnaires earned the distinction of being significantly less expensive to conduct, and they also scored high marks for delivering quality data. However, they also have disadvantages. They usually have significantly lower response rates than other survey modes, unless extensive effort is made to increase response in which case the cost becomes comparable to telephone interviews. Furthermore, it is difficult to implement complex skip patterns in paper-based self-administered surveys.

When considering respondent satisfaction and response rate, the literature review favored in-person interviews, hinting that the more expensive the data collection method, the better it would perform.

With no clear winner, the NAOMS team determined that additional research was needed to select the most appropriate interview mode for the initial population of air carrier pilots. It was determined that the most efficient method to conduct this additional evaluation would be to test the different survey application methods during the field trial as discussed in the report. That way, response rate, data quality, cost, and other key indicators could be compared.

C.2. Memory Organization Literature

NAOMS researchers hoped that a review of literature about memory organization would provide insights about where to set the recall period for pilots and how to structure a questionnaire that would lead to maximum recall of safety-related events. The report and Appendix G speak to this topic in greater detail.

Human memory is not necessarily organized chronologically with one discrete event following another in the order in which they occurred (Sudman, Bradburn, & Schwarz, 1996). Rather, the human mind employs many different types of memory organization approaches (Sedikides & Ostrom, 1988; Srull, 1983). Techniques developed to assess how events or other information are organized in people's memories include: open-ended verbal protocols (also called "think-alouds"); listing; sorting; and speeded recall. These techniques are ordered from the most exploratory to the most confirmatory.

Verbal protocols are best used for assessing the structure of memory in domains where it is unclear in advance what structures might be prevalent, as was true for the NAOMS project, because they rely on open-ended responses. Having discerned likely structures, one can then use the latter techniques to confirm the existence of those structures.

Each of these techniques has advantages and disadvantages, and when used together, they allow researchers to converge on the memory organization typically employed by people who share occupation or other attributes that tend to shape memory organization. Such knowledge would then allow the NAOMS team to design its questionnaire to mirror this organizational pattern. For example, if it was discovered that most pilots store safety-related events by flight phase, questions could be constructed to match that scheme. In that case, the NAOMS survey instrument could instruct respondents to first think about any safety-related incidents that occurred during takeoffs during the last week and prompt them with a list of such possible incidents. Alternatively, if pilots primarily organized their memories according to the focus of attention when the problem occurred, one might first ask them to think about any problems they might have had interacting with ATC during the last month, and prompt them with a list of such possible incidents.

The NAOMS team undertook a simple experimental process to gain insights on pilot memory organization since the literature was bereft of information on this topic. The results of these tests are described in the report.

C.3 References

Dillman, D. A. (1978). Mail and telephone surveys: The total design method. New York: Wiley.

Groves, R. M. (1979). Actors and questions in telephone and personal interview surveys. Public Opinion Quarterly, 43, 191-205.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. Applied Cognitive Psychology, 5, 213-236.

Sedikides, C., and Ostrom, T. M. (1988). Are person categories used when organizing information about unfamiliar sets of persons? Social Cognition, 6, 252-267

Srull, T. K. (1983). Organizational and retrieval processes in person memory. Journal of Personality and Social Psychology, 44, 1157-1170.

Sudman, S.; Bradburn, N. M.; and Schwarz, N. (1996). Thinking about answers: The application of cognitive processes to survey methodology. San Francisco: Jossey-Bass.

Appendix D
Air-Carrier-Pilot Survey Questionnaire

TIME BEGUN .... (MILITARY)..................___ : ___ *(FILLS)*
INTERVIEWER: DATE OF INTERVIEW IS BEING RECORDED AS (START DATE). IS THIS THE CORRECT DATE?
YES..1 NO............ (RECORD DATE OF INTERVIEW)....................0
START DATE ....___ / ___ / ___ MONTH DAY YEAR
....S **TART DATE = 30/90 DAYS BEFORE END DATE**
END DATE..___ / ___ / ___ *(FILLS)* MONTH DAY YEAR **END DATE = DAY** **BEFORE DAY OF INTERVIEW**

SECTION A: BACKGROUND QUESTIONS

INTRODUCTION: For this survey most of the questions will refer to (30/90) days prior to today. Therefore, whenever I say the "last (TIME PERIOD), I am referring to the period from (START DATE) through (END DATE). I am now going to ask you a few questions about the commercial flying that you did during the last (TIME PERIOD).

A1.	During the last (TIME PERIOD), how many **hours** did you fly as a crewmember on commercial aircraft?	# HOURS IN TIME PERIOD.......................................______
	PROMPT IF 30 DAYS>100, 90 DAYS>300: *I'd just like to verify. You said you flew (HOURS A1) hours during the last (TIME PERIOD) as a crewmember on a commercial aircraft. Is this correct?*	*NO*.......................................0 *YES*.............. *(A2)* ....1 *RF*.............. *(A2)* ....7 *DK*.............. *(A2)* 8
A1 *NEW*	*During the last (TIME PERIOD), how many hours did you fly as a crewmember on a commercial aircraft?*	*# HOURS*.......................................______ *RF*.......................................997 *DK*.......................................998

A2. During the last (TIME PERIOD), how many **legs** did you fly as a crewmember on commercial aircraft?

LEGS IN TIME PERIOD.. |__|__|__|

A2.1 During the last (TIME PERIOD), how many of the (#A2) legs you flew involved taking off or landing at an airport outside the United States?

NOTE: THE UNITED STATES MEANS THE 50 STATES AND WASHINGTON DC, BUT DOES NOT INCLUDE US TERRITORIES.

LEGS OUTSIDE U.S.. |__|__|__|

NUMBER OF LEGS IN A2.1 MUST BE LESS THAN OR EQUAL TO LEGS IN A2.

A3. Please tell me the makes, models and series for all of the aircraft you flew commercially as a crewmember during the last (TIME PERIOD)? RECORD VERBATIM IN COLUMN A, THEN ASK PROMPT.

PROMPT A3_A1: Did you fly any other makes, models or series of aircraft commercially during the last (TIME PERIOD)?	*YES ....(ASK PROMPT A3_A2) .... 1* *NO .... (ASK B).......... 0* *RF.......... 7* *DK.......... 8*

PROMPT A3_A2: Please tell me the next aircraft make, model and series you flew commercially as a crewmember during the last (TIME PERIOD)? RECORD IN COLUMN A

A. MAKE/MODEL/SERIES (NOTE; MAKE/MODEL/SERIES DROP DOWN SCREEN INCREASED WITH THIS VERSION)	B. During the last (TIME PERIOD), what **percent** of the (HRS IN A1) did you fly the (MAKE/ MODEL/SERIES)?
1st ______	|__|__|__| %
2nd ______	|__|__|__| %
3rd ______	|__|__|__| %
4th ______	|__|__|__| %
5th ______	|__|__|__| %
6th ______	|__|__|__| %
	THE TOTAL PERCENT OF A3-B SHOULD BE 100.

INTRODUCTION:
During the last (TIME PERIOD), you may have transported passengers or cargo, or conducted other flight operations. We would like to understand what types of operations you flew.

A4. During the last (TIME PERIOD), what percent of the (HRS IN A1) did you fly as a crewmember on flights with revenue passengers?

% WITH REVENUE PASSENGERS .... |__|__|__|

A5. During the last (TIME PERIOD), what percent of the (HRS IN A1) did you work as a crewmember on flights that carried only cargo or freight and did **not** carry revenue passengers?

% CARGO/FREIGHT W/O PASSENGERS .... |__|__|__|

A6. During the last (TIME PERIOD), what percent of the (HRS IN A1) did you work as a crewmember on flights that carried no revenue passengers or cargo, such as maintenance flights, ferry flights, or repositioning flights?

% NO PASSENGER OR CARGO .... |__|__|__|

THE TOTAL PERCENT OF A4, A5, AND A6 SHOULD BE 100.

A. What type of flights were these?

SPECIFY: ____________________

A7. During the last (TIME PERIOD), did you fly a commercial aircraft (READ QUESTIONS)?

	YES	NO	RF	DK
a. as a captain	1	0	7	8
b. as a first officer	1	0	7	8
c. as a flight engineer or second officer	1	0	7	8
d. as a relief pilot	1	0	7	8
e. in any other capacity (SPECIFY)	1	0	7	8

1. What was that capacity?

A7a THROUGH A7e CANNOT ALL BE ANSWERED NO.

SPECIFY: ____________________

INTERVIEWER: CAN INCLUDE CHECK PILOT.

A7.1 Which of the following three categories best describes the number of airplanes currently operated by your airline? Please do not include airplanes operated by code-share partners. READ CATEGORIES.

NOTE: WE ARE ONLY INTERESTED IN AIRPLANES CURRENTLY BEING USED, NOT THOSE IN STORAGE.

PROBE IF PILOT FLEW FOR MORE THAN ONE AIRLINE IN TIME PERIOD: Please tell me the number of airplanes currently operated by the airline **that you flew the most hours for** in the last (TIME PERIOD).

350 airplanes or more....................1
150 to 349 airplanes....................2
149 or less airplanes....................3
RF....................7
DK....................8

A8. Approximately how many **hours** in total have you flown a **commercial aircraft during your career**?

TOTAL HOURS DURING CAREER ....|__|__|__|__|__|

Section B:
Safety Related Events

SECTION B: SAFETY RELATED EVENTS

> **INTRODUCTION:**
> My next questions are about safety related events. In answering these questions, please report only events that **you experienced on a commercial aircraft on which you were a crewmember**. The first of these questions are about **equipment-related events.**

ER1. How many times during the last (TIME PERIOD) did an aircraft on which you were a crewmember divert to an alternate airport or return to land because of an aircraft equipment problem?

EQUIPMENT PROBLEMS .... |__|__|__|

A. What systems caused the diversion or return to land?

SPECIFY: ____

ER2. How many times during the last (TIME PERIOD) did an aircraft on which you were a crewmember experience a spill, fire, fumes, or aircraft damage due to transporting hazardous materials?

HAZMAT .... |__|__|__|

IF 0, SKIP TO ER3.

A. (How many of these [# in ER2] times were the spills, fire, fumes or aircraft damage/Was this spill, fire, fumes or aircraft damage) in the cargo compartment?

IN CARGO COMPARTMENT .... |__|__|__|

> ***THE AMOUNT IN ER2A CANNOT BE GREATER THAN THE AMOUNT IN ER2.***

B. (How many of these [# in ER2] times were spills, fire, fumes or aircraft damage/Was this spill, fire, fumes or aircraft damage) in the passenger compartment?

IN PASSENGER COMPARTMENT .... |__|__|__|

> ***THE AMOUNT IN ER2A AND ER2B COMBINED CANNOT BE GREATER THAN THE AMOUNT IN ER2.***

C. (How many of these [# IN ER2] times were the spills, fire, fumes or aircraft damage/Was the spill, fire, fumes or aircraft damage) caused because the hazardous materials in question were out of compliance with regulations?

OUT OF COMPLIANCE WITH REGULATIONS .... |__|__|__|

> ***THE AMOUNT IN ER2C CANNOT BE GREATER THAN THE AMOUNT IN ER2.***

ER3. How many times during the last (TIME PERIOD) did an aircraft on which you were a crewmember experience a cargo shift

CARGO SHIIFTS .... |__|__|__|

ER4. How many times during the last (TIME PERIOD) did an in-flight aircraft on which you were a crewmember experience uncommanded movements of any of the following devices (READ QUESTIONS)?

a. Uncommanded movements of the elevators? ...... # ELEVATORS .... |__|__|__|

b. Uncommanded movements of the rudder?........... # RUDDER .... |__|__|__|

c. Uncommanded movements of the ailerons? .... # AILERONS .... |__|__|__|

d. Uncommanded movements of the spoilers?......... # SPOILERS .... |__|__|__|

e. Uncommanded movements of the speedbrakes? # SPEEDBRAKERS .... |__|__|__|

f. Uncommanded movements of the trim tabs? .... # TRIM TABS .... |__|__|__|

g. Uncommanded movements of the flaps? .... # FLAPS .... |__|__|__|

h. Uncommanded movements of the slats?............. # SLATS .... |__|__|__|

i. Did any other devices have uncommanded movements during the last (TIME PERIOD)?

YES		1
NO	(SKIP TO ER5)	0
RF	(SKIP TO ER5)	7
DK	(SKIP TO ER5)	8

1. Which devices?

SPECIFY: ____________________

2. **FOR EACH DEVICE LISTED IN ER4i1:** How many times did (DEVICE LISTED IN ER4i1) perform uncommanded movements during the last (TIME PERIOD)? # UNCOMMANDED MOVEMENTS .... |__|__|__|

ER5. How many times during the last (TIME PERIOD) did an inflight aircraft on which you were a crewmember experience smoke, fire, or fumes that originated in any of the following areas (READ QUESTIONS):

A. the engine or nacelle?....................................... # IN ENGINE OR NACELLE .... |__|__|__|

IF 0, SKIP TO ER5B.

1. (Of the [# in ER5A] times there was smoke, fire, or fumes in the engine or nacelle, how many involved/Did the smoke, fire, or fumes in the engine or nacelle involve) electrical components or wiring? # SMOKE/FIRE/FUMES........ |__|__|__|

THE AMOUNT IN ER5A1 CANNOT BE GREATER THAN THE AMOUNT IN ER5A.

B. the flight deck?... # IN FLIGHT DECK .... |__|__|__|

IF 0, SKIP TO ER5C.

1. (Of the [# in ER5B] times there was smoke, fire, or fumes in the flight deck, how many involved/Did the smoke, fire, or fumes in the flight deck involve) electrical components or wiring?

SMOKE/FIRE/FUMES.. |__|__|__|

THE AMOUNT IN ER5B1 CANNOT BE GREATER THAN THE AMOUNT IN ER5B.

C. the cargo hold?..

IN CARGO HOLD .... |__|__|__|

IF 0, SKIP TO ER5D.

1. (Of the [# in ER5C] times there was smoke, fire, or fumes in the cargo hold, how many involved/Did the smoke, fire, or fumes in the cargo hold involve) electrical components or wiring?

SMOKE/FIRE/FUMES.. |__|__|__|

THE AMOUNT IN ER5C1 CANNOT BE GREATER THAN THE AMOUNT IN ER5C.

D. the galley?..

IN GALLEY .... |__|__|__|

IF 0, SKIP TO ER5E.

1. (Of the [# in ER5D] times there was smoke, fire, or fumes in the galley, how many involved/Did the smoke, fire, or fumes in the galley involve) electrical components or wiring?

SMOKE/FIRE/FUMES.. |__|__|__|

THE AMOUNT IN ER5D1 CANNOT BE GREATER THAN THE AMOUNT IN ER5D.

E. elsewhere in the passenger compartment? ......

IN ELECTRICAL COMPONENETS OR WIRING...... |__|__|__|

IF 0, SKIP TO ER5F.

1. (Of the [# in ER5E] times there was smoke, fire, or fumes elsewhere in the passenger compartment, how many involved/Did the smoke, fire, or fumes elsewhere in the passenger compartment involve) electrical components or wiring?

SMOKE/FIRE/FUMES.. |__|__|__|

THE AMOUNT IN ER5E1 CANNOT BE GREATER THAN THE AMOUNT IN ER5E.

F. During the last (TIME PERIOD), how many times did an inflight aircraft on which you were a crewmember experience smoke, fire or fumes that originated other than in the engine or nacelle, flight deck, cargo hold, galley, or passenger compartment?

ORIGINATE OTHER PLACES .... |__|__|__|

1. Where did the smoke, fire or fumes originate? SPECIFY.

SPECIFY: ___________________________

ER6. During the last (TIME PERIOD), how many times did an inflight aircraft on which you were a crewmember experience a precautionary engine shutdown?

PRECAUTIONARY ENGINE SHUTDOWNS............|__|__|__|

ER7. During the last (TIME PERIOD) how many times did an inflight aircraft on which you were a crewmember experience a total engine failure?

TOTAL ENGINE FAILURE ....|__|__|__|

INTRODUCTION:

The following questions relate to **turbulence**.

During the last (TIME PERIOD), how many times did an aircraft on which you were a crewmember (READ QUESTION)?

TU1. Encounter severe turbulence that caused large abrupt changes in altitude, airspeed, or attitude..

CAUSED ABRUPT CHANGES ....|__|__|__|

IF 0, SKIP TO TU2.

A. (Of the [#in TU1] severe turbulence encounters, how many occurred/Did this severe turbulence encounter occur) in I.M.C. conditions? I.M.C. = INSTRUMENT METEOROLOGICAL CONDITIONS

IN IMC CONDITIONS...|__|__|__|

THE AMOUNT IN TU1A CANNOT BE GREATER THAN THE AMOUNT IN TU1.

B. (Of the [# in TU1] severe turbulence encounters, how many occurred/Did this severe turbulence encounter occur) in clear air?

IN CLEAR AIR..|__|__|__|

THE AMOUNT IN TU1A AND TU1B CANNOT BE GREATER THAN THE AMOUNT IN TU1.

TU2. Encounter wake turbulence that resulted in 10 or more degrees of aircraft roll ....

RESULTING IN AIRCRAFT ROLL ....|__|__|__|

INTRODUCTION:

The next few questions are about **weather-related events while airborne**.

During the last (TIME PERIOD), how many times did an aircraft on which you were a crewmember (READ QUESTION)?

WE1. Lack accurate weather information when crewmembers needed it while airborne ....

LACK WEATHER INFORMATION ....|__|__|__|

IF 0, SKIP TO WE2.

A. (Of the [# WE1] times when crewmembers lacked accurate weather information while airborne, how many involved non-U.S. airports or controllers?/ Did this time when crewmembers lacked accurate weather information while airborne involve a non-U.S. airport or controller?)

INVOLVE NON-US AIRPORT OR CONTROLLER.... |__|__|__|

THE AMOUNT IN WE1A CANNOT BE GREATER THAN THE AMOUNT IN WE1.

B. (Of the [# WE1] times when crewmembers lacked accurate weather information while airborne, how many involved ATIS?/Did this time when crewmembers lacked accurate weather information while airborne involve ATIS?)

INVOLVE ATIS .... |__|__|__|

THE AMOUNT IN WE1A AND WE1B COMBINED CANNOT BE GREATER THAN THE AMOUNT IN WE1.

WE2. Fail to receive A.T.C. approval for a request to avoid severe weather...

FAIL RECEIVE ATC APPROVAL .... |__|__|__|

IF 0, SKIP TO WE3.

A. (Of the [# WE2] times crewmembers failed to receive A.T.C. approval to avoid severe weather, how many times was emergency authority invoked in these situations/Was emergency authority invoked in this situation?

EMERGENCY AUTHORITY INVOKED................. |__|__|__|

THE AMOUNT IN WE2A CANNOT BE GREATER THAN THE AMOUNT IN WE2.

WE3. Divert to an alternate airfield because of weather ....

DIVERT TO ALTERNATE AIRFIELD......................... |__|__|__|

WE4. Experience airframe icing that reduced the aircraft's ability to maintain altitude, speed, stability, or directional control..

EXPERIENCE AIRFRAME ICING .... |__|__|__|

WE5. Encounter windshear or a microburst condition that resulted in an airspeed deviation of 15 knots or greater..

ENCOUNTER WINDSHEAR/MICROBURST .... |__|__|__|

WE6. Encounter windshear or a microburst condition that resulted in a windshear avoidance maneuver..........

RESULT IN WINDSHEAR AVOIDANCE .... |__|__|__|

IF A4=0, SKIP TO AC1.

INTRODUCTION:

The next few questions are about **passenger-related events**.

During the last (TIME PERIOD), how many times did an **in-flight** aircraft on which you were a crewmember (READ QUESTIONS):

CP1. Expedite landing or divert to an alternate airport due to a passenger medical emergency ....

DUE TO PASSENGER MEDICAL EMERGENCY .... |__|__|__|

CP2. Expedite landing or divert to an alternate airport due to a passenger disturbance................................ # DUE TO PASSENGER DISTURBANCE .... |__|__|__|

CP3. During the last (TIME PERIOD), how many times did a crewmember leave the cockpit to handle a passenger disturbance on an inflight aircraft on which you were a crewmember .... # CREWMEMBERS LEAVE COCKPIT....................... |__|__|__|

INTRODUCTION:

The next few questions are about **airborne conflicts**.

During the last (TIME PERIOD), how many times did an aircraft on which you were a crewmember (READ QUESTION)?

AC1. Experience a bird strike.. # BIRD STRIKES.. |__|__|__|

AC2. Perform an evasive action to avoid an imminent in-flight collision with another aircraft that was never closer than 500 feet including evasive action in response to a TCAS advisory? .... # EVASIVE ACTIONS .... |__|__|__|

AC3. Experience less than 500 feet of separation from another aircraft while both aircraft were airborne .... # LESS THAN 500 FEET SEPARATION .... |__|__|__|

INTRODUCTION:

The next few questions are about **ground operations**.

During the last (TIME PERIOD), how many times did an aircraft on which you were a crewmember (READ QUESTION)?

GE1. Go off the edge of a runway or taxiway while taxiing.. # GO OFF EDGE RUNWAY/TAXIWAY......................... |__|__|__|

GE2. Collide or nearly collide with a ground vehicle? .... # COLLIDE WITH GROUND VEHICLE .... |__|__|__|

IF 0, SKIP TO GE3.

A. (Of the [# in GE2] near collisions with a ground vehicle, how many occurred/Did this near collision with a ground vehicle occur) while your aircraft was on the ramp, apron or in the gate area? # ON RAMP/APRON/GATE AREA............................... |__|__|__|

THE AMOUNT IN GE2A CANNOT BE GREATER THAN THE AMOUNT IN GE2.

B. (Of the [# in GE2] near collisions with a ground vehicle, how many occurred/Did this near collision with a ground vehicle occur) while your aircraft was on the taxiway? # ON TAXIWAY.. |__|__|__|

THE AMOUNT IN GE2A AND GE2B COMBINED CANNOT BE GREATER THAN THE AMOUNT IN GE2.

C. (Of the [# in GE2] near collisions with a ground vehicle, how many occurred/Did this near collision with a ground vehicle occur) while your aircraft was on the runway?

ON RUNWAY .... |__|__|__|

THE AMOUNT IN GE2A, GE2B, AND GE2C COMBINED CANNOT BE GREATER THAN THE AMOUNT IN GE2.

GE3. Skid, slide, or hydroplane resulting in a significant increase in stopping distance during landing ....

SKID/SLIDE/HYDROPLANE .... |__|__|__|

GE4. Experience a rejected takeoff ....

REJECTED TAKEOFFS .... |__|__|__|

GE5. Go off the edge of a runway while taking off or landing ....

GO OFF EDGE OF RUNWAY .... |__|__|__|

GE6. Go off the end of the runway ....

GO OFF END OF RUNWAY .... |__|__|__|

GE7. Inadvertently enter an active runway ....

ENTER ACTIVE RUNWAY .... |__|__|__|

GE8. Begin takeoff roll while another aircraft occupied or was crossing the same runway ....

TAKEOFF ROLL WITH OCCUPIED RUNWAY .... |__|__|__|

GE9. Land while another aircraft occupied or was crossing the same runway ....

LAND ON OCCUPIED RUNWAY .... |__|__|__|

GE10. Nearly experience a ground collision with another aircraft while both aircraft were on the ground ....

NEAR GROUND COLLISION .... |__|__|__|

IF 0, SKIP TO AH1.

A. (Of the [# in GE10] near collisions with another aircraft, how many occurred/Did this near collision with another aircraft occur) while your aircraft was on the ramp, apron or in the gate area?

ON RAMP/APRON/GATE AREA .... |__|__|__|

THE AMOUNT IN GE10A CANNOT BE GREATER THAN THE AMOUNT IN GE10.

B. (Of the [# in GE10] near collisions with another aircraft, how many occurred/Did this near collision with another aircraft occur) while your aircraft was on the taxiway?

ON TAXIWAY .... |__|__|__|

THE AMOUNT IN GE10A AND GE10B COMBINED CANNOT BE GREATER THAN THE AMOUNT IN GE10.

C. (Of the [# in GE10] near collisions with another aircraft, how many occurred/Did this near collision with another aircraft occur) while your aircraft was on the runway?

ON RUNWAY .... |__|__|__|

THE AMOUNT IN GE10A, GE10B, AND GE10C COMBINED CANNOT BE GREATER THAN THE AMOUNT IN GE10.

INTRODUCTION:

The next few questions are about **aircraft handling-related events.**

During the last (TIME PERIOD), how many times did an aircraft on which you were a crewmember (READ QUESTION)?

AH1. Use some of its reserve fuel as defined by the F.A.Rs ....

USE RESERVE FUEL .... |__|__|__|

AH2. Accept an A.T.C. clearance that the aircraft could not comply with because of its performance limits .... # ACCEPT CLEARANCE NOT COMPLY WITH .... |__|__|__|

AH3. Lose sight of another aircraft from which the aircrew was trying to maintain visual separation

LOSE SIGHT OF AIRCRAFT .... |__|__|__|

IF 0, SKIP TO AH4.

A. (Of the [# in AH3] times an aircraft lost sight of another aircraft, how many occurred/Did losing sight of another aircraft occur) in marginal visual conditions of 3 miles or less?

IN MARGINAL VISUAL CONDITONS .... |__|__|__|

THE AMOUNT IN AH3A CANNOT BE GREATER THAN THE AMOUNT IN AH3.

AH4. Inadvertently land without clearance at an airport with an active control tower .... # LAND W/O CLEARANCE .... |__|__|__|

AH5. Inadvertently begin takeoff roll without A.T.C. clearance at an airport with an active control tower .... # TAKEOFF ROLL W/O CLEARANCE .... |__|__|__|

AH6 Inadvertently deviate from an assigned routing or A.T.C. vector for one minute or more .... # DEVIATIONS .... |__|__|__|

AH7. Experience a tail strike on landing .... # TAIL STRIKES ON LANDING .... |__|__|__|

AH8. Experience a tail strike on takeoff .... # TAIL STRIKES ON TAKEOFF .... |__|__|__|

AH9. Experience a hard landing .... # HARD LANDINGS .... |__|__|__|

AH10. Take off with an out-of-limit center of gravity .... # TAKE-OFF OUT-OF-LIMIT CENTER OF GRAVITY . |__|__|__|

AH11. Take-off overweight .... # TAKE-OFF OVERWEIGHT .... |__|__|__|

AH12. Commence take-off roll with an improper aircraft configuration .... # WITH IMPROPER CONFIGURATION .... |__|__|__|

AH13. Experience an unusual attitude for any reason .... # UNUSUAL ATTITUDE .... |__|__|__|

AH14. Experience a valid stall warning or stick shaker activation .... # STALL WARNING/STICK SHAKER ACTIVATION ... |__|__|__|

AH15. Nearly collide with terrain or a ground obstruction while airborne? .... # NEAR COLLISIONS/GROUND .... |__|__|__|

IF 0, SKIP TO AD1.

INTERVIEWER: INCLUDES BUILDINGS

A. (Of the [# in AH15] near collisions with terrain or a ground obstruction, how many were/Was this near collision with terrain or a ground obstruction)-brought to your attention by A.T.C.?

ATC BROUGHT TO YOUR ATTENTION .... |__|__|

THE AMOUNT IN AH15A CANNOT BE GREATER THAN THE AMOUNT IN AH15.

B. (Of the [# in AH15] near collisions with terrain or a ground obstruction, how many were/Was this near collision with terrain or a ground obstruction) detected through direct sighting of the ground or obstruction?

DETECTED THROUGH DIRECT SIGHTING .... |__|__|

THE AMOUNT IN AH15A AND AH15B COMBINED CANNOT BE GREATER THAN THE AMOUNT IN AH15.

C. (Of the [# in AH15] near collisions with terrain or a ground obstruction, how many were/Was this near collision with terrain or a ground obstruction) detected through activation of G.P.W.S. or E.G.P.W.S.? ....

DETECTED THROUGH GPWS/EGPWS......................... |__|__|

THE AMOUNT IN AH15A, AH15B, AND AH15C COMBINED CANNOT BE GREATER THAN THE AMOUNT IN GE10.

1. (How many of these [# in AH15c] near collisions were/Was this near collision) detected through activation of E.G.P.W.S.?

DETECTED THROUGH ACTIVATION OF EGPWS. |__|__|__|

THE AMOUNT IN AH15C1 CANNOT BE GREATER THAN THE AMOUNT IN AH15C.

INTRODUCTION:

The next few questions are about **altitude deviations**.

How many times during the last (TIME PERIOD) did an aircraft on which you were a crewmember (READ QUESTIONS)?

AD1. Inadvertently deviate from an assigned altitude by more than 300 feet? ....

ALTITUDE DEVIATIONS.. |__|__|__|

IF 0, SKIP TO AD2.

A. (Of the [# in AD1] deviations from an assigned altitude, how many were/Was this deviation from an assigned altitude) in response to a TCAS Resolution Advisory?

IN RESPONSE TO TCAS .... |__|__|__|

THE AMOUNT IN AD1A CANNOT BE GREATER THAN THE AMOUNT IN AD1.

AD2. Descend below Minimum Safe Altitude when you were **not** following A.T.C. radar vectors ....

NOT FOLLOWING ATC RADAR VECTORS............ |__|__|__|

INTRODUCTION:

The next few questions are about **interactions with air traffic control**.

AT1. During the last (TIME PERIOD), how many times was an aircraft on which you were a crewmember unable to communicate with A.T.C. in a time-critical situation because of frequency congestion?

UNABLE TO COMMUNICATE WITH ATC .... |__|__|__|

IF 0, SKIP TO AT2.

These problems may have occurred on the ground, or while airborne in the terminal area, or while en route. I'm going to ask you about each.

A. (Of these [# in AT1] times you were unable to communicate with A.T.C. in a time-critical situation because of frequency congestion, how many occurred/Did the time you were unable to communicate with A. T.C in a time critical situation because of frequency congestion occur) **while on the ground**?

WHILE ON GROUND .... |__|__|__| # TIMES

THE AMOUNT IN AT1A CANNOT BE GREATER THAN THE AMOUNT IN AT1.

B. (Of these [# in ATI1] times you were unable to communicate with A.T.C. in a time-critical situation because of frequency congestion, how many occurred/Did the time you were unable to communicate with A. T.C in a time critical situation because of frequency congestion occur) **while airborne in the terminal area**? ....

WHILE AIRBORNE.. |__|__|__| # TIMES

THE COMBINED TOTALS IN AT1A AND AT1B CANNOT BE GREATER THAN 100.

C. (Of these [# in ATI1] times you were unable to communicate with A.T.C. in a time-critical situation because of frequency congestion, how many occurred/Did the time you were unable to communicate with A. T.C in a time critical situation because of frequency congestion occur) **while en route**?..

WHILE EN ROUTE .... |__|__|__| # TIMES

THE COMBINED TOTALS IN AT1A, AT1B, AND AT1C CANNOT BE GREATER THAN 100.

AT2. How many times during the last (TIME PERIOD) did an aircraft on which you were a crewmember fly at an undesirably high altitude or airspeed on approach due to an A.T.C. clearance ....

NOTE TO INTERVIEWERS: THIS INCLUDES BUT MAY NOT BE LIMITED TO "SLAM DUNK" APPROACHES.

HIGH ALTITUDE OR AIRSPEED .... |__|__|__|

Section C:
In-close Approach Changes

SECTION C: IN-CLOSE APPROACH CHANGES

INTRODUCTION:

My next questions are about **clearance changes received on approach** within 10 miles of the runway threshold **that the flight crew did not request**.

IC1. During the last (TIME PERIOD), how many times did an aircraft on which you were a crewmember receive an unrequested clearance change to runway assignment, altitude restrictions or airspeed within 10 miles of the runway threshold?

UNREQUESTED CLEARANCE CHANGES............|__|__|__|

IF 00, DK OR RF, SKIP TO SECTION D.
IF 01, CONTINUE WITH ROUTE A.
IF 02 OR MORE, SKIP TO ROUTE B.

ROUTE A—ONLY ONE CHANGE

A. Was this unrequested clearance change declined?

YES .... (SKIP TO SECTION D) 001
NO... 000
RF .... (SKIP TO SECTION D) 997
DK .... (SKIP TO SECTION D) 998

B. Did this unrequested clearance change result in (READ QUESTIONS)?

	YES	NO	RF	DK
1. An unstabilized approach	1	0	7	8
2. A go-around or missed approach	1	0	7	8
3. An airborne conflict..................................	1	0	7	8
4. A wake turbulence encounter	1	0	7	8
5. Landing with out-of-limit tailwinds or crosswinds	1	0	7	8
6. Landing on a wrong runway	1	0	7	8
7. Landing long or fast..................................	1	0	7	8
8. Landing without clearance........................	1	0	7	8
9. A conflict on the ground with another aircraft or ground vehicle?........................	1	0	7	8
10. Any other undesirable event after the clearance change?	1 ASK a.	0	7 SKIP TO IC2.	8 SKIP TO IC2.

a. What events occurred?

SPECIFY: ____

SKIP TO IC2.

ROUTE B — TWO OR MORE CHANGES

A. Of the (# IN IC1) unrequested clearance changes, how many, if any, were declined?

\# UNREQUESTED CLEARANCE CHANGES............ |__|__|

IF NUMBER IN IC1A=NUMBER IN IC1, DK or RF, SKIP TO SECTION D.

IF ONLY ONE CHANGE REMAINS, GO TO ROUTE A, IC1B.

THE NUMBER OF UNREQUESTED CLEARANCE CHANGES WAS (NUMBER IC1) SO THE NUMBER OF UNREQUESTED CLEARANCE CHANGES THAT WERE DECLINED HAS TO BE (NUMBER IN IC1) OR FEWER.

B. How many of the accepted clearance changes resulted in (READ QUESTIONS)? IF 01 OR GREATER, ASK C.

THE ANSWERS IN IC1B 1-10 CANNOT BE GREATER THAN IC1 MINUS IC1A.

C. Did (this/any of these) (EVENT) happen in the most recent **accepted** clearance change?

		# CHANGES	YES	NO	RF	DK
1.	An unstabilized approach	|__|__|	1	0	7	8
2.	A go-around or missed approach	|__|__|	1	0	7	8
3.	An airborne conflict	|__|__|	1	0	7	8
4.	A wake turbulence encounter.............................	|__|__|	1	0	7	8
5.	Landing with out-of-limit tailwinds or crosswinds	|__|__|	1	0	7	8
6.	Landing on a wrong runway	|__|__|	1	0	7	8
7.	Landing long or fast..	|__|__|	1	0	7	8
8.	Landing without clearance	|__|__|	1	0	7	8
9.	A conflict on the ground with another aircraft or ground vehicle?..................................	|__|__|	1	0	7	8
10.	Any other undesirable event after the clearance change?	|__|__|	1 ASK a.	0	7	8

IF NONE, SKIP TO IC2.IF $\geq$1, ASK a.

NO/RF/DK: SKIP TO IC2.

a. What events occurred?

SPECIFY:____

INTRODUCTION:

(My next questions are about **this accepted clearance** change that we have been talking about./My next questions are about the **most recent clearance change** that the flight crew **accepted.**)

IC2. At which airport did this event occur?

NAME OF AIRPORT: ________________________________

A. Please tell me the location identifier for (AIRPORT).

AIRPORT LOCATION ID: ____________________________

IC3. **ASK ONLY IF TWO OR MORE MODELS REPORTED IN A3. IF ONLY ONE MODEL, SKIP TO IC4.**

Which model aircraft were you flying when this event occurred, the (LIST MODELS IN A3A)? **CODE MODEL FROM A3A**

NAME/MODEL: ___________________________________

IC4. Were you a crewmember on an F.M.S. or F.M.C. equipped aircraft at the time of this event?

YES 1
NO ....(SKIP TO IC8) 0
RF ....(SKIP TO IC8) 7
DK ....(SKIP TO IC8) 8

A. Was the F.M.S. or F.M.C. that was being used capable of storing multiple routes?

YES 1
NO ....(SKIP TO IC8) 0
RF ....(SKIP TO IC8) 7
DK ....(SKIP TO IC8) 8

B. Are the navigation and communication frequency changes in this aircraft made through the F.M.S. or F.M.C.?

YES 1
NO 0
RF 7
DK 8

IC5. In response to this clearance change, did the flightcrew reprogram or attempt to reprogram the F.M.S. or F.M.C.

YES 1
NO ....(SKIP TO IC8) 0
RF ....(SKIP TO IC8) 7
DK ....(SKIP TO IC8) 8

IC6. When programming changes were made or attempted, (READ QUESTIONS)?

	YES	NO	RF	DK
A. Did the inputs load properly	1	0	7	8
B. Was it possible to complete the programming within available time	1	0	7	8
C. Were all of the programming inputs cross-checked by other crewmembers?	1	0	7	8
D. Were there other programming difficulties	1	0	7	8
	ASK 1.	SKIP TO IC7.		

1. Please describe these difficulties.

SPECIFY:____

IC7.	Overall, did the F.M.S. or F.M.C. assist you in complying with the clearance change?	YES....1 NO....0 RF....7 DK....8

<u>ONLY IF ROUTE B IC1A IS 2 OR GREATER, READ INTRODUCTION:</u>

INTRODUCTION:

Before we continue, I want to remind you that these questions are still about the **most recent** unrequested clearance change within 10 miles of the runway threshold.

IC8.	Was the aircraft on an instrument approach prior to the clearance change?	YES....1 NO....(SKIP TO IC9)....0 RF....(SKIP TO IC9)....7 DK....(SKIP TO IC9)....8
	A. Did this change involve a change from an instrument approach to a visual approach?	YES....1 NO....(SKIP TO IC10)....0 RF....(SKIP TO IC10)....7 DK....(SKIP TO IC10)....8
IC9.	Did this change involve a change from a visual approach to an instrument approach?	YES....1 NO....0 RF....7 DK....8
IC10.	Was the aircraft programmed for an auto-coupled approach at the time of the clearance change?	YES....1 NO....0 RF....7 DK....8 NA....9
IC11.	Did this clearance change the aircraft's runway assignment?	YES....1 NO....(SKIP TO IC12)....0 RF....(SKIP TO IC12)....7 DK....(SKIP TO IC12)....8
	A. Did the runway reassignment involve a change from one runway to another parallel runway	YES....1 NO....0 RF....7 DK....8
IC12.	Did this clearance change the aircraft's altitude assignment?	YES....1 NO....0 RF....7 DK....8
IC13.	Did this clearance change the aircraft's airspeed assignment?	YES....1 NO....0 RF....7 DK....8

<u>ONLY IF ROUTE B IC1A IS 2 OR GREATER, READ INTRODUCTION:</u>

INTRODUCTION:

Once again, before we continue, I want to remind you that these questions are still about the **most recent** unrequested clearance change within 10 miles of the runway threshold.

IC14. In response to this clearance change, did the flightcrew (READ QUESTIONS)?

	YES	NO	RF	DK
A. Change a navigational aid frequency	1 (ASK 1)	0 (SKIP TO B)	7 (SKIP TO B)	8 (SKIP TO B)
1. Confirm the identity of the new navaid	1	0	7	8
B. Change the A.T.C. communication frequency	1	0	7	8
C. Revise the approach briefing	1	0	7	8
D. Change the airplane configuration	1	0	7	8
E. Disconnect any of the automated control systems?	1	0	7	8

IC15. Was the flight crew given a reason for the clearance change?

YES		1
NO	(SKIP TO IC16)	0
RF	(SKIP TO IC16)	7
DK	(SKIP TO IC16)	8

A. Was one of the reasons given (READ QUESTIONS)?

	YES	NO	RF	DK
1. Wake turbulence avoidance	1	0	7	8
2. Maintaining traffic flow and separation	1	0	7	8
3. Providing a runway favorable to your gates	1	0	7	8
4. A change in active runways	1	0	7	8
5. Weather or wind factors	1	0	7	8
6. Noise abatement factors	1	0	7	8
7. A.T.C. equipment problems	1	0	7	8
8. Was any other reason given for the clearance change	1 ASK a	0 SKIP TO IC16	7 SKIP TO IC16	8 SKIP TO IC16

a. What reasons were given?

SPECIFY:______________________________

IC16. Did responding to the clearance change (READ QUESTIONS)?

	YES	NO	RF	DK
A. reduce the quality of cockpit coordination	1	0	7	8
B. reduce situational awareness	1	0	7	8
C. Compromise traffic watch	1	0	7	8
D. Was safety compromised in any other way.	1 ASK 1.	0	7	8

(NO, RF, DK): SKIP TO SECTION D.

1. How was safety compromised?

SPECIFY:______________________________

Section C:
JIMDAT Questions

SECTION C: JIMDAT QUESTIONS

INTRODUCTION:

In the next section, I will be asking you some questions about your flying experience and training as it relates to terminal operations and instrument approaches. As we go forward, please limit you answers to those things that you personally experienced.

JD1. Is the aircraft you flew (most) during the last 60 days equipped with G.P.W.S?

GPWS = ground proximity warning system

NO .... (SKIP TO JD2) 0
YES 1
RF .... (SKIP TO JD2) 7
DK .... (SKIP TO JD2) 8

A. Is it equipped with a terrain display, such as you find in an enhanced G.P.W.S, or Terrain Avoidance Warning System, also known as TAWS (taws)?

NO .... (SKIP TO JD2) 0
YES 1
RF .... (SKIP TO JD2) 7
DK .... (SKIP TO JD2) 8

B. Does your airline require the terrain display to be selected during takeoff at specific airports?

NO OR NEVER .... (SKIP TO JD2) 0
YES OR SOMETIMES 1
RF .... (SKIP TO JD2) 7
DK .... (SKIP TO JD2) 8

C. Does your airline require the terrain display to be selected during descent and landing?

NO OR NEVER 0
YES OR SOMETIMES 1
RF 7
DK 8

D. For times that terrain display is not required, do you usually use it during takeoff?

NO, NOT USUALLY 0
YES, USUALLY 1
RF 7
DK 8

E. For times that terrain display is not required, do you usually use it during descent and landing?

NO, NOT USUALLY 0
YES, USUALLY 1
RF 7
DK 8

F. Has the terrain display experienced a map shift on any aircraft on which you were a crew member?

NO OR NEVER 0
YES OR SOMETIMES 1
RF 7
DK 8

JD2. During the last 60 days, how many times did an aircraft on which you were a crewmember experience a ground proximity warning?

TIME .... |___|___|___|

IF ZERO, SKIP TO JD3

A. Was (this warning/ the most recent of these warnings) valid?

NO .... (SKIP TO JD3) 0
YES 1
RF .... (SKIP TO JD3) 7
DK .... (SKIP TO JD3) 8

B. During this (most recent) warning, did you see the approaching terrain on the terrain display before you heard the aural warning?

NO	(SKIP TO JD3)	0
YES		1
RF	(SKIP TO JD3)	7
DK	(SKIP TO JD3)	8

JD3. During the last 60 days, how many times did which you were a crewmember receive a Minimum Safe Altitude Warning Alert, also known as an MSAW (em-saw) or an altitude awareness call from an A.T.C controller?

TIME ______

IF ZERO, SKIP TO JD4

A. (During the most recent of these events,) What did your aircraft do in response to the warning?

____

____

____

B. (During this most recent A.T.C. warning event,) Did the aircraft have an enhanced G.P.W.S. or T.A.W.S. (taws) installed?

NO	(SKIP TO JD4)	0
YES		1
RF	(SKIP TO JD4)	7
DK	(SKIP TO JD4)	8

GPWS = GROUND PROXIMITY WARNING SYSTEM
TAWS = TERRAIN AVOIDANCE WARNING SYSTEM

1. Did your aircraft also receive a ground proximity warning from this system?

NO	0
YES	1
RF	7
DK	8

JD4. How many times in the last 60 days, did an aircraft on which you were a crewmember fly a non-precision approach?

TIME ______

IF ZERO, SKIP TO JD8

A. (Was this non-precision approach flown in I.M.C? / How many of these non-precision approaches were flown in I.M.C?)

TIME ______

IMC = INSTRUMENT METEOROLOGICAL CONDITIONS

JD5. How many times in the last 60 days did an aircraft on which you were a crewmember fly an un-stabilized non-precision approach where the aircraft was not in landing configuration, on airspeed, or on glide-slope by 1,000 feet I.M.C or 500 feet V.M.C?

MC = METEOROLOGICAL CONDITIONS
VMC = VISUAL METEOROLOGICAL CONDITIONS

TIME .... |__|__|__|

IF ZERO, SKIP TO JD6

A. (During the most recent un-stabilized non precision approach,) What factors contributed to the inability to conduct a stabilized approach?

____

____

____

____

JD6. During the last 60 days, did an aircraft on which you were a crewmember have the choice between flying a constant angle approach or step-down non-precision approach?

NO .... (SKIP TO JD7) 0
YES 1
RF .... (SKIP TO JD7) 7
DK .... (SKIP TO JD7) 8

A. Which did you choose most often, the constant angle or step-down non-precision approach?

CONSTANT ANGLE 1
STEP-DOWN 2
CHOSE BOTH THE SAME 3
RF 7
DK 8

JD7. During the last 60 days, how many times did an aircraft on which you were a crewmember fly a non-precision approach to a runway when glide-slope information was available to you?

TIME .... |__|__|__|

IF ZERO, SKIP TO JD8

A. During (this/the most recent) non-precision approach, did you use the glide-slope information?

NO 0
YES 1
RF 7
DK 8

JD8. (Is the aircraft you fly/Are any of the aircraft you fly) LNAV/VNAV (L-nav/V-nav) capable?

LNAV = LATERAL NAVIGATION
VNAV = VERTICAL NAVIGATION

NO .... (SKIP TO JD9) 0
YES 1
RF .... (SKIP TO JD9) 7
DK .... (SKIP TO JD9) 8

A. Does your airline ever require pilots to use LNAV/VNAV (L-nav/V-nav) to fly constant angle approaches?

NO .... (SKIP TO JD9) 0
YES 1
RF .... (SKIP TO JD9) 7
DK .... (SKIP TO JD9) 8

1. In the last 60 days, how many times did an aircraft on which you were a crewmember use LNAV / VNAV (L-nav/V-nav) to fly constant angle approaches?

TIME .... |__|__|__|

B. During the last 60 days, how many times did an aircraft on which you were a crewmember not fly an LNAV/VNAV (L-nav/V-nav) approach when that option was available?

TIME .... |__|__|__|

IF ZERO, SKIP TO JD9

1. Please explain why the LNAV/VNAV (L-nav/V-nav) approach wasn't flown (during the most recent time that it was available).

____

____

____

JD9. During the last 60 days, was an aircraft on which you were a crewmember equipped to meet Required Navigation Performance standards, sometimes called R.N.P?

NO .... (SKIP TO JD10) 0
YES 1
RF .... (SKIP TO JD10) 7
DK .... (SKIP TO JD10) 8

A. Does your airline choose to use R.N.P?

NO .... (SKIP TO JD10) 0
YES 1
RF .... (SKIP TO JD10) 7
DK .... (SKIP TO JD10) 8

B. How many times in the last 60 days did an aircraft on which you were a crewmember fly an R.N.P approach?

TIME .... |__|__|__|

C. During the last 60 days, how many times did any aircraft on which you were a crewmember not fly an R.N.P approach when that option was available?

TIME .... |__|__|__|

IF ZERO, SKIP TO JD10

1. Please explain why the R.N.P. approach was not flown (most recent time that it was available).

____

____

____

JD10. **IF JD4 = 0, SKIP TO JD11**. During the last 60 days, how many times did an aircraft on which you were a crewmember fly a non-precision approach into an airport without D.M.E.?

TIME .... |__|__|__|

IF ZERO, SKIP TO JD11

DME = DISTANCE MEASURING EQUIPMENT

A. During (this event/the most recent of these events), would D.M.E have improved your ability to land safely?

NO 0
YES 1
RF 7
DK 8

JD11. During the last 60 days, how many times did an aircraft on which you were a crewmember fly an instrument approach into an airport where glide-slope or other ground based vertical angle guidance information was unavailable?

TIME .... |__|__|__|

IF ZERO, SKIP TO JD12

A. During (this approach/the most recent of these approaches), was D.M.E used to calculate the rate of descent for landing?

NO 0
YES 1
RF 7
DK 8

JD12. During the last 60 days, how many times did an aircraft on which you were a crewmember land on a runway without VASI (vasi) or PAPI (papi)?

TIME .... |__|__|__|

VASI = VERTICAL APPROACH SLOPE INDICATOR
PAPI = PRECISION APPROACH PATH INDICATOR

A. During the most recent of these events) would VASI (vasi) or PAPI (papi) have improved the aircraft's ability to land safely?

NO 0
YES 1
RF 7
DK 8

I would now like to ask you some questions about your airline's written standard operating procedures or SOPs.

JD13. Do your airline's written SOPs include Controlled Flight into Terrain prevention, sometimes called C-FIT (C-fit)?

NO 0
YES 1
RF 7
DK 8

JD14. Do your airline's written SOPs talk about how to avoid circumstances that could lead to an in-flight loss of control?

NO 0
YES 1
RF 7
DK 8

JD15. Do your airline's written SOPs talk about how to perform recovery from unusual **attitudes** and departure from controlled flight?

NO 0
YES 1
RF 7
DK 8

JD16. Do your airline's written SOPs talk about how to avoid approach and landing accidents?

NO 0
YES 1
RF 7
DK 8

JD17.	Do your airline's written SOPs talk about how to fly non-precision approaches?	NO .... 0 YES........ 1 RF .... 7 DK........ 8
JD18.	Do your airline's written SOPs require the use of constant angle non-precision approaches when that option is available?	NO .... 0 YES........ 1 RF .... 7 DK........ 8
JD19.	Do your airline's written SOPs talk about how to respond to E.G.P.W.S warnings? EGPWS = ENHANCED GROUND PROXIMITY WARNING SYSTEM	NO .... 0 YES........ 1 RF .... 7 DK........ 8

Now I would like to ask some questions about your recurrent training. By recurrent training I mean training conducted periodically that is designed to maintain your skills and knowledge. CLARIFICATION: This does not include transition or initial training. Recurrent training can include ground school, simulator training sessions, and any training conducted in the aircraft. I am going to read a list of issues. For each issue, please indicate if that topic or issue was covered during your last recurrent training.

JD20.	In what month and year did you receive your most recent recurrent training?	MONTH........ |__|__| YEAR........ |__|__|__|__|
JD21.	Did your most recent recurrent training talk about basic airmanship?	NO .... 0 YES........ 1 RF .... 7 DK........ 8
	A. Did your most recent recurrent training talk about normal approach procedures?	NO .... 0 YES........ 1 RF .... 7 DK........ 8
	B. Did your most recent recurrent training talk about approach briefings?	NO .... 0 YES........ 1 RF .... 7 DK........ 8
	C. Did your most recent recurrent training talk about criteria for initiating go-around and missed approaches?	NO .... 0 YES........ 1 RF .... 7 DK........ 8
	D. Did your most recent recurrent training talk about go-around and missed approach execution?	NO .... 0 YES........ 1 RF .... 7 DK........ 8
	E. Did your most recent recurrent training talk about emergency or abnormal conditions procedures?	NO .... 0 YES........ 1 RF .... 7 DK........ 8

Now I would like to ask you some questions concerning training you may have received addressing controlled flight into terrain, or C-FIT (C-fit), and other issues

JD22. Have you received C-FIT (C-fit) prevention training from your airline?

NO .... (SKIP TO JD23)........................ 0
YES... 1
RF .... (SKIP TO JD23)........................ 7
DK...................................... (SKIP TO JD23)........................ 8

A. In what month and year did you receive your most recent C-FIT (C-fit) prevention training?

MONTH.. |__|__|
YEAR .... |__|__|__|__|

B. Did your most recent C-FIT (C-fit) prevention training talk about minimum obstruction clearance altitudes or MOCA (mo ca)?

NO 0
YES... 1
RF 7
DK... 8

C. Did your most recent C-FIT (C-fit) prevention training talk about minimum enroute altitudes or M.E.A?

NO 0
YES... 1
RF 7
DK... 8

D. Did your most recent C-FIT (C-fit) prevention training talk about grid

MORA = MINIMUM OPERATING RADAR ALTITUDE

NO 0
YES... 1
RF 7
DK... 8

E. Did your most recent C-FIT (C-fit prevention training talk about G.P.W.S or E.G.P.W.S?

GPWS = GROUND PROXIMITY WARNING SYSTEM
EGPWS = ENHANCED GROUND PROXIMITY WARNING SYSTEM

NO 0
YES... 1
RF 7
DK... 8

F. Did your most recent C-FIT (C-fit) prevention training talk about escape maneuvers in response to G.P.W.S or G.P.W.S warnings?

GPWS = GROUND PROXIMITY WARNING SYSTEM
EGPWS = ENHANCED GROUND PROXIMITY WARNING SYSTEM

NO 0
YES... 1
RF 7
DK... 8

G. Did your most recent C-FIT (C-fit) prevention training talk about drift down procedures after engine failure?

NO 0
YES... 1
RF 7
DK... 8

H. Did your most recent C-FIT (C-fit) prevention training talk about maintaining situational awareness?

NO 0
YES... 1
RF 7
DK... 8

I. Did your most recent C-FIT (C-fit) prevention training talk about cockpit resource management, or C.R.M as it relates to C-FIT (C-fit) recovery?

NO 0
YES 1
RF 7
DK 8

NOTE: CRM CAN ALSO = CREW RESOURCE MANAGEMENT

J. How would you rate the quality of the most recent C-FIT (C-fit) prevention training you received from your airline? Would you say it was (READ CATEGORIES)?

EXCELLENT 1
GOOD 2
FAIR 3
POOR 4
VERY POOR 5

JD23. Did you receive training specifically in upset recovery from your airline?

NO .... (SKIP TO JD24) 0
YES 1
RF .... (SKIP TO JD24) 7
DK .... (SKIP TO JD24) 8

A. In what month and year did you receive your most recent training in upset recovery?

MONTH .... |__|__|
YEAR .... |__|__|__|__|

B. Was this training received in a simulator, in a ground school, or both?

SIMULATOR 1
GROUND SCHOOL 2
BOTH 3
RF 7
DK 8

C. How would you rate the quality of the upset recovery training you received? Would you say it was (READ CATEGORIES)?

EXCELLENT 1
GOOD 2
FAIR 3
POOR 4
VERY POOR 5

JD24. Does your airline provide training in Cockpit or Crew Resource Management, sometimes called C.R.M?

NO .... (SKIP TO JD25) 0
YES 1
RF .... (SKIP TO JD25) 7
DK .... (SKIP TO JD25) 8

A. Have you received this C.R.M training?

NO .... (SKIP TO JD25) 0
YES 1
RF .... (SKIP TO JD25) 7
DK .... (SKIP TO JD25) 8

B. Did this C.R.M. training change how you manage the flight deck?

NO 0
YES 1
RF 7
DK 8

C. Do you have suggestions for how the C.R.M training might be improved?

NO 0
YES 1
RF 7
DK 8

D. What suggestions do you have?

____

____

____

JD25. Does your airline have a no-fault missed approach or go-around policy?

CLARIFICATION: No fault means that the airline does not apply disciplinary action or criticize pilots who exercise their authority to exercise a missed approach or go around.

NO .... (SKIP TO JD26).................... 0
YES.. 1
RF .... (SKIP TO JD26).................... 7
DK.................... (SKIP TO JD26).................... 8

A. Would you favor the institution of such policy, oppose it, or neither favor nor oppose it?

FAVOR 1
OPPOSE.. 2
NEITHER FAVOR NOR OPPOSE.................................. 3
RF 7
DK... 8

JD26. During the last 60 days did you perform a missed approach or go around?

NO .... (SKIP TO JD27).................... 0
YES.. 1
RF .... (SKIP TO JD27).................... 7
DK.................... (SKIP TO JD27).................... 8

A. Did you receive any feedback from your airline regarding this missed approach

NO .... (SKIP TO JD27).................... 0
YES.. 1
RF .... (SKIP TO JD27).................... 7
DK.................... (SKIP TO JD27).................... 8

B. Was that feedback positive, negative, or both positive and negative?

POSITIVE 1
NEGATIVE... 2
BOTH POSITIVE AND NEGATIVE 3
RF 7
DK... 8

JD27. Does your airline participate in the safety reporting program called A-SAP (A-sap) also known as the Aviation Safety Action Program?

NO .... (SKIP TO JD28).................... 0
YES.. 1
RF .... (SKIP TO JD28).................... 7
DK.................... (SKIP TO JD28).................... 8

A. Have you been briefed on this A-SAP (A-sap) program?

NO 0
YES.. 1
RF 7
DK... 8

B. Were you told about the general purpose of the A-SAP (A-sap) program?

NO 0
YES.. 1
RF 7
DK... 8

C. Were you told how to submit an A-SA A-sap) report?

NO 0
YES.. 1
RF 7
DK... 8

D. If the situation arises in the future, would you submit an A-SAP (A-sap) report?

NO 0
YES.................... (SKIP TO JD27E) 1
RF .... (SKIP TO JD27E) 7
DK.. (SKIP TO JD27E) 8

1. Why not?

____

____

____

____

E. Do you believe that the confidentiality of A-SAP (A-sap) data is adequately protected?

CLARIFICATION: Confidentiality refers to both the reporter and to the use of the data.

NO		0
YES	(SKIP TO JD27E)	1
RF	(SKIP TO JD27E)	7
DK	(SKIP TO JD27E)	8

1. Why not?

____

____

____

____

F. Are you aware of any positive changes program other than A-SAP (A-sap) for receiving safety reports from pilots?

NO	0
YES	1
RF	7
DK	8

IF ZERO, SKIP TO JD29

JD28. Does your airline have a procedure or program other than A-SAP (A-sap) for receiving safety reports from pilots?

NO	(SKIP TO JD29)	0
YES		1
RF	(SKIP TO JD29)	7
DK	(SKIP TO JD29)	8

A. Are you aware of any positive changes that have resulted from this pilot reporting program?

NO	0
YES	1
RF	7
DK	8

B. Would you favor the establishment of an A-SAP (A-sap) program, oppose it, or neither favor nor oppose it?

FAVOR	1
OPPOSE	2
NEITHER FAVOR NOR OPPOSE	3
RF	7
DK	8

JD29. Does your airline have a Flight Operations Quality Assurance Program, sometimes called FOQA (FO Qua)?

NO	(ASK JD29A)	0
YES	(SKIP TO JD29B)	1
RF	(SKIP TO JD30)	7
DK	(SKIP TO JD30)	8

CLARIFICATION: This is a program at some airlines that analyzes operational data routinely collected from the flight data recorders with concurrence and oversight by the pilot's union or association at that airline.

A. Would you favor the establishment of a FOQA (FO Qua) program at your airline, oppose it, or neither favor nor oppose?

FAVOR	1
OPPOSE	2
NEITHER FAVOR NOR OPPOSE	3
RF	7
DK	8

IF ZERO, SKIP TO JD30

B. Have you been briefed on the program?

NO	0
YES	1
RF	7
DK	8

C. Do you believe that the confidentiality of FOQA (FO Qua) data is adequately protected?

NO	0
YES	1
RF	7
DK	8

CLARIFICATION: Confidentiality refers to both the identity of the pilot flying the aircraft and to the use of the data.

D. Are you aware of any safety improvements that have resulted from the FOQA (FO Qua) program?

NO	0
YES	1
RF	7
DK	8

We are interested in hearing about the safety culture at your airline, as expressed by your senior management. By senior management, we mean the C.E.O., Director of Safety, V.P. for Safety, Director of Flight Operations, and other senior management.

CEO = CHIEF EXECUTIVE OFFICER
VP = VICE PRESIDENT

JD30. Does your airline have a C.E.O. mission statement on safety?
CEO = CHIEF EXECUTIVE OFFICER

NO	0
YES	1
RF	7
DK	8

JD31. Does your airline have a Director of Safety?

NO	0
YES	1
RF	7
DK	8

JD32. Does your airline have a V.P. of Safety?
VP = VICE PRESIDENT

NO	0
YES	1
RF	7
DK	8

JD33. Have you observed a strong commitment to safety among senior management? (This question focuses on behavior.)

NO	(SKIP TO JD34)	0
YES		1
RF	(SKIP TO JD34)	7
DK	(SKIP TO JD34)	8

A. Is this senior management commitment to safety reflected throughout the organization?

NO	0
YES	1
RF	7
DK	8

JD34. If you have a safety concern, do you have a mechanism for bringing that concern to the attention of senior management?

NO	(SKIP TO SECTION D)	0
YES		1
RF	(SKIP TO SECTION D)	7
DK	(SKIP TO SECTION D)	8

A. How effective is this mechanism in reaching senior management? Would you say (READ CATEGORIES)?

EXTREMELY EFFECTIVE	1
VERY EFFECTIVE	2
SOMEWHAT EFFECTIVE	3
NOT VERY EFFECTIVE	4
NOT AT ALL EFFECTIVE	5

Section D:
Questionnaire Feedback

SECTION D: QUESTIONNAIRE FEEDBACK

INTRODUCTION:

I only have a couple more questions and these are about your reactions to the survey we have just done.

D1. How confident are you that you accurately counted all of the safety-related events that I asked you about? Would you say you were (READ QUESTIONS)?

Not confident at all		1
Slightly confident		2
Moderately confident		3
Very confident		4
Extremely confident		5
RF		7
DK`		8

D2. Were any of the questions I asked confusing, poorly worded, or ambiguous?

YES		1
NO	(SKIP TO D3)	0
RF	(SKIP TO D3)	7
DK`	(SKIP TO D3)	8

A. Could you please describe these question problems? RECORD VERBATIM. AT COMPLETION OF INTERVIEW, ENTER QUESTION NUMBER.

QUESTION NUMBER	RECORD VERBATIM

D3. Are there any safety problems happening within the national aviation system that I did not ask about but that you think may be worth asking about in further surveys?

YES		1
NO	(SKIP TO D4)	0
RF	(SKIP TO D4)	7
DK`	(SKIP TO D4)	8

A. What are these problems?

SPECIFY: ______________________________

D4. Do **you** use the internet at home?

YES	1
NO	0
RF	7
DK	8

D5. Do you have any other comments or suggestions about this survey? RECORD VERBATIM.

____

____

____

____

PANEL PASSWORD HINT	TAKES INTERVIEWER TO "NEEDPAS" (PANEL 1ST QTR OR LATER QTR BUT NEVER COMPLETED INTERVIEW) OR PAST PATH (PANEL 2ND QTR OR LATER WHO PREVIOUSLY GAVE PASSWORD).
NEEDPASS: We would like to be able to link the information you give us each time we call. Because we do not link your information with your name, we would like to record an individual password we can use to link your data. May we please have a password that you will repeat to us when we call you again?	AGREED .... 1 REFUSED.......... (ENDINT) 7
PICKPASS: RECORD PASSWORD	TAKES INTERVIEWER TO ENDINT.
ASKFORHINT: Please give us a question that we can use as a hint in case you are unable to remember your password the next time we call. For instance, if you choose the word "RED" as your password, your hint question could be "What is my favorite color?"	RECORD HINT
PASTPATH: At the end of your last interview you gave us a password so we could link your information across quarters. Your hint question was (HINTQUESTION). What was your password? RECORD.	REMEMBERS PASSWORD.......... (REPREATPASS).......... 1 REFUSED.......... (ENDINT) .... 7 CAN'T REMEMBER .... (SUBSPASS) 8
REPEATPASS: RECORD PASSWORD.	IF SUCCESSFUL, TAKES INTERVIEWER TO ENDINT.
IF PASSWORD NOT IN PASSWORD LIST: The word you gave me does not match our list of passwords. Perhaps I spelled it wrong. How do you spell your password? RETURN TO REPEATPASS FIELD AND RECORD PASSWORD AGAIN. IF WORD STILL DOESN'T MATCH AFTER TWO ATTEMPTS, CLICK, SUPPRESS.	IF SUPPRESSED, TAKES INTERVIEWER TO SUBSPASS.

SUBSPASS: Since (you can't remember/we don't seem to have) your previous password, we'd like you to choose another password and hint so we can link your future interviews. May we please have another password and hint that you will repeat to us when we call again?	YES ....(PICKPASS) .... 1 NO ....(ENDINT) 0

ENDINT Again, thank you very much for your time and your help with this survey. Your input will help the aviation industry a great deal to measure the level of safety in the aviation system and will be held in confidence.

IF PANEL MEMBER: We'll be calling again in three months for your (2nd/3rd/last) interview.

QUESTIONNAIRE LENGTH: QUESTIONNAIRE LENGTH (MINUTES) .... |__|__|__|

Appendix E. Memorandum on Recall Period

Authors' Note: The selection of an optimum recall period was one of the key methodological issues studied by the NAOMS team. Some of the factors that guided the memory studies the team conducted were addressed in the following decision document prepared by Battelle and Ohio State University in December 1999 largely based on a review of the literature.

Input to NAOMS Decision on Selecting a Recall Interval to Maximize the Accuracy of Pilot Recollections

Introduction

In the full-scale survey we will ultimately field, we will ask pilots to recall safety-related events that they have experienced. In order to design these questions, one of the fundamental issues we have to resolve is for how long in the past pilots can accurately recall the events that they have experienced (e.g., one week? one month? one year?). For purposes of efficiency, we would like to ask about as long a time interval as possible without compromising accuracy. To help make this decision, we conducted a review of the published literature on the accuracy of recall, and we conducted a study to assess the accuracy of pilots' recall. This memo reports what we found.

In order for pilots to accurately answer our NAOMS survey questions, they must (1) remember that each event they witnessed actually occurred, (2) remember exactly when the events they remember occurred, so they can properly include it in or exclude it from the reference period we ask about, and (3) not believe they remember events that never in fact occurred (called "fabrications"). Therefore, we sought to assess how well people can recall that an event occurred, how accurate people are in recalling the dates of events, and how much they fabricate events. Not surprisingly, occurrence recall and dating become less accurate as the retention interval increases. Therefore, our primary goal in this memo is to ascertain the longest possible recall interval that is not associated with significant compromise in recall accuracy, to inform our decision about the best retention interval for the NAOMS survey.

Dating accuracy can be assessed in two different ways, and we will do both. First, we explore what we call absolute dating error, which refers to how far off a person's recollection of a date is from the truth, regardless of the direction of the error (i.e., ignoring whether the person claims the event occurred more recently than it actually did or that it occurred less recently than it actually did).

Second, we explored the direction of dating errors that people make, focusing on what we refer to as signed dating error, which is positive when people think an event occurred longer ago than it really did and negative when people think an event occurred more recently that it really did. In the memory literature, these two types of error are referred to as "backward telescoping" and "forward telescoping," respectively (Rubin & Baddeley, 1989). Backward telescoping will bias our event counts downward, and forward telescoping will bias our event counts upward, so we need to know when each type of telescoping occurs to inform our decision about recall period.

At present, NAOMS is not intended to ask respondents about the specific details of particular events they witnessed. That is, there are no plans to ask people about circumstances surrounding individual equipment failures they witnessed, for example. Rather, we plan simply to ask people to count up numbers of events. However, past studies of memory have often assessed people's ability to recall details of events they witnessed, and we review this literature to get a sense of how quickly memory fades. It is well-known that memory for details of events fades as time passes, but literature reviews to date have not yet attempted to use the available evidence to inform decisions about optimal recall periods (e.g., Sudman, Bradburn, & Schwarz, 1996). We do so.

As we will document below, the findings of past studies of memory are consistent in supporting some conclusions but are very different in the specifics of findings in many regards. Furthermore, no past study we have located has explored people's ability to remember events that occur in the course of professional work activity that is very involving but is also very repetitive. Therefore, it is difficult to comfortably generalize findings from these past studies to the work that pilots do. As a result, we felt it important to conduct research directly exploring pilots' ability to remember events that occur repeatedly in the course of their professional activities. The new study we describe below did just that.

Literature Review

Criteria for Selecting Studies for the Review

Psychologists and other social scientists have conducted thousands of studies exploring memory, but only a small fraction of these studies apply to the issues of current interest to us. Several criteria were used to select the studies most applicable and useful for our purposes.

To be included in this review, a study had to address the accuracy of memory for specific events or the frequency of personal events that the respondent personally experienced or witnessed. This excluded studies that did not address the accuracy or validity of the memories recalled, studies that tested memory for non-events, such as word lists or abstract symbols, and studies of memory for events that were not personally experienced (e.g., memories of public news events, such as an earthquake that occurred far away from the respondent) or that focused on another person's life events (e.g., a husband answering questions about his wife's medical history).

Additionally, a study had to gauge memory using a free recall or cued-recall paradigm rather than a recognition paradigm. In the standard free recall paradigm, respondents are asked to generate recollections without being given any cues or hints. In a cued-recall paradigm, respondents are provided some help to facilitate remembering (e.g., cue words, a calendar). But in recognition studies, people are presented with descriptions of specific objects or events and are asked to report whether they did or did not occur, were or were not seen previously, and so on. Because the NAOMS surveys will not be able to describe specific events to respondents in this fashion, we focused on studies of free recall and cued recall, which are procedures more like what will be done in NAOMS.

To be included in this review, a study had to involve some method of assessing the accuracy of people's recollections. That is, each study needed to involve use of an objective indicator of the truth of events as a baseline against which to compare people's recollections. For some studies, paper or computer records were checked. For other studies, researchers instigated and observed the events that respondents were later asked to recall. And in other studies, respondents kept diaries or other written records while events occurred, with which later recollections could be compared.

Quite a few studies have gauged reporting accuracy by comparing the aggregation of reports from a representative sample of a population with aggregate figures for the same population derived from institutional records. For example, Gottfredson and Hindelang (1977) compared an aggregation of people's reports of how often they had been victims of crimes with aggregate figures derived from police reports recording crimes. Although such studies are useful for many purposes, they do not address recall at the individual level and so were excluded from this review.

Additionally, studies had to be published in English and relatively easily accessible in order to be included in our review. So, for example, theses, dissertations, and unpublished manuscripts were excluded.

Potentially relevant studies were found through several search methods. The primary method was a computer literature search of published studies in psychology and related fields. This search used keywords (such as "memory" and "recall") and the names of researchers known to have conducted potentially relevant research on memory (e.g., Baddeley and Skowronski). Additionally, some studies were found by searching through the reference lists of all the papers and books we collected.

Overview

General

The above criteria yielded 23 studies to address in this review. The studies involved a diverse set of methods, respondent populations, and memory measures. Some studies included only measures of event occurrence, some included only measures of dating accuracy, and some included only measures of characteristics recall.

The second column of Table E-1 (attached to this appendix) indicates the method employed in each study. In record check studies, respondents were asked to recall events, and written records were checked to assess recall accuracy. In diary studies, respondents kept track of events in a diary and were later asked to recall the events; recollections were checked against the diaries to assess accuracy. In experience sampling studies, respondents carried devices such as electronic pagers that alerted them to complete a small (diary-like) questionnaire at randomly-selected times throughout their days. Respondents were later asked to recall the events they reported on the questionnaires. Case studies involve interviews with a group of people who all experienced the same real-world event (e.g., a robbery). The recollections of these people are then compared to records that exist about the event to check their accuracy.

In laboratory experiments, researchers controlled what respondents experienced, such as by having respondents watch a simulated bank robbery on videotape. Then, researchers could check the accuracy of recollections against the stimulus materials. Field experiments are conducted in natural settings (e.g., restaurants) rather than in a laboratory in order to improve the generalizability of the findings. Some experimental studies used between-subjects designs, in which different respondents provided data for different recall intervals (1-week in the past, 2-

weeks, etc.). Other studies used within-subjects designs, in which the same respondents provided recall data for several recall intervals.

Most respondents were adults, including undergraduate students, professional psychologists, crime victims, bank customers, and employees (see column 3 of Table E-1). Studies involving "one memory researcher" examined the recollections of a single researcher who kept a diary and gauges his or her recall over extended time periods (of up to six years). Most studies were conducted in the U.S., Canada, or Great Britain. No studies examined members of the aviation community as respondents.

A few studies lasted several years, but because the most likely recall intervals for NAOMS surveys is likely to be no longer than one year, data from beyond one year are not addressed in this review.

Measures of Recall Accuracy

Five different measures of recall accuracy were analyzed for this review. First, we review research addressing the degree to which people are able to accurately recall that an event occurred (regardless of the accuracy of dating that event, and regardless of their ability to recall specific details about that event). This is referred to here as occurrence accuracy. Occurrence accuracy was sometimes reported as the percentage of respondents who correctly reported that an event occurred. But occurrence accuracy was also sometimes gauged by the percentage of occurrences that were correctly reported by the entire sample of respondents. Both measures are used in the studies reviewed here. The fourth column of Table E-1 identifies the three studies that assessed occurrence accuracy, and the fifth column indicates what type of event was examined in each.

Second, we review two sets of studies addressing how accurately people can recall the dates of events they have experienced: we report separately research regarding absolute dating accuracy and signed dating accuracy. Both forms of dating accuracy have usually been gauged by the inverse of accuracy, dating error, and this is the index we use in this review. Dating error refers to the deviation between the reported date of an event and the actual date on which it occurred. In this review, signed dating errors are calculated as actual date minus reported date. Therefore, negative deviations (e.g., -2 days) indicate reporting an event as occurring more recently than it actually did (the memory phenomenon known as "telescoping" or "forward telescoping") and positive deviations (e.g., +2 days) indicate reporting an event as occurring further in the past than it actually did (a phenomenon known as "time expansion"). Absolute dating errors are the absolute (or unsigned) values of the deviations, computed by taking the absolute value of the deviation (and disregarding the direction of error). The last column in Table E-1 identifies the 14 studies that gauged dating accuracy, either absolute or signed or both.

Next, we review evidence on the accuracy of recollections of the characteristics of events, which we call characteristics accuracy. Characteristics accuracy has usually been described in terms of the percent of characteristics that were recalled correctly by respondents. In these studies, characteristics accuracy data have usually been quantified in terms of the mean percent of the details of events recalled correctly at different recall intervals. However, some studies instead reported the percent of respondents recalling the details of a given event accurately. The sixth column of Table E-1 identifies the 11 studies that assessed characteristics accuracy, and the seventh column indicates what characteristics were examined.

Finally, we review evidence from two studies on fabrication of event characteristics, which is the only available evidence we have been able to turn up regarding fabrications.

Predictor: Recall Interval

Our primary interest here is in understanding how the length of the recall interval affects recall accuracy. This predictor of accuracy was operationalized in various different ways in different studies. Some studies divided recall interval (and reported accuracy and date deviations) by days, others by weeks, others by months, quarters, 6-month blocks, or other intervals. Results from specific studies are reported in the study's metrics, but occasionally small adjustments were made to allow maximally meaningful comparison.

Findings

Occurrence Accuracy

Although many surveys have asked respondents how often they have performed some behavior, we found only the three studies assessing how recall interval affects accuracy. Key details of two of these studies are provided in Table E-2 (attached to this appendix).[1]

The first column of Table E-2 identifies the study. The second column indicates what the respondents were asked to recall. At the tops of the remaining columns are the recall intervals examined in any one or more studies discussed in this review. For example, data in the "30 Days" column represent the accuracy of respondents recalling events that occurred 30 days before the recall test. Reading across a single row illustrates the results of a single study, showing how longer recall intervals (shown farther to the right) do or do not compromise recall accuracy. Although any one study reported results for only a small portion of these intervals, all intervals are shown in every table of this review to facilitate comparison across tables.

Hagburg (1968) asked respondents to remember (for two recall intervals) the number of times they had attended a class. In this case, a longer recall interval was not associated with lower accuracy. In fact, recall accuracy was slightly lower for events within the last 60 days (48%) than for events that occurred between 60 and 120 days previously (53%; see row 1 of Table E-2).

However, Mathiowitz and Duncan (1988) examined a wider range of recall intervals and did find decline over time (see row 2 of Table E-2). These investigators asked respondents to recall the number of unemployment spells they had experienced. Recall accuracy declined as the recall interval increased from 30 days to 6 months, although there was a surprising surge in accuracy for 9 months. This may be attributable to the seasonal nature of employment and people's enhanced ability to remember events that occurred during circumstances similar to the ones present at the time of recall.

It is also interesting to note that Mathiowitz and Duncan (1988) found absolutely perfect recall at 30 days but less accuracy at all longer intervals. This is one data point suggesting the optimal recall interval for perfect accuracy may be between 30 and 60 days (the next recall interval examined in this study) for some types of repetitive events.

One other study is relevant here but is not listed in Table E-2. Burton and Blair (1991, Study 2) also assessed occurrence accuracy, asking about the number of checks respondents had written in a given interval. These investigators did not report occurrence accuracy in terms of percentage but instead reported the average error in number of checks claimed to have been written. Not surprisingly, respondents made more errors for longer recall intervals. These respondents reported they had written about one too many checks after seven days but almost two too few checks after six weeks.

Summary. Although occurrence accuracy is one of the most important variables we need to consider in our decisions, very little evidence exists to assist us on this point. There is certainly evidence here that recall accuracy declines with increasing recall interval, but it is not clear from these studies along what the optimal interval may be.

Absolute Dating Error

Studies examining absolute dating error have also found a decline in accuracy with increasing recall interval. First, consider two studies that reported either the percentage of respondents who recalled the exact dates of events or the percentage of dates recalled exactly across all respondents in the study. Klein et al. (1999) examined recollections of the date of people's visits to a medical clinic. After approximately 3 weeks, 20% of respondents could recall the exact date, whereas after approximately 6 weeks, only 3%could do so. Likewise, Skowronski et al. (1995) found that people recalled more of the exact dates of activities with friends for shorter recall intervals. The percentages of evenings recalled exactly at 1, 2, 3, 4, 6, 9, and 12 months were 65%, 37.5%, 43%, 36.5%, 14.5%, 16.5%, and 18%, respectively.

Table E-3 (attached to this appendix) shows details of five other studies that included measures of absolute dating error. In this table, the average amount of dating error is reported in days. For example, at the 45-day interval, respondents' date reports were wrong by an average of 11 days in Linton's (1975) study.

Three of these studies documented the expected decline in accuracy with increasing recall interval. For example, Linton (1975) found that she was personally able to date events extremely accurately for recall intervals of less than one week (with an average error of 0 for a two day interval and .8 for a seven day interval), but error rates increased precipitously for longer intervals (to an average of 10.5 days error for events six months in the past).

Rubin and Baddeley (1989, Study 1) asked respondents to recall the dates of research presentations made by scholars who visited their university and found absolute dating error were twice as large for events six months in the past (100 days error) than for events 30 days in the past (50 days error). Skowronski et al. (1991) found an increase in dating error from 2.1 days off at a two-day interval to 11 days off at the 72-day interval in recollections of rare life events.

Skowronski et al. (1995) also found increased dating error for longer recall intervals, but with one glaring exception. As the recall interval increased from 60 days to one year, average dating error increased from 18.3 days error to 62 days, with relatively small deviations from a steadily increasing rate. However, the highest absolute dating error found in this study, 91.7 days, occurred at the 30 day recall interval, which is something of a puzzle.

Finally, Czaja et al. (1994) found no difference in absolute dating accuracy between two relatively long recall intervals. When victims of crimes remembered the dates of those crimes, average error was 27.0 days for a six-month recall interval and 25.4 days for a nine-month interval.

One additional study not listed in Table E-3 complements these results. Baddeley et al. (1978) reported a sizable positive correlation ($r = .57$, $p<.001$) between recall interval and absolute dating error for respondents who recalled dates on which they had previously participated in laboratory experiments. Average dating error increased by an average of 19 days for every 100 days lengthening of the recall interval.

Considering the data from each individual study in isolation allows the strongest scientific integrity to inferences, because all aspects of the recall task were held constant except for the recall interval. Therefore, inferences about the impact of recall interval can be made with confidence. However, each single study examined only a limited set of recall intervals, so it is difficult to use any one study to reach a more general conclusion about the optimal recall interval for accurate recollection.

To address this issue, we conducted a quasi-meta-analysis by combining the results of the studies listed in Table E-3 to generate average figures. Admittedly, this combining entails significant confounding of recall intervals with various other variables, so the results produced should be viewed with significant caution. Specifically, occasionally outlying values should not be taken as seriously as a general trend in the data. Nonetheless, this exercise seems worthwhile. Consequently, we computed the mean absolute dating error for each column in Table E-3 and then graphed the resulting means (see Figure E-1).

In Figure E-1, the x-axis shows each of the recall intervals reported in Table E-3. The y-axis indexes the average absolute dating error in days. Data points and numerical values in the figure are the means computed from all studies for each recall interval. For example, for a 2-day recall interval, two studies yielded absolute dating errors of 0 and 2.1, the average of which, 1.1, is plotted above "2 days" in Figure E-1. All immediately adjacent point are connected by lines, and points separated by recall intervals with no data are not connected by a line (e.g., no studies reported absolute dating error rates at the 180-day interval, so no mean can be computed nor plotted).

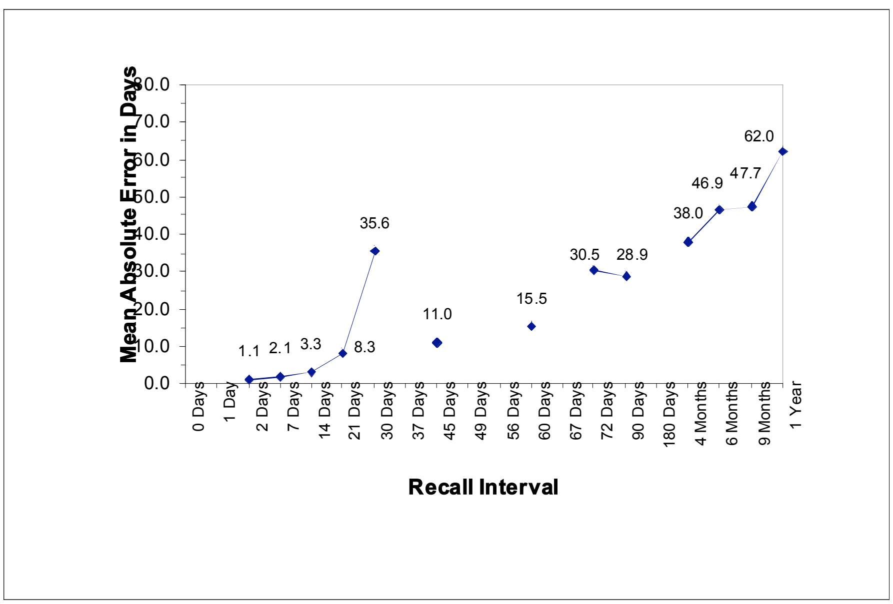


Figure E-1. Absolute Error in Dating Events.

If one ignores the single outlying point for the 30-day interval, Figure E-1 documents the excepted general trend of increasing error with increasing recall interval. Indeed, this pattern is remarkably clear. Absolute dating errors were relatively large (28.9 days error or more) for the longest intervals (of 90 days or longer) and were remarkably small for recall intervals of 14 days and less (e.g., 3.3 days average error or less).

Summary. Absolute dating error increases as the recall interval increases and average dating errors become substantial even for the 21-day recall interval. Therefore, in order to avoid sizable dating error, Figure E-1 suggests that the optimal recall interval for NAOMS may be less than 21 days.

Signed Dating Error

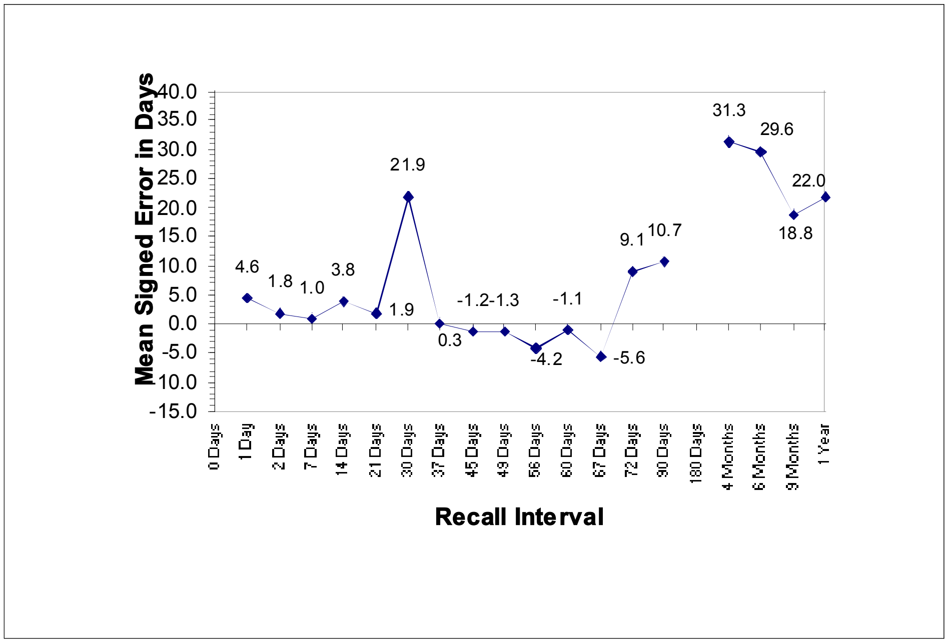
Next, we turn to studies examining signed dating error, documenting the prevalent direction of mistakes in dating events.[2] Eight of ten such studies listed in Table E-4 (attached to this appendix) show a particular pattern: for shorter recall intervals, backward telescoping is generally the rule, whereas for longer recall intervals, forward telescoping appears to be the rule. That is, signed dating errors tend to be positive for short recall intervals and negative for long recall intervals. For example, Betz and Skowronski (1997) found positive signed dating errors for recall intervals between 1 day and 30 days, and negative signed dating errors for recall intervals between 45 and 72 days. Looking across almost every row of Table E-4 reveals the same positive to negative trend.

The only exceptions to this general pattern were reported by Rubin and Baddeley (1989, Study 1) and Skowronski et al. (1995). With intervals of less than one year, both Rubin and Baddeley (1989, Study 1) and Skowronski et al. (1995) found no evidence of forward telescoping. Rubin and Baddeley (1989, Study 1) found essentially no difference in error between 72 days and six months, but the numbers in Table E-4 from Skowronski et al.'s (1995) study are consistent with the assertion that backward telescoping declines in magnitude as the recall interval increases.

Interestingly, both Rubin and Baddeley (1989, Study 1) and Skowronski et al. (1995) did in fact confirm the expected pattern of forward telescoping for long recall intervals, but only for intervals so long as not to appear in Table E-4. Both sets of researchers actually gauged recollection accuracy for recall intervals up to 2.5 years, and these investigators did indeed observe forward telescoping for the longest of these intervals. But because those long intervals fall outside of the range considered in this review, that pattern is not apparent in their contribution to Table E-4.

This same conclusion was supported by another study that used a different analytic approach. Baddeley et al. (1978) found a correlation between recall interval and signed dating error of -.33 ($p<.01$), indicating that respondents recalled dates as progressively more recent the older they actually were (e.g., 10 days earlier at the 90 day interval but 30 days earlier at the one year interval).

Again to yield heuristically value (though potentially misleading) figures, we averaged the values listed in Table E-4 to produce Figure E-2, which is formatted in the same way as Figure E-1. Surprisingly, this figure does not clearly show the pattern seen in most rows of Table E-4, moving from backward telescoping for short recall intervals to forward telescoping for long recall intervals. Instead, it appears that backward telescoping is greatest for the longest and shortest recall intervals, whereas forward telescoping occurs for moderate recall intervals.



<u>Figure E-2. Signed Error in Dating Events.</u>

However, this is an illusion, due entirely to the two outlying studies by Skowronski et al. (1995) and Rubin and Baddeley (1989, Study 1). When we remove the data from these studies, we see the expected pattern with only one exception (see Figure E-3): a relatively smooth progression from backward telescoping for short recall intervals to forward telescoping for long recall intervals.

<u>Implications</u>. Figure E-3 has especially interesting implications regarding the goals of NAOMS. First, note that there appears to be no systematic bias in dating for recall intervals of approximately 40 days. Therefore, if the recall interval specified for the NAOMS questionnaire is 40 days, telescoping should not bias event counts either up or down. Therefore, simply from this viewpoint, a 40-day recall interval might appear to be ideal.

However, the data in Figure E-1 suggest caution about recall intervals of 21 days or longer. Does Figure E-3 indicate that we cannot have our cake and eat it, too? That is, must we reduce accuracy by lengthening the recall interval in order to reduce bias? Fortunately, matters are not quite so bleak. As long as our recall interval is at least 7 days, the sizable dating bias evidenced in Figure E-3 for very recent events (that occurred during the last couple of days) will not distort results, because even the misperceived dates of those events will appear to fall within the question's reference period. Furthermore, the average magnitude of backward telescoping for a recall interval of between 7 days and 21 days is only 2.3 days according to Figure E-3. This means that event counts will be biased downward by about 16% for a 14-day interval (i.e., 2.3 days/14 days) and by about 12% for a 21-day interval (i.e., 2.3 days/21 days). Furthermore, as long as recall interval is held constant across all NAOMS surveys, then this under-counting of events will not bias apparent trends in event rates over time.

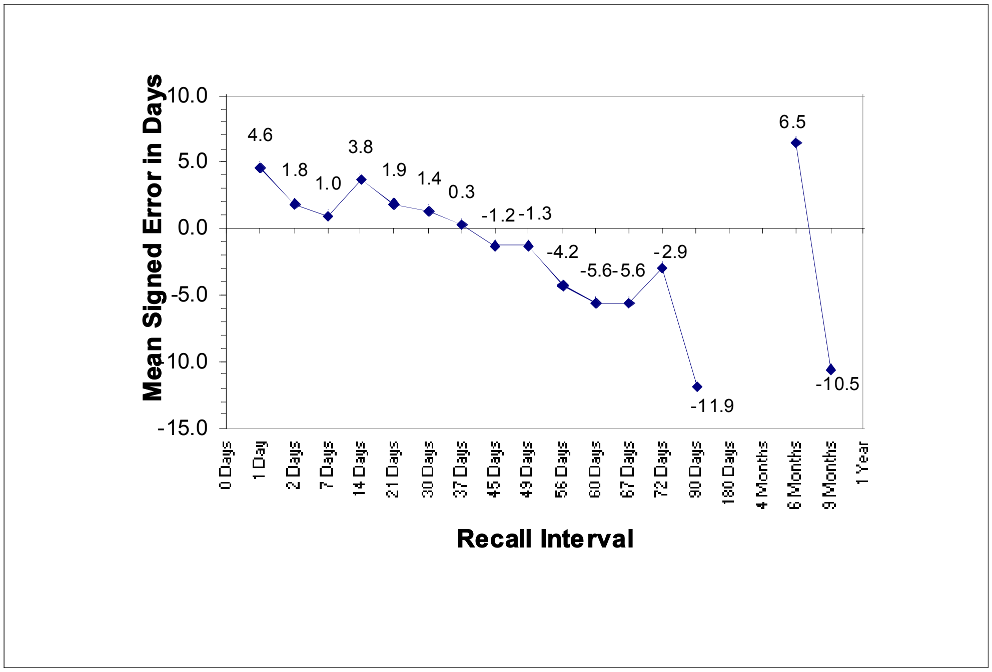


Figure E-3. Signed Error in Dating Events - Excluding Skowronski et al. (1995) and Rubin & Baddeley (1989, Study 1)

Event Characteristics Recall Accuracy

Eleven studies measured the accuracy of respondents' recollections of event characteristics.[3] The first two columns of Table E-5 (attached to this appendix) list such studies and the particular characteristics being recalled. The remaining columns list the proportions of respondents who correctly recalled a characteristic or the percent of characteristics correctly recalled.

In 18 of the 25 rows in this table, the expected pattern is apparent: declining accuracy with increasing recall interval. In four of the remaining rows, there is no change at all in accuracy with increasing recall interval (the third row of Ebbesen & Rienick's (1998) data; both rows of Klein et al.'s (1999) data, and the first row of Yuille & Cutshall's (1986) data). Only three rows show increased accuracy with increased recall interval. However, these increases are generally very small: 1% in the first row of Baddeley et al.'s (1978) data, 3% in the third row of Smith's (1935) data, and 7% in the third row of Bell's (1992) data.

The conclusion that recall accuracy declines as the recall interval becomes longer is also supported by Yarmey and Morris's (1998) study, which used a different metric to assess accuracy. These investigators reported the number of characteristics of an event recalled correctly later; no percentages could be calculated, because the total number of characteristics potentially recallable could not be known. Respondents correctly recalled an average of 15.1 characteristics within minutes of witnessing an event and recalled only 11.6 characteristics on average after a one-week delay.

Whereas some event characteristics were recalled correctly by the vast majority of respondents long after the event had occurred, other characteristics were remember correctly by only a small fraction of respondents very soon after the event. Presumably the nature of the events and characteristics studied contributed to this variability. As a result, Figure E-4 (which is again a rough graphical representation of the data reported in Table E-5) shows tremendous bouncing around and no evidence of a clear trend. Nonetheless, row-wise inspection of Table E-5 makes it clear that decline accuracy with increasing recall interval is the rule here.

Clearly, recall of event characteristics can be quite poor after even short periods of time. This has two important implications for NAOMS. First, our surveys will probably yield more accurate results if they focus on event occurrences rather than asking respondents for detailed descriptions of events. And second, given how low accuracy can be for even short recall intervals, these findings suggest that long recall intervals are especially risky.

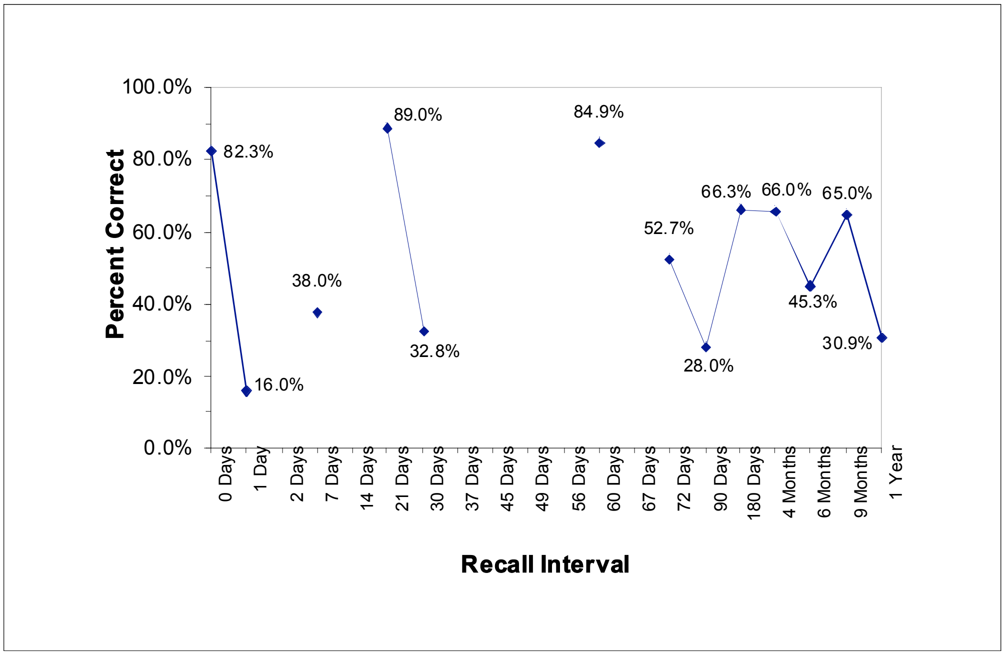


Figure E-4. Percent of Characteristics Recalled Correctly.

Fabrications

Fabrication of event characteristics (or errors of commission) were examined in two studies that we review next. Because the NAOMS surveys will not ask respondents about event characteristics, this sort of fabrication is not directly relevant to our questionnaire. Evidence regarding patterns of characteristic event fabrication could conceivably inform our design decisions, but unfortunately, these particular studies are not useful, as we explain next.

Ebbesen and Rienick (1998) asked respondents to recall details of stories they read or that were read to them. Approximately one to two percent of all details respondents claimed to recall from the stories were in fact fabrications. Further, fabrication rates did not differ among the one day, 7 days, and 30 days time intervals.

In Yarmey and Morris's (1998) study, respondents watched a videotape of a simulated armed bank robbery and later recalled the details of the robbery, after either a few minutes or one week. These researchers did not report the levels of fabrications they observed, but they did report finding no significant difference in the fabrication rates for the two recall intervals.

Summary. These two studies found no evidence of increased fabrication with longer recall intervals. However, neither of these studies examined recall intervals longer than 30 days, so at best, they should be viewed as reassuring only regarding such relatively short intervals.

Conclusions

This review of the existing literature on autobiographic event recall supports a number of general conclusions. First, given how much important and consequential survey research asks respondents to report the frequency of particular events, it is remarkable how little published research has explored individual-level accuracy in respondents' reports. Furthermore, the findings of the research that has been published evidence significant variability across studies. This suggests that the particular recall task and the particular population of respondents may substantially influence recall accuracy. So we certainly cannot walk away from the published literature with a strong recommendation about how to design the NAOMS survey questionnaires.

However, the existing literature does clearly support two general conclusions. First, event occurrence and characteristic recall accuracy declines with increasing recall interval. Therefore, employing a longer recall interval entails the increased risk of inaccurate reporting. Second, dating of events manifests a particular pattern of bias, with backward telescoping characterizing short recall intervals, and forward telescoping characterizing long recall intervals. Furthermore, the particular patterns observed in past studies do suggest that recall intervals of 21 days or less are optimal for maximizing accuracy.

Given the relatively small number of studies available to inform our decision and the clear variability of findings across event domains and respondent populations, it seemed worthwhile to supplement the above literature review with a study expressly examining the accuracy of airline pilots' recall of repeated events that occur in the course of their professional activities. Therefore, a study of pilot recall accuracy was conducted to help guide our decision about the recall interval(s) to be used in the final NAOMS survey.[4]

A Study to Assess the Accuracy of Pilots' Recall

To date, although we are considering other options, we have been most-seriously discussing a one-month recall period for the full NAOMS survey. This interval has been based simply upon our own intuitions and the intuitions of the pilots we have interviewed. The purpose of the study we conducted was to better inform our decision-making on this issue by providing empirical data.

Because it is essential that our measurements be as precise as possible, the possibility that pilots will misremember incidents is of great importance. Our goal in this study was to explore the impact of recall interval on all three sorts of misremembering that are significant threats to the validity of our measurements. First, pilots may forget incidents that occurred. Second, pilots may remember incidents that did occur but may misremember the dates on which they occurred, for example, leading them to report these incidents as having occurred during the "reference period" (i.e., the period during which they are being asked to recall incidents), when in fact they occurred prior to the reference period. Third, pilots may completely fabricate incidents that never occurred (commission errors). We therefore asked pilots to perform a relatively simple memory task, the results of which could be objectively verified against official records, and with which we could gauge the magnitude of these three types of errors for different recall intervals.

Method

For this study, we asked five of the pilots who performed cognitive interviews with us to complete the task of recalling all their flights during the previous 30 days without consulting any written records. Each respondent was given a piece of paper displaying a calendar with a box for each day. In each box, the respondent was asked to write "TO" for each takeoff that occurred and "LD" for each landing that occurred. They were then asked to indicate the city where each takeoff occurred and the city where each landing occurred. These data allow a simple assessment of these pilots' ability to accurately recall the characteristics (i.e., location) and date of one type of routine event.

After completing the recall task, the pilots took an identical calendar home with them and completed it accurately by using their log books. The respondents then mailed these second calendars back to us. We matched the recollections against the actual flight events to assess the accuracy of the pilots' recall.

Results

Figure E-5 shows the total number of actual takeoffs and landings that pilots experienced per week for the four weeks preceding the recollection exercise. "Week 1" means the week immediately before the recall task was performed. "Week 2" means the interval between 8 and 14 days before the recall task was performed. Because we asked about a total of 30 days, Week 4 included 9 days. As can be seen, these pilots experienced a substantial number of takeoffs and landings: 44 during Week 1, 57 during Week 2, 41 during Week 3, and 62 during Week 4.

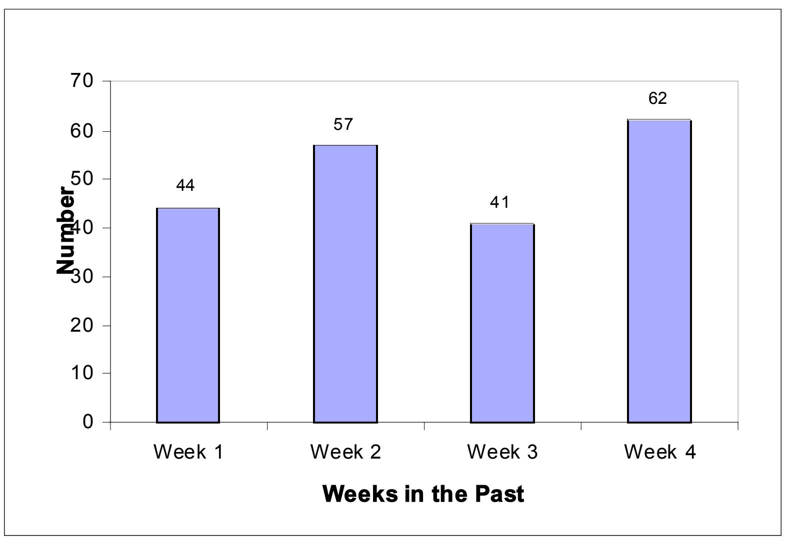


Figure E-5. Number of Actual Takeoffs and Landings by Week

Figure E-6 shows the proportion of these takeoffs and landings correctly recalled per week. The results here are quite striking. The pilots correctly remembered every single event that occurred during Week 1, but accuracy declined dramatically after that week. During Week 2, only 35% of the events were accurately recalled. During Weeks 3 and 4, these figures were 18% and 38%, respectively. This suggests that asking pilots to remember routine events more than one week previously risks significant forgetting and error in our estimates of safety-related occurrences.

This conclusion is reinforced by Figure E-7, which displays the proportion of events that occurred each week that were forgotten by pilots. These figures are simply the mirror images of the figures displayed in Figure E-6. They indicate that more than half the events that occurred during Weeks 2, 3, and 4 were completely forgotten by the pilots.

Independent corroboration of this conclusion comes from Figure E-8, which displays the number of events that pilots said had occurred but in fact had not. Thus, these are purely fictitious events. Although not one such event was reported during Week 1, two such events were reported during Week 2, 8 were reported during Week 3, and 11 were reported during Week 4.

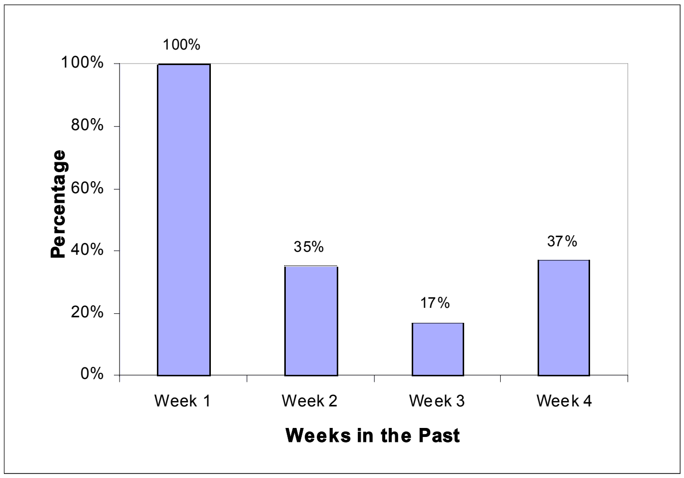


Figure E-6. Percent of Correct Takeoffs and Landings Recalled by Week

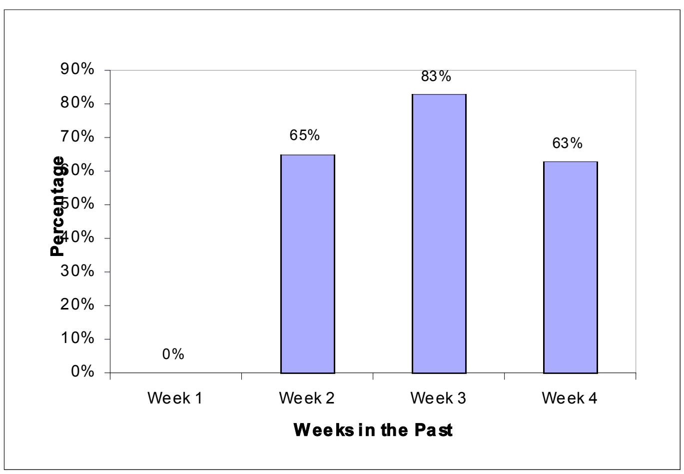


Figure E-7. Percent of Forgotten Takeoffs and Landings Recalled by Week

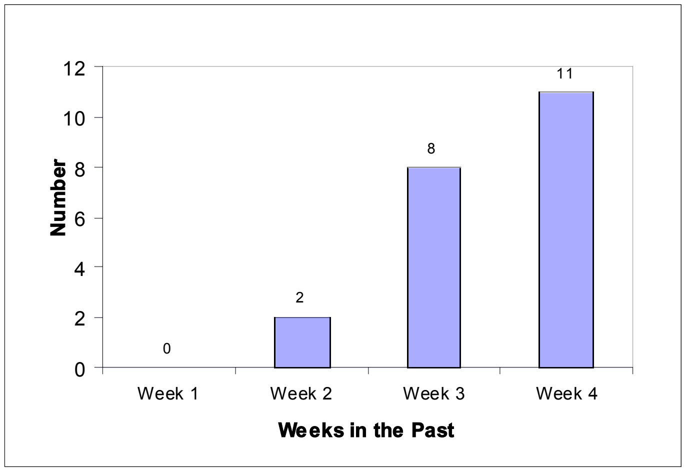


Figure E-8. Number of Fictitious Takeoffs and Landings by Week

Our guess is that these reports are the result of forward telescoping, in which the pilots remembered events as having occurred more recently than they actually did. Thus, we suspect that many of these fictitious events may indeed have happened but took place prior to the reference period of the task. Because forward telescoping is more likely to cause small time displacements rather than long ones, it is most likely to appear near the boundary of the recall period, which in this case is Week 4. This would explain why the most fabrication occurred during this period.

Put together, the results of this study suggest that accurate recollection of relatively routine events is likely to be significantly compromised if pilots are asked to remember such events that occurred more than seven days prior to completing the NAOMS survey questionnaire.

A Final Consideration: Counting vs. Estimating

One final aspect of the memory literature worth considering when specifying our recall interval is work exploring the cognitive strategies that respondents employ when reporting event frequencies. According to much recent work informed by cognitive psychology, when a person is asked to say how many times a particular event occurred during a particular time period, one of two different strategies can be employed: counting or estimating (see Burton & Blair, 1991; Menon, 1993, 1994; Sudman, Bradburn, & Scharz, 1996). Counting involves attempting to remember each individual instance of the event, and totaling up the number retrieved from memory. In contrast, estimating involves using some sort of rule or principle to guess how many times the event occurred, without having to remember each individual instance. For instance, a person can guess how often the event has typically occurred during a one-week period and can then multiply by 12 to yield a total figure for a total figure for a 3-month period.

Most strategies for estimating are based upon counting for some time interval, so any errors inherent in counting are likely to be translated directly into estimates. Furthermore, additional error can appear in estimates if the principles used to estimate are incorrect or biased as well. Therefore, it is preferable to avoid inducing respondents to estimate if they can be induced to count instead, and if the counting task is one they can accomplish reasonably well.

One factor that influences the likelihood of estimating is whether the event occurs relatively rarely—frequencies of rare events are more likely to be counted, whereas frequencies of common events are more likely to be estimated. Because some NAOMS events are thought to occur relatively frequently, there is a substantial risk that respondents will prefer to estimate their frequencies rather than count.

A second factor influencing the decision to count vs. estimate is the distinctiveness of the events to be enumerated. If an event occurs quite similarly on repeated occasions, counting is more difficult, whereas if an event occurs quite differently each time it happens, counting is easier to do. Because some events to be addressed in the NAOMS questionnaire are likely to occur quite similarly each time, this further increases the risk that respondents will prefer to estimate their frequencies rather than count.

Another factor that the existing literature suggests influences the choice of counting vs. estimating is the length of the recall interval. Longer recall intervals make it more difficult for respondents to count, so they resort to estimating. This suggests that we should keep the recall interval as short as possible in order to encourage respondents to count rather than estimate. But unfortunately, the existing literature does not provide good generic guidance on how short the recall interval needs to be to minimize the likelihood of estimating. This suggests that a future study might usefully have respondents think aloud when completing NAOMS questionnaires with varying recall intervals, so we can assess the extent of counting vs. estimating and identify the longest recall interval that allows for minimization of estimation.

Conclusions

Based on the results of a review of the available literature and the results of the study we conducted, we recommend that we explore the possibility of releasing separate representative samples of respondents each week, and that respondents be asked to recall events that occurred only during the seven days prior to their interview or questionnaire completion.

Of course, we will not be able to complete data collection from every respondent during the week of his or her field period. So we will contact some people after the first week of their field period. This means that during the first week of the study, only easy-to-reach respondents will report data to us. The more difficult-to-reach respondents will not provide data until a week or two later. But during those subsequent weeks, easy-to-interview respondents will also be interviewed. This means that during each week after the first week, a mix of easy-to-reach and difficult-to-reach respondents will provide data. This means that there will be a minimal correlation between difficulty of reaching a person and the reference period for which he or she provides recollections. This will assure maximum data validity and reliability, which will allow us to obtain an adequate level of data precision from interviewing fewer people than would be required if there were substantially more measurement error in respondents' recollections.

However, there is a drawback to maximizing recall accuracy in this way: increased necessary sample size and financial cost of data collection. This is so because many events of interest to the NAOMS project are presumed to occur quite rarely. Furthermore, given the nature of commercial pilots' flying schedules, a substantial fraction (perhaps even as many as one-half) of our respondents do not fly during any given week. This further increases the risk of interviewing respondents who report witnessing no safety-related events if the recall interval is only one week. Therefore, to accurately measure the frequencies of rare events, we would need to substantially increase our sample size if the recall interval is made substantially shorter. But almost no pilots do not fly at all during a one-month period, so this somewhat longer recall interval is much more desirable on these grounds.

Therefore, our final decision about recall interval will need to be informed by the particular list of events chosen for inclusion in the final NAOMS questionnaire and the rates at which these events are witnessed by pilots.

Appendix E: Footnotes

[1] The study by Burton and Blair, 1991, addressed occurrence accuracy, but is not included in the table because it focused on raw errors in the number of checks written and not the percent of accurate responses or percent of respondents recalling accurately. Although they cannot be included in the table, the error-related results of this study are mentioned in the text.

[2] The study by Baddeley et al. (1978) also measured signed dating error, but is not included in Table 4 because it did not include data for specific recall intervals. The error-related results that were reported are mentioned in the text.

[3] The study by Yarmey and Morris, 1998, addressed characteristics accuracy but is not included in the table because results were published as raw numbers from which percentages could not be computed. Other results from this study are mentioned in the text.

[4] Although it does not fit the criteria to be included in the review of the effects of time, the findings of Richman and Quiñones (1996) are potentially important to our current goals. In an experimental design, these researchers assessed recall immediately after their participants had either gained more or less experience at a novel activity (practicing three times or only one time constructing models in a production simulation). Importantly, they found, perhaps counterintuitively, that participants with less experience were more accurate than were participants with more experience in their recall of the number of times (events) several tasks had been performed during the simulation. Low-experience participants reported frequencies that deviated on average 1.9 events from reality while higher-experience participants reported frequencies that deviated on average 2.35 events from reality, a statistically significant difference. Because no studies in the review explored experience level, and since the level of experience in this study is actually quite low overall, it isn't clear how these findings might apply to experience levels of aviation personnel, but it may be worth investigating.

Appendix E: References

Baddeley, A. D., Lewis, V., & Nimmo-Smith, I. (1978). When did you last….? In M. M. Gruneberg, P. E. Morris, and R. N. Sykes (Eds.) *Practical aspects of memory*. London: Academic Press.

Bell, C. (1992). Memory for an early school report. In M. A. Conway, D. C. Rubin, H. Spinnler, and W. A. Wagenaar (Eds.) *Theoretical perspectives on autobiographical memory*. Boston: Kluwer.

Betz, A. L., & Skowronski, J. J. (1997). Self-events and other-events: Temporal dating and event memory. *Memory and Cognition, 25,* 701-714.

Brewer, W. F. (1988). Memory for randomly sampled autobiographical events. In U. Neisser, and F. Winograd (Eds.*) Remembering reconsidered: Ecological and traditional approaches to the study of memory*. New York: Cambridge University.

Burton, S. & Blair, E. (1991). Task conditions, response formulation processes, and response accuracy for behavioral frequency questions in surveys. *Public Opinion Quarterly, 55,* 50-79.

Cohen, G. (1996). *Memory in the real world*. East Sussex: Psychology Press.

Christianson, S-A. (1992). Emotional stress and eyewitness memory: A critical review. *Psychological Bulletin, 112,* 284-309.

Czaja, R., Blair, J., Bickart, B., & Eastman, E. (1994). Respondent strategies for recall of crime victimization incidents. *Journal of Official Statistics, 10,* 257-276.

Ebbesen, E. B., & Rienick, C. B. (1998). Retention interval and eyewitness memory for events and personal identifying attributes. *Journal of Applied Psychology, 83,* 745-762.

Gottfredson, M. R., and Hindelang, M. J. (1977). A consideration of telescoping and memory decay biases in victimization surveys. *Journal of Criminal Justice, 5,* 205-216.

Hagburg, E. C. (1968). Validity of questionnaire data: Reported and observed attendance in an adult education program. *Public Opinion Quarterly, 32,* 453-456.

Klein, J. D., Graff, C. A., Santelli, J. S., Hedberg, V. A., Allan, M. J., & Elster, A. B. (1999). Adolescent self-reports of preventive services. *Health Services Research, 34,* 391-404.

Linton, M. (1975). Memory for real-world events. In D. A. Norman & D. E. Rummelhart (Eds.) *Explorations in cognition*. San Francisco: Freeman.

Mathiowetz, N. A., & Duncan, G. J. (1988). Out of work, out of mind: Response errors in retrospective reports of unemployment. *Journal of Business and Economic Statistics, 6,* 221-229.

Morwitz, V. G. (1997). It seems like only yesterday: The nature and consequences of telescoping errors in marketing research. *Journal of Consumer Psychology, 6,* 1997.

Richman, W. L., & Quiñones, M. A. (1996). Task frequency rating accuracy: The effect of task engagement and experience. *Journal of Applied Psychology, 81,* 512-524.

Rubin, D. C., & Baddeley, A. D. (1989). Telescoping is not time compression: A model of the dating of autobiographical events. *Memory and Cognition, 17,* 653-661.

Skowronski, J. J., Betz, A., & Thompson, C. P., & Shannon, L. (1991). Social memory in everyday life: Recall of self-events and other-events. *Journal of Personality and Social Psychology, 60,* 831-843.

Skowronski, J. J., Betz, A., & Thompson, C. P., & Larsen, S. F. (1995). Long-term performance in autobiographical event dating: Patterns of accuracy and error across a two-and-a-half year time span. In A. L. Healy & L. B. Bourne (Eds.), *Acquisition and long-term retention of knowledge and skills: The durability and specificity of cognitive procedures*. Newbury Park, CA: Sage.

Smith, F. F. (1935). The direct validation of questionnaire data. *Educational Administration and Supervision, 21,* 561-575.

Sudman, S. & Bradburn, N. M. (1974). *Response effects in surveys: A review and synthesis*. Chicago: Aldine.

Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco: Jossey-Bass.

Thompson, C. P., Skowronski, J. J., & Betz, A. (1993). The use of partial temporal information in dating personal events. *Memory and Cognition, 21,* 352-360.

Thompson, C. P., Skowronski, J. J., Larsen, S. F., & Betz, A. (1996). *Autobiographical memory: Remembering what and remembering when*. Mahwah, N.J.: Erlbaum.

Thompson, C. P., Skowronski, J. J., & Lee, D. J. (1988). Telescoping n dating naturally occurring events. *Memory and Cognition, 16,* 461-468.

Wagenaar, W. A. (1986). My memory: A study of autobiographical memory over six years. *Cognitive Psychology, 18,* 225-252.

Yarmey, A. D., & Morris, S. (1998). The effects of discussion on eyewitness memory. *Journal of Applied Social Psychology, 28,* 1637-1648.

Yuille, J. C., & Cutshall, J. L. (1986). A case study of eyewitness memory of a crime. *Journal of Applied Psychology, 71,* 291-301.

Table E-1.
Attributes of the Studies Reviewed

Study	Study Type	Respondent Population	Occurrence Recall Analyzed[1]	Occurrence Recalled	Characteristics Recall Analyzed[2]	Characteristics Recalled	Date Recall Analyzed[3]
Baddeley et al. (1978)	Recall task with record check	Experimental panel volunteers			+	Details of previous experimental sessions	+
Bell (1992)	Recall task with record check	Adolescent students			+	Academic performance from when 12 years old	
Betz & Skowronski (1997)	Diary with recall task	Undergraduates					+
Brewer (1988)	Random event sampling with recall task	Undergraduates			+	Life events (e.g., studying, eating)	
Burton & Blair (1991, Study 2)	Recall task with record check	Adult bank customers	+	Number of checks written			
Czaja et al. (1994)	Recall task with record check	Adult crime victims			+	Details of real crimes against them	+
Ebbesen & Rienick (1998)	Field experiment with follow-up recall task	Adults			+	Details of stories read and heard	
Hagburg (1968)	Recall task with record check	Adult union members	+	Class attendance frequency			
Klein et al. (1999)	Recall task with record check	Adolescents & adults			+	Details of clinic wellness checkup	+
Linton (1975)	Long-term diary with follow-up recall tasks	One memory researcher					+
Mathiowitz & Duncan (1988)	Recall task with record check	Adults	+	Unemployment frequency			
Rubin & Baddeley (1989, Study 1)	Recall task with record check	Research psychologists			+	Details of visiting speakers' presentations	+
Skowronski et al. (1991)	Diary with recall task	Undergraduates					+
Skowronski et al. (1995)	Long-term diary with recall task	Undergraduates					+
Smith (1935)	Recall task with record check	Male high school students			+	Details of athletic performance	
Thompson et al. (1988, Study 1)	Diary with recall task	Undergraduates					+
Thompson et al. (1988, Study 2)	Diary with recall task	Undergraduates					+
Thompson et al. (1988, Study 3)	Diary with recall task	Undergraduates					+
Thompson et al. (1988, Study 4)	Diary with recall task	Undergraduates					+
Thompson et al. (1993)	Diary with recall task	Undergraduates					+
Wagenaar (1986)	Long-term diary with follow-up recall tasks	One memory researcher			+	Life events (e.g., visiting friends, eating)	+
Yarmey & Morris (1998)	Experiment with recall task	Female undergraduates			+	Details of a simulated bank robbery	
Yuille & Cutshall (1986)	Case study of crime victims	Adolescent & adult crime witnesses			+	Details of a real robbery & gun battle	

[1] This column identifies studies where occurrence recall accuracy was reported in enough detail to be analyzed in this review.
[2] This column identifies studies where characteristics of recall were reported in enough detail to be analyzed in this review.
[3] This column identifies studies where dating accuracy data were reported in enough detail to be analyzed in this review.

Table E-2.
Occurrence Recall Accuracy for Recall Intervals up to One Year

		Recall Interval																			
Study	Specific Measure	0 Days	1 Day	2 Days	7 Days	14 Days	21 Days	30 Days	37 Days	45 Days	49 Days	56 Days	60 Days	67 Days	72 Days	90 Days	180 Days	4 Months	6 Months	9 Months	1 Year
Hagburg (1968)[1]	Frequency of Attendance												48%					53%			
Mathiowitz & Duncan (1988)[2]	Frequency of Unemployment							100%					93%			86%	73%	63%	55%	80%	

[1] Figures in this row are the mean percent of respondents responding correctly.
[2] Figures in this row are the mean percent of correctly-reported occurrences across respondents who experienced at least one unemployment spell during the previous 30-month period.

Table E-3.
Absolute Dating Error (in Days) for Recall Intervals up to One Year

	Recall Interval[1]																			
Study	0 Days	1 Day	2 Days	7 Days	14 Days	21 Days	30 Days	37 Days	45 Days	49 Days	56 Days	60 Days	67 Days	72 Days	90 Days	180 Days	4 Months	6 Months	9 Months	1 Year
Czaja et al. (1994)																		27.0	25.4	
Linton (1975)[2]			0	.8	2.5		6.5		11						12			10.5		
Rubin & Baddeley (1989, Study 1)														50				100		
Skowronski et al. (1991)[3]			2.1	3.3	4.1	8.3	8.5		11			12.6		11						
Skowronski et al. (1995)							91.7					18.3			45.8		38	50	70	62

[1] The exact interval is reported in the table whenever it was reported, but average intervals are reported when ranges of intervals were combined for analysis by the original investigators.
[2] Figures in this row are rough point estimates based on the ranges of days reported in the original paper: "7", "14", "30", "45", and "90" days, and "6 months" are reported in the original paper as 5-8, 9-16, 17-32, 33-64, 65-128, and 129-256 days, respectively.
[3] Values for 7 days, 21 days, and 45 days are the means of the even-numbered days immediately before and after those days.

Table E-4.
Signed Dating Error (in Days) for Recall Intervals up to One Year

Study	Recall Interval[1]																			
	0 Days	1 Day	2 Days	7 Days	14 Days	21 Days	30 Days	37 Days	45 Days	49 Days	56 Days	60 Days	67 Days	72 Days	90 Days	180 Days	4 Months	6 Months	9 Months	1 Year
Betz & Skowronski (1997)[2]		4.6	4.5	3.8	3.9	3.8	3.3		-.5			-8.0		-10.0						
Czaja et al. (1994)																		6.5	-10.5	
Rubin & Baddeley (1989, Study 1)														45				51		
Skowronski et al. (1995)							83.3					12.5			33.3		31.3	31.3	48.0	22.0
Thompson et al. (1988, Study 1)				.4		1.7		.4		.6	-3.6		-6.8							
Thompson et al. (1988, Study 2)				0		.2		-.1		-1.4	-2.7		-4.0							
Thompson et al. (1988, Study 3)				.5		1.2		.39		-.9	-4.7		-6.1							
Thompson et al. (1988, Study 4)[3]				.5		1.3		.44		-3.5	-5.9									
Thompson et al. (1993)[4] – 10-Week Sample			0	.7	5.4	1.9	.9		-4.6			-6.3								
Thompson et al. (1993)[4] – 14-Week Sample			1	.8	2.0	3.4	0		1.5			-2.5		-3.8	-11.9					

[1] The exact interval is reported in the table whenever it was reported, but average intervals are reported when ranges of intervals were combined for analysis by the original investigators.
[2] These values are from running five-day averages reported in the original paper; raw averages were not reported.
[3] Only recall data for events the participants experienced are reported here, not data for events the participants heard about.
[4] Values in this study for even-numbered intervals (e.g., "14 days") are means of two intervals (e.g., 13 and 14 days), and are presented as they were reported in the original paper; values for odd-numbered intervals (e.g., "7 days") are means of the values for the closest even-numbered intervals (e.g., the values for 6 and 8 days).

Table E-5.
Accuracy of Event Characteristics Recall (Other Than Date) for Recall Intervals up to One Year

		Recall Interval[1]																			
Study	Characteristics Recalled	0 Days	1 Day	2 Days	7 Days	14 Days	21 Days	30 Days	37 Days	45 Days	49 Days	56 Days	60 Days	67 Days	72 Days	90 Days	180 Days	4 Months	6 Months	9 Months	1 Year
Baddeley et al. (1978)*[2]	Length of Session												81%				82%				
	Number of Experimenters												100%				77%				
	Sex of Experimenters												100%				58%				
	General Description of Experimenters												92%				54%				
	General Description of Experiments												97%				59%				
Bell (1992)*	Exam Grades														46%						30%
	Effort Grades														85%						72%
	Teacher Comments														27%						34%
Brewer (1988)*	Actions Given Location Cue				33%			31%					32%								
	Locations Given Action Cue				97%			98%					92%								
Czaja et al. (1994)**	Details of Crime (articles taken, weapon used, etc.)																68%			65%	
Ebbesen & Rienick (1998)*[3]	Details of a Story (type of car mentioned, etc.)[4]		16%		11%			5%													
	Details of a Story (type of car mentioned, etc.)[5]		16%					14%													
	Details of a Story (type of car mentioned, etc.)[6]				11%			11%													
Klein et al. (1999)**	Site of Care						94%												94%		
	Clinician Seen						84%												84%		
Rubin & Baddeley (1989, St. 1)*	Details of Visitors' Talks (speaker's name, topic, etc.)							38%								28%		24%	20%		10%
Smith (1935)**	Best High Jump Score																		26%		21%
	Best Broad Jump Score																		16%		9%
	Best Pull-Up Score																		43%		46%
	Best 100-Yard Dash Score																		29%		24%
Wagenaar (1986)**[8]	Wide Variety of Life Events																		50%		32%
Yuille & Cutshall (1986)*[9]	Descriptions of Actions (e.g., turning and shooting)	82%																82%			
	Descriptions of People (e.g., tall, wore a yellow sweater)	76%																73%			
	Descriptions of Objects (e.g., vehicle color)	89%																85%			

[1] The exact interval is reported in the table whenever it was reported, but average intervals are reported when ranges of intervals were combined for analysis by the original investigators.
* Figures in this row are the mean percent of details recalled correctly.
[2] This study reported only ranges of elapsed time. So "60 days" actually includes recall from 0-60 days in the past, "180 days" includes recall from 121-180 days past.
** Figures in this row are the mean percent of respondents responding correctly.
[3] The "30 day" value actually refers to 28 days.
[4] This study involved a between-subjects design.
[5] This study involved a within-subjects design, with recall being measured both at 1 and 28 days.
[6] This study involved a within-subjects, with recall being measured at 7 and 28 days.
[8] These values are for recall after respondents were given one cue.
[9] "4 months" values are actually 4-5 months.

Appendix F. Pilot Demographics: Distribution of Aviation Operational Personnel and Aircraft in the U.S. and Territories

F1. Introduction

An investigation was undertaken to determine the number and distribution of aviation operational personnel (i.e., pilots, air traffic controllers, flight attendants, mechanics, and dispatchers) working in commercial aviation organizations within the United States as background for the development of NAOMS.

Information for this study was derived from the following sources: McGraw-Hill World Aviation Directory (Winter 1999 edition), The Aviation & Aerospace Almanac 2000, Federal Aviation Administration (FAA), Air Transport Association (ATA), Regional Airline Association (RAA), National Air Traffic Controllers Association (NATCA), Aircraft Owners and Pilots Association (AOPA), Air Line Pilots Association (ALPA), Association of Flight Attendants (AFA), National Business Aviation Association (NBAA), American Federation of Labor-Congress of Industrial Organizations (AFL-CIO), and the Teamsters. In addition, individual aircraft operators were also queried by telephone for missing data.

While Battelle made every effort to ensure that the information included in the report they prepared on this study was current, the accuracy of the information depended on the accuracy of the sources used to compile the results.[17] The results on all of the constituents are presented in Battelle Reference Report (Battelle (2007). This appendix focuses primarily on the information relevant to pilots that was extracted from Battelle's report.

F2. Findings

F2.1 Operational Personnel

Pilots compose the largest group of aviation operational personnel. Mechanics and repairmen represent the second largest grouping, with approximately 33% of the total. Flight attendants, air traffic controllers, and dispatchers represent the smallest groups, with approximately 12% of the total when combined. Table F-1 lists the distribution of these groups. Each operational group has subgroups and distributions unique to its population. The balance of this paper describes these groups and their distributions in greater detail, as well as major aircraft operators and subgroups.

Table F-1. Distribution of Aviation Operational Personnel

Group	*Total*	*% Total*	*Date of Reference*
Active pilots	618,298	55.0%	1998
Mechanics/repairmen	365,484	32.4%	1998
Flight attendants	119,533	10.4%	2000
Air traffic controllers	14,832	1.2%	2000
Dispatchers	11,460	1.0%	1999
Total	**1,129,607**	**100%**	—

[17] With many references, there is a lag between data collection and publication of those data.

F2.1.1. Pilots and Operators

According to the FAA, there were 618,298 active pilots in the civil aviation community in 1998.[18] The types of pilot licenses (often called certificates) include the following: Student, Recreational, Private, Commercial, and Airline Transport Pilot.[19] Table F-2 presents the distribution of pilots by age and by the type of pilot certificate held.

Table F-2. Distribution of Pilots by Age and Certificate Type

Age	*Total*	*Student*	*Recrea-tional*	*Private*	*Commercial**	*ATP***	*Rotorcraft Only***	*Glider Only***	*Flight Instructor**
14-15	*220*	220	0	0	0	0	0	*0*	0
16-19	14,087	10,495	3	3,234	243	0	32	80	57
20-24	36,205	13,427	5	13,430	8,523	313	304	203	4,359
25-29	54,300	13,799	13	16,480	17,122	5,688	904	294	11,516
30-34	68,330	12,964	13	22,866	15,133	15,460	1,289	605	10,462
35-39	82,494	12,914	34	32,006	13,679	21,578	1,241	1,042	9,536
40-44	86,772	10,829	38	39,048	13,464	20,536	1,065	1,792	9,260
45-49	83,012	8,101	30	36,590	14,227	21,144	1,159	1,761	9,590
50-54	70,017	5,265	35	29,929	14,397	18,417	630	1,344	8,594
55-59	48,794	2,994	25	21,060	10,448	13,205	219	843	5,788
60-64	33,362	1,809	17	15,547	8,734	6,538	60	657	4,235
65-69	24,054	1,296	32	13,378	5,982	2,543	41	782	2,573
70-74	13,697	572	16	7,382	4,433	1,274	15	5	1,540
74-79	5,522	188	3	2,416	2,258	652	2	3	818
80 and over	1,395	74	1	636	544	138	0	2	223
Totals	622,261	94,947	265	254,002	129,187	127,486	6,961	9,413	78,551
% of Total	100%	15.30%	0.04%	41%	21%	21%	1.10%	1.50%	14.40%

(Lampl, R., Editor. The Aviation & Aerospace Almanac Edition 2000, New York: McGraw-Hill)

* Includes pilots with an airplane only certificate; also includes those with an airplane and a helicopter and/or glider certificate.

** Glider pilots are not required to have a medical examination; however, totals above represent pilots who received a medical examination.

*** Not included in total active pilots, since a flight instructor rating is added to an existing pilot license.

[18] A pilot needs a current medical certificate for his/her pilot license to be valid. It is estimated that there are approximately 600,000 individuals in the United States with pilot licenses who have current medical certificates. The presence of a valid medical certification likely indicates an active pilot.

[19] U.S. Civil Airmen: Statistics pertaining to both pilot and non-pilot airmen were obtained from the official certification records (Civil Airmen Registry) maintained by the Airmen Certification and Medical Certification Branches of the FAA's Mike Monroney Aeronautical Center in Oklahoma City, OK.

Student Pilots: Requirements for a student pilot certificate are minimal. One must have a current FAA third-class medical certificate (good for 24 months before medical re-certification is required). The minimum age for obtaining a student pilot certificate is 14, but one cannot be certified by the FAA as a pilot until the age of 16. There is no time or upper-age limit for an individual taking flying lessons from an FAA-certified flight instructor (CFI).[20]

Recreational Pilots: A person must be at least 17 years of age to be eligible for a recreational pilot certificate. A recreational pilot may fly no more than one passenger in a light, single-engine aircraft fitted with four or fewer seats. In addition, a recreational pilot is restricted to flying in good-visibility conditions during daylight hours. A recreational pilot is also restricted from carrying passengers for hire and may fly no further than 50 miles from the home airport.[21]

Private Pilots: To obtain a private pilot certificate, one must be at least 17 years of age and hold a third-class medical certificate. In addition, one must pass an approved FAA ground and flight examination. A minimum of 40 hours of flight experience (with an instructor and as a solo student pilot) is required before a private pilot applicant can obtain his/her license. Private pilots may fly at night, carry more than one passenger, and fly in poor-visibility conditions if they are appropriately trained and have an instrument rating.[22] However, a private pilot certificate does not allow a pilot to carry passengers or cargo for hire.[23]

Commercial Pilots: To obtain a Commercial Pilot certificate, one must be at least 18 years of age, have a minimum of 250 hours of flight time, pass an FAA written examination and flight check, and hold a second-class medical certificate (requires medical re-certification every 12 months). An FAA Commercial Pilot certificate allows a pilot to carry passengers or cargo for hire and is the minimum certificate required to be hired as a pilot for an airline or air taxi service. An individual with a Commercial Pilot certificate can serve as pilot in command (PIC) of an air taxi, but cannot serve as PIC of a commuter or air carrier aircraft operating under Part 121 of the Federal Aviation Regulations. An airline transport pilot (ATP) certificate is required under these regulations.[24] Commercial Pilots can act as second in command (SIC) in all Part 135 and 121operations. Commercial Pilots who do not have an instrument rating are limited to flying passengers for hire only in good weather during the daytime.

---

[20] Title 14, Code of Federal Regulations (CFR), Chapter I, Federal Aviation Regulations, Part 61, Subpart C, Student Pilots (1999), Washington, DC: U.S. Government Printing Office.

[21] Title 14, CFR, Chapter I, Federal Aviation Regulations, Part 61, Subpart D, Recreational Pilots (1999), Washington, DC: U.S. Government Printing Office.

[22] An instrument rating is added to private and commercial pilot certificates when a pilot has passed a written, practical, and oral examination that demonstrates that he/she has the ability to control the aircraft in poor-visibility conditions solely by reference to aircraft instrumentation. A pilot is required to have a private or commercial pilot certificate and at least 40 hours of actual or simulated experience flying solely by reference to instruments before he/she is eligible to obtain an instrument rating. Developing the skills required to control the aircraft by reference to instruments requires additional training.

[23] Title 14, CFR, Chapter I, Federal Aviation Regulations, Part 61, Subpart E, Private Pilots (1999), Washington, DC: U.S. Government Printing Office.

[24] Title 14, CFR, Chapter I, Federal Aviation Regulations, Part 61, Subpart F, Commercial Pilots (1999), Washington, DC: U.S. Government Printing Office.

Airline Transport Pilots: An ATP certificate is the highest pilot rating given by the FAA. Every PIC of a commercial aircraft operating under Part 121 and many turbojet aircraft operating under Part 135 of the Federal Aviation Regulations is required to have an ATP rating. The applicant for an ATP certificate must be at least 23 years of age and hold both a commercial pilot certificate and an instrument rating. Total flight experience required for an ATP in a rotorcraft is 1200 flight hours, while 1500 hours are required for an ATP airplane rating. For the ATP pilot license to remain valid, a pilot is required to hold a first-class medical certificate (medical re-certification required every 6 months).[25]

Flight Instructors: To be eligible for a Certified Flight Instructor (CFI) certificate, one must be at least 18 years of age, have a commercial pilot certificate, be instrument-rated, and pass an FAA written examination and an FAA flight check.[26]

F2.2 Aircraft Operators

F2.2.1 Major Air Carriers

The major air carriers are defined as those with gross revenues of more than $1 billion per year. They account for the vast majority of the revenue passenger miles flown each year.[27] The major carriers operate approximately 3,700 aircraft and employ more than 441,000 people, including 54,000 pilots. (See Table F-3.)

Table F-3. Major Air Carriers

Name of Airline	*# of Employees*	*# of Pilots*	*# of Aircraft*	*Pilot Union or Association*
Alaska Airlines	8,596	1,260	85	Air Line Pilots Association
America West Airlines	11,494	1,676	111	Air Line Pilots Association
American Airlines	87,190	9,600	672	Allied Pilots Association
Continental Airlines	34,982	4,097	435	International Association of Continental Pilots
Delta Airlines	68,889	9,495	562	Air Line Pilots Association
Northwest Airlines	47,998	6,305	415	Air Line Pilots Association
Southwest Airlines	27,675	3,400	288	Southwest Airline Pilots Association
TWA Airlines	24,008	2,443	190	Air Line Pilots Association
United Airlines	88,887	10,139	572	Air Line Pilots Association
US Airways	42,104	5,897	393	Air Line Pilots Association
Total	441,823	54,312	3,723	

[25] Title 14, CFR, Chapter I, Federal Aviation Regulations, Part 61, Subpart G, Airline Transport Pilots, (1999), Washington, DC: U.S. Government Printing Office.

[26] Title 14, CFR, Chapter I, Federal Aviation Regulations, Part 61, Subpart H, Flight Instructors, (1999), Washington, DC: U.S. Government Printing Office.

[27] Air Line Pilots Association, Herndon, VA. Lampl, R., Editor, The Aviation & Aerospace Almanac 2000, New York: McGraw-Hill. Weimer, Kent J., Editor. World Aviation Directory, Winter 1999, New York: McGraw-Hill.

F2.2.2 National Air Carriers

The national carriers are defined as those with gross revenues between $100 million and $1 billion annually.[28] The national carriers operate approximately 998 aircraft and employ more than 56,000 personnel, including 9,281 pilots. Attachment F-1 lists the national air carriers with numbers of employees and aircraft.

F2.2.3 Regional Air Carriers

Large regional air carriers have annual gross revenues between $10 million and $99.9 million. A medium regional is a carrier that has annual gross revenues of less than $10 million.[29] Regional air carriers operate 1,781 aircraft and employ more than 40,000 personnel, including 13,323 pilots. These carriers operate smaller airplanes than the major and national airlines and are often referred to as commuter airlines. Attachment F-2 lists the regional air carriers with numbers of employees and aircraft.

F2.2.4 Air Cargo Carriers

The air cargo carriers are a fast-growing segment of the commercial aviation industry.[30] Freight carriers employ approximately 292,000 personnel, including 14,106 pilots, and operate 1,585 aircraft. Attachment F-3 lists airlines that specialize in transporting freight.

F2.2.5 Charter and Non-Scheduled Carriers

The charter and non-scheduled air services, which include air taxi and contract services, are normally small operations that provide on-demand air service to isolated communities around the country. They operate more than 2,000 aircraft and employ more than 17,000 personnel, including 7,179 pilots.[31] Attachment F-4 lists non-scheduled and charter operators.

F2.2.6 General Aviation

General aviation is the largest segment of the aviation industry. Although there is no legal definition of general aviation, it is commonly described as "all civil aviation except that carried out by the commercial airlines or the military." There are more than 183,000 active general aviation aircraft. This number represents 98 percent of the total aircraft in the United States.[32] General aviation also includes a variety of aircraft, including airplanes, helicopters, and gliders.

General aviation aircraft are used for a variety of purposes. According to the National Business Aviation Association, 5,000 U.S. companies have corporate flight departments operating more

---

[28] Air Line Pilots Association, Herndon, VA. Lampl, R., Editor, The Aviation & Aerospace Almanac 2000, New York: McGraw-Hill. Weimer, Kent J., Editor. World Aviation Directory, Winter 1999, New York: McGraw-Hill.

[29] Air Line Pilots Association, Herndon, VA. Weimer, Kent J., Editor, World Aviation Directory, Winter 1999, New York: McGraw-Hill. Lampl, R., Editor. The Aviation & Aerospace Almanac 2000, New York: McGraw-Hill.

[30] Weimer, Kent J., Editor. World Aviation Directory, Winter 1999, New York: McGraw-Hill.

[31] Weimer, Kent J., World Aviation Directory Winter 1999, New York: McGraw-Hill.

[32] Federal Aviation Administration Statistical Handbook 1999, Washington, DC: U.S. Government Printing Office; Aircraft Owners and Pilots Association, Frederick, MD.

than 10,000 aircraft and employing 20,000 pilots.[33] General aviation also includes private flying for pleasure and business, flight instruction, aerial application, aerial observation, photography, fire fighting, police traffic control, and pipeline/powerline surveillance. Table F-4 provides the distribution of general aviation aircraft.[34]

Table F-4. General Aviation Aircraft Distribution

Type	*# of Aircraft*	*% of Total*	*# of Hours Flown*	*% Total*
Corporate	9,652	5.60%	2,548,000	11.00%
Business	25,554	14.90%	3,055,000	13.00%
Instructional	14,568	8.54%	4,156,000	17.50%
Personal	100,839	59.10%	8,116,000	34.20%
Aerial Application	4,215	2.47%	1,210,000	5.10%
Aerial Observation	4,936	2.90%	1,750,000	7.40%
External Load	133	0.08%	172,000	0.75%
Other Work	1,214	0.71%	226,000	0.95%
Air Taxi	3,927	2.30%	1,670,000	7.10%
Sightseeing	1,336	0.78%	323,000	1.40%
Other	4,226	2.48%	640,000	0.30%
Total	170,600		23,866,000	

F2.2.7 Helicopter Air Service

Helicopters provide a variety of services, including air ambulance, pipeline and power line surveys, fire fighting, and police and media reporting. Commercial helicopter operators employ more than 5,000 personnel, including more than 1,700 pilots and operate 1,061 aircraft. Attachment F-5 lists commercial helicopter operators.[35]

[33] National Business Aviation Association Source Book on Aviation 1998, Washington, DC.

[34] U.S. Department of Transportation, the Federal Aviation Administration, the Aircraft Owners and Pilots Association and the National Business Aviation Association. (Both airplanes and helicopters are included in these summary statistics.)

[35] Lampl, R., Editor. The Aviation & Aerospace Almanac 2000, New York: McGraw-Hill. Weimer, Kent J., Editor. World Aviation Directory, Winter 1999, New York: McGraw-Hill.

Attachment F-1. National Airlines[36]

Airline	*State*	*# of Employees*	*# of Pilots*	*# of Aircraft*
Air Tran	GA	*3,500*	*380*	*37*
Air Transport	FL	520	101	30
Air Wisconsin	WI	800	312	28
Aloha	HI	2,249	206	17
American Trans Air	IN	6,000	981	71
Atlantic Southeast	GA	2,762	756	88
Carnival	FL	1,220	234	22
Continental Express	FL	1,820	768	96
Continental Micronesia	GU	2,000	200	19
Hawaiian	HI	2,400	288	23
Horizon Air	CO	3,100	510	62
Mesa	AZ	1,450	1,134	183
Midway	IL	1,000	309	28
Midwest Express	WI	2,223	337	27
Reno Air	NV	2,500	286	30
Simmons	TX	4,400	840	86
Sun Country	AZ	2,010	208	19
Tower	NY	11,800	219	17
Trans States	CA	2,000	682	74
U.S. Airways Shuttle	VA	650	200	12
Western Pacific	CO	1,155	180	17
World	VA	725	150	12
Totals		*56,284*	*9,281*	*998*

[36] Air Line Pilots Association, Herndon, VA. Lampl, R., Editor. The Aviation & Aerospace Almanac 2000, New York: McGraw-Hill. Weimer, Kent J., Editor. World Aviation Directory, Winter 1999, New York: McGraw-Hill.

Attachment F-2. Regional Air Carriers[37]

Name	*State*	*# of Employees*	*# of AC*	*# of Pilots*
Air Midwest/Mesa Air Group	KS	*225*	*12*	*120*
Air Nevada	NV	88	8	16
Air South	SC	550	7	80
Air Sunshine	FL	22	7	12
Alaska Island Air	AK	12	2	4
Alaska Juneau Aeronautics	AK	70	20	27
Allegheny	PA	1,200	41	400
Aloha Islandair	HI	230	6	60
Alpine	UT	55	12	30
American Eagle	CA	1,300	35	350
Arctic Circle Air Service	AK	31	4	20
Arctic Transportation Services	AK	65	15	12
Aspen Mountain Air	TX	250	8	80
Astral Aviation	WI	248	15	150
Atlantic Coast Airlines	VA	1,300	59	590
Atlantic Southeast Airlines	GA	2,762	82	820
Atlantic World Airways	FL	47	3	15
AVI	NV	125	11	50
Aviation Services Ltd./Freedom Air	GU	49	4	12
Baker Aviation	AK	32	5	10
Bemidji Aviation	MN	50	5	15
Bering Air	AK	85	4	16
Big Sky Airlines	MT	75	3	15
Business Express	NH	1,200	39	390
Cape Smythe Air Service	AK	105	8	20
Caribbean Int'l	PR	35	3	12
Casino Express	NV	102	2	15
CCAir	NC	600	26	260
Chautauqua Airlines	IN	320	30	180
Chicago Express	IL	100	10	50
Coastal Air Transport	VI	10	2	4
Colgan Air	VA	140	6	60
Comair Inc.	OH	3,000	96	960
Commutair	NY	340	30	150
Conquest Airlines	TX	141	8	48
Continental Express	TX	1,820	106	1,000

(continued on next page)

[37] Air Line Pilots Association, Herndon, VA. Lampl, R., Editor. The Aviation & Aerospace Almanac 2000, New York: McGraw-Hill. Weimer, Kent J., Editor. World Aviation Directory, Winter 1999, New York: McGraw-Hill.

Attachment F-2. Regional Air Carriers (continued)

Name	*State*	*# of Employees*	*# of AC*	*# of Pilots*
Corporate Air	MT	320	82	109
Corporate Express	TN	150	8	60
East Coast Aviation Services/Executive Airlines	NY	43	3	18
Eastwind Airlines	NC	120	2	20
Empire Airlines	ID	250	48	60
ERA Aviation	AK	160	87	120
Executive Airlines	PR	1,446	18	180
Express Airlines 1	GA	900	58	500
Flagship Airline	TN	2,126	68	680
Flamenco Airways	PR	65	4	20
Flying Boat Inc/ Pan Am Air Bridge	FL	45	5	20
40 Mile Air	AK	25	4	6
Frontier Airlines	CO	700	10	100
Frontier Flying Service	AK	85	4	30
Grand Canyon Airlines	AZ	40	5	25
Great Lakes Airlines	IA	1,400	57	560
Gulfstream Int'l Airlines	FL	500	23	200
Haines Airways	AK	40	7	7
Hyannis Air Service	MA	200	5	50
Island Airlines	MA	53	2	4
Jettrain	PA	120	2	20
Jib/Action Airlines	CT	11	2	4
Kenmore Air Harbor	WA	65	7	30
Ketchikan Air Service	AK	75	3	10
LAB Flying Service	AK	75	2	6
Laker Airways	FL	175	4	40
Larry's Flying Service	AK	60	2	15
Las Vegas Airlines	NV	28	2	8
Mahalo Air	HI	300	7	70
Mesa Air Group	NM	2,500	100	1,134
Mesaba Aviation	MN	1,800	74	750
Mission Aviation Fellowship	CA	295	6	60
Nations Air Express	GA	102	3	30
New England Airlines	RI	15	2	8
Olson & Sons	AK	29	3	10
Pacific Island Aviation	MP	130	7	50
Pan American World Airways	FL	600	4	60
Paradise Island Airlines	FL	125	5	40
Peninsula Airways	AK	350	16	40

(continued on next page)

Attachment F-2. Regional Air Carriers (continued)

Name	*State*	*# of Employees*	*# of AC*	*# of Pilots*
Pennsylvania Aviation	PA	50	2	10
Piedmont Airlines/ Sub US Airways	MD	1,750	49	490
Planet Airways	FL	10	1	6
Prestige Airways	VA	150	5	50
Proair	WA	175	2	20
PSA Airlines	OH	900	25	250
Redwing Airways	MO	7	6	5
Reeve Aleutian Airways	AK	240	5	40
Rich Int'l Airways	FL	1,100	10	100
Samoa Aviation	AS	65	3	15
Skagway Air Service	AK	10	12	10
Skywest Airlines	UT	2,100	69	690
Southcentral Air	AK	28	7	10
Spirit Airlines	MI	400	13	130
Springdale Air Service	AR	24	23	18
Sunshine Airlines	CA	18	7	12
Tanana Air Service	AK	14	8	8
Tatonduk Outfitters	AK	28	5	18
Trans Air	HI	52	5	14
Tristar Airlines	NV	165	1	8
UFS	MO	400	9	90
Vanguard Airlines	MO	568	8	80
Vieques Air Link	VI	53	2	10
Village Aviation	AK	50	8	7
Virgin Air/Air St. Thomas	VI	22	4	8
West Isle Air	WA	26	4	7
Wright Air Service	AK	25	2	6
Yute Air Alaska	AK	180	1	4
Totals		40,917	1,781	13,323

Attachment F-3. Cargo-Only Airlines[38]				
Name	*State*	*# of Employees*	*# of Pilots*	*# of Aircraft*
ABX Air/Airborne Express	OH	6,800	*1,100*	*109*
Air Cargo Carriers	WI	140	60	19
AirPac Airlines	WA	30	6	2
American Int'l Airways	MI	3,000	300	54
Ameriflight	CA	550	120	83
Amerijet	FL	512	150	16
Atlas Air	NY	610	676	22
BankAir	SC	85	40	30
Bax Global	CA	630	407	37
BigHorn Airways	WY	16	6	4
Burlington Air Express	CA	6,300	250	18
Business Air Inc.	VT	40	21	13
Capital Cargo Int'l Airlines	FL	21	6	2
Challenge Air Cargo	FL	800	30	4
Custom Air Transport	FL	90	20	5
DHL Airways	CA	10,000	487	33
Emery Worldwide	CA	10,020	454	62
Evergreen Int'l	OR	475	295	20
Express One Int'l	TX	455	258	27
Federal Express	TN	145,000	5,833	505
Fine Airlines	FL	1,000	120	15
Florida West Int'l Airways	FL	71	11	2
Gemini Air Cargo	DC	250	90	6
IFL Group (Corporate Express)	MI	51	18	8
Kitty Hawk Air Cargo	TX	500	300	46
Merlin Express	TX	130	64	38
Mid-Atlantic Freight	NC	120	35	20
Million Air	FL	150	27	5
Mountain Air Cargo	NC	280	120	69
Northern Air Cargo	AK	220	70	14
Polar Air Cagro	CA	600	184	13
Regional Express	ID	175	12	7
Relient Airlines	MI	150	36	14
Renown Aviation	CA	115	30	5
Suburban Air Freight	NE	60	30	6
UPS Airlines	KY	103,000	2,321	193
USA Jet Airlines Inc.	MI	320	75	25
Westair Inc.	CA	60	44	34
Totals		292,826	*14,106*	*1,585*

[38] Air Line Pilots Association, Herndon, VA. Lampl, R., Editor. The Aviation & Aerospace Almanac 2000, New York: McGraw-Hill. Weimer, Kent J., Editor. World Aviation Directory, Winter 1999, New York: McGraw-Hill.

Attachment F-4. Charter and Non-Scheduled Carriers[39]

Name	*State*	*# of Employees*	*# of AC*	*# of Pilots*
A&M Aviation/CNS	IL	*15*	*12*	*24*
Aberdeen Flying Service	SD	8	1	2
Ace Aerial Service/CNS	CA	3	10	20
ACM Aviation/CNS	CA	40	9	20
Action Airlines/CNS	CT	36	8	16
Adirondack Flying Service/CNS	NY	8	6	8
Aero Air/CNS	OR	75	4	6
Aero Freight/dba Aero Executive/CNS	TX	20	8	16
Aero Tech Flight Service/CNS	AK	30	17	25
Aeroflite/CNS	IL	15	5	10
Agile Air Service/CNS	NH	5	2	6
Air Alpha/CNS	OH	5	1	3
Air America/CNS	MI	15	4	5
Air Cargo Carriers	WI	92	10	25
Air Carriage/CNS	CA	5	4	10
Air Charter of Virginia/CNS	VA	14	5	10
Air Charter Service/CNS	PA	12	3	8
Air Midway/CNS	NE	5	3	4
Air Molokai	HI	30	3	6
Air Nevada Airlines	NV	88	11	33
Air San Luis/CNS	CA	18	4	8
Air Service Int'l/CNS	CA	90	14	22
Air Trek/CNS	FL	12	2	3
Airmotive Enterprises Inc./CNS	MN	10	4	6
Airspect Inc./CNS	OH	5	6	4
Airstar Int'l Airlines/CNS	FL	13	1	4
Alexander Aviation Inc./CNS	MN	5	8	25
Alpine Air Charter/CNS	IL	5	1	2
American Flag Airlines, Inc./CNS	FL	7	2	4
American Flight Services/CNS	DC	12	8	12
Ameristar Jet Charter, Inc/CNS	TX	45	4	18
Archway Aviation Inc./CNS	MO	10	8	6
Aroostook Aviation Inc./CNS	ME	15	4	6
Atlantic Aviation Flight Service Inc./CNS	NJ	35	8	16
Aviation Methods, Inc./NS	CA	300	50	225
Aviation Resources Ltd./Valley Aircraft/CNS	ND	45	7	25

(continued on next page)

[39] Lampl, R., Editor. The Aviation & Aerospace Almanac 2000, New York: McGraw-Hill. Weimer, Kent J., Editor. World Aviation Directory, Winter 1999, New York: McGraw Hill.

Attachment F-4. Charter and Non-Scheduled Carriers (continued)				
Name	*State*	*# of Employees*	*# of AC*	*# of Pilots*
Aviex Jet, Inc./CNS	TX	*24*	*10*	*12*
Baron Enterprises/CNS	OH	12	7	7
Basco Flying Service Inc./CNS	PA	16	9	9
Basler Airlines/CNS	WI	32	7	7
Bay Air Flying Service/CNS	FL	30	4	4
Beaver Aviation Services Inc./CNS	PA	80	3	3
BeckAir Co. Inc./CNS	IN	8	2	2
Bird Air Fleet, Inc./CNS	NH	16	7	7
Blackhawk Air Service/CNS	IL	7	4	4
Bluffton Flying Service Co./CNS	OH	8	10	4
Bowman Aviation Inc./CNS	IN	125	10	6
Bridgeford Flying Service/CNS	CA	30	25	20
Brooks Seaplane Service/CNS	ID	13	1	2
Bullock Charter Inc./CNS	MA	5	1	2
Bun Air Corp./CNS	PA	11	8	8
Business Jetsolutions/CNS	TX	250	14	25
Capital Aviation Corp./CNS	ND	6	5	5
Casper Air Service/CNS	WY	60	10	20
Central Air Service/CNS	MT	3	5	3
Central Air Southwest/CNS	OK	25	4	8
Champion Air/CNS	MN	200	6	30
Channel Islands Aviation/CNS	CA	30	19	22
Charter Jet Int'l/CNS	CO	42	4	10
Chester Country Aviation/CNS	PA	36	3	5
Cheyenne Charter Inc./CNS	IN	7	3	3
Clay Lacy Aviation/CNS	CA	58	12	24
Clintondale Aviation/CNS	NY	75	3	10
Coastal Air Services/CNS	CT	15	11	12
Commercial Aviation Corp./CNS	OH	16	9	8
Condor Enterprise Inc./CNS	IL	8	2	4
Consolated Airways Inc./CNS	IN	11	3	6
Corporate Jets, Inc./CNS	PA	700	35	140
Croporate Airways/CNS	FL	10	7	7
Crossings Aviation/CSN	WA	30	6	14
Crossjet, Inc./CNS	DC	6	2	2
Davisair, Inc./CNS	PA	20	6	14
Deland Aviation/CNS	FL	10	11	4
Denison Aviation, Inc./CSN	IA	7	7	3
Dodson Int'l Air/CNS	GA	22	8	11
Don Davis Aviation/CNS	KY	18	5	6
Downeast Flying Service/CNS	ME	10	1	2

(continued on next page)

Attachment F-4. Charter and Non-Scheduled Carriers (continued)				
Name	*State*	*# of Employees*	*# of AC*	*# of Pilots*
EL Aero Services/CNS	NV	4	13	3
Elmira Aeronautical Corp/CNS	NY	20	1	3
Encore Int'l Airways/CNS	WA	16	2	4
Executive Flight/CNS	WA	10	3	5
Executive Fliteways, Inc./CNS	NY	25	11	15
F.I.T. Aviation/CNS	FL	15	43	10
Falcon Aviation/CNS	SD	6	8	5
Falwell Aviation/CNS	VA	20	7	10
Flight Int'l/CNS	VA	125	21	84
Flight One Inc./CNS	MI	11	3	6
FlightStar Corp./CNS	IL	55	5	20
Gibson Aviation/CNS	MD	12	4	6
Global Air Charter/CNS	FL	65	6	30
Grand Aire Express/CNS	MI	170	24	100
Gunnison Valley Aviation/CNS	CO	10	5	6
Hansen Flying Service/CNS	MI	16	6	8
Hart Enterprises/CNS	ID	7	3	3
Havre Flying Service/CNS	MT	5	1	2
Holman's Transportation Systems/CNS	AL	4	1	2
Hutcherson Air Service/CNS	TX	10	5	5
Iliamna Air Taxi/CNS	AK	12	11	7
International Aviation/CNS	FL	73	5	36
Int'l Jet Aviation Services/CNS	CO	32	12	25
Island Air Charters/CNS	FL	5	3	3
Jaax Flying Service/CNS	CA	3	3	2
Jackson Hole Aviation/CNS	WY	24	5	12
Jet Aviation Int'l/CNS	FL	424	47	225
Jet Charter Inc./CNS	NJ	10	4	4
Jet East, Inc./CNS	TX	86	4	24
Jet Services/CNS	NJ	18	1	4
Jim Air/CNS	AK	8	4	4
Kaiser Air Inc./CNS	CA	130	3	20
Katmai Air/CNS	AK	15	7	6
Kenai Air/CNS	AK	2	4	1
Lake Mead Air/CNS	NV	13	23	10
Lakeland Aviation Co./CNS	WI	7	1	2
Lane Aviation/CNS	OH	162	4	30
Logan & Reavis Air/CNS	OR	9	5	5
Lumanair/CNS	IL	35	2	8
Lynch Flying Service/CNS	MT	55	22	25
Lynstar Aviation/CNS	NJ	30	3	15

(continued on next page)

Attachment F-4. Charter and Non-Scheduled Carriers (continued)				
Name	*State*	*# of Employees*	*# of AC*	*# of Pilots*
Magnus Aviation/CNS	WI	32	6	20
Martin Aviation/CNS	CA	70	10	45
Mayo Aviation/CNS	CO	50	14	28
Meeker Airport/CNS	CO	3	4	2
Miami Air Int'l/CNS	FL	329	5	55
Mid-Coast Air Charter/CNS	TX	10	4	6
Midstate Aviation/CNS	WA	25	15	18
Miller Flying Service/CNS	TX	18	2	6
Mobile Air Center/CNS	AL	45	4	6
Monterey Airplane Co./CNS	CA	9	3	3
Mountain Air Services/CNS	ME	3	7	2
Mountain Bird Inc./CNS	ID	12	9	7
National Jets/CNS	FL	60	7	30
Navajo Aviation/CNS	CA	45	5	20
New Mexico Flying Service/CNS	NM	20	16	10
North American Airlines/CNS	NY	160	3	30
Omni Air Express/CNS	OK	50	3	30
Orco Aviation/CNS	CA	25	4	5
PAB Aviation/CNS	PA	20	3	3
Pacific Flights/CNS	OR	17	4	7
Panama Aviation/CNS	FL	8	5	4
Pensacola Aviation Center/CNS	FL	49	12	15
Phoenix Air/CNS	GA	140	23	98
Prime Airborne/CNS	NY	10	7	6
Pro-Flite of Vero/CNS	FL	54	31	38
Pronghorn Aviation/CNS	CA	3	2	1
Redtail Aviation/CNS	UT	7	9	5
Rhoades Aviation/CNS	IN	65	22	44
Richmor Aviation/CNS	NY	180	20	100
Ross Aviation/CNS	NM	126	6	65
RSVP Jet/CNS	CA	4	1	2
Ryan Int'l Airlines/CNS	KS	698	41	410
Scenic Airlines/CNS-Sub Sky West	AZ	45	22	30
Schaefer Air Service/CNS	CA	250	3	20
Seneca Flight Operations/CNS	NY	22	6	12
Servair/CNS	ND	8	8	6
Sierra Nevada Airways/CNS	NV	18	4	12
Sierra Pacific Airlines/CNS	AZ	30	1	15
Silver Ranch Airpark/CNS	NH	5	2	2
Sky Aviation/CNS	WY	5	7	6
Skybird Aviation/CNS	CA	6	1	2

(continued on next page)

Attachment F-4. Charter and Non-Scheduled Carriers (continued)				
Name	*State*	*# of Employees*	*# of AC*	*# of Pilots*
Southeast Airmotive/CNS	NC	25	15	17
Southwest Aviation/CNS/Midwest Aviation	MN	10	6	5
Sportsflight Airways/CNS	AZ	125	1	15
Star Airlines/CNS	OH	30	2	10
Star Aviation/CNS	SD	9	4	4
Sternair/CNS	TX	20	6	8
Sugarpine Aviatiors/CNS	CA	3	2	2
Summit Aviation/CNS	NT	20	3	12
Sun Pacific Int'l/CNS	AZ	50	2	24
Sunbird Aviation/CNS	MT	20	8	14
Sundance Helicopter/CNS	NV	32	11	12
Superior Aviation/CNS	MI	80	21	38
T.S.P.I./CNS	OK	25	10	15
Taft Air/CNS	NJ	10	3	4
Taquan Air Service/CNS	AK	21	3	12
Telford Aviation/CNS	ME	50	16	37
Thunderbird Airways/CNS	TX	7	12	5
Towle Enterprises/CNS/Twin Air Service	FL	30	4	8
Trans Northern Airways/CNS	FL	25	10	15
Trans-Florida Airlines/CNS	FL	19	5	6
Transit Aviation of Lake Charles/CNS	LA	28	7	14
Tri-Star Aviation/CNS	VA	4	4	3
Tulip City Air Service/CNS	MI	38	5	10
Umiat Enterprises/CNS	AK	7	3	2
Universal Airways/CNS	TX	10	2	2
Vee Neal Aviation/CNS	PA	10	14	8
Victoria Aviation Services/CNS	TX	15	8	8
Viscount Air Services/CNS	AZ	250	4	60
Ward Air/CNS	AK	10	4	6
Wayfarer Aviation/CNS	NY	60	15	30
West Coast Air Charter/CNS	TX	9	4	3
Weyerhaeuser Co. Aviation/CNS	WA	80	16	49
Wiggins Airways/CNS	MA	105	35	70
Wild Blue Yonder/CNS	ID	7	4	6
World Aircraft/Spares Corp./CNS	FL	18	22	15
Wren Air/CNS	AK	7	3	5
Totals		9,004	1,566	3,713

Attachment F-5. Commercial Helicopter Services[40]				
Helicopter Air Service	*State*	*# of Employees*	*# of AC*	*# of Pilots*
Advance Life Support Emergency Rescue Team	MT	4	1	3
Aero-Copters Inc.	WA	6	3	4
Air Logistics of Alaska	AK	44	14	20
Air Methods	CO	250	34	130
Aircoastal Helicopters	FL	6	5	5
Allied Helicopter Service	OK	8	11	5
Arctic Air Service	AK	12	3	4
Arrowhead Helicopters	AZ	2	1	2
Astrocopters	CA	8	5	5
Aviation Services Unlimited	NY	6	2	3
Joe Brigham	NH	5	5	5
Cane Air	LA	5	1	3
Carson Services	PA	75	6	12
Cascade Helicopters	WA	22	9	10
Central Helicopters	MT	5	1	2
Classic Helicopter Corp.	WA	13	10	10
Columbia Helicopters	OR	800	40	70
Crew Concepts	ID	5	3	3
Diamondback Aviation Services	AZ	10	3	5
ERA Helicopters	AK	485	90	110
Erickson Air Crane	OR	60	16	20
Evergreen Helicopters	OR	140	63	80
Evergreen Helicopters of Alaska	AK	9	5	6
Falcon Helicopters	CO	4	3	3
Fetsko Aviation Sales & Transportation	PA	6	2	3
Fly Wright Corp.	WA	5	3	4
Fostaire Helicopters	IL	10	8	6
Geo-Seis Helicopters	CO	75	26	33
Heli-Cab Helicopter Services	TX	5	1	3
Helicopter Consultants of Maui	HI	98	13	20
Helicopter Minit Men	OH	20	4	6
Helicopter Services Inc.	TX	12	13	10
Helicopters Inc.	IL	60	37	43
Heliflight	FL	15	12	13
Heli-Jet Corp.	OR	25	7	12
Helinet Corp.	CA	18	6	9
Helistream Inc.	CA	16	11	11

(continued on next page)

[40] Lampl, R., Editor. The Aviation & Aerospace Almanac 2000, New York: McGraw-Hill. Weimer, Kent J., Editor. World Aviation Directory, Winter 1999, New York: McGraw-Hill.

Attachment F-5. Commercial Helicopter Services (continued)				
Helicopter Air Service	*State*	*# of* Employees	*# of AC*	*# of Pilots*
High Tech Applications	WV	6	2	3
Hillsboro Aviation	OR	65	27	30
Horizon Helicopters	CA	10	3	5
Houston Helicopters	TX	38	29	25
Industrial Helicopters	LA	60	14	20
Interstate Helicopters	OK	3	3	3
Keystone Flight Services	PA	71	18	25
Landells Aviation	CA	8	4	5
Liberty Helicopters	NJ	45	11	14
Maritime Helicopters	AK	9	3	5
McMahon Helicopter Services	MI	12	8	9
Metro Aviation	LA	85	21	25
Metropolitan Helicopter Services	NJ	4	1	2
Mid Valley Helicopters	OR	5	4	4
Midwest Helicopter Airways	IL	15	4	6
Miller-Crestar Helicopters	PA	2	2	2
National Helicopter Service & Engineering Company	CA	14	10	12
New England Helicopter	NY	2	2	2
New York Helicopter	NY	150	4	20
Norcross Helicopter	NY	5	3	3
Papillon Grand Canyon Helicopters	WA	100	20	30
Petroleum Helicopters	LA	1,850	242	642
Redding Air Service	CA	10	4	6
Reforestation Services Inc.	OR	18	1	4
Rogers Helicopters	CA	25	23	23
Royale Helicopter Service	PA	3	3	3
Sacramento Executive Helicopters Inc.	CA	15	4	6
St. Louis Helicopter Airways	MO	70	17	25
Salaika Aviation	TX	15	4	8
San Joaquin Helicopters	CA	90	19	25
Shier Aviation	CA	6	6	6
Sky Helicopters	TX	4	5	4
Skyhawk Helicopter Service	UT	5	2	4
South Sea Helicopter Corp.	HI	25	3	9
Southeast Mississippi Air Ambulance District	MS	7	1	3
Suncoast Helicopters	FL	15	9	12
U.S. Helicopters Inc.	NC	22	17	18
Versatile Helicopters Inc.	OK	9	8	8

(continued on next page)

Attachment F-5. Commercial Helicopter Services (continued)

Helicopter Air Service	*State*	*# of Employees*	*# of AC*	*# of Pilots*
West Florida Helicopters Inc.	FL	9	2	3
Western Helicopter Services	OR	15	8	8
Whirl-Away Helicopters Inc.	IN	10	5	5
Wolfe Air Aviation	CA	6	5	5
Zebra Air Inc.	TX	4	3	3
Totals		5,291	1,061	1,773

Appendix G. NAOMS Air-Carrier-Pilot Survey Sample Size by Individual Survey Item

An analysis was made of sample size requirements for each event addressed in the questions of Section B of the air-carrier-pilot survey using the records of the first 8,000 interviews. This study was done to complement and supplement a similar study that had been conducted based on the data obtained during the field trial. From the field trial, the decision was made to conduct 8,000 interviews/year. The analysis reported here was undertaken to confirm that with approximately 8,000 interviews, NAOMS would be able to detect a shift from year to year of 20% with 95% confidence for a substantial number of the events studied.

The preliminary analysis estimated upper-bound sample sizes by simplifying much of the information collected in the survey. The unit of observation used was the respondent rather than aggregate hours or legs from which a rate might be calculated. We assumed that pilots had equal time 'at risk' for each of the events and that each pilot had an equal chance to be selected in the sample. Also, this simple analysis did not distinguish captains, first officers, or other flight-crew positions or circumstances that might affect the probability of an event being observed.

Sample sizes were calculated to detect year-to-year differences in proportions, i.e. the proportion of pilots reporting equipment problems or the proportion of pilots reporting bird strikes. This approach allowed straightforward statistical techniques to be used, specifically those described in Snedecor and Cochran (1989) (Section 7.13-Sample size for comparing two proportions, pp. 129–133) and Kahn (1989) (Difference Between Binomial Parameters, pp. 27–33).

According to formulas in the two references, only the sample proportions ($\hat{p}$ and 1- $\hat{p}$), the selected Type I and Type II errors, and the desired change to be detected affect the calculations. Sample sizes were calculated to detect a 20% change, in the proportion of pilots reporting each event addressed in items of the survey. Significance levels of 0.05, 0.10, and 0.20 were calculated. These sample sizes are presented as Table G-1, along with the number of pilots giving a response to each item, the number of pilots reporting one or more events, the percent ($\hat{p}$ *100) of pilots reporting the event, and the number of pilots that would involve a 20% change. Each event question is presented in Table G-1 as it appears on the survey questionnaire. Results were obtained for 88 event questions. Of these, for 47 events (53%), NAOMS would be able to detect 20% rate shifts with 95% confidence should such changes occur with 8,000 or fewer interviews per year.

It is believed these calculations represent upper bounds for sample size requirements. They ignore substantial information that the survey collects, such as the number of hours flown during the recall period and the actual number of events reported, that more detailed statistical measures can apply. More precise estimates that reduce variance and increase power may be derived through special applications of generalized linear models such as Poisson regression that can model underlying rates instead of simple characteristics of the survey respondents.

References:

Snedecor, G.W. and Cochran, W.G. 1989 *Statistical Methods*. Iowa State University Press.
Kahn, H.A. and Sempos, C.T. 1989 *Statistical Methods in Epidemiology*. Oxford University Press.

Section B Question	Number of Pilots Reporting	Number of Pilots Reporting 1+ Event	Percent of Pilots Reporting 1+ Event	+/- 20% Change in Number of Pilots Reporting 1+ Event	Sample Size Power=.20		
					Alpha=.05	Alpha=.10	Alpha=.20
ER1	*7996*	*427*	*5.3*	*85.4*	*6998*	*5539*	*4021*
ER2	7998	22	0.3	4.4	143135	113295	82242
ER2A	21	16	76.2	3.2	123	98	71
ER2B	22	2	9.1	0.4	3948	3125	2268
ER2C	19	7	36.8	1.4	677	536	389
ER3	7982	150	1.9	30.0	20614	16317	11844
ER4A	7996	102	1.3	20.4	30555	24185	17556
ER4B	7994	43	0.5	8.6	73002	57783	41945
ER4C	7997	43	0.5	8.6	73030	57805	41961
ER4D	7995	47	0.6	9.4	66764	52846	38361
ER4E	7995	18	0.2	3.6	174964	138490	100530
ER4F	7993	24	0.3	4.8	131092	103763	75322
ER4G	7996	37	0.5	7.4	84926	67221	48796
ER4H	7993	32	0.4	6.4	98220	77744	56435
ER4I	7997	133	1.7	26.6	23344	18477	13413
ER4I2_1	133	133	100.0	.	.	.	.
ER4I2_2	4	4	100.0	.	.	.	.
ER4I2_3	1	1	100.0	.	.	.	.
ER4I2_4	1	1	100.0	.	.	.	.
ER5A	7998	42	0.5	8.4	74787	59196	42971
ER5A1	42	2	4.8	0.4	7896	6250	4537
ER5B	7997	37	0.5	7.4	84936	67230	48802
ER5B1	31	24	77.4	4.8	115	91	66
ER5C	7999	19	0.2	3.8	165818	131250	95275
ER5C1	18	4	22.2	0.8	1382	1094	794
ER5D	7994	184	2.3	36.8	16758	13264	9629
ER5D1	176	57	32.4	11.4	824	652	474
ER5E	7995	126	1.6	25.2	24657	19516	14167
ER5E1	122	21	17.2	4.2	1899	1503	1091
ER5F	7998	89	1.1	17.8	35084	27770	20159
ER6	7998	68	0.9	13.6	46041	36443	26454
ER7	7999	19	0.2	3.8	165818	131250	95275
TU1	7995	450	5.6	90.0	6620	5240	3803

Table G-1. Sample Sizes to Compare Two Proportions

(continued on next page)

Table G-1. Sample Sizes to Compare Two Proportions *(continued)*

Section B Question	*Number of Pilots Reporting*	*Number of Pilots Reporting 1+ Event*	*Percent of Pilots Reporting 1+ Event*	*+/- 20% Change in Number of Pilots Reporting 1+ Event*	*Sample Size Power=.20*		
					Alpha=.05	*Alpha=.10*	*Alpha=.20*
TU1A	*449*	*242*	*53.9*	*48.4*	*338*	*267*	*194*
TU1B	448	242	54.0	48.4	336	266	193
TU2	7980	1360	17.0	272.0	1922	1521	1104
WE1	7953	1152	14.5	230.4	2331	1845	1339
WE1A	1151	336	29.2	67.2	958	758	550
WE1B	1146	268	23.4	53.6	1293	1024	743
WE2	7988	395	4.9	79.0	7589	6007	4361
WE2A	394	57	14.5	11.4	2334	1848	1341
WE3	7995	665	8.3	133.0	4352	3445	2500
WE4	7994	62	0.8	12.4	50510	39980	29022
WE5	7978	1709	21.4	341.8	1448	1146	832
WE6	7995	283	3.5	56.6	10759	8516	6182
CP1	6445	365	5.7	73.0	6576	5205	3779
CP2	6443	69	1.1	13.8	36471	28868	20955
CP3	6445	310	4.8	62.0	7813	6184	4489
AC1	7985	1107	13.9	221.4	2453	1942	1409
AC2	7990	903	11.3	180.6	3099	2453	1780
AC3	7973	299	3.8	59.8	10133	8020	5822
GE1	7998	9	0.1	1.8	350455	277396	201363
GE2	7991	235	2.9	47.0	13030	10314	7487
GE2A	235	198	84.3	39.6	74	58	42
GE2B	235	45	19.1	9.0	*1667*	1319	958
GE2C	235	0	0.0	.	.	.	.
GE3	7996	291	3.6	58.2	10454	8274	6006
GE4	7996	288	3.6	57.6	10567	8364	6071
GE5	7997	1	0.0	0.2	3156861	2498750	1813853
GE6	7998	3	0.0	0.6	1052155	832813	604542
GE7	7997	31	0.4	6.2	101452	80302	58292
GE8	7981	104	1.3	20.8	29903	23669	17181
GE9	7985	81	1.0	16.2	38525	30494	22136
GE10	7995	95	1.2	19.0	32831	25987	18864
GE10A	94	21	22.3	4.2	1372	1086	789
GE10B	95	67	70.5	13.4	165	131	95

(continued on next page)

Table G-1. Sample Sizes to Compare Two Proportions *(continued)*

Section B Question	*Number of Pilots Reporting*	*Number of Pilots Reporting 1+ Event*	*Percent of Pilots Reporting 1+ Event*	*+/- 20% Change in Number of Pilots Reporting 1+ Event*	*Sample Size Power=.20*		
					Alpha=.05	*Alpha=.10*	*Alpha=.20*
GE10C	95	7	7.4	1.4	4963	3929	2852
AH1	7981	764	9.6	152.8	3729	2952	*2143*
AH2	7988	305	3.8	61.0	9945	7872	5714
AH3	7983	1916	24.0	383.2	1250	990	718
AH3A	*1906*	*329*	*17.3*	65.8	1892	1498	1087
AH4	7993	46	0.6	9.2	68207	53988	39190
AH5	7995	5	0.1	*1.0*	*630898*	*499375*	*362498*
AH6	7986	441	5.5	88.2	6755	5347	3881
AH7	7995	4	0.1	0.8	788722	624297	453180
AH8	7996	9	0.1	1.8	350368	277326	201312
AH9	7995	154	1.9	30.8	20102	15911	11550
AH10	7989	25	0.3	5.0	125769	99550	72264
AH11	7991	24	0.3	4.8	131059	103737	75303
AH12	7996	59	0.7	11.8	53111	42039	30516
AH13	7992	69	0.9	13.8	45334	35883	26048
AH14	7991	112	1.4	22.4	27774	21984	15958
AH15	7996	4	0.1	0.8	788820	624375	453236
AH15A	4	0	0	.	.	.	.
AH15B	4	3	75.0	0.6	132	104	76
AH15C	4	1	25.0	0.2	1184	938	681
AH15C1	0	0	.	.	.	.	.
AD1	7994	399	5.0	79.8	7515	5948	4318
AD1A	399	32	8.0	6.4	4528	3584	2602
AD2	7991	21	0.3	4.2	149838	118601	86093
AT1	7969	2321	29.1	464.2	961	760	552
AT1A	2318	613	26.4	122.6	1098	869	631
AT1B	2320	1686	72.7	337.2	148	118	85
AT1C	2319	845	36.4	169.0	689	545	396
AT2	7966	2938	36.9	587.6	676	535	388

Appendix H. Workshop 1: Agenda and Participants

NAOMS held Workshop 1 on the developing program on May 11, 1999, in Alexandria, Virginia. Its purpose was to acquaint stakeholders with the nature of the program and its methods and to enlist their support in implementing the program. During this first workshop, the 76 participants formed breakout groups to address questions posed to them and to provide their feedback, comments, and recommendations to the NAOMS team.

This appendix includes the workshop agenda and attendance list, as well as the questions posed and feedback from workshop participants

Workshop 1 Agenda

8:00 to 8:30 A.M.................*Registration*

8:30 to 9:00 A.M. ....*Linda Connell* will provide project overview with focus on policy issues and the reason for the workshop.

900 to 9:40 A.M..................*Bob Dodd* will introduce the project goals and tasks completed to date. He also will focus on the tasks to be completed in FY99. This session will include a high-level discussion of the experiment; planned accomplishments; and the outcome, including analysis products.

9:40 to 10:00 A.M...............*Break and Questions*

10:00 to 10:45 A.M.............*Jon Krosnick* will speak about survey research methods and the development of the survey instrument. His primary goal is to describe the strengths and weaknesses of survey research methodology, using examples where possible. Jon will describe how we developed an instrument that was reliable and valid.

10:45 to 11:15 A.M.............*Joan Cwi* will discuss the process of applying the survey, emphasizing that the process will be anonymous and that we hope to work with the stakeholders to facilitate the process.

11:15 to 11:45 ....*Linda Connell* will speak again, setting up the workshop and speaking specifically to the sensitivity issues that she perceives might be a problem.

11:45 to 1:00 P.M. ....*Lunch*

1:00 to 1:15 P.M. ....*Introduction to Workshop Activities*

1:15 to 3:15 P.M. ....*Working Groups*

3:15 to 3:30 P.M. ....*Break*

3:30 to 5:00 P.M. ....*Work Group Summaries and Discussion*

Workshop 1 Attendance List

First Name	*Last Name*	*Organization*	*Office*	*Job Title*
Mark	Anderson	Virginia Polytechnic Institute and State University		Associate Professor
Mac	Armstrong	United Airlines		
Henry	Armstrong	Federal Aviation Administration	Rotorcraft Directorate	Manager
Julie	Austin	United Airlines		
Susan	Baker	Johns Hopkins Injury Prevention Research Center		Professor
Jim	Blancahrd	Embry Riddle Aeronautical University		Professor
Phil	Boyer	Aircraft Owners and Pilots Association		President
Bill	Bozin	Air Transport Association		
Mads	Brandt	Teledyne Controls		Director, Flight Data Systems
Joseph	Breen	Transportation Research Board National Research Council		Senior Program Officer Aviation
Malcom	Brenner	National Transportation Safety Board		
Jan	Brett-Clark	Federal Aviation Administration		
R. Thomas	Buffenbarger	The International Association of Machinists & Aerospace Workers		President
Phillippe	Burcier	Airbuse Industries		Operational Prevention and Safety Assurance
Brigadier General Charles M.	Burke	U.S. Army Safety Center		
Kim	Cardosi	U.S. Department of Transportation		Engineering Psychologist
Rick	Cassell	Rannoch Corp.		
Terry	Clark	Alaska Airlines		

First Name	*Last Name*	*Organization*	*Office*	*Job Title*
Walt	Coelman	Regional Airline Association		
Steve	Corrie	Federal Aviation Administration	Office of System Safety	
James	Deimler	Flight Data Company		Regional Manager
Thomas	Diefiore	Federal Aviation Administration	Aviation Safety Division	
Eleana	Edens	Federal Aviation Administration	Human Factors Dividsion	
William	Edmunds	Airline Pilots Association		
Carolyn	Edwards	Federal Aviation Administration	Office of System Safety	
Jack	Enders	Enders and Associates		President
Ray	Fenster	Association of Flight Attendants		
George	Finelli	National Aeronautics and Space Administration	Aviation Safety Program Office	Aviation Safety Program Office
Charles	Fluet	Federal Aviation Administration	Office of Integrated Safety Analysis	Deputy Director
Roy	Fox	Bell Helicopter		
Mike	Gallagher	Federal Aviation Administration	Transport Airplane Directorate	Manager
Daniel	Garland	Embry Riddle Aeronautical University	Department of Human Factors and Systems	Chair
Major General Francis C.	Gideon	Air Force Safety Center		
Curtis	Graeber	Boeing Commercial Airplane Group	Human Engineering	Chief
Christopher	Hart	Federal Aviation Administration	ASY-1: Office of System Safety	Assistant Administrator
Chuck	Hedges	Federal Aviation Administration	Office of System Safety	
Captain Mike	Holtom	Meridian	Senior Manager, Safety	

First Name	Last Name	Organization	Office	Job Title
Charles	Huettner	National Aeronautics and Space Administration	Aviation Safety Research	Director
Mike	Kennedy	Pratt & Whitney		
Dr. James	Kuchar	Massachusetts Institute of Technology	Department of Aeronautics and Astronautics	
Carl	Kuwitzky	SouthWest Airlines Pilots' Association		Chairman, Air Safety Committee
Bruce	Landsberg			
John	Lauber	Airbus Industries		Airbus Training Center
Captain Richard	LaVoy	Allied Pilots Association		President
Nancy	Leveson	Massachusetts Institute of Technology		Hunsaker Visiting Professor of Aeronautical Information Engineering
Guohua	Li	Johns Hopkins University Hospital		
Bernard	Loeb	National Transportation Safety Board		Director, Office of Research and Engineering
Thomas	Longridge	Federal Aviation Administration	Data Management and Analysis Section	Aviation Research Psychologist, Supervisor
Nancy	Mathiowetz	Joint Program in Survey Methodology		Assistant Professor
Stuart	Matthews	Flight Safety Foundation		President
John	McCarthy	Naval Research Laboratory		Manager, Scientific and Technical Program
Michael	McNally			
Tom	McSweeny	Federal Aviation Administration	Office of Regulation and Certification	Associate Administrator

First Name	*Last Name*	*Organization*	*Office*	*Job Title*
John	O'Brien	Air Line Pilots Association		Director, Engineering and Air Safety
John	Olcott	National Business Aircraft Association		President
Jay	Pardee	Federal Aviation Administration	Engine & Propeller Directorate	Manager
Dave	Patterson	MacFadden and Associates		Senior Consultant
Ben	Phelps	National Air Traffic Controllers Association		Safety Coordinator
Tom	Poberezny	Experimental Aircraft Association		President
Jacques	Press	Federal Aviation Administration	FAA Technical Center	
Ronald	Robinson	Boeing Commercial Airplane Group		Director, Airplane Safety
Paul	Russell	Boeing Commercial Airplane Group	Airplane Safety Engineering	Chief Engineer,
Stewart	Schreckengast	The Mitre Corporation		Ph.D.
David	Schroeder	Federal Aviation Administration	Human Resources Research Division	Manager
Ronald	Simmons	Federal Aviation Administration	Human Factors Division AAR-100	
Catherine	Simonne-Jondot	Airbus Industrie		Group Manager, In-Service Data Collection
Stan	Smith	National Transportation Safety Board	Data Systems	Manager
Captain Ed	Soliday	United Airlines	Safety & Security	Vice President
Jeremy	Sprung	Sandia National Laboratories		
Larry	Sukut	Alaska Airlines		

First Name	Last Name	Organization	Office	Job Title
Ronald	Swanda	General Aviation Manufacturers Association		Vice President Operations
Dr. Jay	Swink			Senior Technical Specialist in Crew Systems Technology
Robert	Toenniessen	Federal Aviation Administration, ASY-100	NASDAC	Manager
Robert	Vandal	Flight Safety Foundation		Executive Vice President
Ron	Wojnar	Federal Aviation Administration	Transport Airplane Directorate	Manager ANM-100
Richard	Wright	Helicopter Association International		

Questions Posed and Feedback from Workshop Participants

This section presents the discussion questions posed to the workshop participants and provides a bulletized summary of their responses.

Group A

Question: Do we have the courage to act on the data we collect?

Group Discussion:

- Maybe
- Could be heroes or villains
- The data will not be ignored
- But, reasonable people may disagree about the actions it does (or does not) motivate

Question: Could we reform existing data sources instead?

Group Discussion:

- In some cases, yes
- But, many existing data collection efforts are passive, not active and statistically designed
- NAOMS should avoid redundancy unless that redundancy serves to validate

Other Comments:

- Not linking causal factors to events is a mistake
  - —Will lead to data that cannot be analyzed

- Need to explain why some questions relate to the past 30 days and others to the past year
  —Not obvious from survey design
  —Could be off-putting to respondents
- Need to recognize the limitations of the survey
  —Not create expectations that cannot be met
- Causal information may be at wrong level of detail
  —Is not sufficient to support intervention strategies

Group B

Question: Do you have specific suggestions regarding the conduct of the field trial?
Group Discussion:
- Important to experiment with sampling during trial (i.e., medical-based)
- Obtain more feedback
- Additional survey needed to encourage questions/comments on the original survey

Question: Have we adequately addressed issues surrounding data sensitivity and use?
Group Discussion:
- There needs to be a statement in the first paragraph stating the confidentiality agreement
- Link it to ASRS, to show a pattern of confidentiality
- Get AOPA and ATA to endorse it early and clear on the front page
- Needs an endorsement letter to assure the aviation community of its support
- Overall consensus was that the issues have been addressed adequately

Question: Could you suggest ways of improving the proposed data collection process?
Group Discussion:
- Recommend surveying on a monthly basis; need to look at 30-day data to identify trends
- Is there a core questionnaire for all groups; or is this survey tailored to each group (i.e., flight attendants, pilots, etc.)?
- Send out quarterly reports
- Date the survey it so you can refer back to it in 30-day increments
- Why not 60-to 90-day increments?

Question: What can we do to maximize participation and response in the field trial and beyond?
Group Discussion:
- Needs to be clarified that this is not a duplication of information
- Need to give follow up data to close the loop:
  —Post card
  —Web site
  —Callback or Directline Publications
- Persons being surveyed need to see that this survey has had a direct impact on future survey tools

Question: Do you have suggestions for ways to improve the survey instrument?
Group Discussion:

Flight Experience Section
- Shorten form
- Use standard categories: light, medium, and heavy
- Define "Other Aircraft" column
- Flight Experience matrix needs to be more specific
- Where do unscheduled aircraft fit in?
- Demographic information should be 60+ instead of 60-65

Safety Events Section (Previous 12 Months)
- Text introduction should insert the word "observed" after "flight experienced"
- Needs to be directed to more than just pilots and flight crewmembers
- Aircraft Equipment Problems:
  - —"Experienced an engine fire" does not capture the consistency; there can be a large range of engine fires
  - —Needs to be clarified more on the severity of what aircraft problems would lead to the return to land or diversion
- Actual Or Potential Loss of Control:
  - —Encountered wake turbulence that induced 45 or more degrees of roll; needs to be changed to 30
- Airborne Conflicts:
  - —Not just "nearly" collided, but "actually" collided
  - —Expand bird strike to include volcanic ash, hail, etc.
  - —Take the word "residual" out; maybe replace it with "horizontal" or "vertical"
- ATM Problems:
  - —Write out the acronym
  - —Take out the word nearly; it is redundant
  - —Need to be more specific on whether the ATC clearance you received that resulted in a near collision with terrain or a ground obstruction was followed, corrected, or not heard

Safety Events Section (Previous 30 Days)
- Wrong Place, Wrong Time:
  - —Landed without a clearance is too vague. This happens all day with non-tower airports

Safety Events Section (Contributing Factors and Positive Factors)
- Page too busy
- Should not be an opinion-based questionnaire, we should stick with occurrences and events
- Needs to be more specific
- Review this page more to see if it adds value
- Comments and details section should be broken into the following two categories:
  - (1) What could be done to enhance survey?
  - (2) Other comments and details

Question: How might we formalize industry, government, and professional organization participation in continuing NAOMS development?
Group Discussion:
- Create a committee
- Define who the implied users are that would benefit from the data
- Determine what information would be essential to each group to have a better focus
- Data may be too soft; "fringe" data

Question: Would an advisory panel be appropriate?
Group Discussion:
- They believed it was a premature step
- Maybe a user group

—Doesn't need to include all people that attended workshop
—Should be separate from ASRS Subcommittee
—If NASA is internally assessing this, why form a committee at all?

Other Comments:
- Be careful not to collect data we have no use for; concentrate on specifics

Group C

Question: Do you have specific suggestions regarding the conduct of the field trial?
Group Discussion:
- If our goal is to increase response rate, then cover letter from union would be helpful

Question: Have we adequately addressed issues surrounding data sensitivity and use?
Group Discussion:
- Feel it has been pretty well covered
- But, when would the database be released so that single incidents couldn't be matched to other databases?
- Desperate for this type of information in a timely manner
- But also will want to link it to other databases for validation, etc.

Question: Could you suggest ways of improving the proposed data collection process?
Group Discussion:
- Recommend surveying on a monthly basis; need to look at 30-day data to identify trends
- The survey would have to be altered
- Cost concerns
- Scantron or automated response for data collection would be helpful
- Eventual resolution of Web site survey issues

Question: What can we do to maximize participation and response in the field trial and beyond?
Group Discussion:
- Need to make a commitment to offer feedback to those that are completing the survey
- Suggest a Web site
- Data should be accessible in a timely fashion

- Produce articles to be disseminated to industry for publication in internal documents (also could use Callback and DirectLine)

Question: Do you have suggestions for ways to improve the survey instrument?

Group Discussion:

Flight Experience Section

- Remove word "air carrier" from spanners
- Use standard classifications of type of operations (column 1)—suggest using FAR 119; needs to be dynamic/flexible with changing FARs
- Use standard categories: light, medium, and heavy
- Define "Other Aircraft" column
- Add lighter-than-air, rotorcraft, etc., now
- As pilot gets older, the flight time may not be as accurate—100 hours doesn't mean as much when you have 15,000 hours
  - —A range of flight hours might be better (e.g., 10,000-12,000 hrs.)

Safety Events Section (Previous 12 Months and Previous 30 Days)

- Instead of using Number of Occurrences column, suggest two columns: one for 30 days, one for 12 months for all questions
- Recommend use of Jon's categories from his presentation
- Need to document the objectives of each question and how we can use the data
- Would like to gather less serious or "precursor" information to know if something is about to happen (e.g., deviated due to icing)
- Use a fixed reference period (e.g., March 1999 instead of "last 30 days") so data can be compared over time and with other data sources
  - —Does this compromise confidentiality? Overall, the questions will collect good and useful data

Safety Events Section (Contributing Factors and Positive Factors)

- Overall, what is learned from this part if we can't link anything up?
  - —Suggest reformulating this page to target risk areas and understand where to do further research
- As a pilot, I find it sort of difficult (might be the layout). Ask for the most significant rather than circle all that apply
- May want to eliminate the use of the word "aircraft design" and just leave it as "problems"

Question: How might we formalize industry, government, and professional organization participation in continuing NAOMS development?

- Just ask
- Identify organizations that are doing safety analysis

Question: Would an advisory panel be appropriate?

- In the beginning of the program and to help identify topical questions
- Maybe a user group

Other Comments

- Suggestion to get the BTS involved

Appendix I. FAA Comments after Workshop 1 and NASA Responses (February 20, 2000)

The FAA Office of System Safety conducted a survey of FAA staff members who attended the 11 May 1999 NAOMS workshop. Their inputs, and NASA responses to those inputs, are summarized here. The questions are numbered and listed with the Comments (C) as in the original document from the FAA. NASA's Responses (R) follow the comments where appropriate.

Q1. What was your general reaction to the National Aviation Operational Monitoring Service (NAOMS) concept presented by NASA representatives at the May 11, 1999 workshop (Holiday Inn, Alexandria, VA)?

C1.1 I thought that the general concept had merit. However, I felt that additional effort was needed to determine precisely the type of information required that would provide the basis for decision-making to enhance safety. Consideration needs to be given to narrowing the focus of the questionnaire.

R1.1 Input on questionnaire content was derived from a number of different sources, including:

- Information provided during focus groups with active air carrier pilots
- Analyses of ASRS and NASDAC incident and accident data resources
- Results from background research on current safety issues and relevant safety information sources, and
- Review by industry and government safety representatives.

It should also be remembered that the field trial participants will be asked about their opinions on the questionnaire including question content and formatting.

The survey instrument has a tri-part structure. The first part collects risk exposure information (hours and legs flown, etc.). The second part collects information on incident occurrences for trending and high-level safety indication. The third part contains focus questions that can be made more or less detailed depending on the specific information needs of decision makers. Parts 1 and 2 are needed to calculate event rates. Part 3 is needed to evaluate technology impacts and to address topical questions of interest to the aviation community.

C1.2 With a sound survey design and support of all major segments of the aviation community, the proposal could provide high-level information for assessing the general safety status of the national aviation system. However, to significantly affect the accident rate, much more detailed information than that currently contained in the draft proposal will be required.

R1.2 The NAOMS questionnaire is a compromise between the specificity needed to produce meaningful safety information and the brevity needed to obtain a high survey response rate. It is important to remember that one of the primary goals of NAOMS is to provide early, high-level indications of emerging safety concerns. NAOMS data will reveal presence of increasing or decreasing rates of events in a quantitative manner but may not pinpoint their exact nature. Additional comparison to other data sources may need to be accomplished, or additional investigations may be required, to understand why the observed trends are occurring. The latter will provide the detailed information needed to develop interventions that address the emerging safety issue, but the NAOMS data will help guide these efforts toward interventions that are data driven.

C1.3 The NAOMS concept could perhaps be an effective component of a well thought out integrated approach to a coordinated safety program that would help identify precursors to accidents and incidents. However, things need to be addressed from a systematic perspective by identifying the "problem space," current data collection efforts, and the data/information gaps. An assessment needs to be made of how this concept fits in with other initiatives and data collection programs to determine if NAOMS is the best way to fill in the gaps.

R1.3 This issue was considered during the design of the NAOMS process and the questionnaire. Background research was conducted to evaluate current aviation (and proposed) safety data or information sources, their strengths and weaknesses, and what gaps needing to be addressed. The results from this analysis were used during the design of the field trial questionnaire.

More generally, the NAOMS development process was assisted by more than two decades of experience operating the ASRS as well as consultations with FAA, ALPA, HAI, and other industry representatives. The ASRS has fielded many thousands of data requests over that time period. Its experience suggested that there were at least two significant gaps in available aviation safety information. The first involved numbers (quantitative value). There is a paucity of hard numbers—scientifically collected and reliable—characterizing the frequency of unwanted safety events. The other related to impacts of new technologies on aviation safety. Information relating to technological impacts tends to be variable. Severe side effects are generally reported. But, lesser issues often failed to surface in a timely way or available data are not at the level of detail sought by technology implementers. NAOMS was designed to fill both of these gaps.

C1.4 Sounds like a good idea, however, question the value of this program versus other possibilities, e.g. APMS.

R1.4 See Response 1.3. It is our opinion that no single program is adequate to address aviation safety issues due to the system's complexity and dynamic characteristics. NAOMS proposes to provide an accurate, reliable quantitative measure of safety events in the system from the experiences of the people working in it. Other programs contribute to other aspects of any quantitative and qualitative safety effort.

C1.5 I was very positively impressed by the presentation of the data system. It appears to have the rudiments of a very valuable addition to the aviation safety data set if it is implemented properly. The actual draft form that was reviewed by participants in the afternoon session was unfortunately highly inconsistent with the presentations made in the morning. If this is the form of the survey that is implemented, it will not live up to the expectations created during the morning sessions and presentations.

R1.5 The questionnaire used in the field trial was modified to be in concert with the concepts discussed in the morning presentations at the workshop. The field trial itself is expected to produce many additional changes in the instrument as a consequence of respondent feedback.

C1.6 As analysts, we are always seeking quality aviation safety data. When concrete data do not exist, surveys can provide some useful insight into the effectiveness of our aviation domain. Thus, I do support the NAOMS concept. Three major concerns that I have with the NAOMS concept, as presented, are:

a. the frequency of the data collection
b. the consistency of data being collected. (Will the questionnaire be constantly undergoing change/modification? For comparison purposes, we do need to standardize the questionnaire.)
c. response rate.

R1.6 Frequency of Data Collection: The field trial is designed to evaluate the optimum recall period (which will influence the frequency of data collection) as will the response rate achieved for the various modes of questionnaire application (mail, telephone and face-to-face interviews). The working hypothesis is that a monthly survey application may the best option.

Consistency of Data Collected: The NAOMS program from its inception was (and is) designed to use a "core" set of questions that will be applied consistently over time so meaningful trends can be developed.

Response Rate: The field trial will determine whether NAOMS achieves the necessary response rate for statistical accuracy. The target figure is 70%. The lowest acceptable figure is approximately 50 percent.

Q2. How effective do you believe the proposed NAOMS will be in providing information to meet the measurement objectives described in the introductory briefing, i.e.

- **Better, more comprehensive numbers to help reach the safety improvement goal (80% reduction in fatal accident rate in 10 years)**
- **Better and more rapid feedback on technological and procedural change**
- **Escape from event-driven safety policy, and**
- **Create a data-driven basis for safety decisions.**

C2.1 If the authors do not provide significant changes in the format of the questionnaire and aspects of the approach to data gathering, I think that the outcomes will not provide the necessary information to have a significant impact on aviation safety. The critical feature will be whether the information gathered will be of sufficient detail to permit the identification of specific intervention strategies. As reviewed, the information would be too global to have the specificity needed to identify appropriate intervention strategies. For example, it did not appear that the questionnaire adequately addressed safety issues in US airspace versus flights involving oceanic or international airspace. If used effectively, questionnaires can provide rapid feedback on the impacts of technological and procedural changes. However, this requires a highly specific and focused questionnaire.

R2.1 The questionnaire was modified in response to the comments from the first NAOMS industry workshop. As discussed earlier, one of the primary goals of NAOMS is to provide early indicators of safety concerns before they result in undesirable outcomes. The primary method for this will be the tracking of trends for safety issues identified in the NAOMS questionnaire. If unusual or changing trends are noted, additional investigation will be required to understand why the observed trends are occurring. It is this resulting investigation and evaluation that will provide the detailed information needed to design interventions to reduce the accident potential.

If the standard questions in the NAOMS questionnaire do not address an area of concern, then questions can be developed and added to the questionnaire in the "topical questions" section (questions that are not routinely asked in the "core" portion of the questionnaire). These questions would not be asked over an extended period of time but only for a few cycles of the questionnaire's application. These questions may be much more detailed than the core questions.

The last point to remember is that the initial NAOMS industry workshop, the field trial of the questionnaire, and the second NAOMS industry workshop are all designed to evaluate the proposed NAOMS process and the questionnaire. It is hoped that any concerns about question specificity or subject areas will be discovered through that process.

C2.2 If the questions remain at the high, broad level presented in the draft survey instrument, it may be difficult to get information that is detailed enough to affect the accident rate. For example, it appears that in the survey of air carrier pilots, no information will be collected by airport.

R2.2 At this point, specific questions concerning airports are not included. NAOMS has been advised that this information may lead to some level of identification. In regard to asking more specific information, please look at Response 1.2. The general observation is true, however, that NAOMS will not be useful as a carrier-, aircraft-, or location-specific surveillance system. It is thought that other Federal programs (e.g., SPAS) fill these requirements.

C2.3 The survey may be effective in providing feedback on technological changes, and perhaps procedural changes, to the extent that they are implemented at a national level.

- *Re better, more comprehensive numbers, think it is questionable whether NAOMS will succeed. Concerned about how you factor out biases.*
- *Re better and more rapid feedback, presupposes that people will respond to the survey.*
- *Re escape from event-driven safety policy…perhaps.*
- *Re create a data-driven basis for safety decisions…will have data, but question the validity of the data due to its subjective nature.*

R2.3 Biases / Subjectivity. Two approaches have been used to minimize the impact of biases and subjectivity. First, the questionnaire was designed to remove questions concerning the reporter's opinion. Instead, the focus was on actual operational experience. This modification occurred after the first NAOMS industry workshop. Secondly, the field trial was designed to evaluate the influence of reporter biases so that they might be better understood and mitigated.

C2.4 Each of these objectives can be furthered significantly if NAOMS is implemented consistent with the presentations made about it during the morning session. If not, determining its value will be problematic.

R2.4 See Response 1.5.

C2.5 It does appear that the questionnaire might provide us with some insight into potential safety issues, thus we should have better numbers in determining how well we are attaining our safety goal. If the data are of high quality and complete, then I could see the FAA moving into a more data-driven basis (pro-active) for safety decisions. However, this does not mean the end for an event-driven safety policy, but it will help minimize the tendency for the FAA to be in a reactive mode.

R2.5 NAOMS agrees with these perceptions and is optimistic about a positive contribution to the safety process.

C2.6 Since it isn't clear what problems are being solved by this data collection effort, it isn't clear that the right questions are being asked. For example, while there were questions in the draft survey instrument by aircraft weight categories, there were no questions that would provide information on new generation aircraft with glass cockpits versus legacy aircraft types.

R2.6 The questionnaire was revised after the first NAOMS industry workshop to collect information on aircraft make and model. Also, please see Responses 1.2 and 1.3.

Q3. Are there aspects of the NAOMS concept that you like? Please specify.

C3.1 I think that the use of questionnaires can provide relevant data for planning safety interventions. Significant planning was evident in developing an appropriate approach to the questionnaire.

C3.2 Plans to include all segments of the aviation community in the NAOMS effort.

C3.3 NASA appears to have assembled a well-qualified development team to work on this initiative.

C3.4 Getting the industry together to define safety issues is important, the question is what is the best way to do it.

C3.5 The "proposed" survey would be open to new evaluations of incident "cause" over and above those developed in the past. (The draft document didn't fulfill this promise.)

C3.6 First, I believe that NASA (Linda Connell, Mary Connors, et al) has assembled a quality team. Battelle, OSU, Dodd and Associates have vast experience when it comes to aviation safety. (I have to admit, I have never heard of CPHRE until the meeting.) Second, submitting a standard survey to all types of aviation employees is a good idea. I would be interested in compiling and comparing the results by user type, e.g., pilot, ATC, mechanic, flight attendant, etc. Third, the de-identification is a must; you wouldn't receive any useful information without this.

C3.7 Like the idea of reaching out to people involved in the day-to-day operations of the system. Such an approach may provide opportunities to identify precursors and allow us to get ahead of some of the emerging safety issues. The questions that remain are: (1) are there other ways of getting at the information, and (2) how is one method valid over others?

Q4. What are your greatest concerns about the NAOMS proposal?

C4.1 In general, the specificity of much of the data being gathered was insufficient to permit the development of appropriate intervention strategies. There was little indication of how the questionnaire was designed to handle information involving safety concerns on international flights versus those in US airspace. As proposed, the questionnaire also did not adequately address issues associated with the implementation of new technologies and possible training issues associated with new technologies and procedures. It would be my strong recommendation that the questionnaire become more focused prior to distribution. I also had some concerns associated with the response categories proposed for various items. The use of "always" and "never" has been shown, in some contexts, to modify the typical response distribution. While the questionnaire appeared to be designed to link flight time and other variables to the types of incidents reported, the structure was insufficient to permit direct comparison of those issues. This was one of several concerns raised during the subgroup meetings. Additional work was also *required to define and categorize data regarding type of aircraft flown and the equipage of the aircraft.*

R4.1 Specificity of data: See Response 1.2.

International airspace: During the background research for NAOMS, international airspace was not an issue that caused extensive concern for U.S. air carrier pilots. If it is an issue that needs to be evaluated, then perhaps it could be a "special topic" for a subsequent NOAMS survey.

New technologies: NAOMS was designed from the beginning to incorporate questions on special topics, including new technologies and procedures. In the field trial, two special topics (MEL Items and In-Close Approach Changes) were evaluated.

Response categories: The opinion related response categories were dropped from the questionnaire used in the field trial.

Aircraft flown: The questionnaire was revised to obtain information on the aircraft make and type.

C4.2 The development of NAOMS is based on the premise that FOQA will not be in place and operating any time in the near future. Would it not be more productive if NASA were to focus the resources toward the development of software tools that can be applied to FOQA and ASAP data? The CFIT JSIT is preparing to include this proposal in their Detailed Implementation Plan.

R4.2 NAOMS is not designed to replace any current or planned data collection or analysis systems. It was designed from its inception to be complimentary to these systems. However, NAOMS does not assume that one or more carriers will not have operational FOQA programs. It recognizes, however, that (a) FOQA data can address only a limited range of aviation safety questions, (b) FOQA data are highly proprietary and may not be available to the government for this purpose, and (c) even if the data were available, they would only describe the operations of selective fleets flown by selective air carriers. NAOMS wishes to address a broad array of safety issues, on a system-wide level, in a statistically robust fashion. FOQA data obviously cannot meet these requirements at least in the near- to medium-term.

C4.3 We don't want to burden the system with yet another data gathering system. It was pointed out to NASA that several of the events being monitored by the survey would already require written reports due to the nature of the event. It is important that this should not be redundant to existing programs.

R4.3 NAOMS data collection will avoid redundancies. If reliable quantitative information is already available in a safety area, NAOMS will not seek duplicative information. It is NAOMS' perception, however, that many types of aviation safety events are not well quantified.

C4.4 The ASRS program has only enough funding to process a small portion of the reports they receive. Shouldn't this be a priority before initiating a new program?

R4.4 NAOMS is not being funded with ASRS monies. The NASA Aviation Safety Program (AvSP) research and development dollars are funding this effort.

C4.5 There are no current plans to review other databases in conjunction with the NAOMS data. This would mean that NAOMS is a stand-alone system and does not benefit from the other safety programs currently underway.

R4.5 See Responses 1.1 and 1.3. NAOMS benefit from prior database reviews conducted by the FAA Office of Aviation Safety as well as others. NAOMS disagrees with the perception that it will be a stand-alone system that does not benefit from other programs. In fact, there will be a strong synergy. NAOMS will create a numeric framework that can be used to calibrate other data sets and put them in larger perspective.

C4.6 The plan doesn't contain a mechanism to "change" the system once a safety trend has emerged.

R4.6 This is true. Much as with other safety data gathering systems, NAOMS' responsibility will be to collect data, convert it to useful information, and bring that information to the attention of persons with operational authority. Whether or not these aviation authorities ultimately use the information developed by NAOMS is a matter to be decided by them.

C4.7 During the planning phase, NASA held meetings with ALPA, ATA, HAI, Boeing, etc. regarding the draft questionnaire, but apparently left the FAA off their list of people to see. So essentially, this was the first time that the FAA had an opportunity to provide input to the questionnaire, when in fact, we are one of the major customers.

R4.7 The NAOMS project proposal was briefed very early in the process to many FAA offices (e.g. ASY-1, ASY-300). A kick-off workshop for the entire AvSP element (then titled "System-Wide Monitoring and Data Analysis") was held at NASA-HQ in November 1997. The NAOMS program was subsequently briefed at numerous meetings of the ASRS Advisory Subcommittee the following year with two FAA representatives in attendance. Thus, the May 1999 NAOMS workshop was the last formal briefing opportunity on the NAOMS program that the FAA attended.

C4.8 Little, if any, information gathered in the survey address automation issues. I think with the changing levels of automation in the cockpit that this is a major area to look at.

R4.8 See Response 1.2 and 4.1.

C4.9 Concerned that the current proposed survey would not be able to obtain information at a detailed enough level to affect accident rates. I question the ability of the current proposal to adequately capture information on aviation events (many of which are rare occurrences) and then accurately extrapolate sample results to population levels. For example, suppose you were trying to determine the number of runway incursions that occurred at specific airports across the nation during a particular year. Suppose that you use a sample size of 10% of the air carrier pilot population who conducted flights covering 200 of the nation's approximately 350 towered airports. Would you not only be able to estimate the number of air carrier pilots involved in runway incursions (at the national level) and the number of incursions at those 200 airports serviced by pilots constituting your 10% sample? You would not be able to say anything about the number of events that occurred at the remaining 150 airports not serviced by pilots in the survey? (Incidentally, it doesn't appear that the current proposal will capture any information by airport location).

R4.9 The NAOMS sample size will probably range between five and ten thousand observations per year. This will allow NAOMS to make inferences about trends involving rare events (defined by NAOMS to be once-in-a-career pilot experiences) with substantial certainty at a national scale. This is in keeping with NAOMS role as a high-level surveillance system that will map progress towards the national 80% aviation safety improvement objective. However, NAOMS data will not be able to make reliable inferences about rare events at particular locales.

C4.10 Concerns about plans to administer the survey through special interest groups who may attempt to sway their members to respond in ways that could bias the survey results.

R4.10 This possibility cannot be dismissed, but attempts to bias survey results in that fashion may be detectable through various means. Also, experience suggests that such biasing influences cannot be sustained over an extended time period, e.g., they may cause a "blip" in the data but not a long-term trend.

C4.11 Concerns about how you maintain the momentum? How do you get respondents to respond year after year?

R4.11 Survey application designs that will minimize declining participation are being considered for the fully operational NAOMS if this is discovered to be a problem. However, the pilot population is large enough that any given individual is likely not to be asked twice. This may be more of a challenge with the much smaller air traffic controller population.

C4.12 Concerns about method for developing questions to be asked in the survey.

R4.12 NASA will make the final decision about what questions to include in the NAOMS survey and their exact phrasing. However, NASA will engage in extensive consultations with all segments of the aviation community as it develops topical questions for the NAOMS survey instrument. Its goal is for survey instruments to be as inclusive and neutral as possible.

C4.13 Concerns about how to factor out biases re political/economic agendas?

R4.13 See Response 4.10.

C4.14 Concerned that NASA will implement a version of the survey only slightly different from the one shown to the participants at the conference. This will potentially nullify the benefits that it might otherwise provide.

R4.14 The questionnaire has been significantly modified since the first workshop. It is also the goal of the second workshop, and subsequent discussions to obtain input from all interested parties on the questionnaire content and design. The field trial has also generated much valuable feedback that is being used to enhance the instrument design.

C4.15 How does this proposal fit in with other efforts such as CAST, FOQA, etc. and is it the best way to fill in any gaps in existing information? What information would be provided as an adjunct to other information sources? Also concerned that this approach does not appear to be addressing the issues of aircraft with advanced automation versus legacy aircraft types.

R4.15 See Responses 1.3 and 4.5.

C4.16 Need to identify and define the metrics that NAOMS is interested in measuring. Without this, how does anyone know if NAOMS is collecting, via the survey, the 'right' data? It was unclear to me as to whether a complete set of metrics had been defined.

R4.16 NAOMS intent is to develop metrics that are useful to mid- and high-level decision makers in both the governmental and private sectors of the aviation community. Accordingly, NAOMS has developed questions that relate to prominent, widely recognized safety issues. NAOMS understanding of these issues is informed by more than two decades of experience operating the ASRS, and also by regular interactions with the aviation community through NASA advisory committees and other mechanisms. The NAOMS Workshops also provide a very important reality check in this regard.

Finally, NAOMS asked respondents during the field trial whether it was asking questions that were relevant to aviation safety. The overwhelming majority answered in the affirmative.

C4.17 Another issue was the timeframe. If I recall properly, the user would be required to respond to questions regarding events that took place in the most recent 30 days and most recent 12 months. I wonder (from a human factors perspective) the statistical importance/significance of data that is based on human recall over a 12-month period.

R4.17 One of the goals of the field trial was to evaluate what the most accurate recall period might be. The field trial evaluated time periods from one week to twelve months. The upcoming Workshop will report on the outcome of this research.

C4.18 RE: Questionnaire: Some of the questions posed (e.g., on aircraft equipment problems) may not be necessary. Turnbacks, engine damage, fire or smoke events are required to be submitted to the FAA. The acronym "ATM" should be spelled out.

R4.18 There may be some redundancy between NAOMS topics and other data systems. This is by design since many of these other data systems appear to suffer from under-reporting. However, NAOMS will not collect data when strong pre-existing data resources exist. "ATM" will be spelled out.

C4.19 In the flight experience section, need to remove "other aircraft" column. There is a row entitled "GA/other."

R4.19 This portion of the questionnaire was revised.

C4.20 Regarding the contributing factors section, I have serious reservations as to the reliability and utility of the collected data/information. As an example, I don't believe pilots have sufficient knowledge in determining whether a contribution factor could have been a design flaw.

R4.20 This section of the questionnaire was dropped.

Q5. What changes would you suggest to NASA to improve or revise the NAOMS approach described at the workshop?

C5.1 A number of my concerns were addressed during the workshop. I am not including all of those concerns in this discussion. Many of the changes needed are addressed above. Recommend reviewing the response alternatives ("always" and "never"). Item revisions are needed to introduce greater specificity in the response. For example, "Received ATC clearance that nearly resulted in a near collision with terrain or a ground obstruction." To ensure that all respondents have the same frame of reference in responding to that item it would require that the parameter of nearness with terrain or ground be defined (e.g. within x feet). Reduce time to respond to items concerning flight experience in the past 12 months. Since you are asking for approximations, why not consider asking pilots to place percentages in the various categories rather than indicating specific hours (e.g. 500 hours last 12 months, with 90% wide body and 10% GA/other). I also feel that in dealing with a sample of pilots the large number of categories in the matrix (type of operation versus type of aircraft) will be too large. Recommend inclusion of questions concerning equipage of the aircraft flown. Are automation-related concerns more likely in aircraft with certain display configurations, etc? Information is needed to determine whether problems were most common in US or other airspace. Need to develop a way of linking events to contributing factors.

R5.1 See Responses 1.2, 1.6, 2.6 and 4.1.

C5.2 Rather than using the membership lists of various special interest organizations, use national population databases, such as the Airmen Registry, for the sampling frame so that a truly representative sample can be selected.

R5.2 This is the approach that was used in the field trial.

C5.3 Develop a briefing or some materials that would describe to the various stakeholders how the information that is collected would be used.

R5.3 This will be developed as the program matures. The NAOMS is intended to be an industry/gov't program. The continued involvement and advice is required and will likely take the form of an Advisory Committee, Executive Council, or Steering Group of stakeholder membership.

C5.4 Describe approaches that you plan to use to validate information that will be collected in the survey.

R5.4 This information will be presented at the second NAOMS industry workshop.

C5.5 More cooperation with FAA & industry in development of survey instrument.

R5.5 See Response 4.12.

C5.6 Would like to see a thorough analysis of the cost of this approach versus other approaches, e.g. APMS.

R5.6 As mentioned earlier, NAOMS is designed to supplement these programs, not replace them. The NAOMS survey process was the only viable method identified for achieving the statistically reliable national measures of aviation safety sought by this initiative.

C5.7 Would like more clarity on what would be done with the information resulting from the survey.

R5.7 It is intended that NAOMS survey results will be a regular input to mid- and high-level decision makers in NASA, the FAA, and the aviation industry. It is hoped that NAOMS' products will be used by government and industry officials when they develop safety program plans and make safety investment decisions. NAOMS data will help these organizations monitor aviation safety trends; track the effectiveness of safety interventions; and evaluate the impacts of new aviation technologies and procedures.

C5.8 It is very important that any information gathered about causes be linked back to events. The form that was circulated at the conference completely failed to ensure this. Such data would probably be useless in helping to identify, in advance, any serious potential causes of accidents or incidents.

R5.8 The questionnaire has been revised to focus on events. The causal portion of the instrument was deleted after NAOMS concluded that this information is better obtained through other data resources.

C5.9 Would have been great to have the draft questionnaire in advance. Perhaps I would have circulated the questionnaire to my staff for comments. Would have been better prepared to participate in the afternoon session.

R5.9 The questionnaire will be made available before the second NAOMS industry workshop.

C5.10 Stop what you are doing and let's look at things from a systematic perspective – define the problem space, identify current data collection efforts, then identify gaps. Let's examine what is being done and why. Start with a small core group of NASA/FAA folks to develop a "straw man" integrated approach. Add to this group slowly.

R5.10 See Response 1.3. The NAOMS industry workshops along with input from the NASA AvSP Executive Committee (FAA, NASA, industry associations) have served as a review and advisory panel throughout the development process. The NAOMS program will proceed in deliberate, incremental steps beginning with an air carrier pilot implementation in FY2000.

Q6. Other Comments:

C6.1 One of the reasons cited for the NAOMS concept in lieu of using FOQA data is that it will be many years before all the major U. S. carriers have FOQA programs. (There are 10 U.S. carriers (excluding cargo carriers that are considered major according to the Bureau of Transportation Statistics (BTS) definition of major, i.e. $1 billion or more in revenues). Airlines that currently have FAA-approved FOQA programs include Alaska, Continental, Delta, TWA, United, and US Airways. According to the Air Transport Association, a number of additional carriers are working on FOQA programs and they expect as many as 10 carriers to have FOQA programs by the end of the year.

R6.1 See Response 4.2.

C6.2 If you have other workshops, provide more information prior to the meeting, e.g. the draft survey instrument was first provided to participants at the May workshop.

R6.2 Advance materials will be made available.

C6.3 It is important to get feedback from the industry on safety issues, but don't know if this is the best process.

R6.3 It is the best process that NASA has been able to identify but NASA is always open to new possibilities.

C6.4 Think NASA is taking their charge seriously and applaud them for their effort.

R6.4 Thank you.

Appendix J. Workshop 2 Agenda and Participants

Workshop 2 was held on March 1, 2000, in Washington, D.C. Its purpose was to update stakeholders on progress being made toward NAOMS implementation, especially the results of the field trial. During the workshop, the participants formed breakout groups and provided comments, questions, and recommendations to the NAOMS team.

This appendix contains the workshop agenda and attendance list, as well as feedback from workshop participants to issues posed for discussion by the NAOMS team.

Workshop 2 Agenda

8:00 to 8:30 A.M.............Registration

8:30 to 9:00 A.M.............Welcome and Opening Comments; Introduction of NAOMS Team & Workshop Goals. *Linda Connell*, NASA Project Manager

9:00 to 10:00 A.M...........Project Background: Goals, Development and Experimental Work, Questionnaire Development. *Robert Dodd*, Sc.D., Project Manager, Dodd and Associates

10:00 to 10:15 A.M.........Break

10:15 to 11:00 A.M.........Conducting the NAOMS Field Trial. *Joan Cwi*, Ph.D., Battelle

11:00 to Noon ....Field Trial Findings: Mode Effects and Recall Periods. *Jon Krosnick*, Ph.D., Ohio State University

Noon to 1:00 P.M............Lunch

1:00 to 2:00 P.M. ....Field Trial Findings: Feedback from Participants. *Elisa Ingebretson*, Research Scientist, Battelle

2:00 to 2:30 P.M. ....Next Steps. *Linda Connell*, NASA Project Manager

2:30 to 3:00 P.M. ....Break

3:00 to 4:45 P.M. ....Discussions

4:45 to 5:00 P.M. ....Summary and Closing Comments. *Linda Connell*, NASA Project Manager

5:00 P.M..........................Adjourn

Workshop 2 Attendance List

First Name	Last Name	Organization	Office	Job Title
Ralph	A'Harrah	NASA	Office of Aerospace Technology	Goal Manager, Aviation Safety
Jim	Burin	Flight Safety Foundation		Director of Technical Programs
Doug	Carr	NBAA	Domestic Operations	Manager
Linda	Connell	NASA ARC; 262-7		Director Aviation Safety Reporting System
Mary	Connors	NASA ARC; 262-4	NASA Aviation Safety Program	
Joan	Cwi	Battelle		Director of Survey Operations
Robert	Dodd	Dodd and Associates		Principal Investigator
Bill	Edmunds	ALPA		Human Performance Specialist
Ray	Fenster	Fenster	Information Overload Corporation	
Charles	Fluet	Federal Aviation Administration	Office of Integrated Safety Analysis	Deputy Director
Michael	Ganley	Airbus Industrie of North America		
Larry	Hackler	Federal Aviation Administration AAR-424	Technical Center	
Charles	Harrison	Federal Aviation Administration ASW-110	Rotorcraft Directorate	
Chris	Hart	Federal Aviation Administration, ASY-1	Office of System Safety	Assistant Administrator for System Safety
Chuck	Hedges	Federal Aviation Administration, ASY-300	Office of System Safety	Manager, Systems Safety Engineering & Analysis Division

First Name	Last Name	Organization	Office	Job Title
Priscilla	Hospers	Battelle	ASRS	
Elisa	Ingebretson	Battelle	ASRS	Research Scientist
Mike	Jobanek	Florida Technical		Aviation Domain Consultant
Ray	King	HQ Air Force	Safety Center	AFSC/SEPR
Jon	Krosnick	Ohio State University	Department of Psychology (Social)	
Mike	Lewis	NASA Langley Research Center	Aviation Safety Program Office	
Harkey	Mayo	FAA ASY-100	Office of System Safety	Data Systems Manager
Tom	Nesthus	Federal Aviation Administration	Civil Aeromedical Institute	
Albert	Prest	Air Transport Association		
Loren	Rosenthal	Battelle		
Mike	Schanck	General Aviation Manufacturers Organization		Safety Affairs and Operations Manager
Vincent	Schultz	NASA Langley Research Center		Program Manager
Nan	Shellabarger	Federal Aviation Administration	Office of Aviation Policy and Plans	
Michael	Silver	Ohio State University	Department of Psychology (Social)	
Stan	Smith	National Transportation Safety Board	Data Systems	Manager
Lee	Snowberger	Conwal		Program Manager
Arthur	Salomon	Federal Aviation Administration, APO-110	NASA Aviation Safety Program	
Irv	Statler	NASA	NASA Aviation Safety Program	
Bruce	Tesmer	Continental Airlines	Flight Crew Performance	Captain, Manager

First Name	*Last Name*	*Organization*	*Office*	*Job Title*
Jim	Varsel	International Association of Machinists and Aerospace Workers		Assistant Airline Coordinator
Carla	Winkler	International Association of Machinists and Aerospace Workers		
Dick	Wright	Helicopter Association International	Safety and Flight Operations	Director
Brien	Wygle	Aerospace Industries Association, Retired	Boeing, Subcommittee	Chairman ASRS

Feedback from Workshop Participants

This section presents the discussion issues posed to the workshop participants and provides a bulletized summary of their responses.

Survey Content

- In Section A, regarding the potential inclusion of International Operations, item A3 should be redesigned to avoid errors. A distinction should be made between domestic and international flying, especially regarding ATC and language problems with international ATC.
- Consider adding autorotation/emergency procedures if you are going to look at rotorcraft operations.
- Consider adding “execute emergency procedure” to list of events.
- Consider adding autorotation to list of events for helicopter pilots.
- Regarding FC4 and 5 (sterile cockpit), different companies and the government have different regulations about sterile cockpit and flight time/duty time restrictions. Clarify in the questionnaire what is being asked about.
- Consider having a question that doesn't constrain the respondent to a recall period but instead allows him to report on any life-changing event that may have occurred in his career.
- Make sure unions are involved in the development of items for all future questionnaires.
- Was a fault tree analysis used to look at causal factors? Use a “fault tree approach” to identify item types for future surveys.

- How flexible are the responses allowed to be? For example, if the recall period is four weeks, but a pilot experienced something 4.5 weeks ago that he/she wants to report, how can the pilot report that?
- Include government flight operations (FAA, etc.) in future surveys.
- What does "engine exhaust" refer to in the Main Events section?
- Consider adding a question about "loss of situational awareness."

Data Protection

- How is NASA going to protect the data?
- When will data be released?
- Will the data be indefinitely confidential?
- The FAA's new Advisory Circular could potentially cover the pilots. NASA could protect the data as a "research instrument" for a while. Others could help NASA analyze the data when it is ready for manipulation.
- How did NASA decide on the specific MEL and ICAC questions and sections?
- Consider using the safer skies model for topical sections.
- CAST could help to develop ideas for topical sections. Consider making a presentation to CAST.

Other Comments

- The FAA's General Aviation survey work could help the NAOMS team, and vice versa.
- Will there be more workshops in the out years?

Appendix K. Field Trial Results

Authors' Note: In April 2000, after all data had been collected, the NAOMS team produced the following final report on field trial results. The results pointed the way toward implementation of the full air- carrier (AC) survey.

NAOMS Field Test Results: Implications For Full Project Implementation

Introduction

The National Aviation Operations Monitoring Service (NAOMS) field trial was designed to evaluate the feasibility of collecting primary data on aviation safety events from air-carrier pilots. The field trial enabled evaluation of issues needing resolution before a full-scale survey for air carrier pilots could be initiated. These issues included the rate of response, the accuracy and quality of the information collected, most effective recall period, necessary sample size, and the projected cost. Central to all these issues is the selection of the mode of survey application: mail or telephone.

This document provides a summary of the key issues associated with selecting the appropriate mode for full-scale implementation of NAOMS for the air-carrier-pilot community and the appropriate recall period. Each issue is addressed separately and the implications for each mode discussed. A summary matrix of the characteristics of each of these topics by mode is provided at the end of this document.

K1. Mode Selection Factors

K1.1 Completion Rates

One of the more important dimensions of selecting a mode for survey application is the rate of completion to the questionnaire (that is, the number of people who complete the questionnaire from the pool of eligible respondents). Usually, higher completion rates are better since the basis for conducting a sample based survey is the desire to apply the findings to a larger total population. Generally speaking, a completion rate has to exceed 70% for the findings to be accepted as representative of the total population. Lower completion rates may indicate that a significant portion of the sample (those who chose not to respond) may differ markedly from those who did respond. If so, generalization to the full population from a sample with inadequate responses may be erroneous. This was a concern in the NAOMS field trial. If completion rates were low, it might have been due to the fact that those pilots were more prone to safety problems and not willing to admit this fact to the researchers. This did not turn out to be a problem.

Table K-1 presents the response rates by telephone and mail modes attained during the field trial. It should be noted that these completion rates did not occur as a result of the first contact with the survey respondents. For most respondents, more than one request was required before a successful interview was completed regardless of mode. Additional contact was required because the pilots did not respond to earlier requests, had scheduling conflicts, lost the original mailing, etc.

Table K-1. Completion Rates by Mode		
	Mail	*Telephone*
Response rate	*70%*	81%

K1.2 Data Quality

The quality of data collected is also an important consideration when evaluating what mode of survey should be selected. A high completion rate is not of much value if questions are not answered accurately. There are a number of approaches to evaluating data quality, each of which is presented below.

K1.2.1 Time for Questionnaire Completion

Evaluation of the time needed to complete the interview is a relative measure of data quality. The underlying assumption is that the more time a respondent takes to complete a questionnaire, the better the quality of the resulting data. Table K-2 provides the key data.

As can be seen, the average time to complete the telephone interview took 12 minutes more (70% more) than the mail mode. Some of this difference may be due to the need for the respondent to listen and then assimilate what the interviewer asked in the telephone interview versus the ability of the respondent to quickly read the question in the mail interview. It is unlikely, however, that this explanation explains all the difference in average completion time for the two modes. The lesser amount of time needed to complete the interview when conducted by mail may be indicative of pilots working through the questionnaire quickly, thereby paying less attention to questions or spending less time trying to accurately recall the events.

Table K-2. Questionnaire Mean Completion Time		
	Mail	*Telephone*
Completion time	*17 minutes*	29 minutes

K1.2.2 Missing Responses

Another way to evaluate the quality of data reported is to look at the number of missing responses for the questionnaire. Table K-3 presents the percentage of respondents that did not complete at least one question in the questionnaire by mode.

Table K-3. Percentage of Respondents Who Failed to Complete at Least One Question		
	Mail	*Telephone*
One or more missing answers	*4.8%*	0.0%

The lack of any missing answers for the telephone mode is due to the fact that each question, when read by the interviewer, requires a response. Since most of the responses to the questions in this survey appropriately received the response of '0' (for incidents that did not occur during the recall period) it is easy to see how respondents would be tempted to skip quickly across questions in the instrument. This would also explain why the mail version of the questionnaire took so much less time to complete than the telephone version. In the mail version, the pilots did not have anyone prompting them to slow down and think about each answer. In contrast, the interviewers during the telephone interview asked the pilot each question in turn. The pilot did not know what question came next so he or she had to listen to the question to understand its meaning and then think to develop a response.

K1.2.3 Total Number of Events and Total Hours Flown in the Recall Period

The observed relationship between the reported number of events and the total hours flown in the recall period also provides insight into data quality. If the questionnaire is capturing accurate responses from pilots about the frequency of events they experience, then pilots with more flight time should experience and report a proportionately greater number of events than those pilots who flew fewer hours.

Several quantitative analyses were conducted looking at the association between the number of events reported and the number of hours flown during the recall period. For all such analyses, one would expect to see a positive relationship between the variables if the data are valid (more flight hours should result in greater number of events reported). This tight relationship is evidenced by a higher coefficient of regression (COR). Higher CORs indicate stronger relationships between the data. Table K-4 shows the pertinent findings.

Table K-4. Demonstrated Association between Number of Events Experienced and Hours Flown

Mode	*Unstandardized Coefficient of Regression*	*Significance*	*Number of Respondents*
Mail	*.086*	*p<.001*	223
Telephone	*.136*	*p<.001*	220

Associations for both modes were positive, indicating that pilot reports of event frequencies corresponded to experience during the recall period. However, data from the telephone mode showed a somewhat higher degree of association between number of events reported and the number of hours flown indicating greater consistency in the data.

K1.2.4 Total Number of Events and Number of Days in Recall Period

By similar logic, one would expect respondents who were asked to use longer recall periods to report proportionately more events than those asked to use shorter periods. Table K-5 shows the relationships found in the data. Once again, the relationship was positive with both modes, but it was considerably stronger with the telephone mode. This suggests that telephone respondents were working harder to recall events accurately for the longer recall periods.

Table K-5. Demonstrated Association between Number of Pilot Reported Events and Recall Period

Mode	*Unstandardized Coefficient of Regression*	*Significance*	*Number of Respondents*
Mail	*0.190*	*p<.001*	228
Telephone	*0.265*	*p<.001*	220

K1.3 Recall Period

Another key objective of the NAOMS field trial was to determine the appropriate recall period. There were two primary competing considerations. First, as recall period lengthens, the memory of events weakens. Shorter recall periods promote quality.[41] Second, longer recall periods favor the recollection of more events permitting more events to be uncovered from fewer respondents. Longer recall periods promote cost savings. Respondents were asked to use a variety of recall periods ranging from one week to six months during the field trial. The resulting data have helped the NAOMS team find the point that provides a reasonable balance between these two considerations.

K1.4 Reporter Confidence

One way to address the effect of the various recall periods used during the field trial was to ask the respondents how confident they were in the answers the had provided. The results are summarized in Table K-6.[42]

[41] NAOMS did some earlier experimental work with a group of air-carrier pilots who were asked to recall the number of landings made in the previous month. It was found that shorter recall periods were more accurate. For routine events like recalling number of landings, accuracy fell off sharply after one week.

[42] This table was derived from analysis of pilots who completed the survey by either telephone or by mail. Face-to-face interview results were not included.

Table K-6. Respondent Confidence in Recollection Accuracy by Recall Period			
Recall Period	*Extremely Confident*	*Very Confident*	*Moderate to No Confidence*
1 week	62%	36%	*4%*
2 weeks	58%	34%	8%
4 weeks	47%	39%	14%
2 months	36%	49%	15%
4 months	34%	47%	19%
6 months	29%	44%	27%

It can be seen that the confidence the pilots have in their ability to accurately recall events dropped markedly as the recall periods got longer. However, at 60 days, 85% of respondents still indicated that they were either 'Extremely' or 'Very' confident in their recall. This suggests that a 60- or 90-day recall period may strike the best balance between quality considerations and the need to operate NAOMS in an economical manner.

K1.5 Pilot Comments

The questionnaire allowed pilots to offer free-form observations about the questionnaire and the interview process. One recurring suggestion was that the recall period be increased. The majority of these comments came from pilots who were assigned the one- or two-week recall periods. They felt that these periods were too short for them to report events they remembered experiencing that fell outside the recall window.

K2. Costs

The NAOMS field trial provided an opportunity to strengthen earlier estimates of the cost of doing this work. Cost elements include:

- Project management and administration including OMB interactions
- Time devoted to industry/labor interactions to build and maintain program support
- Development and testing of survey instruments including topical sections
- Development and maintenance of a database to hold survey results
- Survey data collection
  - Mailings and postage (self-administered)
  - Interviewer training and interview time (telephone)
  - Establishing and maintaining respondent tracking programs[43]
- Data analysis
- Deliverables preparation

[43] NAOMS will track whether or not a respondent has completed a written survey or participated in a telephone survey session. It will *not* maintain any record of the actual responses provided by any participant.

NAOMS is intended to serve multiple constituencies including air carrier pilots, GA pilots, air traffic controllers, and others. NAOMS costs will increase in a linear fashion with the addition of each new participant group. Modest savings in administration, training and database development areas will be realized when new groups are added. However, these savings will be fully offset by the costs of engaging the new participants in the NAOMS project, particularly development of a customized survey instrument. Accordingly, serving two participant groups will be twice as expensive as serving one group, and so on. Table K-7 provides high-level estimates of the cost of conducting NAOMS as an ongoing production for a single participant group. The estimates shown cover all direct and allocated costs, but they do not include contractor fees.

Table K-7. Estimated Cost of a Fully Operational NAOMS Program for One Participant Group (1999 Dollars)

Cost Estimate	*Estimate*	*Comments*
Project management and administration and Industry/Labor interactions	$125K	
Development and testing of survey instruments	$100K	Assumes 4 sets of topical questions developed and tested on 100 respondents each year
Data system maintenance and administration	$50K	
Data collection, telephone	$408K	4,800 completed interviews @ $85
Data collection, self-administered	$322K	4,800 completed questionnaires @ $67
Data analysis and deliverables preparation	$200K	Assumes quarterly reports and an annual report
Total (before fee): telephone	*$883K*	
Total (before fee): self-administered	*$797K*	

K2.1 Other National Data Collection Systems

Telephone interviewing is the preferred method for many government survey programs. Most of the other long-term government data gathering efforts use the face-to-face mode despite its higher cost to maximize data quality. The underlying rationale is that improved data quality is worth the higher data-collection costs. Examples of ongoing surveys that use the telephone mode include:

- Survey of Income and Program Participation (Census Bureau) 1984–
- Consumer Expenditure Surveys (Census Bureau) 1968–
- Annual Housing Surveys (Census Bureau) 1973–
- Consumer Attitudes and Behavior (SRC) 1953–
- Health and Nutrition Examination Surveys (NCHS) 1959–

- National Health Interview Surveys (NCHS) 1970
- American National Election Studies (NSF) 1948–
- Panel Study of Income Dynamics (NSF) 1968–

Many firms compete to provide support services to the government for these programs. Examples include the Gallup Organization, Westat, SPSS Services, Research Triangle Institute, and Mathematica, to name just a few.

K3. Conclusions and Recommendations

The NAOMS field trial was a highly successful undertaking that shed light on many methodological issues. The following paragraphs summarize the NAOMS team's recommendations for implementing the full NAOMS system. The recommendations are based on the field trial results, input from senior survey methodologists, and aviation safety domain experts.

K3.1 Survey Mode

The weight of the evidence proceeding from the NAOMS field trial strongly suggests that telephone is the preferred NAOMS data-collection mode. All data indicators suggest that data collected by phone will be of substantially higher quality and will have few inappropriate outlier values (due to question misinterpretations, etc.) that have the potential for confounding NAOMS data analyses. The literature also suggests that the telephone mode will consistently yield better quality data than self-administered surveys. This is the reason that most federal agencies that have implemented long-term survey data collection efforts have chosen face-to-face or telephone modes.

K3.2 Recall Period

The literature on survey methodology and theoretical considerations favor shorter recall periods when accuracy is a paramount concern. On the other hand, longer recall periods would be expected to result in higher observation rates and potentially more economical data collection. It is clear from the NAOMS field trial data that data accuracy declined as recall periods were extended. The fall-off in participant confidence in the accuracy of their responses was particularly noticeable when the recall period was lengthened from two to four weeks. However, when the recall period was further extended from one to four months, the decline in respondent confidence was relatively small. In fact, more than 80% of respondents said that they were "extremely or very" confident in their inputs when a four-month recall period was used.

Since NAOMS research has been inconclusive on this issue, a split design is recommended for the first year of implementation. Under this design, half of all respondents would be asked to use a 30-day recall period; the other half would be asked to use a 90-day period. The data would then be evaluated at the end of one year. If the longer 90-day recall period does not appear to materially compromise data quality, it should be adopted since it is the more economical approach. Otherwise, the 30-day recall period would be preferred.

K3.3 Random versus Panel Design

The field trial itself did not address the issue of random versus panel designs. The literature indicates that a purely random approach is statistically optimal. It is usually easier to administer random designs as well. However, the domain experts on the Team tend to prefer the panel approach. The rationale underlying the NAOMS effort is that the aviation community—pilots, controllers, mechanics, flight attendants, and others—are a highly professional and generally well educated group who can be enlisted as active monitors of NAS safety. It is further believed that enrollment in NAOMS panels will cause participants to become even more acute observers of aviation system safety.

These competing considerations also suggest that a split design would be desirable in Year 1 of the NAOMS implementation. Approximately half of data collection could be accomplished with respondents chosen on a purely random basis from the sample pool and the other half with respondents who agreed to join a NAOMS panel. Each panel respondent[44] would be asked to enroll for one-year period with the expectation that he/she would be asked to participate in four surveys spaced at three-month intervals.

K3.4 Sample Size

An annual sample size of 4,800 observations (400 per month) is recommended. While a larger sample size would give both greater precision and accuracy, a sample of 4,800 should be sufficient to detect relatively modest downward or upward trends in the occurrence of infrequently occurring aviation safety events.

[44] Some participants in the first year would be asked to enroll for several additional quarters so that one-fourth of panel participants could be replaced each quarter beginning in Year 2.

Appendix L. OMB Submission

NAOMS, like all significant government data collection efforts, is required to secure Office of Management & Budget (OMB) approval prior to the commencement of data collection activities. The NAOMS project team submitted the following document to OMB in June 2000, in support of its request to commence active data survey operations.

APPLICATION FOR OMB CLEARANCE

This submission requests approval for a data-collection process that will be used to help evaluate national aviation safety through the establishment of a survey-based methodology. The system will be called the National Aviation Operations Monitoring Service (NAOMS) NASA will operate NAOMS on behalf of government and industry.

NAOMS will use information provided by the "first-line" operators of the aviation system to measure and monitor aviation safety; namely, the pilots, air traffic controllers, mechanics and flight attendants who routinely operate aircraft and provide support services.

While NAOMS is envisioned to be an ongoing process involving many aviation stakeholder groups, this submission requests approval for collecting safety information from commercial air- carrier-pilots for a period of one year. Workshops, focus groups, and field trials have been conducted to help ensure the feasibility of NAOMS and to lay groundwork for its design. Many questions have been answered, but some issues are not yet fully resolved. The latter will be addressed during the first year of the survey effort.[45] NASA will make revised submissions to OMB thereafter for continued operation and development of NAOMS.

NASA, as a research agency with an exemplary reputation in aviation safety and human factors research, will conduct the survey. The trust and confidence engendered from their lengthy history of successful research efforts and proven protection of confidentiality will provide the basis for participation by aviation personnel in this important project.

Justification

I.A. Explain Why the Data Collection is Necessary

To improve aviation safety one must be able to measure safety in both relative and absolute terms. To know whether the actions taken to improve safety are working, one must be able to measure safety trends over the near- and long-term. Reliable measurements, such as these, in the current National Aviation System (NAS) are limited and often absent. This is a critical problem for both aviation policy makers and technology developers. It constitutes a critical technical challenge for the nation's safety efforts both in the government and the aviation industry.

[45] The primary methodological issue relates to whether a cross-sectional or panel design is the most effective and efficient approach for this project.

The White House Commission on Aviation Safety has articulated the nation's safety goal, which is to improve NAS safety by 80% over the next decade[46] This proposed data-collection process provides the essential tool for measuring the effectiveness of interventions and improvements intended to meet that goal.

I.B. Indicate by Whom and for What Purpose the Information Will be Used

Each year many new aviation safety investments are proposed. Many other proposals are carried over from previous years. Some of these proposed investments involve new technologies; others entail improvements in training, facilities, equipment, and procedures. Managers in government and industry must sort through these alternatives and make hard choices. These managers want this to be a rational, data-driven process. This requires numbers that quantify the safety risks these investments are expected to reduce, numbers that reveal trends portending future safety problems, and still more numbers that measure the effectiveness of past safety investments.

Further, the aviation safety efforts underway to meet the goals of the White House Commission on Aviation Safety need accurate numbers on the frequency and causes of safety events. Additionally, the NASA's Aviation Safety Program (AvSP), as well as other government and industry efforts, are developing technologies and proposing procedural changes that are intended to improve aviation safety. NAOMS' metrics will be used to determine how well these safety investments have worked when they are implemented into operating environment.

The information developed by NAOMS will be used by NASA's AvSP managers to evaluate the progress of their efforts to improve aviation over the next decade. Information will also be made available to other Government agencies such as the Federal Aviation Administration (FAA), Department of Transportation (DOT) and the National Transportation Safety Board (NTSB). Private organizations such as universities and aviation industry associations will also be provided information if they so desire. It is anticipated that these and other academic and private sector organizations will use NAOMS information to evaluate their own safety programs and performance and to help develop well-informed safety policies.

I.C. Whether, and to What Extent, the Data Collection Will Use Automated, Electronic, Mechanical, or Other Technological Methods for Data Collection

Interviews will be conducted using Computer Assisted Telephone Interviewing (CATI). This technology ensures that all skip patterns included in the survey are automatically and correctly followed, provides pre-established prompt cues to interviewers, and provides a clean data file immediately after the termination of the interview. Prior data collection efforts have determined that there was a qualitative advantage to the CATI versus self-administered method. In addition, standardized interviewer prompts will be developed within the CATI system to address issues discovered during the field trial.

A management information system (MIS) developed by Battelle (the contractor that will be employed by NASA to conduct the proposed data collection) will be used to track mailing schedules (advance letters), the release of new names, appointments, and final dispositions in the CATI center. The system will provide study data to the NAOMS survey managers who will use the MIS to maximize the survey response rate and ensure data are collected efficiently and with minimum burden to respondents.

---

[46] *Final Report to President Clinton, White House Commission On Aviation Safety*, Vice President Gore, Chairman, Washington D.C. February 12, 1997.

In future years, Web interviewing techniques will be carefully explored. NASA needs to reassure itself that responses obtained over the Web have the levels of quality and completeness necessary to the NAOMS mission.

I.D. Describe Efforts to Identify Duplication

A significant effort was undertaken to identify and evaluate other aviation safety data-collection efforts currently in progress or planned. Two government-industry workshops were conducted on the proposed NAOMS effort. One of the central purposes of these workshops was to identify viable alternative data sources that would obviate the need for NAOMS. None were identified. The following paragraphs further explain why a new aviation safety data resource is needed.

I.D.1. Characteristics Desired in the NAOM Aviation Safety Measurement System

To achieve the desired functionality, NAOMS aviation safety metrics must be:

- ACCURATE – So that policy makers, technology developers, and the other data users would have sufficient confidence to act on the data they have been given;
- COHERENT – So that data are conducive to seeing the big safety picture and understanding the key forces and trends influencing NAS safety;
- PRECISE – So that decision-makers have enough operational details to develop credible responses to identified safety deficiencies;
- QUANTITATIVE – So that decision-makers can move beyond qualitative assessments of safety issues and formally assess risk;
- STABLE – So that data users do not perceive “trends” or “shifts” that are really statistical artifacts resulting from inadequate or inconsistent data collection methods;
- COMPREHENSIVE – So that the metrics address commercial, corporate and general aviation operations; on large- and small-scales; during all-airborne and ground phases; in all relevant environments including flight decks, aircraft cabins, control facilities, maintenance hangars, and other operational environments;
- HUMAN-CENTERED – So that the data help to illuminate the human performance issues that are central to aviation safety and to the ultimate success or failure of aviation technologies;
- TOPICAL – So that the data address the issues that are of prime interest to policy makers, technology developers, and other members of the aviation community;
- ACCESSIBLE – So that political, legal, or commercial considerations do not inhibit practical application of the data to current safety issues.

Many valuable sources of aviation safety data currently exist, but none have all of the virtues outlined above. The main area where existing data sets fail is in the quantification of NAS safety and safety trends. The current absence of solid quantitative information places a significant burden on the aviation decision-makers who must assess risk and allocate scarce resources among competing safety investments.

I.D.2. Deficiencies in Existing Sources of Aviation Safety Data

One might hope that existing data sources, or some combination of those sources, would provide the needed quantitative information. Several potential data sources were identified:

- Accident Data, Commercial Flight Operations
- Aviation Incident Data

• Air Carrier Digital Flight (FOQA) Data
• Air Traffic Control Radar Tapes

The following passages examine these data sets and explain why the information they contain is highly complementary, but not fully commensurate with the needs outlined earlier.

l.D.2.1 Commercial Air Carrier Accident Data[47]

Major air transport accidents are mercifully rare events in both the statistical and practical senses. Major U.S. air carrier operators (FAR part 121) in scheduled service experienced an average of just 3.4 fatal accidents over the 1982-1997 time period.[48] Many millions of flight operations were safely accomplished during this same time frame. Rare events, like air carrier accidents, are often characterized by Poisson distributions. If the accidents of Part 121 operators really are an outcome of a Poisson process with a mean of 3.4, one would expect these data to display a great deal of year-to-year volatility with observed values generally ranging between zero and seven. As Figure 1 indicates, actual air-carrier-accident data display just such traits. It is obvious that such accident data are too coarse and volatile to form the basis of a comprehensive NAS safety measurement system.

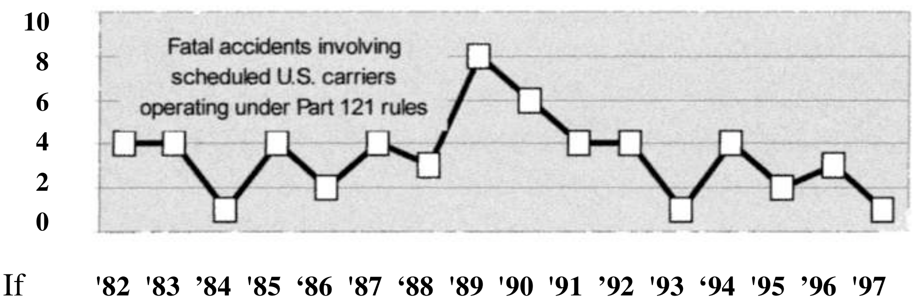


Figure 1. Commercial Air Carrier Data are Too Coarse and Volatile to Measure Short- and Medium-Term Safety Trends

l.D.2.2. Aviation Incident Data and Other Event Data

Aviation safety incidents can be defined as any unsafe event or situation that occurs within the NAS but does not culminate in an accident. Aviation safety incidents are far more common than accidents, and would be expected to more reliably measure safety trends. Aviation incidents come to light in a variety ways. One important type of incident data is the incident report such as those submitted to the NASA Aviation Safety Reporting System (ASRS), to the FAA’s NAIMS database, and to various other FAA, NTSB, and air carrier data-collection systems.

These incident reports are mainly composed of free-form textual descriptions of undesirable occurrences. These reports are often highly topical, shedding light on current safety problems. The ASRS has demonstrated that such data can be used to provide powerful insights into human performance and the dynamics of safety events, the qualitative side of safety data.

The difficulty is that the incident data collected by NASA, FAA, NTSB, and others suffer from unmeasured reporting biases. These data-collection systems are unable to calculate what portion of the total universe of safety events is reported to them each year. They do not know whether near-midair collisions are more likely to be reported than altitude deviations; or whether carrier X’s pilots are more likely to report than carrier Y’s; or whether an unusually large number

[47] This discussion relates to data describing commercial aviation accidents in the U.S.
[48] The source is the National Transportation Safety Board (NTSB)

of controller reports coming from region Z results from a greater awareness of the incident reporting programs or to poorer system performance.

The data of every incident reporting program suffers from the same statistical limitations. This does not negate the value of incident data for explaining safety events and human performance issues, but it does make them relatively limited source of quantitative safety information.

l.D.2.3. Air Carrier Digital Flight (FOQA) Data

European air carriers have routinely collected and analyzed flight data for many decades; U.S. air carriers are beginning to do the same. Modern air transports record thousands of parameters that are available for collection and analysis via the on-board digital flight data recorders. Virtually every control surface, control position, instrument reading, display mode, and switch position can be monitored and recorded by these recorders.

Enormous amounts of digital flight data may eventually be collected under air carrier flight operational quality assurance (FOQA) programs. However, it will probably be many years before all major U.S. air carriers have such programs, and many more years before smaller air carriers and GA operators institute FOQA programs. When these data do become available, they will be among the most proprietary safety data possessed by aircraft operators. They will be protected for reasons of public relations, competition, and fear that they may be used in the course of litigation.

Digital flight data are an unexcelled quantifier of aircraft flight performance, but the range of issues they can address is limited by the nature of the sensing devices. These data may ultimately be used to make inferences about aircrew performance, but it is unlikely that they can be used to measure the performances of individual crewmembers. The inferences that can be made will generally document human behavior but not explain it. Moreover, flight data recorders do not currently capture contextual information vital to interpreting observed behaviors (e.g., operative ATC clearances).

l.D.2.4. ATC Radar Data

The FAA routinely records the radar tracks of aircraft within airspace under its control. These radar data have many of the virtues and weaknesses of digital flight data. They measure the gross movements of aircraft in controlled airspace. However, the parameters collected by ATC radar tapes are relatively few in number compared to the flight data recorders on board modern aircraft; they provide relatively little data about participants who do not participate in the air traffic control system, and even less about human performance issues. (The latter include the plans, thoughts, expectations, and actions of the people who operate the aviation system; the efficacy of human-technology interfaces; and the effects of training, schedules, workload, and other factors on human performance.) ATC radar data would probably be suitable for measuring total system safety performance in a gross sense, but the range of safety issues that could be addressed would be very limited for the reasons outlined above.

l.D.3. Reliably Quantifying NAS Safety

Each of the data sources described above contributes to the total body of information available about NAS safety performance. This body lacks a skeleton—a framework for assembling and arranging the information so that it makes sense in a total, integrated way. A comprehensive, structured approach is needed to measure NAS safety. This structure should first define key safety metrics in accordance with the government and industry needs and then measure NAS performance against those metrics. These data would add perspective to the many other kinds of aviation safety data collected. They would be invaluable for assessing safety risks and for achieving a rational allocation of safety investment resources.

Accordingly, NASA proposes to undertake the development of NAOMS for surveying the operators of the national aviation system - its pilots, controllers, mechanics, and flight attendants—to obtain critical quantitative data on NAS safety performance. NAOMS would, in effect, view the aviation-operating environment through the eyes of the many thousands of persons who make it work. These persons would be regularly asked about their operational exposure; the rates at which they experience safety events; and their experiences with new technologies and procedures. NAOMS would also address the human performance issues that are crucial to understand in relation to new aviation technologies and their success or failure.

These survey data will be collected voluntarily in accordance with a rigorous statistical design. A well-designed, carefully implemented system will produce data that have all of the ideal attributes enumerated at the beginning of this submission: accurate, coherent, precise, quantitative, stable, comprehensive, human-centered, topical, and accessible. Most importantly, the data collection effort would be focused on precisely those topics that are of greatest importance to government and industry.

I.E. Data Collection Impacts On Small Businesses or Other Small Entities

There are no small entities involved with this study. Potential respondents for this study include approximately 5,000 commercial airline pilots who will be randomly selected from the Federal Aviation Administration's Airman's Medical Certification database that includes a list of certified pilots (N~700,000) operating in the United States.

I.F. Describe Consequences if Collection is Not Conducted or is Conducted Less Frequently

As described above, failure to collect these data will result in maintaining the status quo that is an inability to accurately measure the progress in meeting the aviation safety goals as described in the White House Commission on Aviation Safety report published in 1997, and thus improved public safety. It is anticipated that policy makers will wish to review progress towards safety goals on at least quarterly basis (preliminary evaluations). They will need definitive safety measurements annually. The NAOMS sample design has been structured in accordance with these considerations. Reducing the planned data collection frequency (described in more detail later in this submission) will invalidate the underlying statistical design and reduce or eliminate NAOMS ability to generalize findings from the sample to the air carrier pilot population at large.

I.G. Describe Special Circumstances that would Cause Information to be Collected in a Manner…

<u>1.G.1 More Often than Quarterly</u>

NAOMS will not ask respondents to provide information more often than once per quarter.

<u>I.G.2. Responding Fewer than 30 Days</u>

NAOMS will not ask respondents to respond in fewer than 30 days.

<u>I.G.3. Submission of More than an Original and Two Copies of a Document</u>

NAOMS will not ask respondents to submit any documents.

<u>I.G.4. Requires Respondents to Retain Records for More than 3 Years</u>

NAOMS will not ask respondents to retain records for the purpose of this study.

<u>I.G.5. That Cannot be used to Generalize to the General Population</u>

The data collected by NAOMS will not be generalized to the general public. However, it will be generalized to the air carrier population, which is the intent of this survey.

I.G.6. Using a Statistical Data Classification that has not been Reviewed and Approved by OMB

NAOMS will not be using any standardized data classification schemes. Data classification will categorize the unique events reported by air carrier pilot. This is necessary to support the analytical needs of NAOMS described elsewhere. NASA will bear the full cost entailed by prescribed classification coding scheme. The classification-coding scheme will not shift agency information costs to respondents.

I.G.7. That Includes a Pledge of Confidentiality

The participants in this data collection effort will be assured of data confidentiality. For more on this see section I.J. below.

I.G.8. Requires Respondents to Submit Proprietary Trade Secrets or Confidential Information

NAOMS respondents will not be asked to submit proprietary trade secrets, but they will be asked to provide confidential information. All information will be recorded anonymously in manner that prevents its linkage to any given respondent.

I.H. Provide a Copy of Federal Register Requesting Comments on Data Collection

[To Be Inserted by NASA]

I.I. Explain Decision to Provide Payment or Gift for Participation

NAOMS does not intend to provide payment or other incentives to respondents.

I.J. Describe Any Assurance of Confidentiality if Based on Statue, Regulation, or Agency Policy

The data being collected are potentially sensitive since respondents will be reporting safety-related events in which they are involved. Data will be collected anonymously to prevent the disclosure of respondent identities. The CATI interview management system is programmed so that there is no linkage between the respondent’s database and the completed interview. Once completed, the completed interview will be electronically submitted without any linkage to identifiers. The interviewer will note in the system that the respondent has completed the interview in order to prevent re-contacting them, but the respondent ID will not be linked to the recorded questionnaire responses.

In this current proposal, pilots will be notified in an advance letter and consent statement that the study is anonymous. A brief explanation will be given about how NASA and its supporting contractor will assure this anonymity. In addition to anonymity of reporting, NASA and its supporting contractor will put the following confidentiality protection measures in place:

- The identity of respondents will not be revealed to anyone outside of the study staff.
- The data presented in reports and publications will be in aggregate form only.
- The respondent will be assured that participation is completely voluntary and in no way affects their employment.

6

All interviewers and project staff will sign a confidentiality statement attesting that they will maintain the privacy of all information collected through the conduct of this study. The confidentiality statement includes non-disclosure of any information learned about respondents or the data obtained. It further requires interviewers to decline interviewing a respondent he or she knows personally.

I.K. Provide Justification for Any Questions of Sensitive Nature Including Why the Questions Are Necessary, the Specific Uses of the Information, and the Explanation Given to the Respondents

Respondents will be asked questions about safety events that happened during the conduct of aviation operations. Potentially, some of these events could have been due to operator error, and thus be self-implicating. Leakage of such information could lead to serious consequences for the pilot such as punishments, fines, notations in their personnel records, or other penalties. Such information is central to NAOMS purpose: the measurement of aviation safety levels and trends. The data are not available elsewhere. Accordingly, these questions must be asked. But, they will be posed in the most diplomatic manner possible. As noted earlier, no records will be kept that link respondents' identities to the responses they provided.

I.L. Provide Estimates of the Hour Burden of the Collection of Information

We will be employing a split-design during the first year of NAOMS data collection in accordance with Figure 2. The design will employ two distinct recall periods (one month and three months) and two respondent selection processes (cross cross-sectional and panel). Panel participants will be surveyed on a quarterly basis. The split-design is intended to resolve methodological issues that were addressed, but not fully answered, during pre-survey research.

		Design	
		Cross-Sectional	Panel
Recall period (months)	1	2,000	500
	2	2,000	500

Figure 2. Number of Respondents

<u>I.L1. The Number of Respondents</u>

A total of 7,142 randomly sampled individuals will be invited to participate the survey. Based on pre-survey research, we expect a response rate of 70%. This will result in net participation by 5,000 individuals over the course of the year. Roughly 1,400 of these persons will be asked to become member of a panel that will be surveyed once per quarter over the course of the year. The balance will be interviewed just once during the survey period. We expect the combined process to yield 8,000 responses total over the one-year survey period.

I.L2. Frequency of Response

The pilots in the cross-sectional sample will be interviewed once. Panel members will be interviewed quarterly for one year, for a total of four times. The commercial air pilot respondents will have roughly a 5% chance of being selected in any given year.

I.L.3. Annual Hour Burden and an Explanation of How the Burden Was Estimated

The total time per completed interview is 45 minutes, including 30 minutes to complete the interview, 15 minutes to read the advance letter, screen for eligibility, and have informed consent read to them. The average time per response is slightly lower for panel members since they will be participating four times over the year and will not need to repeat some of the initial screening steps. Pilots who are screened ineligible have a burden of 10 minutes for reading the advance letter and having the screening conducted. Table 1 displays the results of the burden calculation. The total annual burden is estimated as 5,907 hours.

Table 1. Estimate of Annual Hour Burden

Type of Respondent	Number of Respondents	Frequency of Response	Average Time Per Response (Hours)	Annual Hour Burden
Cross-sectional	4,000	1	0.75	3,000
Panel	1,000	4	0.56	2,240
Decline to participate	2,142	n/a	0.17	364
Ineligibles	1,785	n/a	0.17	303
Subtotal				5,907

I.L.4. Estimates of Annualized Costs to Respondents Using Appropriate Wage Categories

Based on the average salary of $100,000 for air carrier pilots who work on average 1,500 hours per year (including training, preflight activities, and other work activities ancillary to flying), the average cost per hour for respondents will be $67 ($100,000 / 1500 = $67). The total estimated cost burden is $395,769 ($67 x 5,907 = $395,769).

I.M Provide an Estimate of Other Total Annual Cost Burden to Respondents or Record Keepers

The data collection entails no additional cost burden to respondents or record keepers.

I.N. Provide Estimates of Annualized Costs to the Federal Government

Table 2 provides an estimate of the annualized costs to the Federal government for this data-collection effort. These estimates are based on previous survey work for other government agencies and have been divided by costs elements/tasks to the federal government. The total estimate for all data collection activities is $1,013K for the first year.

Table 2. Estimated Cost of Data Collection to the Federal Government

Cost Element	Estimate	Comment
Project Management & Administration including Interactions with Stakeholders	$125K	
Development and Testing of Survey Instruments	$100K	Assumes four sets of topical question developed and tested on 100 respondents each year
Data system maintenance and administration	$50K	
Data Collection by Telephone	$680K	@ $85 per interview
Contractor Fee	$58K	@ 6.1%
Total Cost plus Fee	$1,013K	

I.O. Outline Plans for Tabulation and Publications

Current plans include one summary publication for the initial year. As the data collection continues for the pilots, and other operational personnel such as air traffic controllers are added, it is likely that quarterly reports and annual reports will be developed.

For this first year, the annual report will provide estimates of aviation safety event rates for commercial air carrier pilots. These event rates will be expressed in terms of the numbers of events per flight hour and per landing. Because these are sensitive data, their interpretation and publication will be accomplished in coordination with an aviation government-industry steering committee.

Collection of Information Employing Statistical Methods

II.A. Describe the Respondent Universe and Sampling Methods to Be Used. Indicate Expected Response Rates for the Whole.

Figure 2, presented earlier, describes the split-design that will be used during the first year of the survey effort. The design elements differ in terms of the sampling process (cross-sectional vs. panel) and recall period (1 vs. 3 months).

II.A.1 Respondent Universe

The study universe, or target population, includes all commercial airline pilots in the United States, a group comprised of roughly 100,000 individuals. For the purposes of NAOMS, the target sample is limited to pilots who have flown commercially during the recall period.

II.A.2 Sampling Methods

The airline pilot sample will be extracted from the FAA's Airman's Medical Certification database, which is publicly available. Commercial pilots are required to update their medical certification every six months, so this data source remains relatively current. Pilots will be randomly selected from this database from among those who designate themselves as U.S.-based, commercial aviation pilots flying multi- engine planes. Telephone numbers are not available in the database. So, the latter will be obtained through Telematch, a service that matches names and addresses with telephone numbers. If telephone numbers are not available through Telematch or addresses are not current, the pilot will be located using other conventional methods including: National Change of Address and Directory Assistance searches.

II.A.3 Cross-Sectional vs. Panel Design

Survey methodology literature suggests that a purely cross-sectional study approach is usually statistically optimal. It is usually easier to administer cross-sectional designs as well. However, the domain experts on the research team are interested in potential benefits of the panel approach. The rationale underlying the NAOMS effort is that the aviation community—pilots, controllers, mechanics, and flight attendants—is comprised of highly professional and well-educated persons who can be en enlisted as active participants in the ongoing evaluation of NAS safety. It is further believed that enrollment in NAOMS panels will encourage participants to become even more acute observers of aviation system safety.

These competing considerations also suggest that a split design would be desirable in Year 1 of the NAOMS implementation. One half of data collection will be accomplished using a cross-sectional sample, one-twelfth of which would be interviewed each month. The other half will employ a panel design. Each panel respondent would be asked to enroll for a one-year period with the expectation that he/she would be asked to participate in four surveys spaced at three-m-month intervals.

II.B. Describe the Procedures for the Collection of Information

II.B.1. Statistical Methodology for Stratification and Sample Selection

No sampling stratification will be done during this first year of the NAOMS process.

Table 3 describes the sampling frame. The frame size is designed to compensate for respondents who are ineligible (pilots who did not fly a commercial aircraft during the recall period), who refuse to participate, for whom current location or telephone number information cannot be obtained, or cannot be contacted. Based on the field trial results (discussed more fully later in this document), we expect roughly 20% of the persons contacted to be ineligible. We further expect that at least 70%of those who are eligible will elect to participate in the survey. Thus, it is anticipated that 5,000 out of the initial 8,928 persons who are contacted by NAOMS would be eligible and would agree to participate.

Potential respondents will be randomly assigned to either the cross-sectional or panel groups. Persons in the cross- sectional group will be further randomly assigned to one of the twelve months in the survey period. Members of the cross-sectional group will be interviewed once within 90 days of their assigned interview-release date. Panel members will be interviewed shortly after their release date, and once each quarter thereafter.

Table 3. Sampling Frame

Persons		Number of Persons	Balance
Initially Contacted		8,928	8,928
Estimates based on field trial findings	Ineligible	1,786	7,142
	Refusals	857	6,285
	Unlocatable	643	5,642
	Non-response	642	5,000
NAOMS Participation			5,000

As Table 4 indicates, NAOMS expects to receive a total of 8,000 completed surveys. Four thousand of these will come from the Cross-sectional group. The balance will come from the panel members each of whom will participate four times in the NAOMS survey effort.

Table 4. Number of Responses

Group	Number of Participants	Responses per Participant	Completed Surveys
Cross-sectional	4,000	1	4,000
Panel	1,000	4	4,000
Total	5,000		8,000

II.B.2.Estimation Procedures

The goal of NAOMS is to estimate rates at which particular types of events occur in the course of commercial air travel. Each respondent will be asked to report how many occurrences of each event type or category he/she experienced during the recall period, and respondents will also report how many hours and how many legs (takeoff / landing cycles) they flew commercially during the recall period. Therefore, NASA and its supporting contractor will be able to compute the total number of times each event occurred during the recall period per hour flown and per flight leg. These are, in effect, event frequencies normalized for exposure. They are the key indicators that NAOMS will use to measure NAS safety levels and trends.

II.B.3 Degree of Accuracy Needed for the Purpose Described in the Justification

The goal of NAOMS is to be able to detect changes in the operation of the commercial air travel system from one year to the next. Because safety-related events vary in nature and frequency across the calendar year due to weather and climate fluctuation, as well as the volume of air travel, data collection, must occur throughout the year. However, statistics will be computed for

The level of accuracy required is one that will permit reliable detection of a 20% change from one year to the next. That is, NASA and its supporting contractor must be able to detect a 20% decrease (or increase) in event rates with a high degree of certainty and have confidence in the accuracy of the measurement. Pilots may experience some of the events being measured only once every ten years or less (e.g., a near mid-air collision). While such events are far more frequent than accidents, a substantial sample size is still required to measure their frequency within operationally useful confidence intervals (roughly, +/- 20%). Power analyses, using classical statistical methods and simulations, determined that the samples sizes proposed in this document are the minimum required to reliably detect such change.

NAOMS' primary reason for using a split-design during this first year of the data collection is to determine whether a 3-month recall period will yield responses of acceptable quality, i.e., whether there ether was a significant diminution of quality relative to a one-month recall period. The field trial results were ambiguous in this regard. The split-design will also explore the use of cross-sectional vs. panel approaches. NAOMS' intent is to ask panel participants to maintain brief notes regarding the their aviation safety experiences between survey applications, thus elevating response quality.

II.B. 4. Unusual Problems Requiring Specialized Sampling Procedures

Specialized sampling procedures will not be implemented.

II.B. 5. Any Use of Periodic (Less Frequent than Annual) Data Collection Cycles to Reduce Burden

Data collection will occur at the least frequent intervals possible in order to provide sufficiently valid data. Because statistics must be generated to describe air travel system safety for each year, data collection must occur throughout the year. It is an anticipated that the policy makers who use NAOMS data will wish to see preliminary data on, at least, a quarterly basis as well as more definitive results at year-end.

II.B. 6. Information on Data Collection Procedures

Battelle will conduct all data collection activities for this study under contract with NASA. Battelle's Centers for Public Health Research (CPHRE) survey operations unit has over twenty years of experience in survey data collection procedures. This particular study will use computer-assisted telephone interviewing (CATI).

Telephone interviewing is a popular mode of conducting data collection because most people have telephones, it is cost effective, and data can be collected quickly. With the advent of computer-assisted-telephone interviewing (CATI), additional advantages have been introduced in term of quality control. CATI questionnaires have many built-in features including: (1) predefined terms and definitions that the interviewer can read to help respondents with questions; (2) range of value, skip pattern, and logic checks that greatly reduce respondent or interviewer error in administration; and (3) elimination of the need for data entry for receipt control or data cleaning. It also allows telephone supervisors to monitor interviewer work in real time by silently listening in on the interviewer. The supervisor can see what interviewer is recording in the computer and can compare it with what he/she heard.

CATI interviews will be administered from Battelle's Telephone Center located in Baltimore, Maryland. In the field trial, an 81% completion rate for CATI interviewing was obtained, and no data were lost due to respondent or interviewer error.

II.B. 6.1. Interviewer Training

Interviewers will be given sixteen hours of training led by the data collection Project Director and other key staff.

Battelle supervisors, interviewers, and editors will attend the training.

Interviewer training will consist of introducing the study, its background and purpose. There will also be an introduction to aviation and aircraft terminology. The interviewers will be trained on the questionnaire, the advance letters, confidentiality, and various administrative forms and procedures used in the study. During training, there will be group role-playing to familiarize the interviewers with the questionnaire and the types of responses they might receive from respondents. Each interviewer will then conduct a one-on-one interview with a member of training team. Interviewers will return to the office over the next few days to conduct a "certification" interview, in which they will have to demonstrate they were competent in all the procedures involved in the study.

After being certified, CATI interviewers will begin interviewing from the Battelle Telephone Center. Interviewers working out of the Telephone Center will have their work silently monitored by CHPRE validation staff. The introductory telephone script read to pilots will include the fact that supervisors may monitor the call for quality assurance purposes.

There will be periodic trainings for new interviewers to replace staff lost to attrition.

II.B. 6.2. Implementation

The following implementation tasks will occur on the schedule outlined below.

Day1: A cover letter on NASA stationary is mailed describing the purpose of the study, confidentiality provisions, and the identity of the funding agency. The letter will describe what participation means for both the Cross-sectional and Panel sample members.
Day 3-90: Respondents are called to screen for eligibility and to pre-arrange appointments to conduct the interview. A call scheduler is built into the CATI system. Whenever an appointment is made, it recorded in the scheduler and will automatically come up as an interviewer assignment at the appointed time. Calls will be made from 9am to 10pm seven days a week and across all time zones.
Day 60: A reminder letter will be mailed to respondents who have not yet been interviewed explaining that interviewing will soon be ending. The letter will indicate that it is important to talk with him/her to establish eligibility and find out whether or not they want to participate.
Panel Sample: Follow-up interviews for Panel Sample members will be conducted around the 3-, 6-, and 9-month anniversary of the first interview. A letter will be mailed to the pilots a week before their anniversary date reminding them that an interviewer will be calling again.

II.C. Methods to Maximize Response Rates

The field trial results indicate there should not be a problem achieving a 70% completion rate. Nevertheless, measures have been incorporated to maximize the response rates. Pilots frequently mentioned how important the NASA name was in their decision to participate. The NASA name/logo will be used on all materials and in verbal introductions. The advance letter serves the dual purpose of introducing the study and establishing its legitimacy to the pilots. It will have the telephone number of a high-level person well versed in the issues most likely to arise. The advance letter will be marked with "address correction requested" so bad addresses can be updated and new letters can be sent out. Attempts will be made to locate pilots whose addresses are not current. Finally, interviews will be scheduled at times that are most convenient for respondents.

Most importantly, as noted earlier, we intend to use a multi-layered set of prompts and reminders to achieve the maximum response rate practical from our population.

II.D. Describe Tests of Procedures and Methods

Numerous tests were made of procedures and methods prior to finalizing the interview contents and the design of the full-survey. Development of the questionnaire followed normal survey research experimental protocols. Then a field trial was conducted to test the design. Each of these steps is described below.

II.D.1. Focus Groups

Four different confidential focus groups with nine active air carrier pilots were conducted. These air carrier pilots were randomly selected from different pilot sub-populations including international air carrier pilots, domestic air carrier pilots and regional commuter airline pilots. An experienced focus group facilitator conducted each session. The discussions were recorded and transcribed. Identifying information for each individual was n removed from the transcript and the recording tapes destroyed.

The goal of these focus groups was to obtain a listing of safety events that these individual pilots had experienced. Personal opinions on individual safety concerns were not solicited. The findings were then used to begin development of a draft questionnaire. The focus group responses were substantially augmented by input from the NASA Aviation Safety Reporting System's (ASRS) incident database.

II.D. 2. One-on-One Interviews

Volunteer active airline pilots were interviewed one-on-one by project researchers to test the validity of the list of events developed from the focus groups and ASRS inputs. Each event derived from the focus groups was presented to the pilot and discussed. Through this process, the relative perceived safety significance of each of the events was determined.

Additionally, the pilots participating in the one-on-one interviews provided information that was used to help in the organization of the draft questionnaire. Each pilot took a stack of 96 cards, each with a safety-related incident described on it derived from the focus groups, and sorted them into groupings of events that seem seemed to be related to one another. Then the pilots wrote descriptions of what each grouping was about. This process helped determine how the pilots organized and categorized safety events. This information was used in designing the format and sequence of questions at in the draft questionnaire.

II.D. 3. Pilot Recall Accuracy

One of the most significant decisions in designing the NAOMS questionnaire is specifying the length of time during which respondents will be asked to recall events in their recent pasts. Because it is essential that measurements be as precise as possible, the possibility that pilots will misremember the frequency, nature, or timing of events is of great importance. Based upon the cognitive psychology literature on memory and the experience of survey researchers seeking to document events, three sorts of misremembering are significant threats to the validity of NAOMS measurements. First, respondents may forget events that occurred. Second, respondents may remember events that did occur but may misremember the dates on which they occurred, leading respondents to report these events as having occurred during the "reference period" (i.e., the period during which they are being asked to recall events), when in fact they occurred prior to the reference period. Third, respondents may imagine events that never occurred. The recall period for the NAOMS survey instrument must keep these potential errors at acceptable levels.

Rare, emotionally arousing events are likely to be remembered well, but events that occur frequently and are of less severity are more at risk for being e misremembered. For starters, it is important to know how long pilots can accurately recall highly routine events. A small-scale pretest study was conducted to help make this determination. In this study, pilots were asked to perform a relatively simple memory task (remembering the number of landings that they had recently performed) the results of which could be objectively verified against official records. This research indicated that pilots had highly accurate recall of these routine events for one week after their occurrence. Accuracy declined thereafter. The question remained whether pilots could accurately recall more unusual occurrences, like safety events, for a longer period of time. This question was addressed during the NAOMS field trial.

II.D.4. Field Trial

A field trial was conducted during the November 1999 - February 2000 time frame. It was designed to test mode (self-administered, CATI, in-person); recall period (one week, two weeks, four weeks, two months, four months and six months); and alternative questionnaire structures (four). This design resulted in 72 cells for the pilot population that would be the initial subjects of this survey effort.

The field trial used air carrier pilot respondents. The sample was randomly drawn from the FAA's Airman's Medical Certification database. Sampled pilots were then randomly assigned to interviewing cell. Telematch was used to obtain current telephone numbers.

NAOMS received 627 responses, averaging 8.7 completes per cell. Response rates by mode were 70% for self-administered questionnaires (SAQ) and 81% for computer-assisted-telephone interviews (CATI). In-person interviewing was cancelled after reaching a 59% completion rate because it became clear this mode would be too expensive to implement on a nationwide scale.

II.D.4.1. Field Trial Data Quality Assessment

The quality of data collected is also important consideration when evaluating what mode of survey should be selected. A high response rate is not of much value if questions are not completed accurately. There were a number of approaches to evaluating data quality based on the field trial results, each of which is presented below.

Completion Time for Questionnaire. Evaluation of the time needed to complete the interview is a relative measure of data quality. The underlying assumption is that the more time a respondent takes to complete a questionnaire, the better the quality of the resulting data.

As can be seen in Table 5, the average time to complete the telephone interview took 12 minutes more (40% more) than the mail mode. Some of this difference may be due to the need for the respondent to listen and then assimilate what the interviewer asked in the telephone interview versus the ability of the respondent to quickly read the question in the mail interview. It is unlikely, however, that this explains all the difference between the two modes. The lesser amount of time needed to complete the self-administered surveys is probably indicative of pilots working through the questionnaire quickly, paying less attention to questions and spending less time trying to accurately recall the events, than respondents who participated over the telephone.

Table 5. Questionnaire Mean Completion Time (Minutes)

	Mail	Telephone
Completion Time	17	29

Missing Responses. Another way to evaluate the quality of data reported is to look at the number of missing responses for the questionnaire. Table 6 presents the percentage of respondents that did not complete at least one question in the questionnaire by mode.

Table 6. Respondents Who Failed to Complete at Least One Question

	Mail	Telephone
One or More Missing Answers	4.8%	0.0%

The lack of any missing answers for the telephone mode is due to the fact that each question, when read by the interviewer, requires a response. Since most of the responses to the NAOMS field trial questions appropriately received the response of '0' (for event or events that did not occur during the reference period) it is easy to see how respondents would be tempted to skip quickly across questions in the survey instrument. This would also explain why the mail version of the survey instrument took so much less time to complete than the telephone version. In the mail version, the pilots did not have anyone prompting them to slow down and think about each answer. In contrast, the CATI interviewers asked the pilot each question in turn. The pilot did not know what question came next so he or she had to listen to the question to understand its meaning and then think to develop a response.

Total Number of Events and Total Hours Flown in the Recall Period. The observed relationship between the reported number of events and the total hours flown in the recall period also provides insight into data quality. If the questionnaire is capturing accurate responses from pilots about the frequency of events they experience, then pilots with more flight time should experience and report a proportionately greater number of events than those pilots who flew fewer hours.

Several quantitative analyses were conducted looking at the association between the number of events reported and the number of hours flown during the recall period. For all such analyses, one would expect to see a positive relationship between the variable if the data are valid (more flight hours should result in greater number of events reported). The higher the correlation, the tighter the relationship. Table 7 shows the pertinent findings.

Table 7. Demonstrated Association Between Number of Events Experienced and Hours Flown

Mode	Unstandardized Coefficient of Regression	Significance
Mail	.086	p<.001
Telephone	.136	p<.001

Associations for both modes were positive, indicating that pilot reports of event frequencies corresponded to experience during the recall period. However, data from the telephone mode showed a somewhat higher degree of association between number of events reported and the number of hours flown indicating greater consistency in the data.

Total Number of Events and Number of Days in Recall Period. By similar logic, one would expect respondents who were asked to use longer recall periods to report proportionately more events than those asked to use shorter periods. Table 8 shows the relationships found in the data. Once again, the relationship was positive with both modes, but it was considerably stronger with the telephone mode. This suggests that telephone respondents were working harder to recall events accurately for the longer recall periods.

Table 8: Demonstrated Association Between Number of Pilot Reported Events and Recall Period

Mode	Unstandardized Coefficient of Regression	Significance
Mail	.190	p<.001
Telephone	.265	p<.001

Optimum Recall Period. Another key objective of the NAOMS field trial was to determine the appropriate recall period. There were two primary competing considerations. First, as recall period lengthens, the memory of events weakens. Shorter recall periods promote quality. Second, longer recall periods favor the recollection of more events permitting more events to be uncovered from fewer respondents. Longer recall periods promote cost savings.

Respondents were asked to use a variety of recall periods ranging from one week to six months during the field trial. At the end of the survey, respondents were asked how confident they were in the answers they had provided. The results are summarized in Table 9[49] It can be seen that the confidence the pilots have in their ability to accurately report events dropped as the recall

[49] This table was derived from analysis of pilots who completed the survey either by telephone or by mail. Face-to-face interview results were not included.

periods got longer. However, at 60 days, 85% of respondents still indicated that they were either 'Extremely' or 'Very' confident in their event reporting. This suggests that a 60- or 90-day recall period may provide the best balance between quality considerations and the need to operate NAOMS in an economical manner.

Table 9: Respondent Confidence in Reported Accuracy by Recall Period

Recall Period	Extremely Confident	Very Confident	Moderate to No Confidence
1 Week	62%	36%	4%
2 Weeks	58%	34%	8%
4 Weeks	47%	39%	14%
2 Months	36%	49%	15%
4 Months	34%	47%	19%
6 Months	29%	44%	27%

II.D.4.2. Costs

NAOMS wishes to achieve its data-collection objectives in the most cost-efficient manner possible without compromising quality. This is why tests with various recall periods and data collection modes were conducted during the field trial. It was determined that telephone collection costs would be 15 to 25% higher per completed response than self-administered questionnaires. Despite this higher cost, NAOMS has elected to use the telephone mode for quality considerations. Every quality indicator in the field trial suggested that phone interviews yielded more complete and consistent results than self-administered surveys.

II.E. Provide the Names and Telephone Numbers of Individuals Consulted on Statistical Aspects of the Design, Name of the Unit within NASA, Contractor(s), Grantee(s) or Other Person(s) Who Will Actually Collect and/or Analyze the Information for NASA

Battelle and its supporting contractors worked with NASA personnel to design the study protocol.

Linda Connell, M.A., NASA Ames Research Center, [(650) 960-6059] is the NASA Project Manager for NAOMS and the Level III Lead for NASA AvSP Extramural Monitoring.

Mary Connors, PhD., NASA Ames Research Center, [(650) 604-6114] is the Branch Chief for System Safety Research and Level III Lead for NASA AvSP Modeling and Simulation.

From Battelle's Mountain View, CA Operations, Loren Rosenthal, M.S., is the Battelle Vice President Aviation Safety & Efficiency, and Elisa Ingebretson, B.S., is a Battelle Research Scientist. Both are available at (650) 969-3969.

From Battelle's Center for Public Health Research (CPHRE), Joan Cwi, PhD, is Project Director of Data Collection and Louise Glezen, M.A., is Study Leader. Both are available at (410) 377-5660.

From Battelle's Pacific Northwest National Laboratories (PNNL), Tom Ferryman, PhD, is a Senior Statistician on the NAOMS project.

Consultants include Robert Dodd, PhD, Principle Investigator, and Mike Jobanek, M.S., from Dodd & Associates. Dr. Dodd and Mr. Jobanek are available at (410) 923-6086.

From Ohio State University's Department of Psychology and Political Science, Jon Krosnick, PhD [(614) 292-3496] and Michael Silver, M.S. [(614) 292-1714] provide survey methodology expertise.

Battelle and its supporting contractors will co collect and analyze the data for NASA. Battelle's Survey Operations Unit (within CPHRE) will conduct data collection under the direction of Dr. Cwi. The team of Rosenthal, Dodd, Krosnick and Cwi will analyze data, with assistance from Dr. Tom Ferryman from Battelle’s Pacific Northwest Laboratories on statistical design.

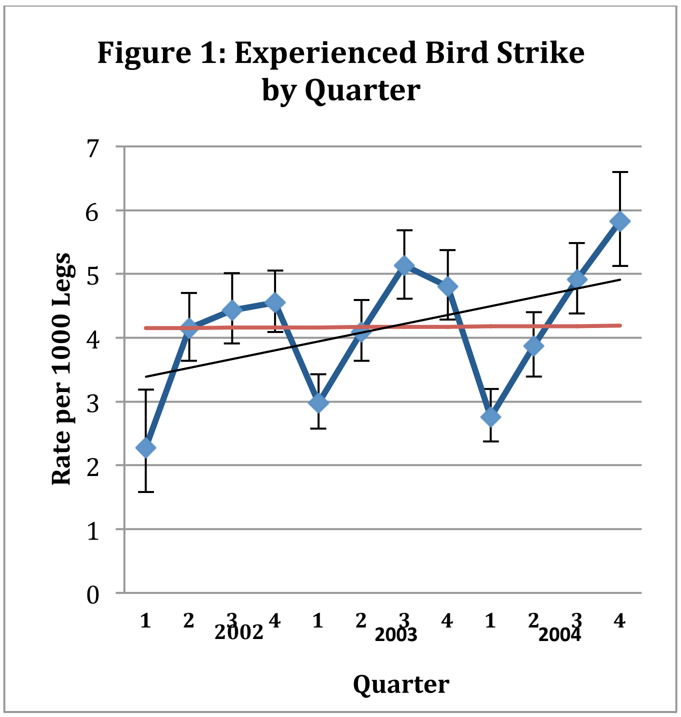
Appendix M. Results of Analysis

The results for the 43 events for which there were sufficient number of reports during the collection period of 3 years to allow reliable statistical analysis are presented here.

<u>Questions about airborne conflicts</u>

Across the study period, a mean rate of 4.24 bird strikes per 1,000 legs was reported. The lowest rate was observed in the first quarter of 2002 (2.28) and the highest rate was observed in the fourth quarter of 2004 (5.83). The reported rate of bird strikes increased across the observation period (b=.020, Wald $X^2(1)$ =8.227, p=.004; see Figure 1).[50] However, this trend is dependent on a usually low rate of reported bird strikes during the first quarter 2002 and an unusually high rate of reported bird strikes during the final quarter of 2004. If these quarters are removed from the data, the observed increase is not significant (Wald $X^2(1)$=.005, p=.945; red line).

AC1: Experience Bird Strike

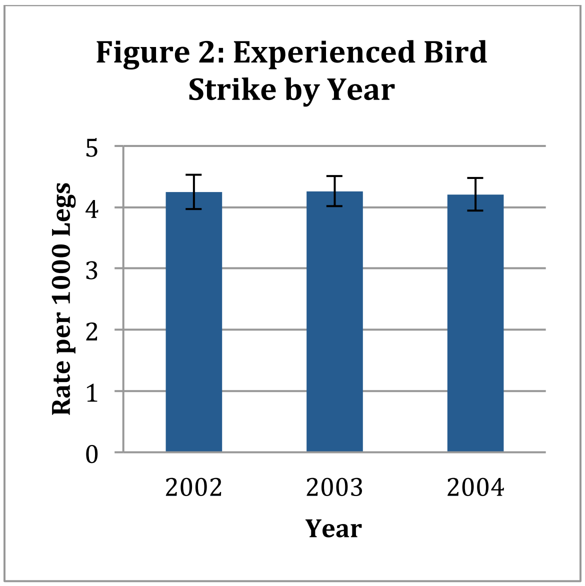


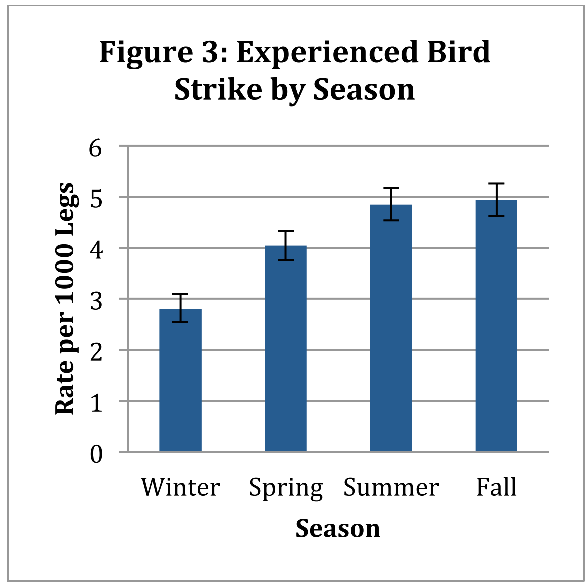
[50] Of the 18358 reports obtained, 19 (0.10%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 139.42 per 1,000 legs compared to 5.66 per 1,000 legs for the other cases. When the suspect observations are included, little change in the trend was observed (but see main text). There was a significant increase in the rate of reported bird strikes across the observation period (b=.024, Wald $X^2(1)$=11.270, p=.001). However, the overall fit of the model decreased (Model with outliers: Pearson $X^2(18356)$=21946.95, ratio=1.196; Model without outliers: Pearson $X^2(18327)$=20144.24; ratio=1.099). These cases were omitted from the remaining analyses.

Year/Season

However, there was a seasonal pattern (Wald $X^2(3)$ =70.351, p<.001; see Figure 3). The rate of reported bird strikes was highest in the summer and fall (which did not differ: Wald $X^2(1)$=0.752, p=.386), lower in the spring than in the summer (Wald $X^2(1)$=8.858, p=.003) and still lower in the winter than in the spring (Wald $X^2(1)$= 30.199, p<.001). No year by season interaction was observed (Wald $X^2(6)$ =8.242, p=.221).

If the first and last quarters are omitted from the analysis, no substantial changes result. The change across years is not statistically significant (Wald $X^2(2)$=1.478, p=.478). There is a significant seasonal pattern (Wald $X^2(3)$=58.437, p<.001) and no year by season interaction (Wald $X^2(4)$=1.712, p=.789). The rate of reported bird strikes is highest in the summer and fall (which do not differ: Wald $X^2(1)$=0.270, p=.603), lower in the spring than in the summer (Wald $X^2(1)$=8.858, p=.003) and still lower in the winter than in the spring (Wald $X^2(1)$= 24.711, p<.001).



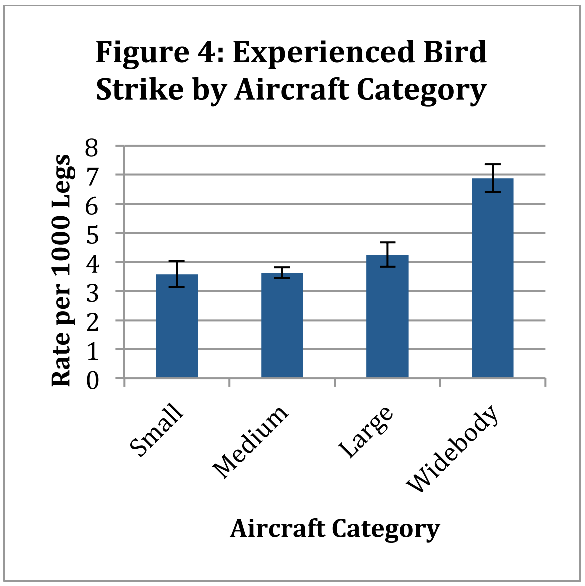


Aircraft Category

There were significant differences in the rates of reported bird strikes between aircraft categories (Wald $X^2(3)=140.781$, $p<.001$; see Figure 4). The highest rates of bird strikes were reported by pilots of wide-body aircraft. Pilots of large aircraft reported significantly lower rates of bird strikes than did pilots of wide-body aircraft (Wald $X^2(1)=43.724$, $p<.001$). Pilots of medium aircraft reported lower rates than did pilots of large aircraft but this difference was only marginally significant (Wald $X^2(1)=3.197$, $p=.074$). The rates of reported bird strikes did not differ between medium and small aircraft categories (Wald $X^2(1)=.011$, $p=.915$).

<u>Interactions</u>

Although the full interaction model could not be properly specified, there was no evidence of interactions between aircraft category and year/season in this (ΔLLR $X^2(33)=25.98$, $p=.802$) or smaller models.

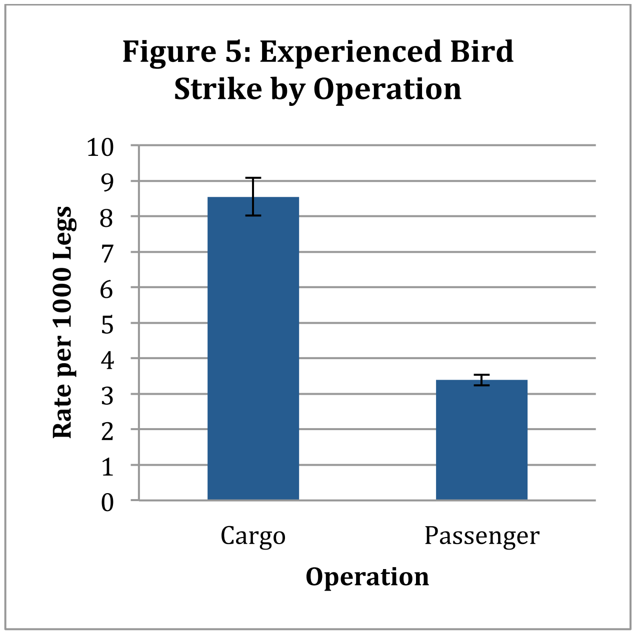


Operation

Pilots of aircraft engaged in cargo operations reported significantly higher rates of bird strikes than did pilots of passenger aircraft (Wald $X^2(1)=384.387$, $p<.001$).

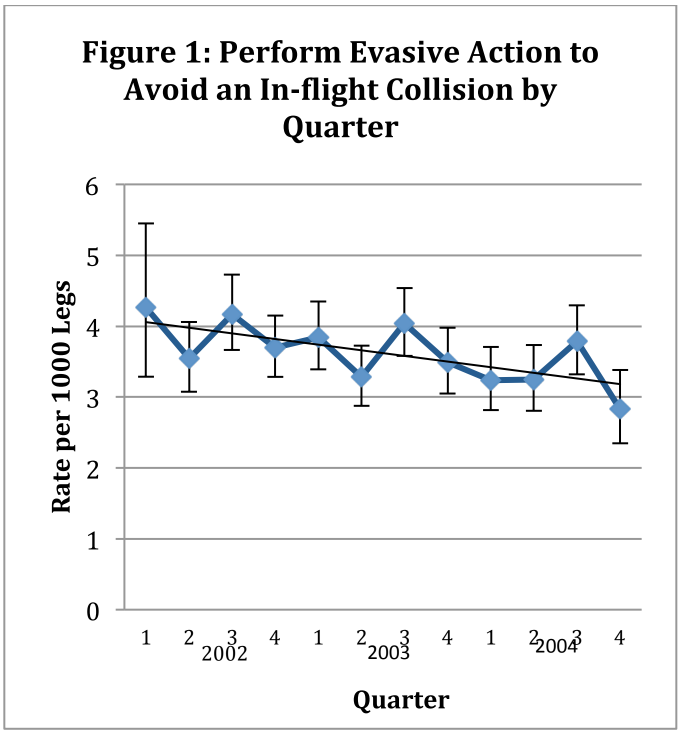
<u>Interactions</u>

There were no interactions between operation and year/season (ΔLLR $X^2(11)=11.042$, $p=.440$).



AC2: Perform Evasive Action to Avoid In-flight Collision

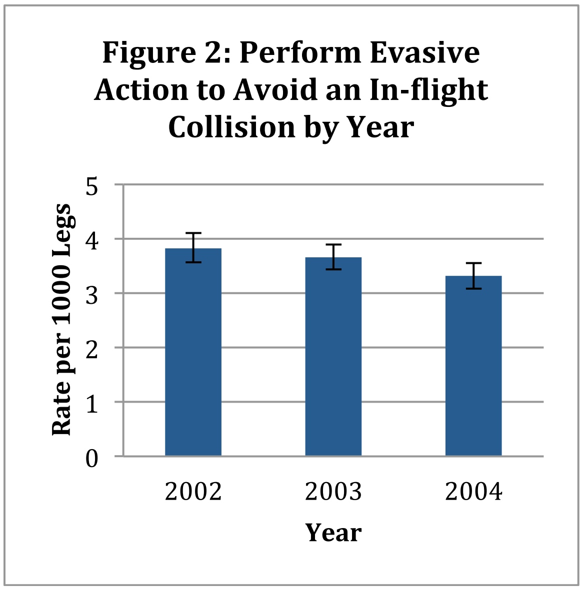
Across the study period, a mean rate of 3.60 evasive actions to avoid in flight collisions per 1,000 legs was reported. The lowest rate was observed in the fourth quarter of 2004 (2.83) and the highest rate was observed in the first quarter of 2002 (4.27). The reported rate of evasive actions to avoid in-flight collisions decreased across the observation period (b=-.018, Wald $X^2(1)$ =6.093, p=.014; see Figure 1).[51]

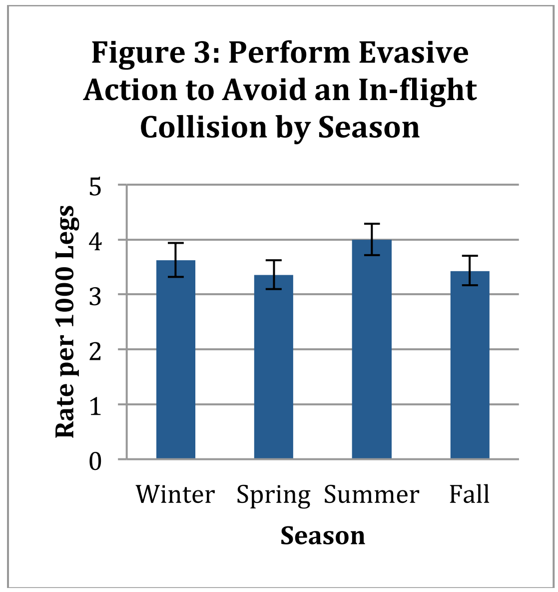


[51] Of the 18,354 reports obtained, 42 (0.23%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 249.96 per 1,000 legs compared to 3.94 per 1,000 legs for the other cases. When the suspect observations are included, little change in the trend was observed. There was a significant decrease in the rate of evasive actions to avoid in-flight collisions across the observation period (b=-.014, Wald $X^2(1)$=3.208, p=.073). However, the overall fit of the model decreased (Model with outliers: Pearson $X^2(18351)$=22598.388, ratio=1.231; Model without outliers: Pearson $X^2(18309)$=19253.608; ratio=1.052). These cases were omitted from the remaining analyses.

Year/Season

There was a significant linear decrease across years (Wald $X^2(1)=7.840$, p=.005) in the reported rate of these events (year: Wald $X^2(2)=8.500$, p=.014; see Figure 2). There was also a seasonal pattern (Wald $X^2(3)=10.961$, p=.012; see Figure 3). The highest average reported rate of evasive actions to avoid in-flight collisions was in the summer. However, the average reported rate for winter was only slightly lower and not significantly different (Wald $X^2(1)=.732$, p=.392). The reported rates for fall (Wald $X^2(1)=8.323$, p=.004) and spring (Wald $X^2(1)=6.868$, p=.009) were significantly lower than in the summer. No other differences between seasons were statistically significant. There was no year by season interaction (Wald $X^2(6)=2.601$, p=.857).





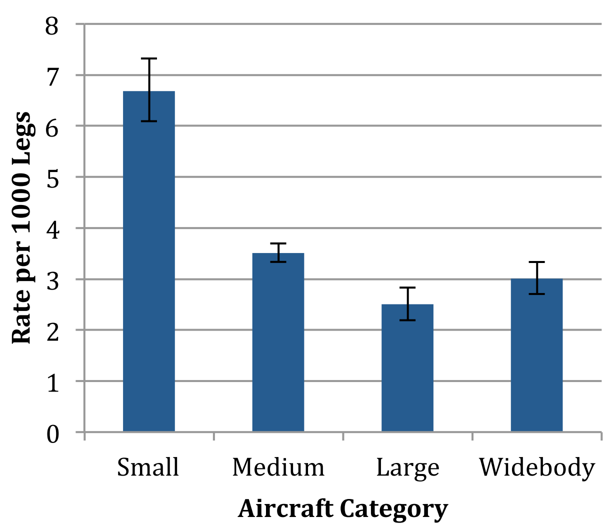
Aircraft Category

There were significant differences between aircraft categories in the reported rates of these events (Wald $X^2(3)=173.372$, $p<.001$; see Figure 4). Pilots of small aircraft reported substantially higher rates of taking evasive actions to avoid in-flight collisions than did pilots of other aircraft (Wald $X^2(1)=85.986$, $p<.001$). Pilots of medium aircraft reported higher rates of these events than did pilots of larger aircraft (Wald $X^2(1)=25.807$, $p<.001$). Pilots of large aircraft reported the lowest rate of evasive actions to avoid in-flight collisions; however, this rate was only slightly lower than that reported by pilots of wide-body aircraft (Wald $X^2(1)=3.659$, $p=.056$).

<u>Interactions</u>

No interactions between aircraft category and year/season were observed (Δ LLR $X^2(33)=23.202$, $p=.897$).

Figure 4: Perform Evasive Action to Avoid an In-flight Collision by Aircraft Category

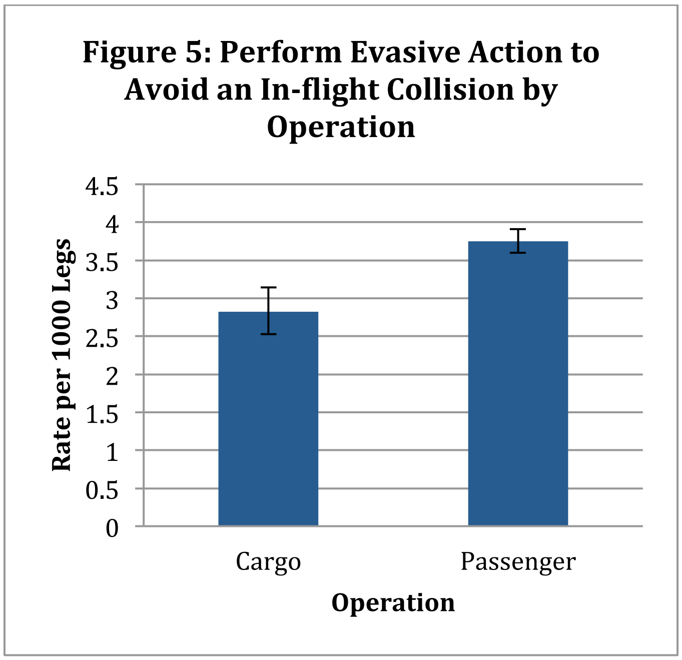


Operation

Pilots of passenger aircraft reported significantly higher rates of these events than did pilots of cargo aircraft (Wald $X^2(1)=17.811$, $p<.001$; see Figure 5).

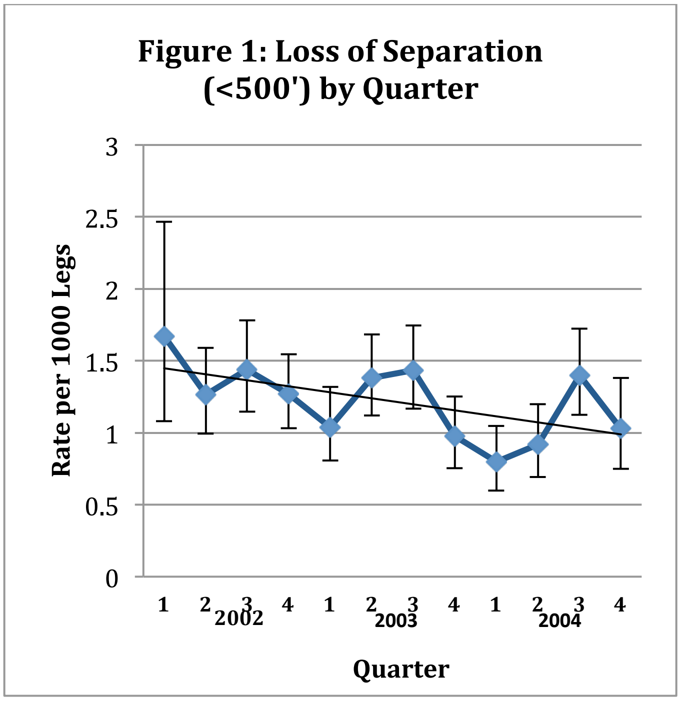
Interactions

No interactions between type of operation and year/season were observed (ΔLLR $X^2(11)=11.564$, $p=.397$).



AC3: Loss of Separation

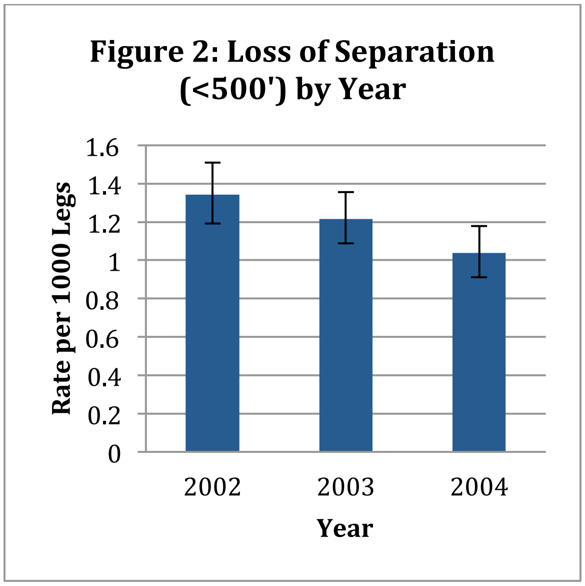
Across the study period, a mean rate of 1.20 events of loss of separation while airborne per 1,000 legs was reported. The lowest rate was observed in the first quarter of 2004 (0.80) and the highest rate was observed in the first quarter of 2002 (1.67). The reported rate of loss of separation while airborne decreased across the observation period (b=-.027, Wald $X^2(1)$ =5.246, p=.022; see Figure 1).[52]

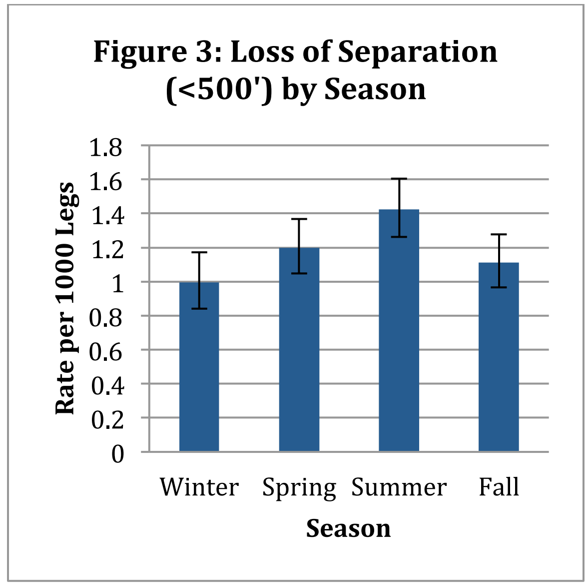


[52] Of the 18,354 reports obtained, 43 (0.23%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 261.77 per 1,000 legs compared to 1.32 per 1,000 legs for the other cases. When the suspect observations are included, the observed decrease in the reported rate of loss of separation remains (b=-.021, Wald $X^2(1)$=3.726, p=.054). However, the overall fit of the model decreased (Model with outliers: Pearson $X^2(18352)$=52849.376, ratio=2.880; Model without outliers: Pearson $X^2(18309)$=24891.475; ratio=1.360). These cases were omitted from the remaining analyses.

Year/Season

There was a significant change over years in the rate of reported events (Wald $X^2(2)=9.364$, $p=.009$). The yearly average rate of reported events decreased linearly over the observed period (linear Wald $X^2(1)=8.849$, $p=.003$; see Figure 2). There were also significant changes across seasons (Wald $X^2(3)=8.721$, $p=.033$). The reported rate of events was higher in the summer than during the other seasons (Wald $X^2(1)=6.141$, $p=.013$; see Figure 3). No other differences between seasons were observed. There was no year by season interaction (Wald $X^2(6)=9.006$, $p=.173$).



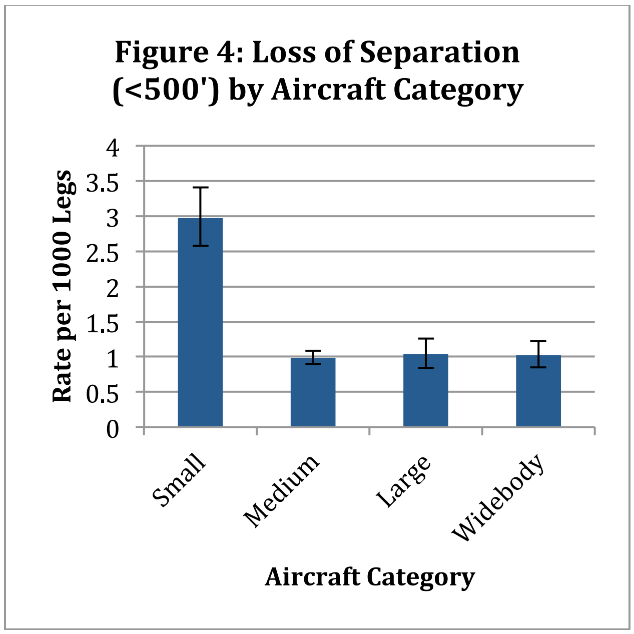


Aircraft Category

There were significant differences in the rate of reported events across aircraft categories (Wald $X^2(3)=141.374$, $p<.001$). Pilots of small aircraft reported significantly higher rates of events than did pilots of other aircraft (Wald $X^2(1)=58.625$, $p<.001$). No other differences between aircraft categories were observed.

<u>Interaction</u>

There were no significant interactions between aircraft category and year/season (ΔLLR $X^2(33)=44.396$, $p=.089$).

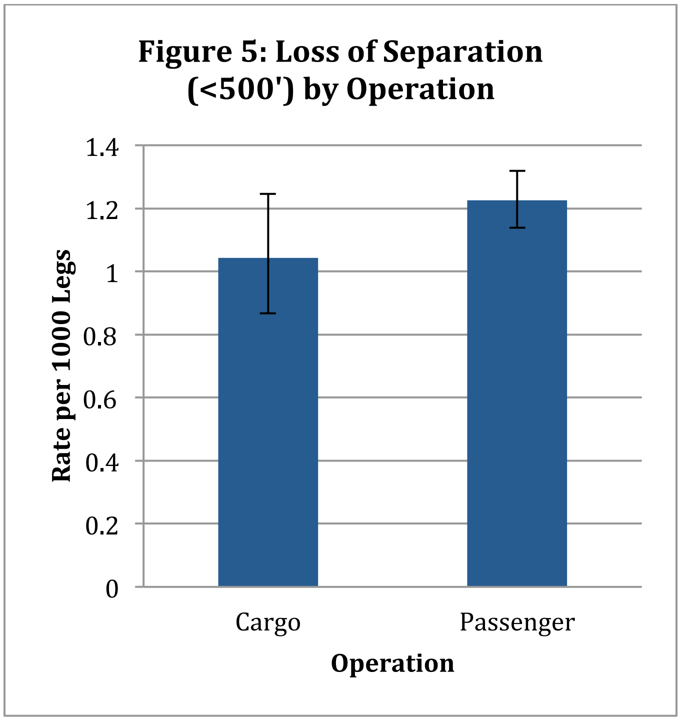


Operation

Pilots of passenger aircraft reported slightly higher rates of these events than did pilots of cargo aircraft, but this difference was not statistically significant at conventional levels (Wald $X^2(1)=3.048$, p=.081).

Interactions

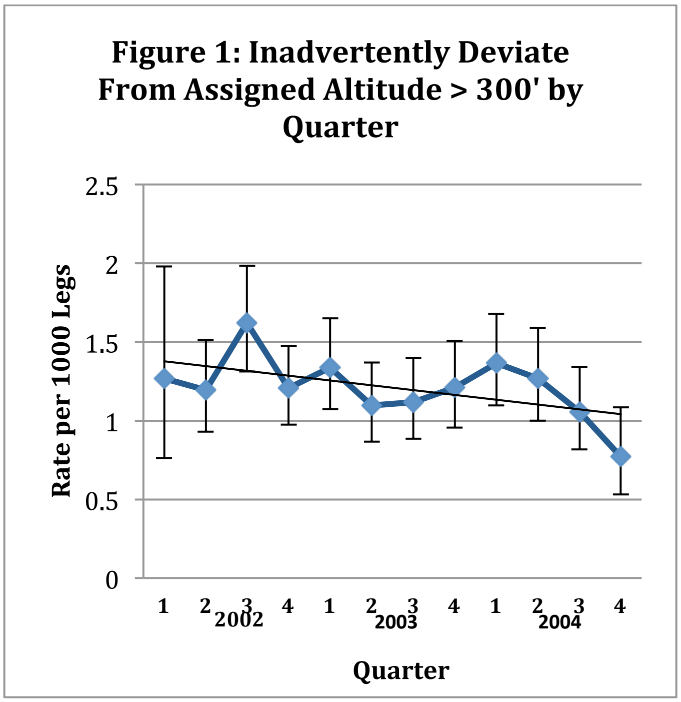
No interactions between type of operation and year/season were observed (Wald $X^2(11)=10.20$, p=.512).



Questions about altitude deviations

AD1: Inadvertently Deviate from an Assigned Altitude > 300 Feet

Across the study period, a mean rate of 1.21 events per 1,000 legs was reported in which a crew inadvertently deviated from an assigned altitude by more than 300 feet. The lowest rate was observed in the fourth quarter of 2004 (0.77) and the highest rate was observed in the third quarter of 2002 (1.62). The rate of reported deviations decreased linearly across the observation period (b=-.026, Wald $X^2(1)$ =4.637, p=.031; see Figure 1).[53]

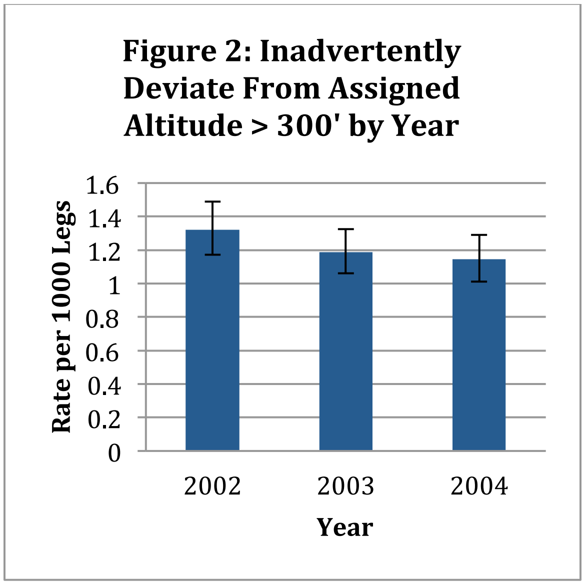


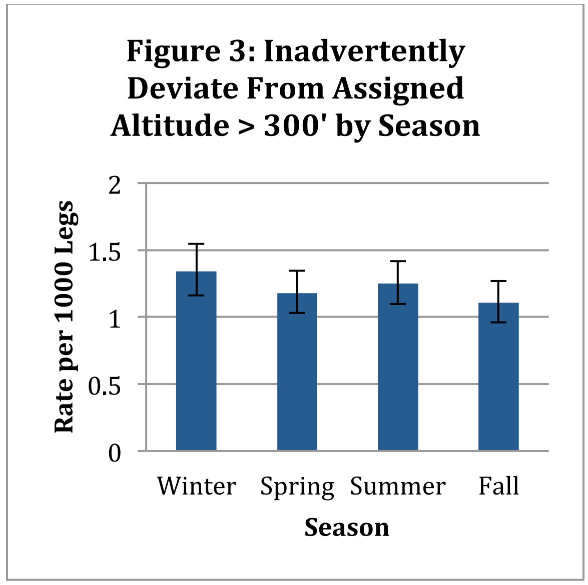
---

[53] Of the 18,372 reports obtained, 47 (0.26%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 138.35 per 1,000 legs compared to 1.34 per 1,000 legs for the other cases. When the suspect observations are included, little change in the trend was observed. The rate of reported events decreased across the observation period (b= -.030, Wald $X^2(1)$=6.628, p=.010). However, the overall fit of the model decreased (Model with outliers: Pearson $X^2(18370)$=23905.764, ratio=1.301; Model without outliers: Pearson $X^2(18323)$=19660.872; ratio=1.073). These cases were omitted from the remaining analyses.

Year/Season

Taking into account seasonal variations, the mean rates of these events decreased linearly across years (b=-.231, Wald $X^2(1)$=4.692, p=.030); however, the omnibus test of differences between years was not significant (Wald $X^2(2)$ =3.204, p=.201; see Figure 2). No differences between seasons were observed (Wald $X^2(3)$ =4.237, p=.237; see Figure 3). No year by season interaction was observed (Wald $X^2(6)$ =10.336, p=.111).

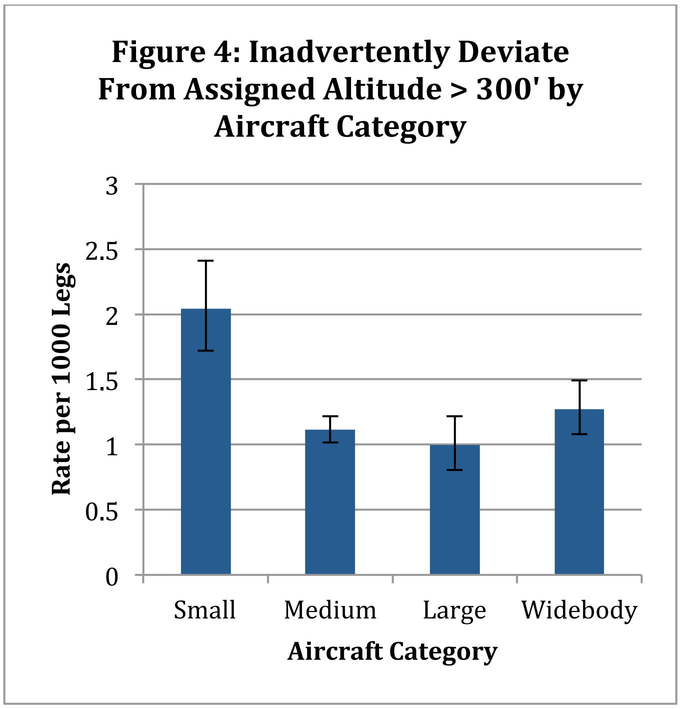




Aircraft Category

There were significant differences in the rate of these events between aircraft categories (Wald $X^2(3)=41.482$, $p<.001$). Pilots of small aircraft reported higher rates of inadvertent deviations than did pilots of larger aircraft (Wald $X^2(1)=23.018$, $p<.001$; see Figure 4). No other differences between aircraft types were observed.

<u>Interactions</u>
No interactions between aircraft category and year/season were observed (ΔLLR $X^2(33)=35.254$, $p=.362$).[54]



---

[54] Note: The full model could not be properly estimated given the large number of parameters and the pattern of the data. Smaller models were fit and the same conclusion obtained.

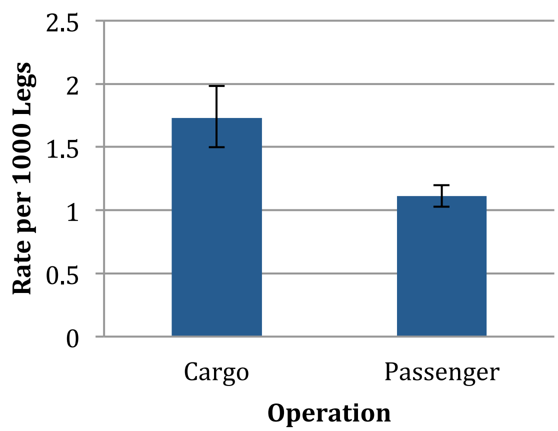
Operation

Pilots of cargo aircraft reported significantly more inadvertent deviations than did pilots of passenger aircraft (Wald $X^2(1)=27.199$, $p<.001$; see Figure 5).

Interactions

No interactions between operation and year/season were observed (ΔLLR $X^2(11)=10.278$, $p=.506$).

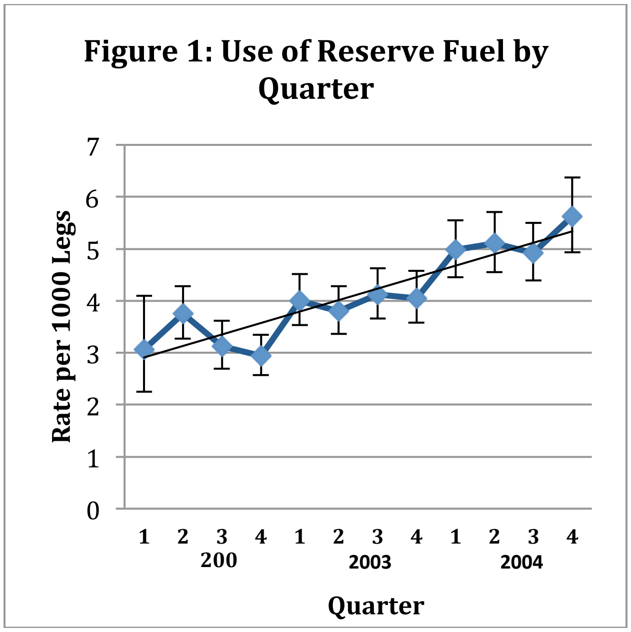
Figure 5: Inadvertently Deviate From Assigned Altitude > 300' by Operation



Questions about aircraft-handling related events

AH1: Use of Reserved Fuel

Across the study period, a mean rate of 4.13 events in which reserve fuel was used per 1,000 legs was reported. The lowest rate was observed in the fourth quarter of 2002 (2.94) and the highest rate was observed in the fourth quarter of 2004 (5.62). The reported rate of use of reserve fuel increased linearly across the observation period (b=.057, Wald $X^2(1)$ =61.459, p<.001; see Figure 1).[55]

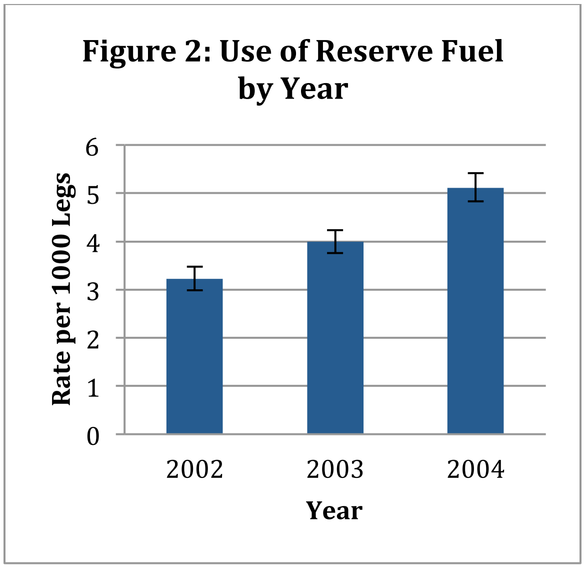


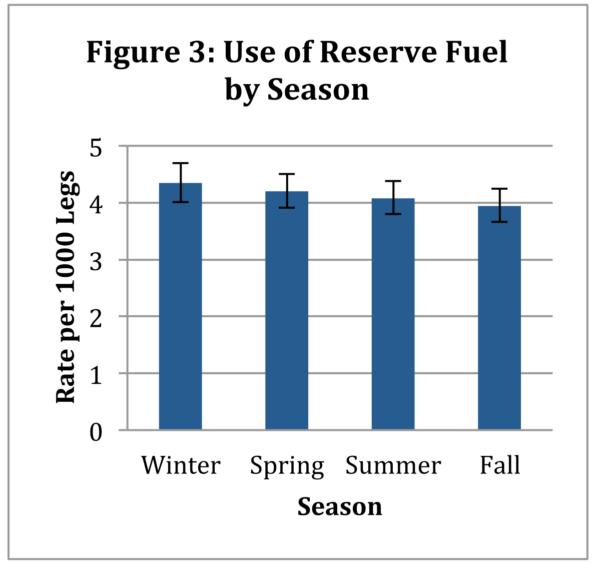
---

[55] Of the 18,350 reports obtained, 57 (0.31%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 361.74 per 1,000 legs compared to 5.44 per 1,000 legs for the other cases. When the suspect observations are included, little change in the trend was observed. The rate of reported events demonstrated a linear increase across the observation period (b=.066, Wald $X^2(1)$=89.194, p<.001). However, the overall fit of the model decreased (Model with outliers: Pearson $X^2(18348)$=42953.235, ratio=2.341; Model without outliers: Pearson $X^2(18291)$=31541.373; ratio=1.724). The outlier cases were omitted from the remaining analyses.

Year/Season

The rate of events increased across the years in the sample (Wald $X^2(1) = 60.909$, $p < .001$; omnibus test: Wald $X^2(2) = 61.132$, $p < .001$; see Figure 2). There was no seasonal pattern (Wald $X^2(3) = 1.729$, $p = .631$; see Figure 3) or year by season interaction (Wald $X^2(6) = 8.456$, $p = .207$).





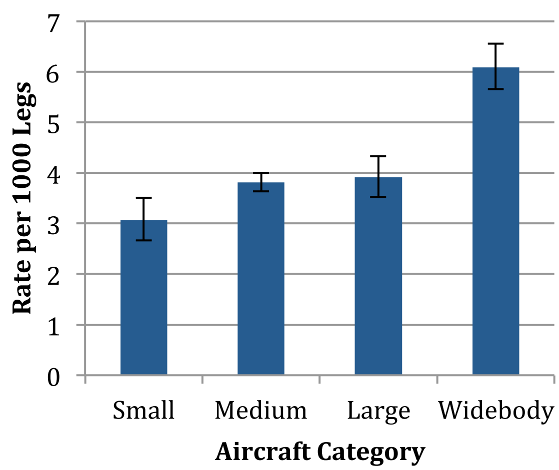
Aircraft Category

The rates of these events differed by aircraft category (Wald $X^2(3)=86.382$, $p<.001$). Overall, pilots of wide-body aircraft reported higher rates of using reserve fuel than did the pilots of smaller aircraft (Wald $X^2(1)=66.221$, $p<.001$). Pilots of small aircraft reported the lowest rates, significantly less than pilots of medium aircraft (Wald $X^2(1)=3.948$, $p=.047$). Pilots of medium and large aircraft did not differ in the rate of events that they reported (Wald $X^2(1)=0.050$, $p=.822$; see Figure 4).

Interactions

The observed temporal patterns differed by aircraft category (ΔLLR $X^2(33)=79.142$, $p<.001$). The observed linear increase across years in the reported rate of use of reserve fuel was statistically significant for medium (Wald $X^2(1)=46.146$, $p<.001$), large (Wald $X^2(1)=10.104$, $p=.006$), and wide-body aircraft (Wald $X^2(1)=11.816$, $p=.001$). Although small aircraft demonstrated a general increase in the rate of these events across the observation period, there was considerable variation across quarters and the overall pattern across years was not statistically significant (Wald $X^2(2)=1.465$, $p=.481$).

Figure 4: Use of Reserve Fuel by Aircraft Category

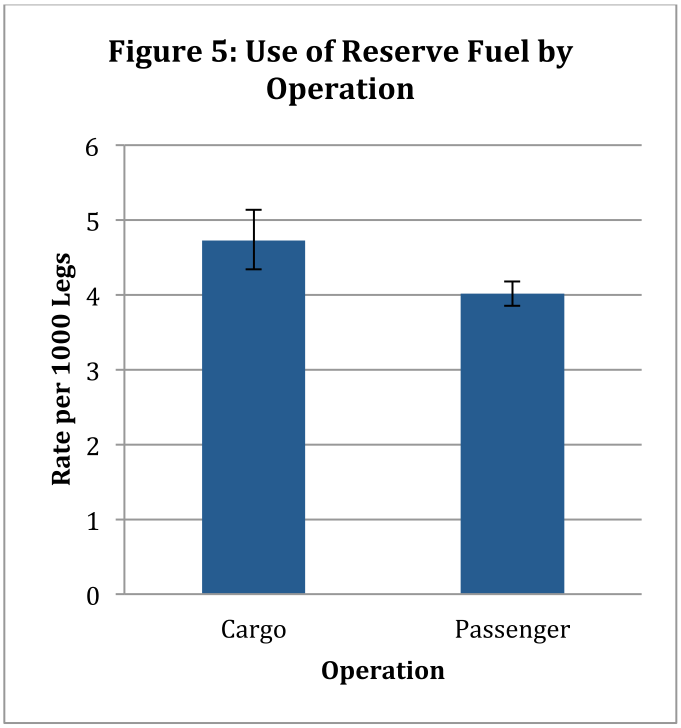


Operation

In general, pilots flying cargo operations reported a higher rate of use of reserve fuel than did pilots flying passengers (Wald $X^2(1)=6.824$, p=.009; see Figure 5).

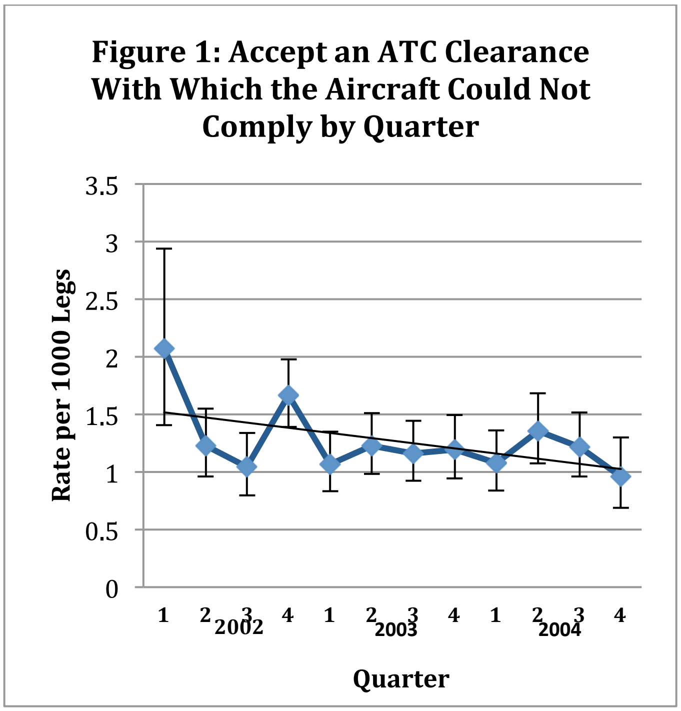
Interactions

There were significant interactions between type of operation and the year/season patterns (ΔLLR $X^2(11)=35.104$, p<.001). The linear increase across years in the reported rates of the use of reserve fuel described above was statistically significant only for passenger aircraft (Wald $X^2(1)=65.133$, p<.001). No change across years was observed for cargo aircraft (Wald $X^2(2)=1.970$, p=.373).



AH2: Accepted an ATC Clearance with Which the Aircraft Could Not Comply Due to Performance Limitations

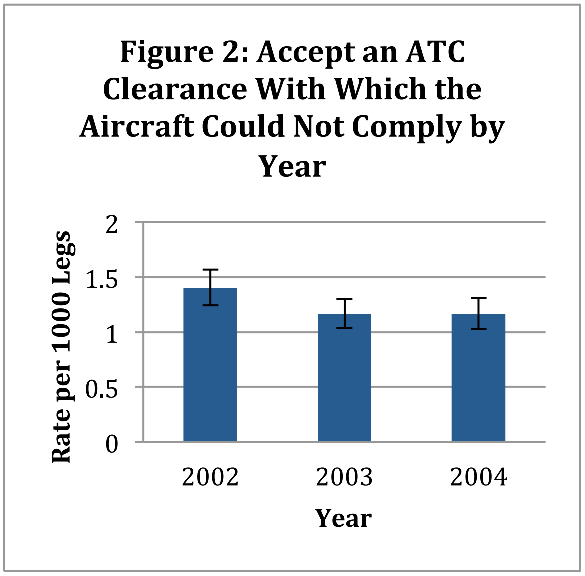
The rate of instances in which pilots reported accepting ATC clearances with which the aircraft could not comply due to performance limitations did not change linearly across the observation period (b=-.024, Wald $X^2(1)$ =2.020, p=.155; see Figure 1).[56] Across the study period, a mean rate of 1.23 events per 1,000 legs was reported. The lowest rate was observed in the fourth quarter of 2004 (0.96) and the highest rate was observed in the first quarter of 2002 (2.07).

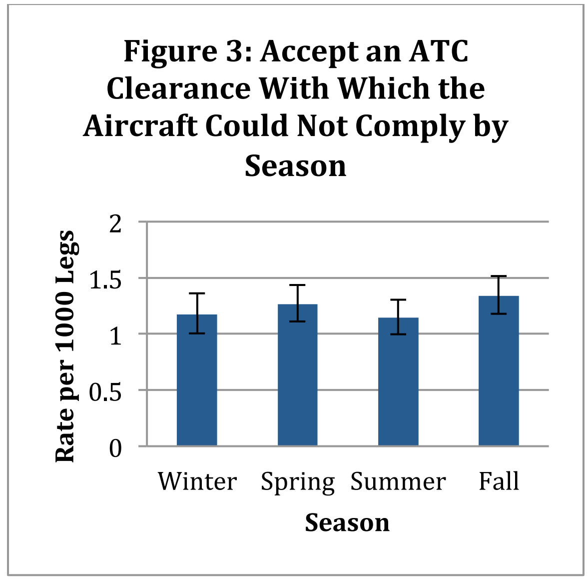


[56] Of the 18,364 reports obtained, 67 (0.36%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 246.99 per 1,000 legs compared to 1.52 per 1,000 legs for the other cases. When the suspect observations are included, little change in the trend was observed. The rate of reported events decreased across the observation period (b=-.008, Wald $X^2(1)$=0.249, p=.618). However, the overall fit of the model decreased (Model with outliers: Pearson $X^2(18362)$=24983.998, ratio=1.361; Model without outliers: Pearson $X^2(18295)$=16228.589; ratio=0.887). These cases were omitted from the remaining analyses.

Year/Season

No differences between years were observed (Wald $X^2(2)$ =3.181, p=.204; see Figure 2). No differences between seasons were observed (Wald $X^2(3)$ =0.179, p=.981; see Figure 3). No year by season interaction was observed (Wald $X^2(6)$ =7.784, p=.254).



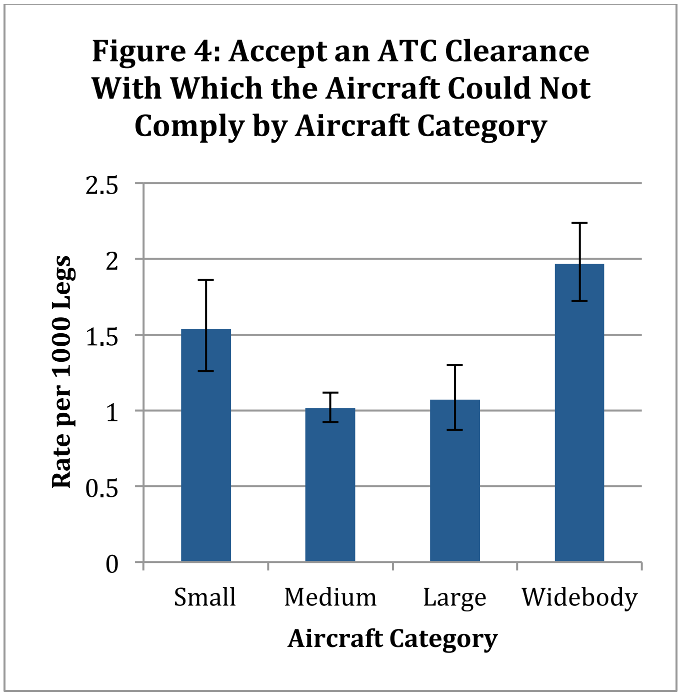


Aircraft Category

There were significant differences between aircraft categories (Wald $X^2(3)=23.738$, $p<.001$). Pilots of small aircraft and wide-body aircraft reported higher rates of accepting ATC clearances that the aircraft could not perform than did pilots of medium and large aircraft (Wald $X^2(1)=12.783$, $p<.001$; see Figure 4). No other differences between aircraft types were observed.

Interactions

No interactions between aircraft category and year/season were observed (ΔLLR $X^2(33)=25.588$, $p=.687$).

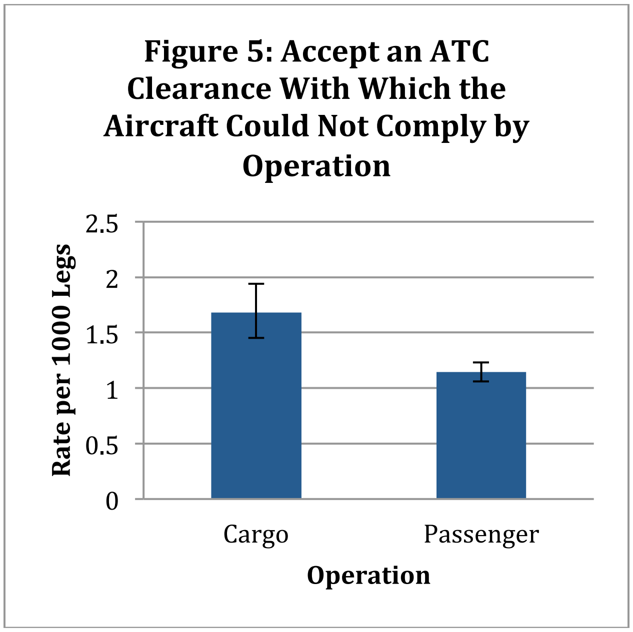


Operation

Pilots of cargo aircraft reported significantly higher rates of these events than did pilots of passenger aircraft (Wald $X^2(1)=17.647$, $p<.001$).

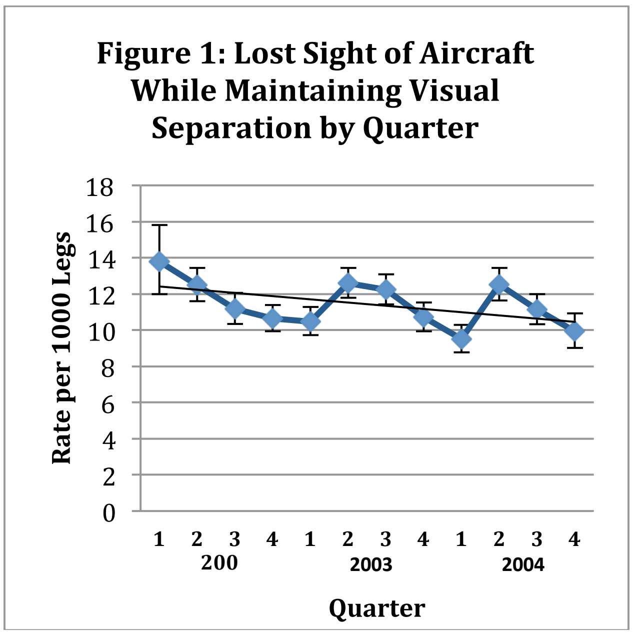
<u>Interactions</u>

There was a significant interaction between operation and year/season (ΔLLR $X^2(11)=20.278$, $p=.042$). Pilots of passenger aircraft reported a significant decrease across years in the rates of these events (Wald $X^2(1)=12.751$, $p<.001$). For cargo aircraft, no significant effects of year (Wald $X^2(2)=3.113$, $p=.211$), season (Wald $X^2(3)=1.200$, $p=.753$, or year by season were observed (Wald $X^2(6)=6.624$, $p=.357$).



AH3: Lost Sight of Another Aircraft While Attempting to Maintain Visual Separation

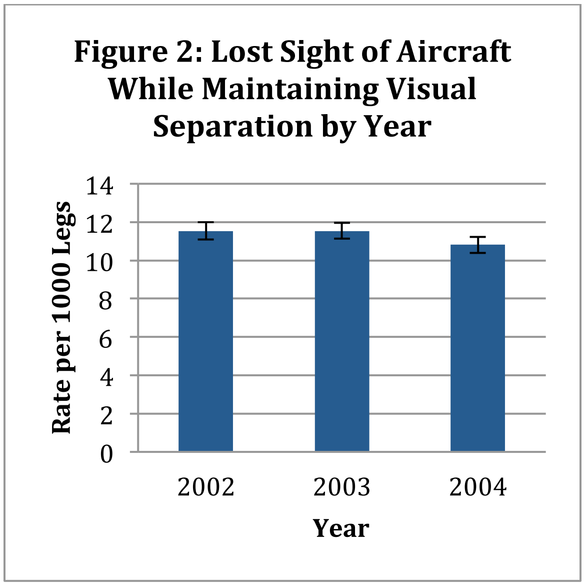
Across the study period, a mean rate of 4.27 events of loss of visual contact with another aircraft while attempting to maintain visual separation from that aircraft per 1,000 legs was reported. The lowest rate was observed in the first quarter of 2004 (9.51) and the highest rate was observed in the first quarter of 2002 (13.79). The reported rate of loss of visual contact with another aircraft while attempting to maintain visual separation from that aircraft decreased across the observation period (b=-.014, Wald $X^2(1)$ =5.166, p=.023; see Figure 1).[57] Although the first quarter 2002 rate was unusually high, the observed decrease is not due to this anomalous quarter. The observed rate does not change appreciably when this quarter is not considered (b=-.012, Wald $X^2(1)$ =3.599, p=.058).

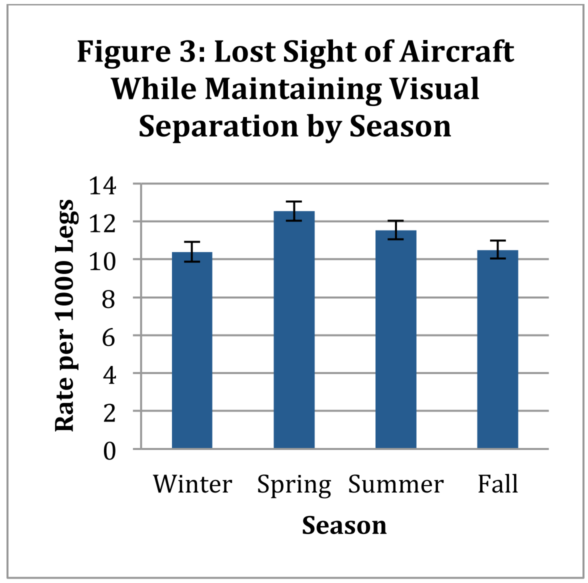


[57] Model Pearson X2(18295) = 18922.350; ratio = 1.034. Of the 18,374 reports obtained, 50 (0.27%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 484.02 per 1,000 legs compared to 13.21 per 1,000 legs for the other cases. When the suspect observations are included, little change in the trend was observed. These cases were omitted from the remaining analyses.

Year/Season

The rate of these events differed across years (Wald $X^2(2)=9.760$, p=.008). Overall the rate of events in 2002 did not differ from the rate in 2003 (Wald X^2 (1) =0.875, p=.350) but both differed from the lower rate observed in 2004 (Wald $X^2(1)$ =7.680, p=.006; see Figure 2). A seasonal pattern was also observed (Wald $X^2(3)$ =11.581, p=.009). In general, the reported rates of events were higher in the spring and summer than in the winter or fall (Wald $X^2(1)$ =7.282, p=.007; see Figure 3). There was no year by season interaction (Wald $X^2(6)$=8.113, p=.230).



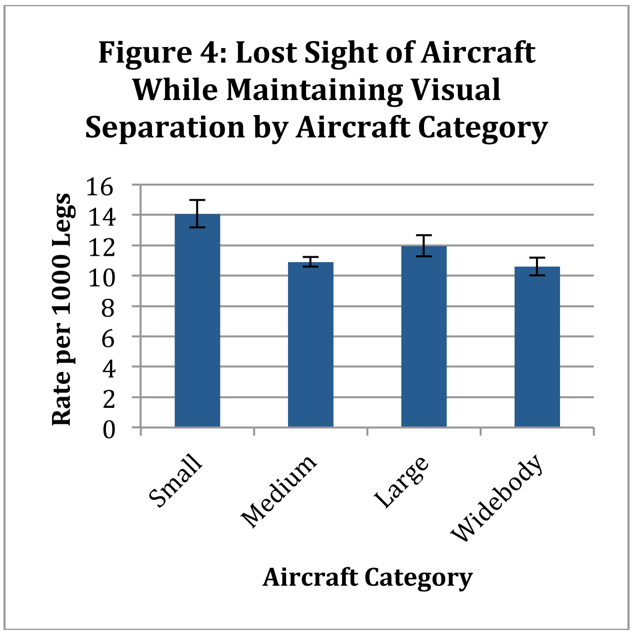


Aircraft Category

There were differences between aircraft categories in the rate of loss of visual contact with another aircraft while attempting to maintain visual separation (Wald X^2 (3)=38.717, p<.001; see Figure 4). Pilots of small aircraft reported significantly higher event rates than those reported by pilots of larger aircraft (Wald X^2 (1)=21.331, p<.001). Pilots of wide-body aircraft reported lower rates of these events than did pilots of the other aircraft (Wald X^2 (1)=28.586, p<.001). The reported rate of events did not differ between medium and large aircraft (Wald X^2 (1)=0.517, p=.472).[58]

Interactions

There were no interactions between aircraft category and year/season (ΔLLR X^2(33)=36.008, p=.330).



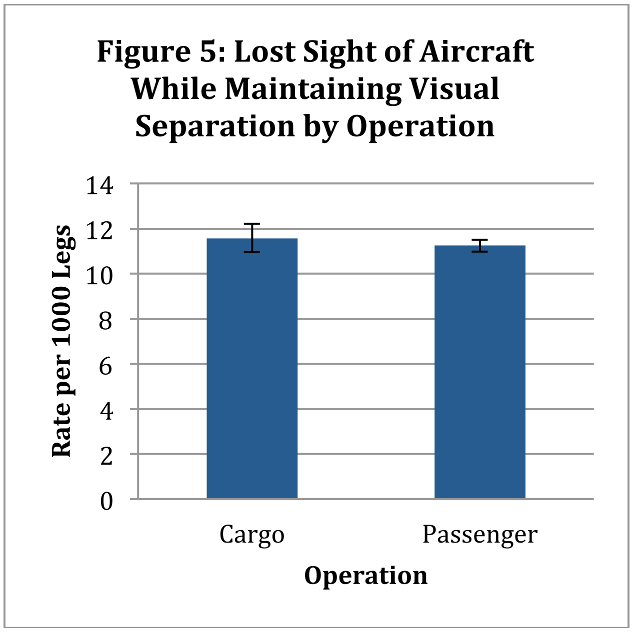
[58] Note: The figures presented in the text display the number of events/number of legs together with the associated 95% Poisson confidence intervals. The analyses reported in the text were conducted on the reported rates provided by each individual pilot. Occasionally, as in this instance the mean of the reported rates is substantially different from the mean events/leg. Hence, the pattern depicted in the figure may depart from the results reported in the text.

Operation

There were no difference between pilots of aircraft engaged in cargo operations and pilots of aircraft engaged in passenger operations in the reported rates of these events (Wald $X^2(1)=0.540$, $p=.462$; see Figure 5).

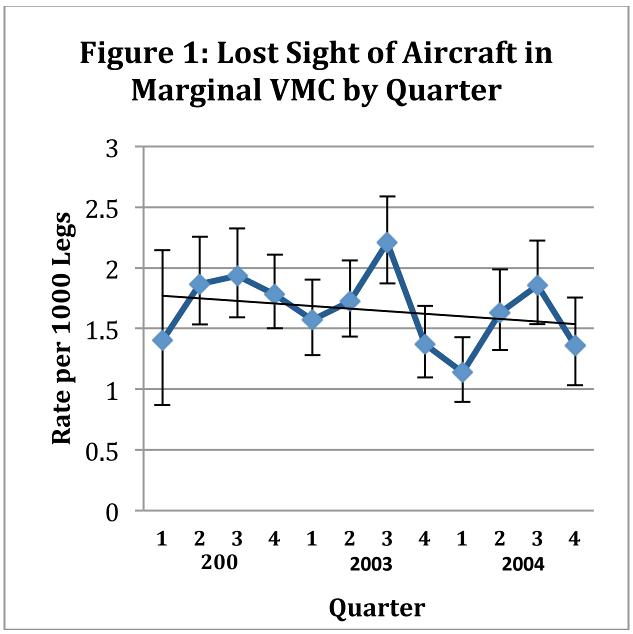
Interactions

No significant interaction between operation and year/season was observed (ΔLLR $X^2(11)=11.736$, $p=.384$).



AH3A: Lost Sight of Another Aircraft While Attempting to Maintain Visual Separation in Marginal Visual Conditions

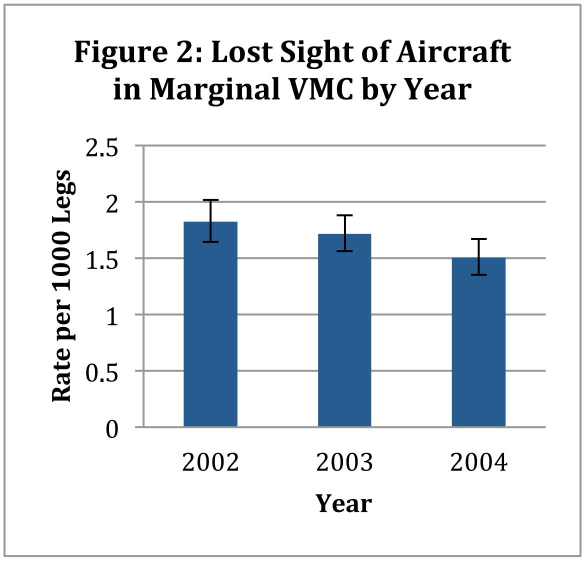
Across the study period, a mean rate of 1.68 events of loss of visual contact with another aircraft while attempting to maintain visual separation from that aircraft in marginal VMC per 1,000 legs was reported. The lowest rate was observed in the first quarter of 2004 (1.14) and the highest rate was observed in the third quarter of 2003 (2.21). The reported rate of loss of visual contact with another aircraft while attempting to maintain visual separation from that aircraft decreased across the observation period. This decrease was marginally significant (b=-.019, Wald $X^2(1) = 3.549$, p=.060; see Figure 1).[59]

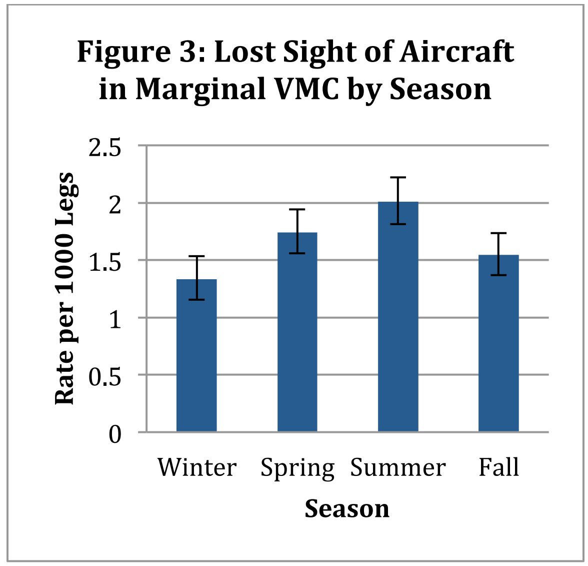


[59] Of the 18,334 reports obtained, 52 (0.28%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 255.95 per 1,000 legs compared to 1.75 per 1,000 legs for the other cases. When the suspect observations are included, the linear decrease across quarters becomes significant (b=-.028, Wald $X^2(1)=8.765$, p=.003). However, the fit of the model decreased (Model without outliers Pearson $X^2(18280)=31162.357$, ratio=1.705; Model with outliers: Pearson $X^2(18332)=53360.646$, ratio=2.911). These cases were omitted from the remaining analyses.

Year/Season

There was no significant difference in the rate of these events across years (Wald $X^2(2)=4.495$, $p=.106$). A seasonal pattern was observed (Wald $X^2(3)=17.619$, $p=.001$). The reported rates of these events were highest in the spring and summer and lowest in the winter. The mean reported rate in the winter was significantly lower than in the spring (Wald $X^2(1)=5.759$, $p=.016$) and summer (Wald $X^2(1)=14.517$, $p<.001$) but not significantly less than in the fall (Wald $X^2(1)=.885$, $p=.347$; see Figure 3). There was no year by season interaction (Wald $X^2(6)=4.643$, $p=.590$).





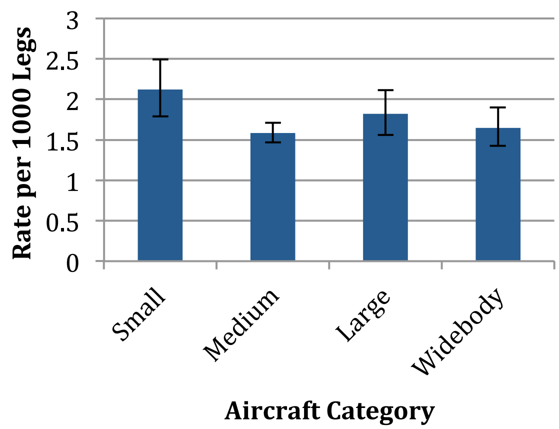
Aircraft Category

There were differences between aircraft categories in the rate of loss of visual contact with another aircraft while attempting to maintain visual separation in marginal visual conditions (Wald X^2 (3)=9.055, p=.029; see Figure 4). Pilots of small aircraft reported significantly higher event rates than did pilots of medium aircraft (Wald X^2 (1)=6.471, p=.011) or wide-body aircraft (Wald X^2(1)=5.531, p=.019). The reported rate of events did not differ between small and large aircraft (Wald X^2 (1)=2.176, p=.140). No other differences between aircraft were significant.

Interactions

A test of the full interaction model could not be performed. A partial model revealed no interactions between aircraft category and year/season (ΔLLR X^2(15)=17.798, p=.273).

Figure 4: Lost Sight of Aircraft in Marginal VMC by Aircraft Category

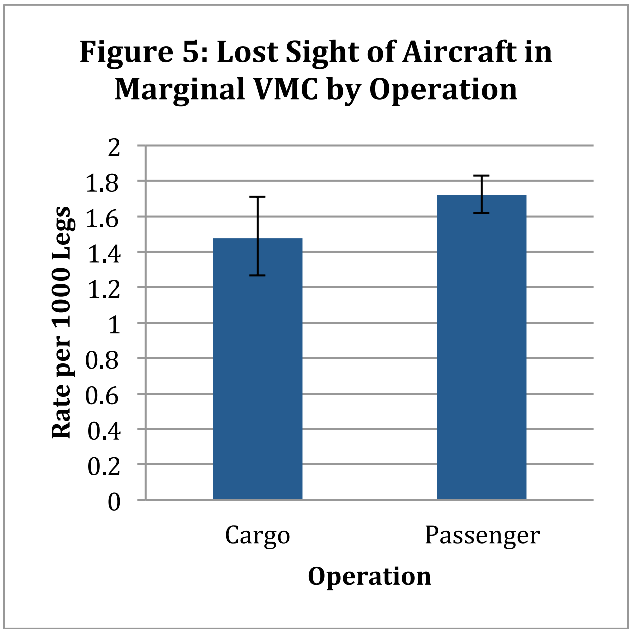


Operation

There were no differences between cargo and passenger operations in the reported rates of these events (Wald $X^2(1)=3.590$, $p=.058$; see Figure 5).

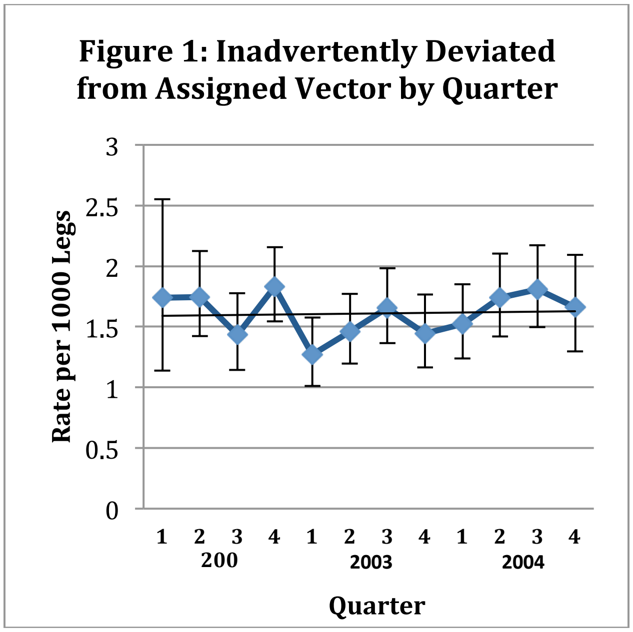
<u>Interactions</u>

A test of the full interaction model could not be performed. A partial model revealed no significant interaction between operation and year/season (ΔLLR $X^2(5)=9.366$, $p=.095$).



AH6: Inadvertently Deviated From Assigned Vector

Across the study period, a mean rate of 1.60 events of inadvertent deviations from assigned vectors per 1,000 legs was reported. The lowest rate was observed in the first quarter of 2003 (1.27) and the highest rate was observed in the fourth quarter of 2002 (1.83). The rate of reported inadvertent deviations from assigned vectors did not change across the observation period (b=.003, Wald $X^2(1)$ =0.077, p=.782; see Figure 1).[60]



[60] Of the 18,361 reports obtained, 43 (0.23%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 274.09 per 1,000 legs compared to 1.79 per 1,000 legs for the other cases. When the suspect observations are included, little change in the trend was observed. No change in the rate of reported events was observed (b=-.008, Wald $X^2(1)$=.391, p=.532). However, the overall fit of the model decreased substantially (Model with outliers: Pearson $X^2(18359)$=35009.661, ratio=1.907; Model without outliers: Pearson $X^2(18316)$=19054.112; ratio=1.044). The outlier cases were omitted from the remaining analyses.

Year/Season

There were no differences across years (Wald $X^2(2) = 3.140$, p=.208; see Figure 2) or seasons (Wald $X^2(3) = 1.330$, p=.722; see Figure 3) and no year by season interaction (Wald $X^2(6) = 3.688$, p=.719).

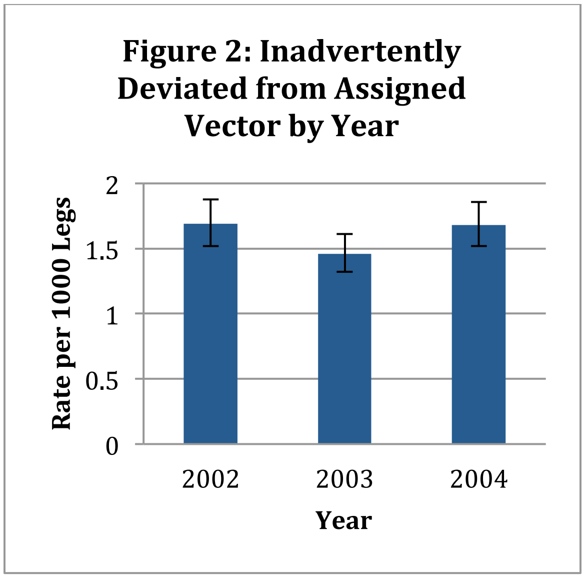


Figure 2: Inadvertently Deviated from Assigned Vector by Year

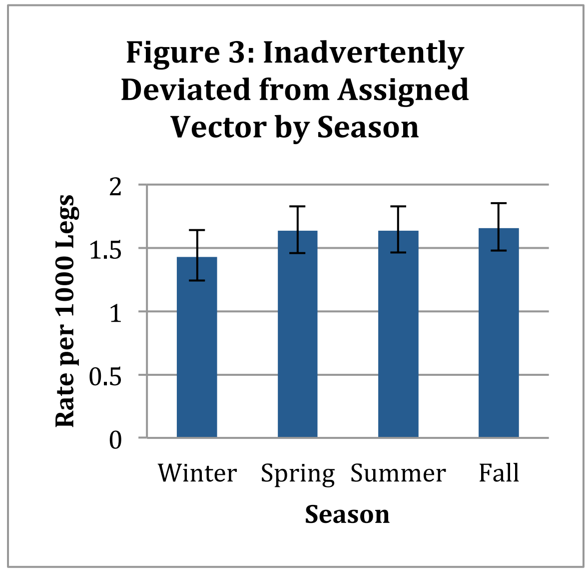


Figure 3: Inadvertently Deviated from Assigned Vector by Season

Aircraft Category

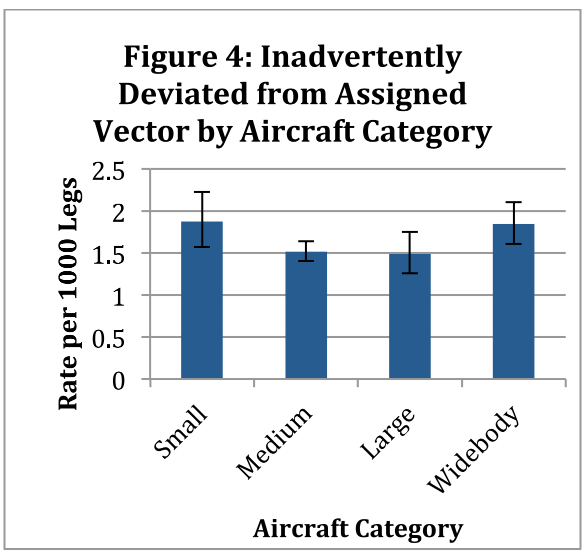


Figure 4: Inadvertently Deviated from Assigned Vector by Aircraft Category

There were no significant differences between aircraft categories in the reported rates of these events (Wald $X^2(3) = 5.924$, p=.115; see Figure 4).

Interactions

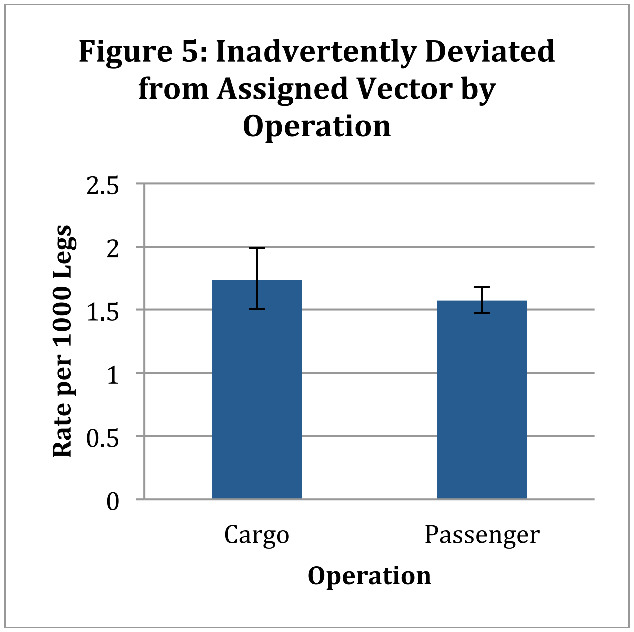
There were no interactions between aircraft category and year/season ($\Delta X^2(33) = 33.396$, p=.448).

Operation

The type of operation did not affect the rate of reported events (Wald $X^2(1)=1.328$, p=.249; see Figure 5).

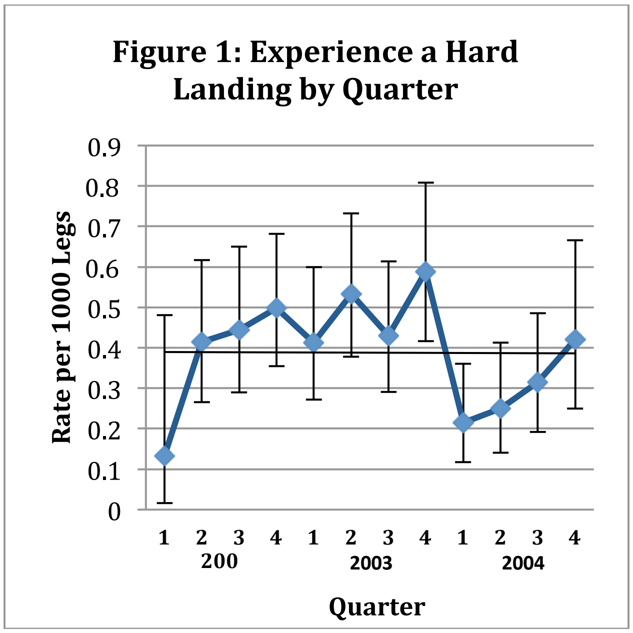
Interactions

There were no interactions between operation and year/season (ΔLLR $X^2(11)=6.574$, p=.832).



AH9: Experience Hard Landing

The rate of instances in which pilots reported hard landings did not change linearly across the observation period (b=-.029, Wald $X^2(1)$ =2.227, p=.136; see Figure 1).[61] Across the study period, a mean rate of 0.41 events per 1,000 legs was reported. The lowest rate was observed in the first quarter of 2002 (0.13) and the highest rate was observed in the fourth quarter of 2003 (0.59).



[61] Of the 18,375 reports obtained, 28 (0.15%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 238.13 per 1,000 legs compared to 0.49 per 1,000 legs for the other cases. When the suspect observations are included, little change in the trend was observed. The rate of reported events did not change linearly across the observation period (b=-.033, Wald $X^2(1)$=3.617, p=.057). However, the overall fit of the model decreased (Model with outliers: Pearson $X^2(18373)$=68785.737, ratio=3.744; Model without outliers: Pearson $X^2(18345)$=28527.071; ratio=1.555). These cases were omitted from the remaining analyses.

Year/Season

Significant differences between years were observed (Wald $X^2(2)$ =11.992, p=.002; see Figure 2). The rate of hard landings in 2004 was significantly lower than in the previous years (Wald $X^2(1)$ =4.947, p=.026). The mean rate of hard landings in 2003 was higher than in 2002, but this difference was only marginally statistically significant (Wald $X^2(1)$ =3.714, p=.054). However, these differences are a product of a more complex pattern (see Figure 1). Only the first quarter of 2002 demonstrates a substantially lower rate than the corresponding quarter in 2003. In addition, after a drop in the rate of hard landings in the first quarter of 2004 the rate climbs steadily back towards the mean for 2003.

Significant differences between seasons were observed (Wald $X^2(3)$ =8.581, p=.035; see Figure 3). The rate of reported hard landings in winter was significantly lower than in the other quarters (Wald $X^2(1)$=8.757, p=.003), which did not differ (Wald $X^2(1)$=1.121, p=.290). No year by season interaction was observed (Wald $X^2(6)$ =3.414, p=.755).

Figure 2: Experience a Hard Landing by Year

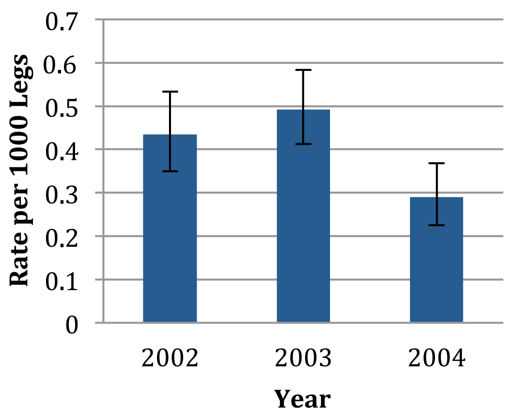
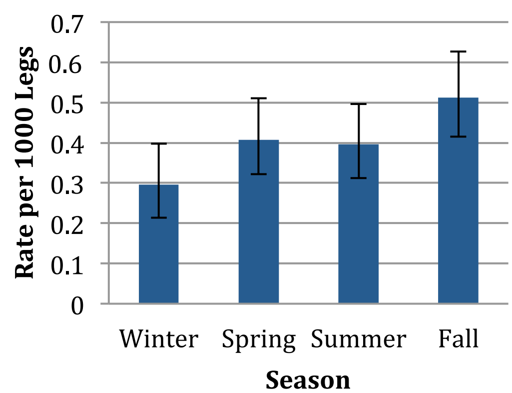


Figure 3: Experience a Hard Landing by Season



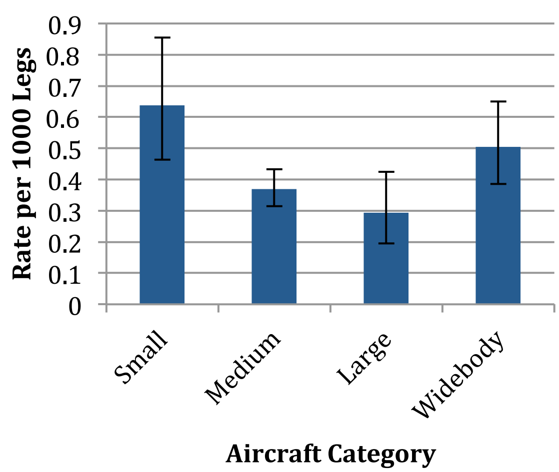
Aircraft Category

There were significant differences in reported rates of hard landings between aircraft categories (Wald $X^2(3)=13.742$, p=.003; see Figure 4). Pilots of medium and large aircraft reported lower rates of hard landings than did pilots of small and wide-body aircraft (Wald $X^2(1)=11.062$, p=.001). No other differences were significant.

Interaction

Full interaction models could not be estimated. Investigations of year/season patterns within each category were hampered by the small number of events. The general yearly pattern described above was observed in each aircraft category. It was significant for the medium aircraft (Wald $X^2(1)=7.580$, p=.006) and nearly significant for wide-body aircraft (Wald $X^2(1)=3.349$, p=.067). No significant season effects were observed.

Figure 4: Experience a Hard Landing by Aircraft Category

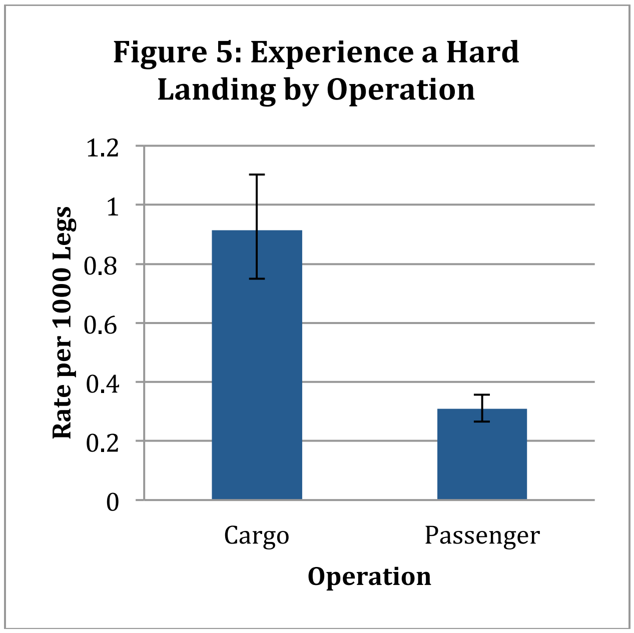


Operation

Pilots of cargo aircraft reported a higher rate of hard landings than did pilots of passenger aircraft (Wald $X^2(1)=73.260$, $p<.001$; see Figure 5).

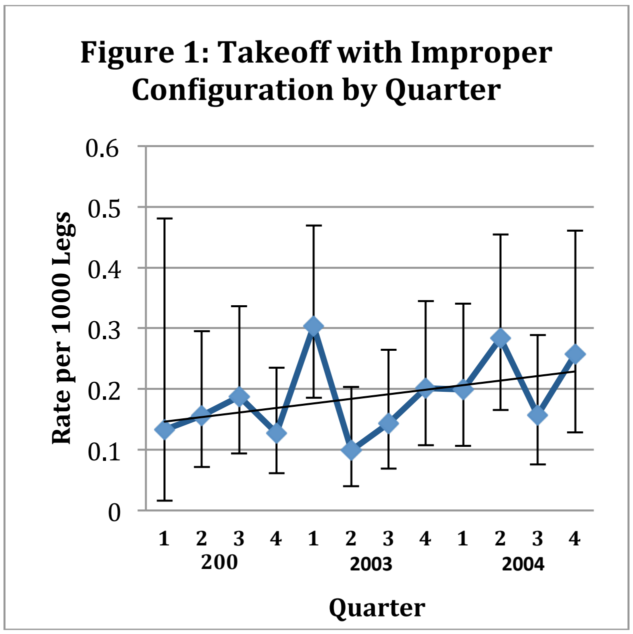
Interaction

Due to convergence problems, a full interaction model could not be specified. Smaller interaction models revealed no significant interactions (LLR $X^2(5)=3.962$, $p=.5549$).



AH12: Takeoff with Improper Configuration

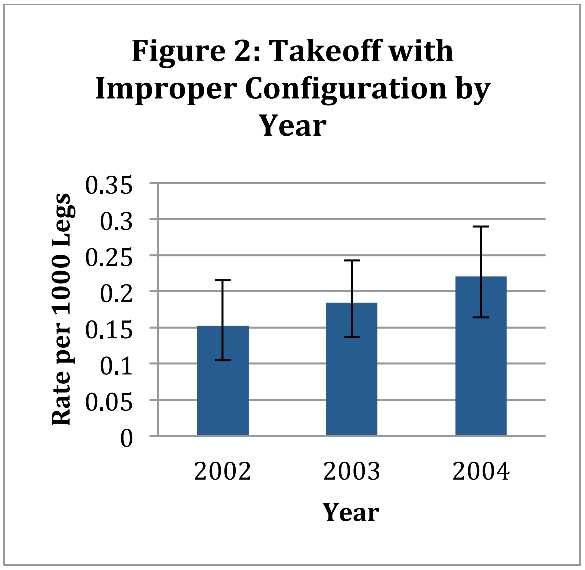
The rate of instances in which pilots reported taking off with improper configurations did not change linearly across the observation period (b=.035, Wald $X^2(1)$ =1.494, p=.222; see Figure 1).[62] Across the study period, a mean rate of 0.19 events per 1,000 legs was reported. The lowest rate was observed in the second quarter of 2003 (0.10) and the highest rate was observed in the first quarter of 2003 (0.30).

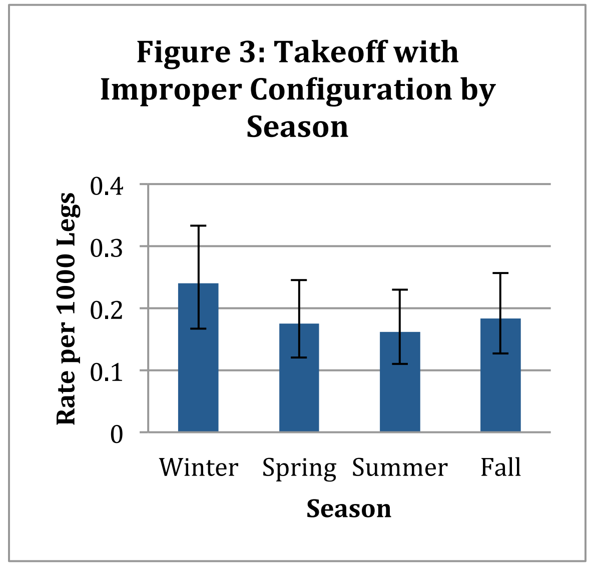


[62] Of the 18,373 reports obtained, 42 (0.23%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 96.36 per 1,000 legs compared to 0.14 per 1,000 legs for the other cases. When the suspect observations are included, little change in the trend was observed. The rate of reported events did not change across the observation period (b=0.018, Wald $X^2(1)$=0.594, p=.441). However, the overall fit of the model decreased (Model with outliers: Pearson $X^2(18371)$=38016.518, ratio=2.069; Model without outliers: Pearson $X^2(18329)$=15447.432; ratio=0.843). These cases were omitted from the remaining analyses.

Year/Season

No statistically significant differences between years were observed (Wald $X^2(2) = 2.306$, p=.316; see Figure 2). No differences between seasons were observed (Wald $X^2(3) = 0.732$, p=.866; see Figure 3). No year by season interaction was observed (Wald $X^2(6) = 9.081$, p=.169).



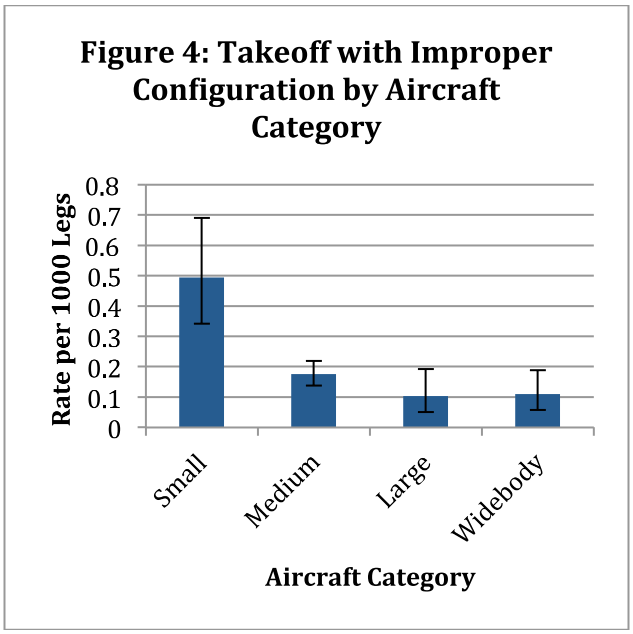


Aircraft Category

Rates of takeoffs with improper configurations varied by aircraft category (Wald $X^2(3)=35.255$, $p<.001$; see Figure 4). Pilots of small aircraft reported higher rates than did pilots of larger aircraft (Wald $X^2(1)=15.424$, $p<.001$). Pilots of medium aircraft reported higher rates of these events than did pilots of large and wide-body aircraft (Wald $X^2(1)=5.390$, $p=.020$). The rates of events reported by pilots of large aircraft did not differ from the rates of events reported by pilots of wide-body aircraft (Wald $X^2(1)=0.026$, $p=.873$).

<u>Interactions</u>

No interactions between aircraft category and year/season were observed (ΔLLR $X^2(33)=46.426$, $p=.061$).[63]



---

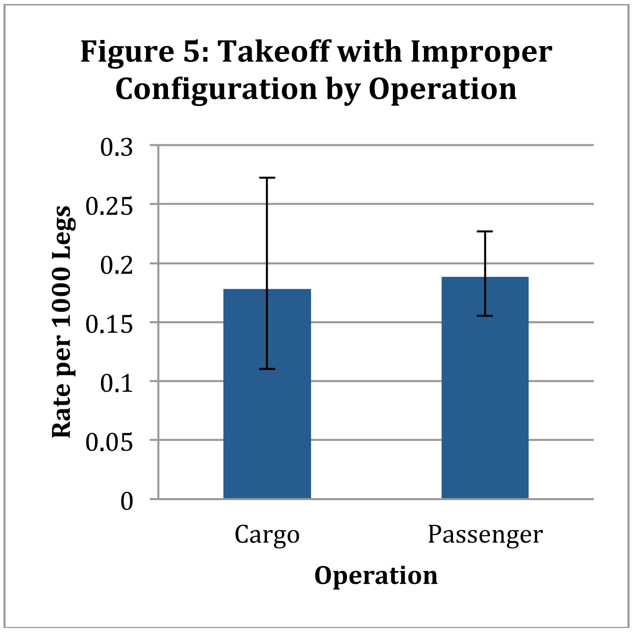
[63] The full factorial model did not fulfill convergence criteria. However, smaller models that did fulfill criteria were fit to the data as well and these also did not demonstrate any interaction effects.

Operation

There were no significant differences between cargo and passenger operations in the reported rates of these events (Wald $X^2(1)=0.048$, $p=.826$).

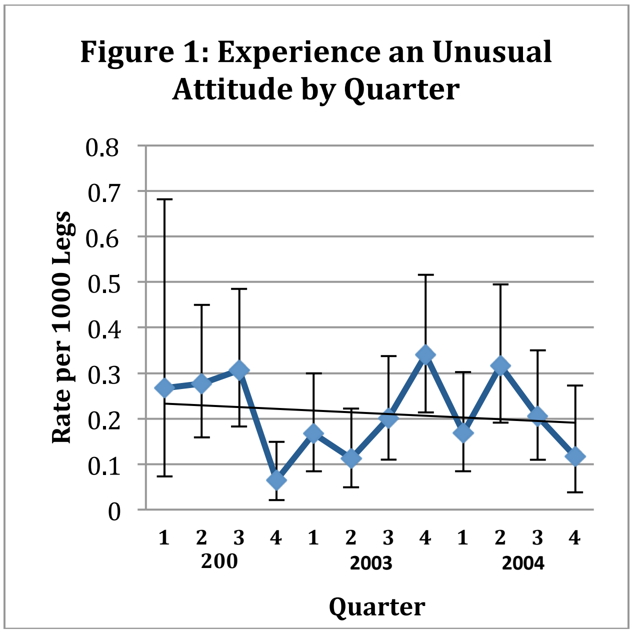
Interactions

Interactions between year/season and operations could not be fit due to the very small number of events (21) reported by pilots of cargo aircraft and the distribution of these events across years and seasons.



AH13: Experience an Unusual Attitude

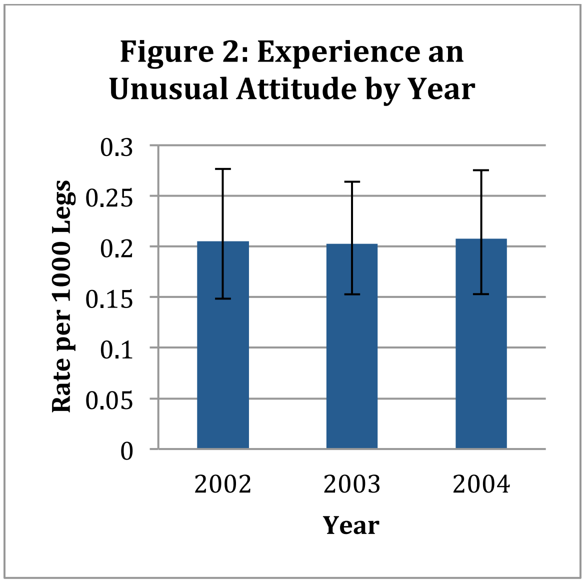
The rate of reported instances of unusual attitudes did not change linearly across the observation period (b=.001, Wald $X^2(1)$ =0.001, p=.980; see Figure 1).[64] Across the study period, a mean rate of 0.21 events per 1,000 legs was reported. The lowest rate was observed in the fourth quarter of 2002 (0.064) and the highest rate was observed in the fourth quarter of 2003 (0.34).

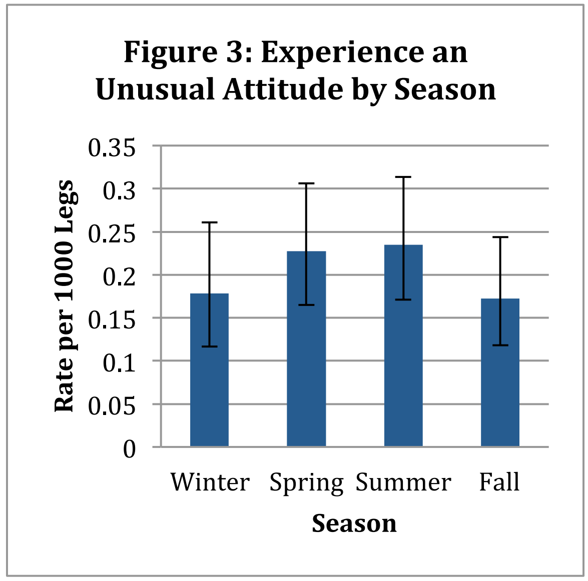


[64] Of the 18,372 reports obtained, 32 (0.17%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 115.21 per 1,000 legs compared to 0.18 per 1,000 legs for the other cases. When the suspect observations are included, little change in the trend was observed. The rate of reported events did not change linearly across the observation period (b=-.016, Wald $X^2(1)$=0.489, p=.484). However, the overall fit of the model decreased (Model with outliers: Pearson $X^2(18370)$=42713.391, ratio=2.325; Model without outliers: Pearson $X^2(18338)$=17638.572; ratio=0.962). These cases were omitted from the remaining analyses.

Year/Season

No differences between years (Wald $X^2(2)$ =0.023, p=.989; see Figure 2) or seasons (Wald $X^2(3)$ =3.945, p=.267; see Figure 3) were observed. However, there was a significant year by season interaction (Wald $X^2(6)$ =22.343, p=.002), indicating that the rates of these events varied across seasons and this seasonal pattern changed across years. However, there was no clear pattern in these shifts.



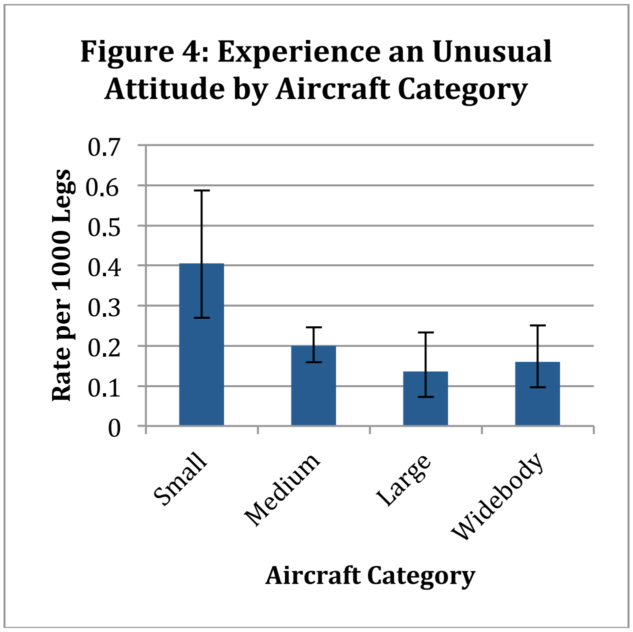


Aircraft Category

The rate of these events differed by aircraft category (Wald $X^2(3)=14.257$, p=.003). Pilots of small aircraft reported higher rates of unusual attitudes than did pilots of larger aircraft (Wald $X^2(1)=7.965$, p=.005). No other differences between aircraft categories were observed.

<u>Interactions</u>

No interactions between aircraft category and year/season were observed (ΔLLR $X^2(33)=40.662$, p=.169).[65]



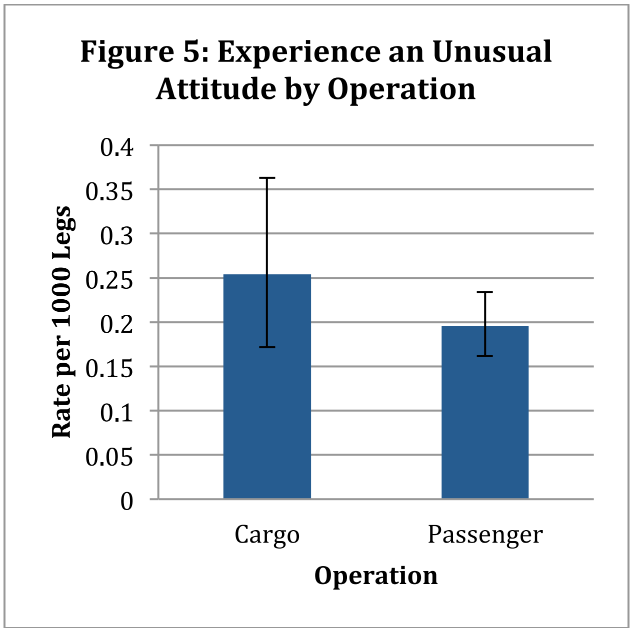
[65] The full factorial model did not fulfill convergence criteria. However, smaller models that did fulfill criteria were fit to the data as well and these also did not demonstrate any interaction effects.

Operation

No differences due to type of operation were observed (Wald $X^2(1)=1.823$, $p=.177$; see Figure 5).

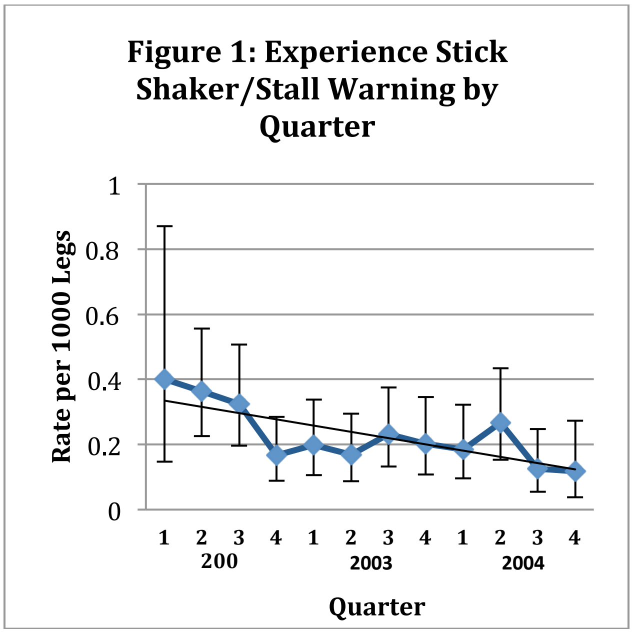
Interaction

There were no significant interactions between operation and year/season (Δ LLR $X^2(11)=11.536$, $p=.340$).



AH14: Experience Stick Shaker/Stall Warning

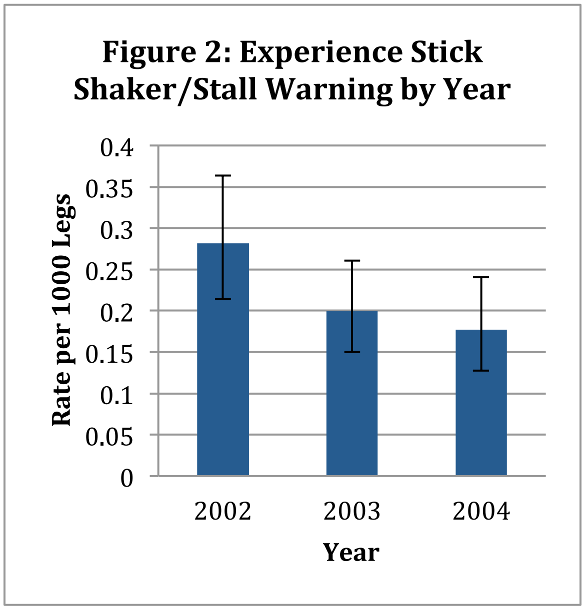
The reported rate of stick shaker/stall warning events decreased linearly across the observation period (b=-.072, Wald $X^2(1)$ =7.295, p=.007; see Figure 1).[66] Across the study period, a mean rate of 0.22 events per 1,000 legs was reported. The lowest rate was observed in the fourth quarter of 2004 (0.12) and the highest rate was observed in the first quarter of 2002 (0.40).

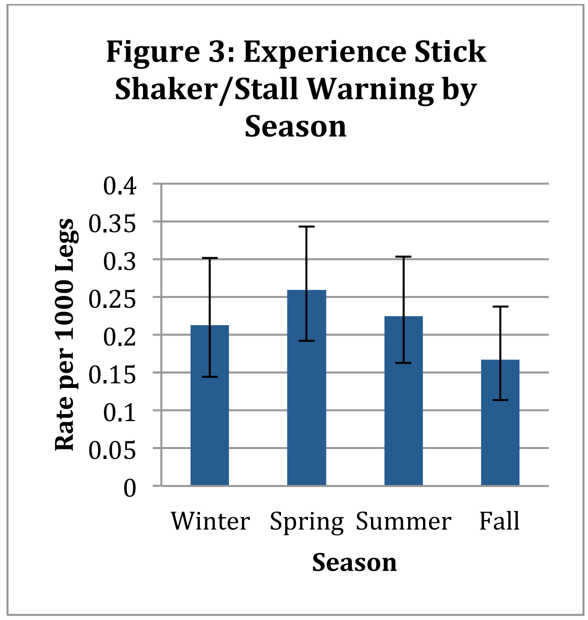


[66] Of the 18,356 reports obtained, 50 (0.27%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 122.33 per 1,000 legs compared to 0.21 per 1,000 legs for the other cases. When the suspect observations are included, the observed trend remains in the same direction, but it becomes non-significant (b=-.019, Wald $X^2(1)$=0.774, p=.379) and the overall fit of the model is much poorer (Model with outliers: Pearson $X^2(18354)$=44093.694, ratio=2.402; Model without outliers: Pearson $X^2(18304)$=19946.043; ratio=1.090). These cases were omitted from the remaining analyses.

Year/Season

Significant differences between years were observed (Wald $X^2(2) = 7.374$, p=.025; see Figure 2). The rate of stick shaker/stall warning experiences decreased linearly across the years in the sample (Wald $X^2(1) = 4.962$, p=.026). No differences between seasons (Wald $X^2(3) = 4.024$, p=.259) and no year by season interaction were observed (Wald $X^2(6) = 6.240$, p=.397).





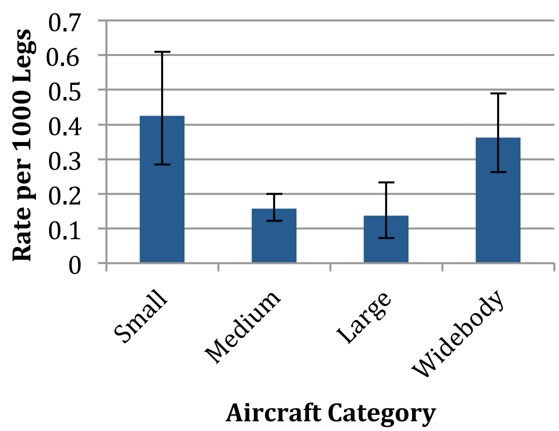
Aircraft Category

There were significant differences in reported rates of stick shaker/stall warning events between aircraft categories (Wald $X^2(3)=31.032$, $p<.001$; see Figure 4). Pilots of medium and large aircraft reported lower rates of these events than did pilots of small and wide-body aircraft (Wald $X^2(1)=20.061$, $p<.001$). No other differences between categories were observed.

<u>Interaction</u>

A significant interaction between aircraft category and year/season was observed (ΔLLR $X^2(33)=49.294$, $p=.034$). However, given the small number of events and the uneven distribution of those events, neither the full interaction model nor models fit to each category separately could be properly fit. Smaller models did not demonstrate significant interactions.

Figure 4: Experience Stick Shaker/Stall Warning by Aircraft Category

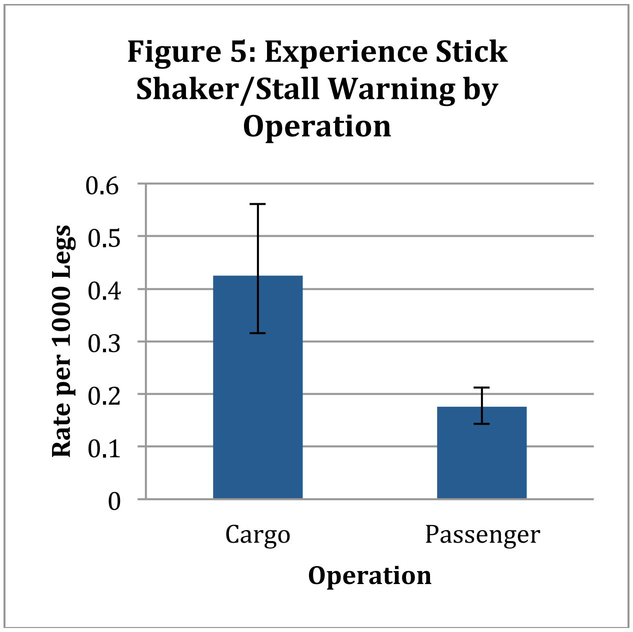


Operation

Pilots of cargo aircraft reported significantly higher rates of these events than did pilots of passenger aircraft (Wald $X^2(2)=26.765$, $p<.001$).

Interactions

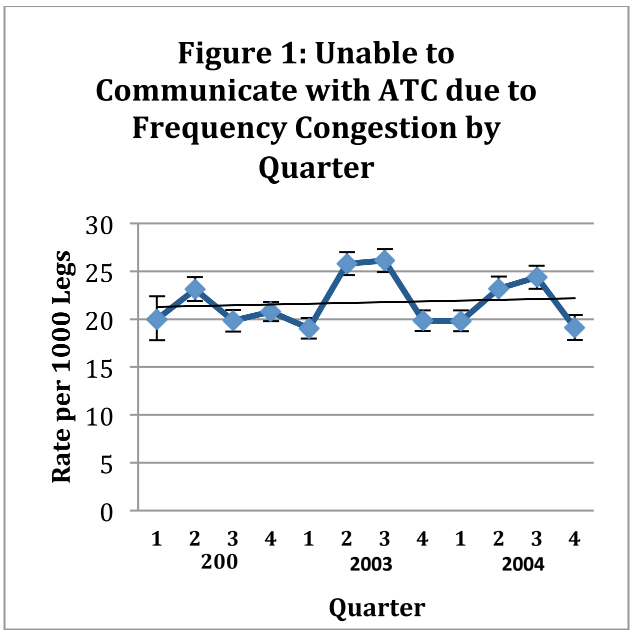
No interactions between operation and year/season were observed (ΔLLR $X^2(11)=4.594$, $p=.949$).



Questions about interactions with air traffic control

AT1: Unable to Communicate with ATC Due to Frequency Congestion

Across the study period, a mean rate of 21.97 events in which pilots were unable to communicate with ATC due to frequency congestion per 1,000 legs was reported. The lowest rate was observed in the first quarter of 2003 (19.02) and the highest rate was observed in the third quarter of 2003 (26.09). The rate of reported instances in which pilots were unable to communicate with ATC due to frequency congestion did not change linearly across the observation period (b=-.001, Wald $X^2(1)$ =0.060, p=.807; see Figure 1).[67]



[67] Of the 18,355 reports obtained, 47 (0.26%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 1084.41 per 1,000 legs compared to 26.11 per 1,000 legs for the other cases. When the suspect observations are included, little change in the trend was observed (b=.000, Wald $X^2(1)$=.001, p<.997). However, the overall fit of the model decreased (Model with outliers: Pearson $X^2(18353)$=26708.736, ratio=1.455; Model without outliers: Pearson $X^2(18305)$=20657.881; ratio=1.129). The outlier cases were omitted from the remaining analyses.

Year/Season

There were no differences across the years (Wald $X^2(2)$ =0.923, p=.630; see Figure 2), but there was a significant seasonal pattern (Wald $X^2(3)$ =17.256, p=.001; see Figure 3). The rate of these events was higher in the spring and summer than in the winter and fall (Wald $X^2(1)$ =17.641, p<.001). The rate of events during the fall did not differ from that during the winter (Wald $X^2(1)$ =0.426, p=.514) and the rate during the spring did not differ from that during the summer (Wald $X^2(1)$ =0.108, p=.742). There was no year by season interaction (Wald $X^2(6)$ =9.339, p=.155).

Figure 2: Unable to Communicate with ATC due to Frequency Congestion by Year

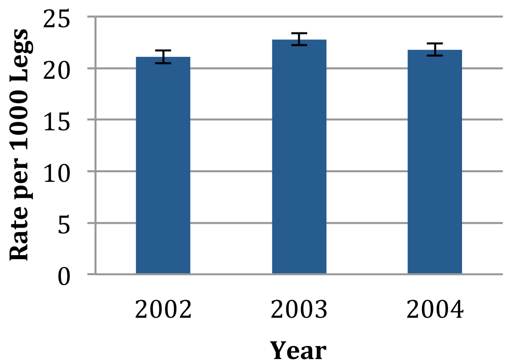
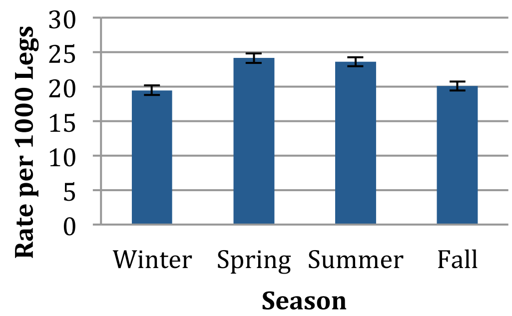


Figure 3: Unable to Communicate with ATC due to Frequency Congestion by Season

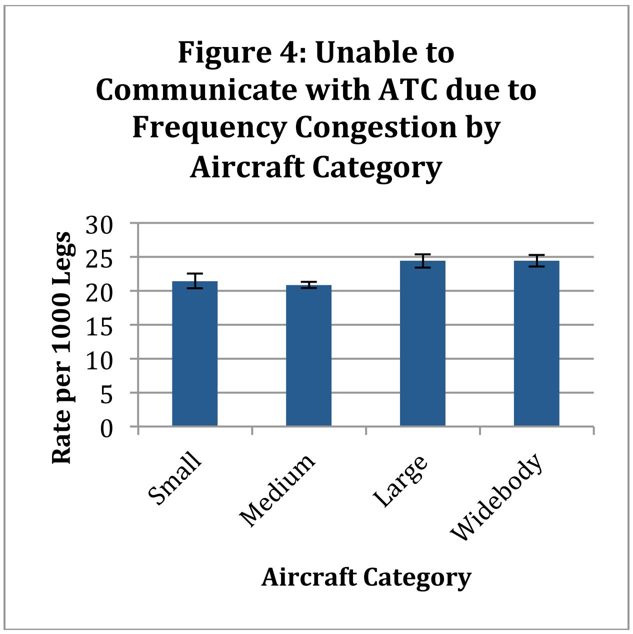


Aircraft Category

There were marginally significant differences between aircraft categories (Wald $X^2(3)$= 7.748. p=.052; see Figure 4). Pilots of medium sized aircraft reported significantly lower rates of these events than did pilots of wide-body aircraft (Wald $X^2(1)$=4.889, p=.027) and nearly significantly lower rates than did pilots of large aircraft (Wald $X^2(1)$=3.654, p=.056). No other differences between aircraft categories were statistically significant.

Interactions

There was a significant interaction between year/season and aircraft category (ΔLLR $X^2(33)$ =53.488, p=.013). This was due largely to the higher variability in the rates of these events reported by pilots of small aircraft. There is no interaction between aircraft category and year/season when small aircraft are eliminated from the analysis (ΔLLR $X^2(21)$ =32.043, p=.080). For small aircraft, the seasonal pattern varied by year (Wald $X^2(6)$=16.765, p=.010).

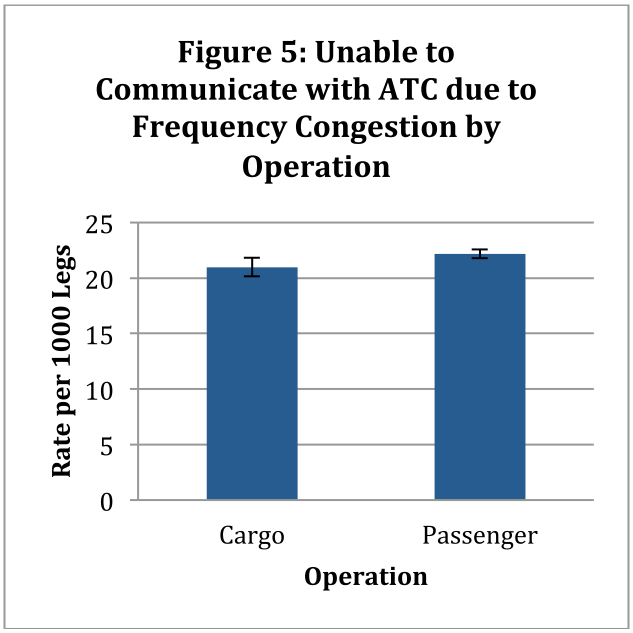


Operation

The rates of reported failures to contact ATC due to frequency congestion did not differ between pilots of passenger aircraft and pilots of cargo aircraft (Wald $X^2(1)=0.475$, p=.491; see Figure 5).

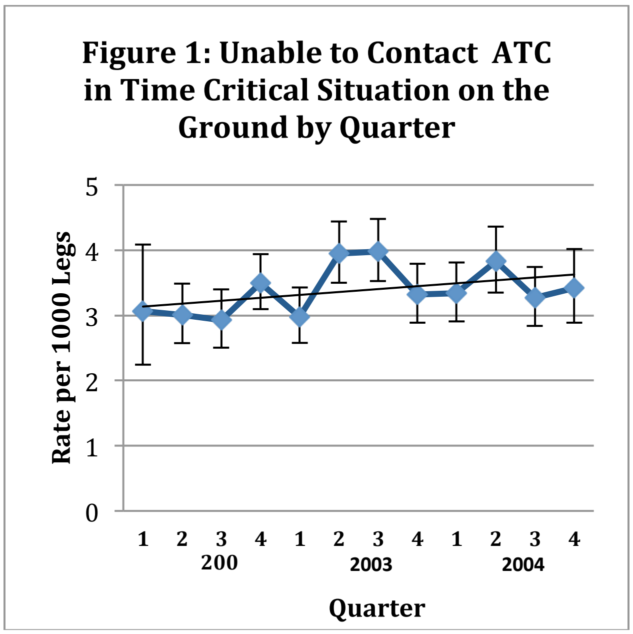
Interactions

There were no interactions between type of operation and year/season (ΔLLR $X^2(11)=9.21$, p=.603).



AT1A: Unable to Contact ATC in Time Critical Situation on the Ground

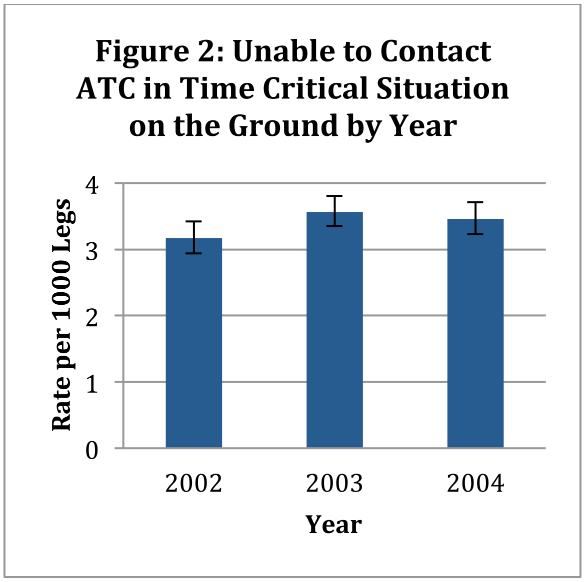
Across the study period, a mean rate of 3.42 events in which aircraft were unable to contact ATC in time critical situations while on the ground per 1,000 legs was reported. The lowest rate was observed in the third quarter of 2002 (2.93) and the highest rate was observed in the third quarter of 2002 (3.98). The rate of reported these events did not change linearly across the observation period (b=.008, Wald $X^2(1)$ =0.955, p=.328; see Figure 1).[68]

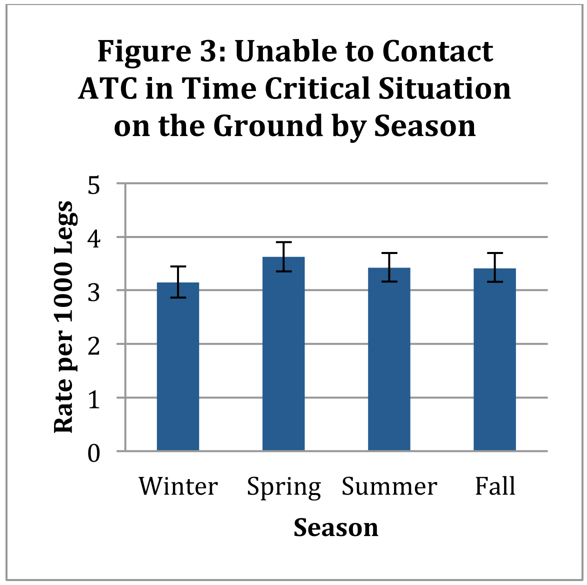


[68] Of the 18,357 reports obtained, 52 (0.28%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution. The mean reported event rate amongst these cases was 447.22 per 1,000 legs compared to 3.66 per 1,000 legs for the other cases. When the suspect observations are included, a significant change in the trend was observed (b=.020, Wald $X^2(1)$=7.448, p=.006). However, the overall fit of the model with outliers was very poor (Pearson $X^2(18355)$=80064.030, ratio=4.362) and substantially worse than the fit of the model without outliers (Pearson $X^2(18303)$=41559.360; ratio=2.271). The outlier cases were omitted from the remaining analyses.

Year/Season

There were no significant differences in the rates of these events across years (Wald $X^2(2) = 2.751$, $p = .276$; see Figure 2) or seasons (Wald $X^2(3) = 5.263$, $p = .154$; see Figure 3). There was no year by season interaction (Wald $X^2(6) = 6.069$, $p = .416$).



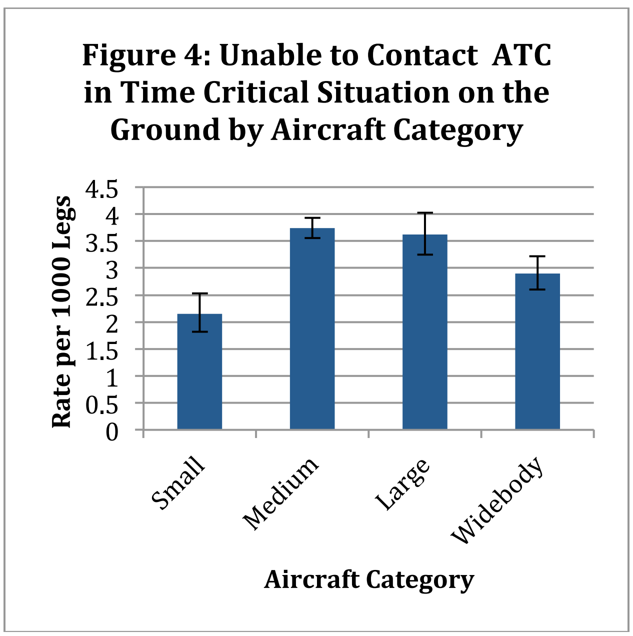


Aircraft Category

The reported rates of these events differed by aircraft category (Wald $X^2(3)=43.564$, p<.001; see Figure 4). Pilots of small aircraft reported the lowest rates of these events and significantly lower rates than did pilots of wide-body aircraft (Wald $X^2(1)=5.361$, p=.021). In turn, pilots of wide-body aircraft reported significantly lower rates than did pilots of medium aircraft (Wald $X^2(1)=22.814$, p<.001) and large aircraft (Wald $X^2(1)=7.918$, p=.005). The rates of these events reported by pilots of medium aircraft did not differ from the rates reported by pilots of large aircraft (Wald $X^2(1)=0.507$, p=.476).

Interactions

There was a significant interaction between aircraft category and year/season (ΔLLR $X^2(33)=102.184$, p<.001). In several years/seasons the pattern of rates differed from that described above. However, there was no clear pattern to these exceptions. Furthermore, the pattern described above was observed during most quarters. In 9/12 quarters, the mean rate of events reported by pilots of small aircraft was lower than that reported by pilots of wide-body aircraft. In 8/12 quarters, the mean rate of events reported by pilots of wide-body aircraft was lower than that reported by pilots of large aircraft. In 11/12 quarters the mean rate of events reported by pilots of wide-body aircraft was lower than that reported by pilots of medium aircraft.

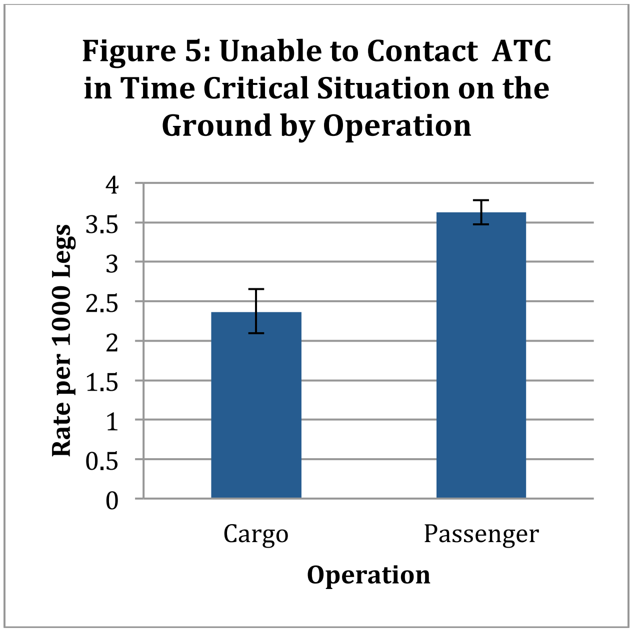


Operation

Pilots of cargo aircraft reported significantly lower rates of these events than did pilots of passenger aircraft (Wald $X^2(1)=34.916$, $p<.001$).

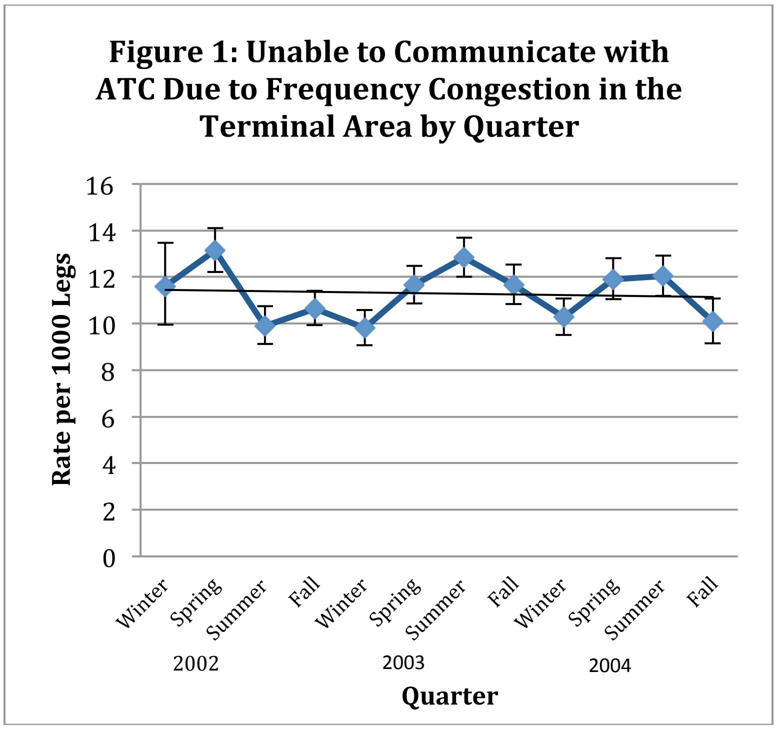
Interactions

There were significant interactions between operation and year/season (ΔLLR $X^2(11)=42.908$, $p<.001$). The mean rates of these events reported by cargo pilots were higher than the mean rate of events reported by passenger pilots during two quarters (spring 2002, winter 2003). During all other quarters, the pattern described above was observed.



AT1B. Unable to Communicate with ATC Due to Frequency Congestion in the Terminal Area

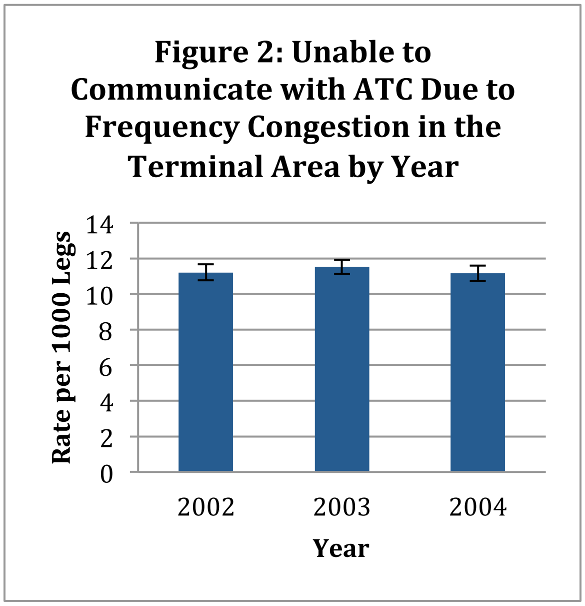
Across the study period, a mean rate of 11.3 events of failures to communicate with ATC due to frequency congestion in the terminal area per 1,000 legs was reported. The lowest rate was observed in the first quarter of 2003 (9.81) and the highest rate was observed in the second quarter of 2002 (13.14). The rate of reported failures to contact ATC due to frequency congestion in the terminal area did not change linearly across the observation period (b=-.004, Wald $X^2(1)$ =0.297, p=.586; see Figure 1).[69]

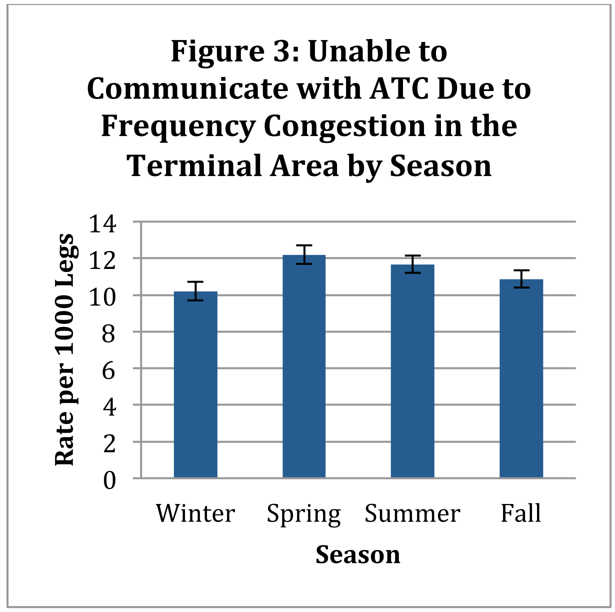


[69] Of the 18,350 reports obtained, 53 (0.29%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 646.35 per 1,000 legs compared to 13.62 per 1,000 legs for the other cases. When the suspect observations are included, little change in the trend was observed. No change in the rate of reported events was observed (b=.003, Wald $X^2(1)$=.243, p=.622). However, the overall fit of the model decreased substantially (Model with outliers: Pearson $X^2(18347)$=24645.037, ratio=1.343; Model without outliers: Pearson $X^2(18294)$=19364.898; ratio=1.054). The outlier cases were omitted from the remaining analyses.

Year/Season

No differences in the event rates by year (Wald $X^2(2)=.730$, p=.694; see Figure 2) or season (Wald $X^2(3)=6.069$, p=.108; see Figure 3) were observed. No year by season interaction was observed (Wald $X^2(6)=6.921$, p=.328).



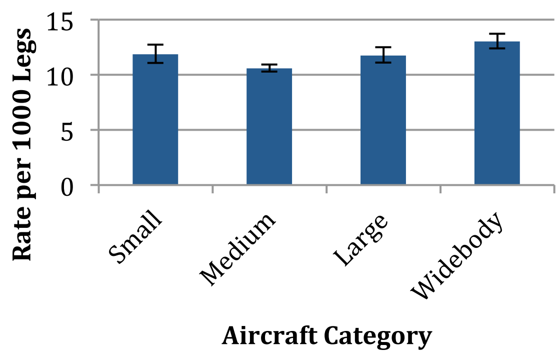


Aircraft Category

There were differences across aircraft categories in the rate of reports of being unable to contact ATC due to frequency congestion in the terminal area (Wald $X^2(3)=13.348$, p=.004; see Figure 4). Pilots of medium aircraft reported fewer events than did pilots of small (Wald $X^2(1)=4.689$, p=.030) or wide-body aircraft (Wald $X^2(1)=9.863$, p=.002). Pilots of large aircraft reported slightly more events than did pilots of medium aircraft, but this difference was not statistically significant at conventional levels (Wald $X^2(1)=2.137$, p=.144). No other differences between aircraft categories were observed.

Interactions. A significant interaction between year/season, and aircraft category was observed (ΔLLR $X^2(33)=48.705$, p=.038). This was due to statistically significant changes across quarters in the rate of reported events from pilots of small aircraft (Wald $X^2(6)=18.927$, p=.004). No other aircraft category demonstrated a significant change over time. The observed changes in the rates of reported events in small aircraft did not appear to follow any consistent pattern.

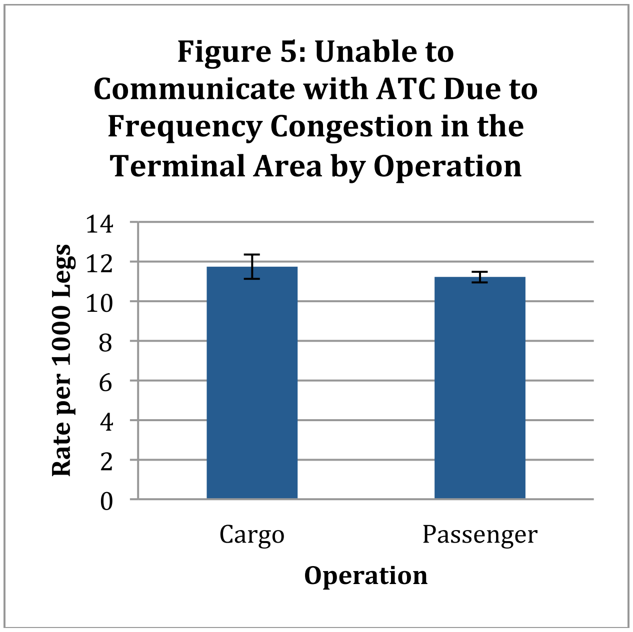
Figure 4: Unable to Communicate with ATC Due to Frequency Congestion in the Terminal Area by Aircraft Category



Operation

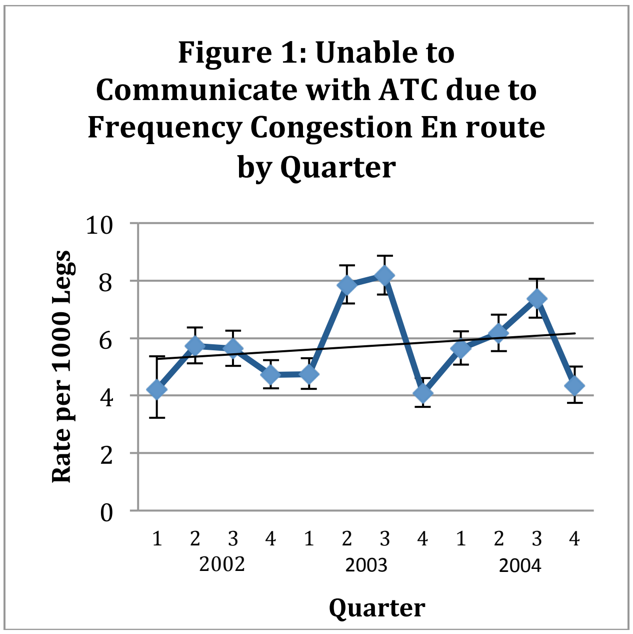
The reported rates of these events did not differ by type of operation (Wald $X^2(1)=2.339$, p<.126; see Figure 5).

Interactions. No interactions between type of operation and year/season were observed (ΔLLR $X^2(33)=9.014$, p=.999).



AT1C: Unable to Communicate with ATC Due to Frequency Congestion En route

Across the study period, a mean rate of 5.88 failures to communicate with ATC due to frequency congestion en route per 1,000 legs was reported. The lowest rate was observed in the fourth quarter of 2003 (4.08) and the highest rate was observed in the third quarter of 2003 (8.17). The rate of reported failures to communicate with ATC due to frequency congestion while en route did not change linearly across the observation period (b=.008, Wald $X^2(1)=1.635$, p=.201; see Figure 1).[70]



[70] Of the 18,352 reports obtained, 57 (0.31%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution. The mean reported event rate amongst these cases was 477.96 per 1,000 legs compared to 6.67 per 1,000 legs for the other cases. When the suspect observations are included, little change in the trend was observed. There was no change in the rate of reported events across the observation period (b=-.006, Wald $X^2(1)=1.032$, p=.310). However, the overall fit of the model decreased (Model with outliers: Pearson $X^2(18350)=57526.932$, ratio=3.135; Model without outliers: Pearson $X^2(18293)=37081.227$; ratio=2.027). These cases were omitted from the remaining analyses.

Year/Season

There were significant differences in the rate of these events between years (Wald $X^2(2)=7.061$, $p=.029$). The rate of reported events was higher in 2003 than in 2002 (Wald $X^2(1)=7.401$, $p=.007$; see Figure 2). The rates in 2003 and 2004 did not differ (Wald $X^2(1)=1.062$, $p=.303$). There were also significant differences between the seasons (Wald $X^2(3)=84.224$, $p<.001$). The rates of reported events were higher in the spring and summer than in the winter or fall (Wald $X^2(1)=75.648$, $p<.001$). No differences between spring and summer (Wald $X^2(1)=1.493$, $p=.222$)or between winter and fall (Wald $X^2(1)=1.523$, $p=.217$) were observed. There were significant differences in the seasonal pattern across years (Wald $X^2(6)=19.236$, $p=.004$; see Figure 3). However, this was apparently due to small (but statistically significant) shifts in the seasonal pattern, not dramatic changes in the pattern.

Figure 2: Unable to Communicate with ATC due to Frequency Congestion En route by Year

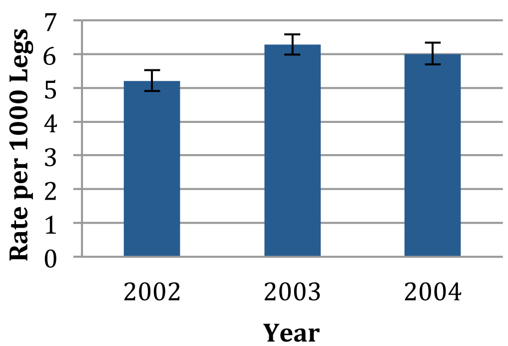
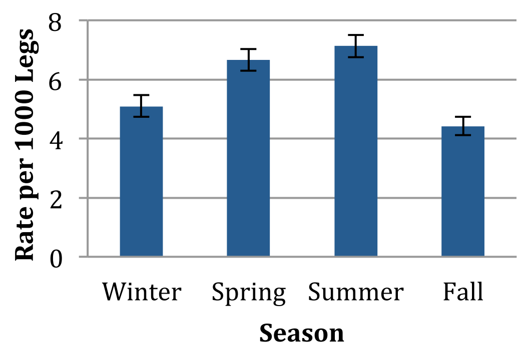


Figure 3: Unable to Communicate with ATC due to Frequency Congestion En route by Season



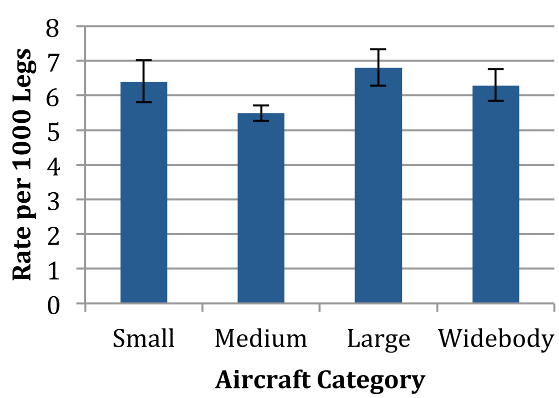
Aircraft Category

There were significant differences between aircraft categories in the rate of these events (Wald $X^2(3)=20.058$, $p<.001$). Pilots of medium aircraft reported significantly lower rates of these events than did pilots of other aircraft (Wald $X^2(1)=17.954$, $p<.001$; see Figure 4). No other differences between aircraft categories were observed.

<u>Interactions</u>

There were significant differences in the year/season patterns by aircraft category (ΔLLR $X^2(33)=89.058$, $p<.001$). For small aircraft, there was a significant seasonal pattern (Wald $X^2(3)=23.298$, $p<.001$) that changed across years (Wald $X^2(6)=35.422$, $p<.001$). For these aircraft, there were no differences between the seasons in 2002 (Wald $X^2(3)=2.719$, $p=.437$). In 2003 and 2004, the rates of these events during the spring and summer were higher than during the winter and fall (2003 Wald $X^2(1)=4.976$, $p=.026$; 2004 Wald X2(1)=7.962, p=.005) and the rate during the winter was higher than during the fall (2003 Wald $X^2(1)=6.059$, $p=.014$; 2004 Wald $X^2(1)=5.649$, $p=.017$). For medium aircraft, the rates of these events during the spring and summer are higher than the winter and fall during each year (2002: Wald $X^2(1)=15.159$, $p<.001$; 2003 Wald $X^2(1)=38.493$, $p<.001$; Wald $X^2(1)=17.616$, $p<.001$). For large aircraft, there was no seasonal effect during 2002 (Wald X2(3)=2.885, p=.410). During 2003, the rates of these events were higher during the spring and summer than during the winter and fall (2003 Wald $X^2(1)=9.052$, $p=.003$). During 2004, the rates of these events during the fall were lower than during the other seasons (Wald $X^2(1)=12.559$, $p<.001$), which did not differ. For wide-body aircraft, there were differences between the seasons during 2002 (Wald $X^2(3)=17.319$, $p=.001$) and 2004 (Wald $X^2(3)=8.533$, $p=.036$) but the pattern varied each year. For these aircraft there were no differences between the seasons during 2003 (Wald $X^2(3)=3.314$, $p=.346$).

Figure 4: Unable to Communicate with ATC due to Frequency Congestion En route by Aircraft Category

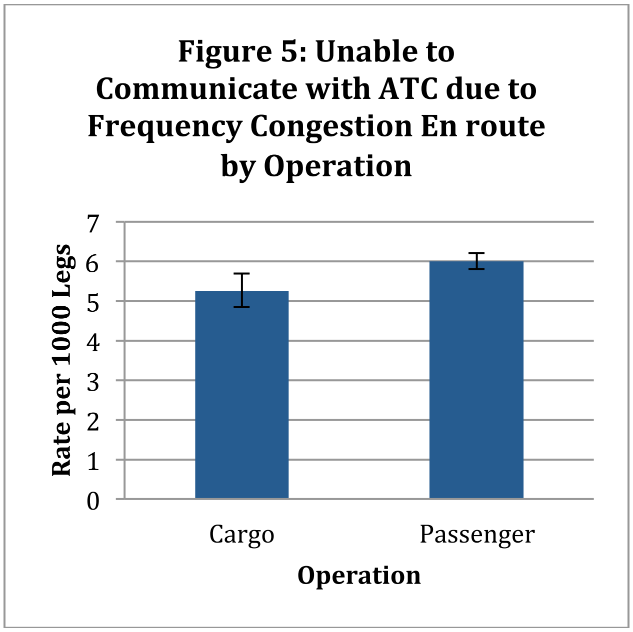


Operation

Pilots of passenger aircraft reported higher rates of these events than did pilots of cargo aircraft (Wald $X^2(1)=4.571$, $p=.033$; see Figure 5).

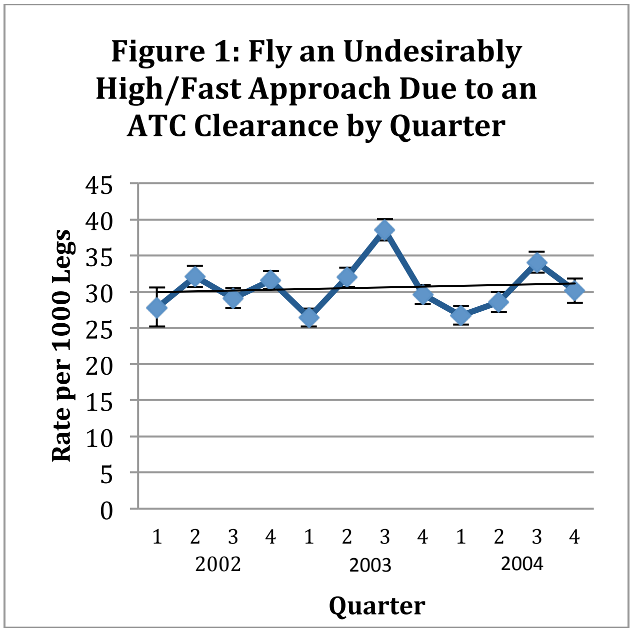
Interactions

There was a marginally significant interaction between operation and year /season (ΔLLR $X^2(11)=19.41$, $p=.054$). For passenger operations the seasonal pattern described above was generally apparent. However, for cargo operations, the rates of these events during spring appeared to be consistently higher than during the other quarters with summer also demonstrating a similarly high rate only during 2004.



AT2: Fly an Undesirably High/Fast Approach Due to an ATC Clearance

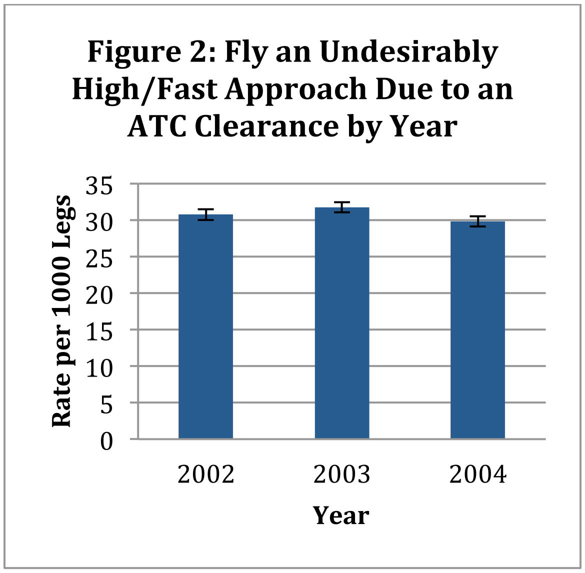
Across the study period, a mean rate of 30.84 instances in which pilots flew an undesirably high/fast approach per 1,000 legs was reported. The lowest rate was observed in the first quarter of 2004 (26.41) and the highest rate was observed in the third quarter of 2003 (38.53). The rate of reported undesirably high/fast approaches due to ATC clearances did not increase or decrease linearly across the observed quarters (b=-.005, Wald $X^2(1)$=.930, p=.335).[71]

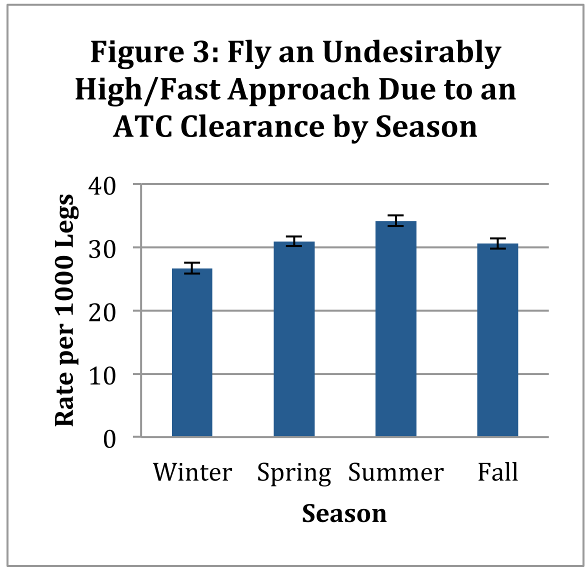


[71] Of the 18,342 reports obtained, 34 (0.19%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 930.31 per 1,000 legs compared to 38.02 per 1,000 legs for the other cases. When the suspect observations are included, little change in the trend was observed. The rate of these events did not change across the observation period (B=-.006, Wald $X^2(1)$=1.543, p=.214). However, the overall fit of the model decreased somewhat (Model with outliers: Pearson $X^2(18339)$=20957.247, ratio=1.143; Model without outliers: Pearson $X^2(18305)$=19314.558; ratio=1.055). The outlier cases were omitted from the remaining analyses. These cases were omitted from the remaining analyses.

Year/Season

The rate of reported events did not change over the observed years (Wald $X^2(2)=3.664$, p=.160; see Figure 2). However, there was a significant seasonal pattern (Wald $X^2(3)=15.045$, p=.002; see Figure 3). The rate of reported events was lower in the winter than during the other quarters (Wald $X^2(1)=12.200$, p<.001), which did not differ. This pattern did not vary significantly by year. There was no year by season interaction (Wald $X^2(6)=7.808$, p=.252).

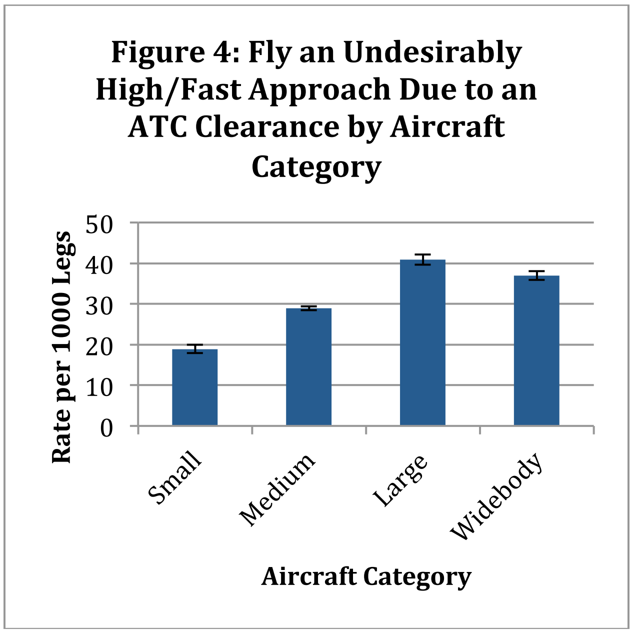




Aircraft Category

The rates of reported events differed across aircraft categories (Wald $X^2(3)=94.954$, $p<.001$; see Figure 4). Each of the aircraft categories differed from the others ($p\leq.001$). Pilots of small aircraft reported the lowest rate of instances in which they flew an undesirably high or fast approach due to an ATC clearance. Pilots of large aircraft reported the highest rate.

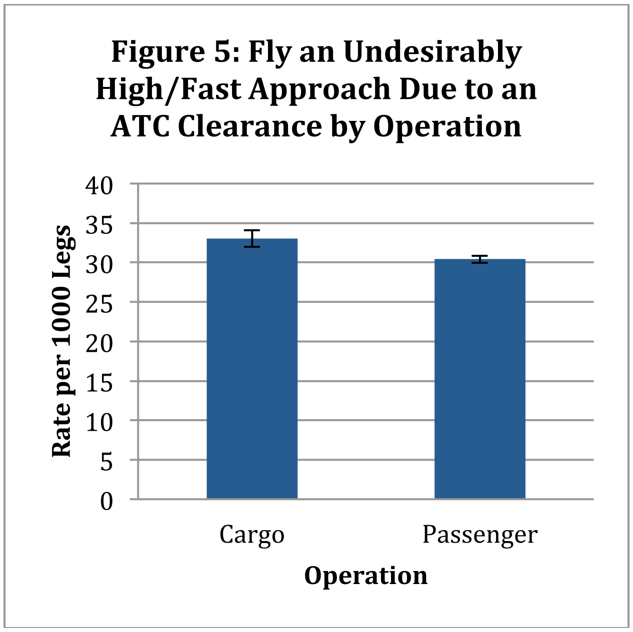
Interactions. No interaction between aircraft category and year/season was observed (ΔLLR $X^2(33)=30.282$, $p=.603$).



Operation

Pilots of cargo aircraft reported higher rates of these events than did pilots of passenger aircraft (Wald $X^2(1)=6.714$, p=.010; see Figure 5).

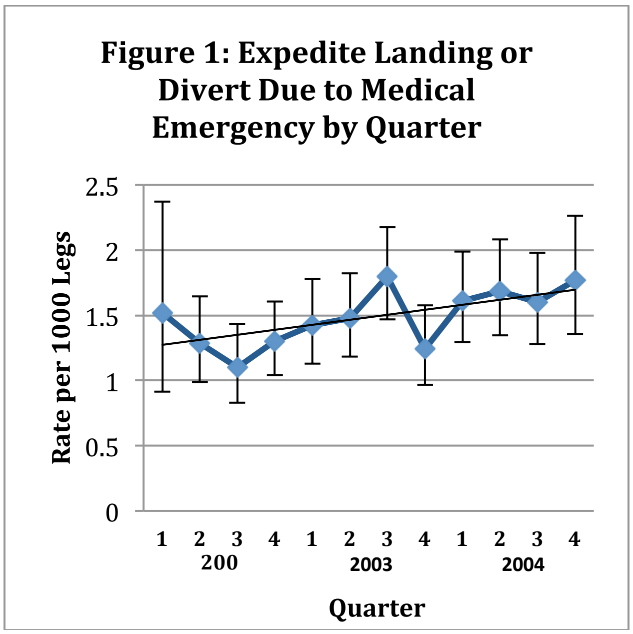
Interactions. No interaction between operation and year/season was observed (ΔLLR $X^2(11)=13.638$, P=.254).



Questions about passenger-related events

CP1: Expedite Landing or Divert Due to Medical Emergency

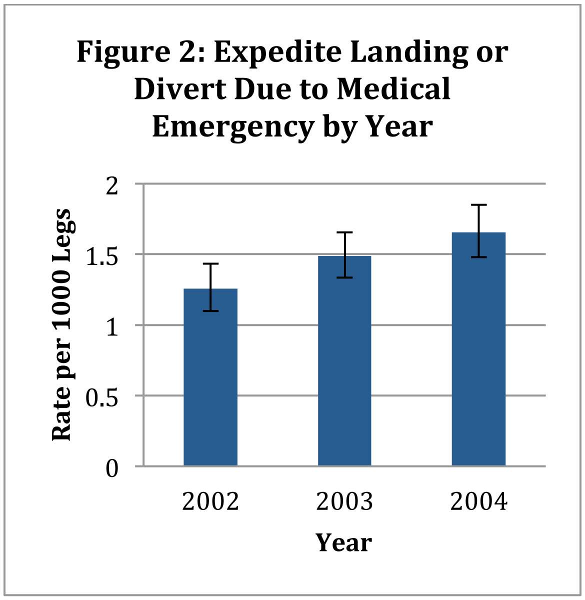
The reported rates of expedited landings or diversions due to medical emergencies increased linearly across the observation period (b=.035, Wald $X^2(1)$ =8.079, p=.004; see Figure 1).[72] Across the study period, a mean rate of 1.47 events per 1,000 legs was reported. The lowest rate was observed in the third quarter of 2002 (1.10) and the highest rate was observed in the third quarter of 2003 (1.80).

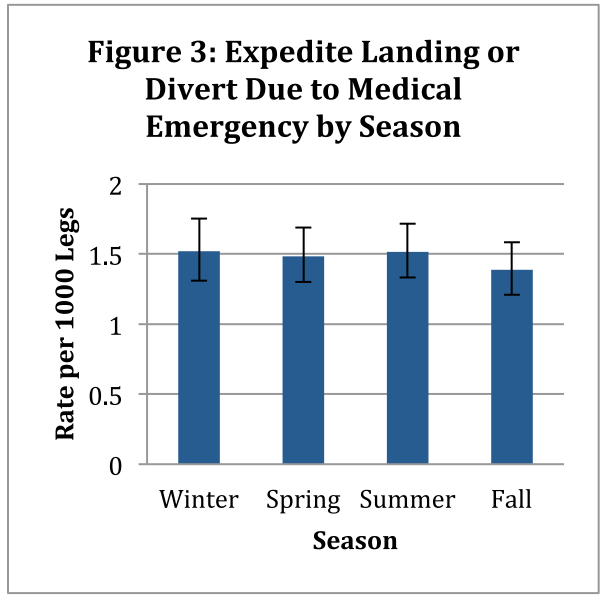


[72] Of the 14,282 reports obtained, 48 (0.34%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 176.08 per 1,000 legs compared to 1.93 per 1,000 legs for the other cases. When the suspect observations are included, little change in the trend was observed. The rate of reported events increased linearly across the observation period (b=.040, Wald $X^2(1)$=11.389, p=.001). However, the overall fit of the model decreased (Model with outliers: Pearson $X^2(14280)$=23014.080, ratio=1.612; Model without outliers: Pearson $X^2(14232)$=17047.351; ratio=1.198). These cases were omitted from the remaining analyses.

Year/Season

Significant differences between years were observed (Wald $X^2(2) = 6.482$, p=.039; see Figure 2). The rate of these events increased linearly across the years in the sample (Wald $X^2(1) = 6.619$, p=.010). No differences between seasons were observed (Wald $X^2(3) = 0.268$, p=.966; see Figure 3). No year by season interaction was observed (Wald $X^2(6) = 6.781$, p=.342).



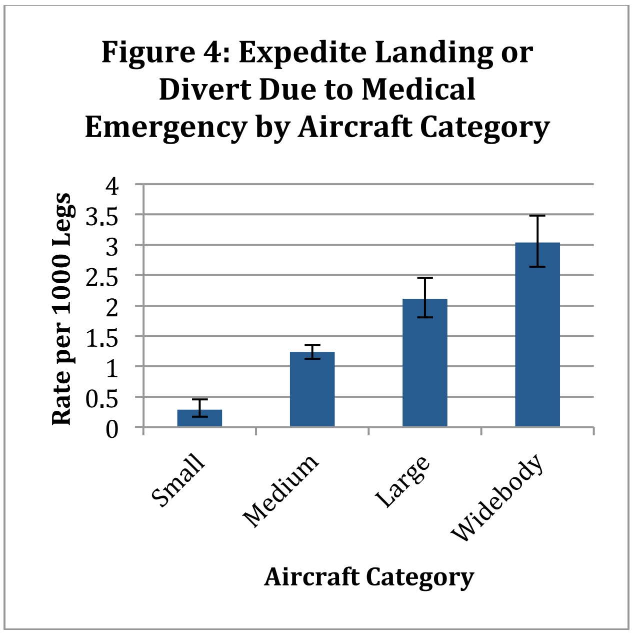


Aircraft Category

There were significant differences between aircraft categories in reported rates of these events (Wald $X^2(3)=146.180$, $p<.001$; see Figure 4). Pilots of small aircraft reported the lowest rates of these events. Pilots of wide-body aircraft reported the highest rates of these events. All differences between aircraft categories were statistically significant ($p<.001$).

Interaction

No interactions between aircraft category and year/season were observed (ΔLLR $X^2(33)=42.742$, $p=.119$).

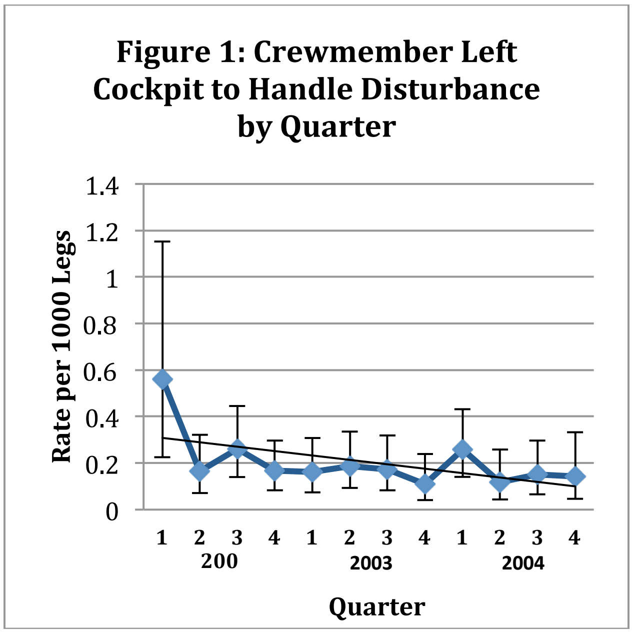


Operation

Cargo operators did not report any instances of diversions or expedited landings due to medical emergencies.

CP3: Crewmember Left Cockpit to Handle Disturbance

The rate of instances in which pilots reported hard landings did not change linearly across the observation period (b=-.056, Wald $X^2(1)$ =3.060, p=.080; see Figure 1).[73] Across the study period, a mean rate of 0.18 events per 1,000 legs was reported. The lowest rate was observed in the fourth quarter of 2003 (0.11) and the highest rate was observed in the first quarter of 2002 (0.56).



[73] Of the 14,280 reports obtained, 53 (0.37%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 126.50 per 1,000 legs compared to 0.29 per 1,000 legs for the other cases. When the suspect observations are included, the linear decrease in the rate of reported events became significant (b=-.056, Wald $X^2(1)$=4.877, p=.027). However, the overall fit of the model decreased (Model with outliers: Pearson $X^2(14278)$=43979.033, ratio=3.080; Model without outliers: Pearson X^2(14225=25453.191; ratio=1.789). In addition, when the first quarter is eliminated from the analysis the negative trend becomes non-significant (Wald $X^2(2)$=2.675, p=.102). The outlier cases were omitted from the remaining analyses.

Year/Season

No significant differences between years (Wald $X^2(2)=5.263$, p=.072; see Figure 2) or seasons (Wald $X^2(3)=7.214$, p=.065) were observed. There was no year by season interaction (Wald $X^2(6)=4.618$, p=.594).

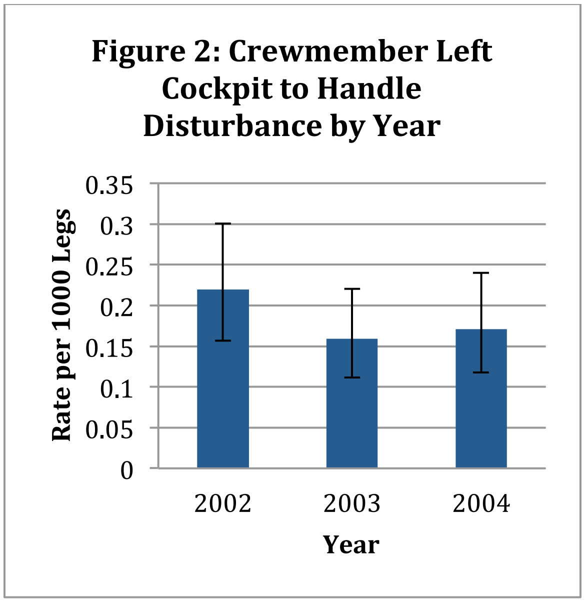


Figure 2: Crewmember Left Cockpit to Handle Disturbance by Year

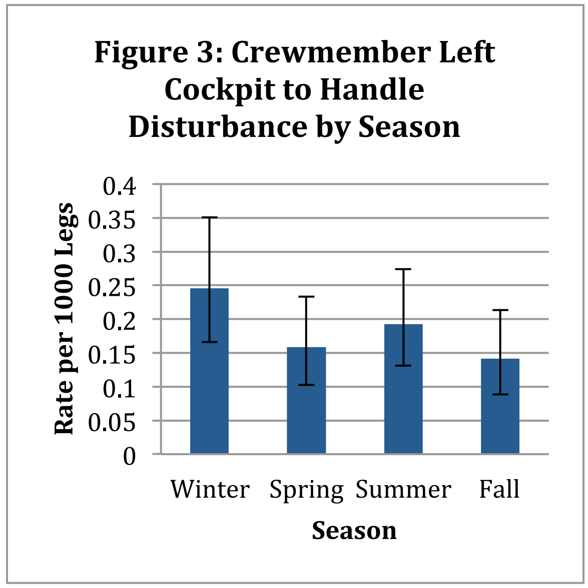


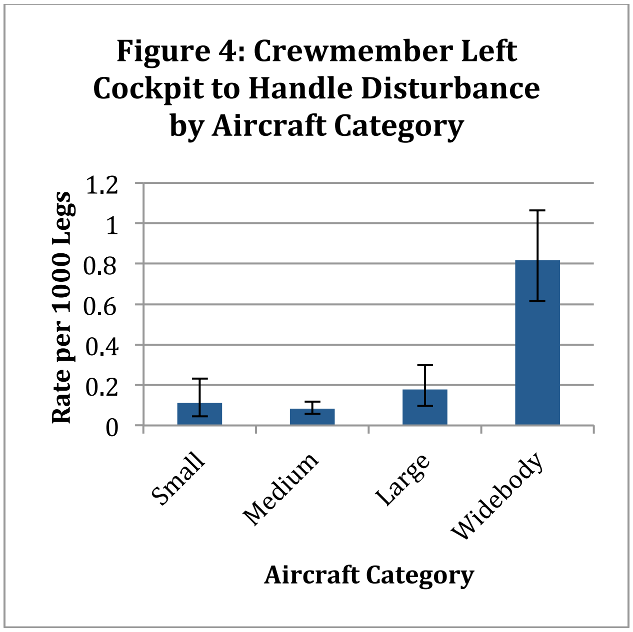
Figure 3: Crewmember Left Cockpit to Handle Disturbance by Season

Aircraft Category

There were significant differences in the rates of these events between aircraft categories (Wald $X^2(3)=112.191$, $p<.001$; see Figure 4). Pilots of wide-body aircraft reported higher rates of leaving the cockpit to handle disturbances than did pilots of smaller aircraft (Wald $X^2(1)=34.479$, $p<.001$). No other differences were statistically significant.

Interactions

Models with complete interaction terms could not be properly specified. Based on partial models, there was no evidence of interactions between aircraft category and year/season.



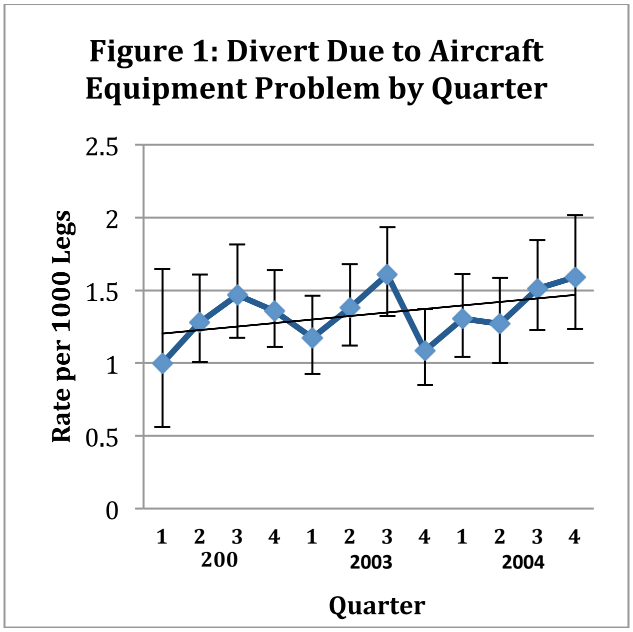
Operation

Only two events of leaving the cockpit to handle disturbances were reported by pilots of cargo aircraft.

Questions about equipment-related events

ER1: Divert Due to Equipment Problem

The rate of instances in which pilots reported diverting due to aircraft equipment problems did not change linearly across the observation period (b=.010, Wald $X^2(1)$ =0.822, p=.365; see Figure 1)[74]. Across the study period, a mean rate of 1.35 events per 1,000 legs was reported. The lowest rate was observed in the first quarter of 2002 (1.00) and the highest rate was observed in the third quarter of 2003 (1.61).

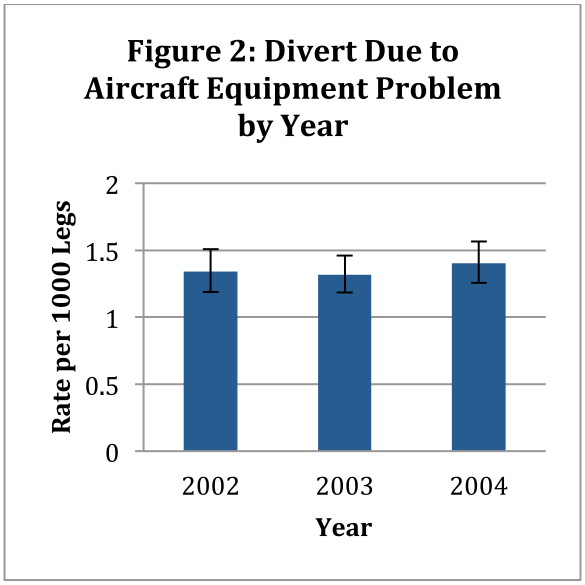


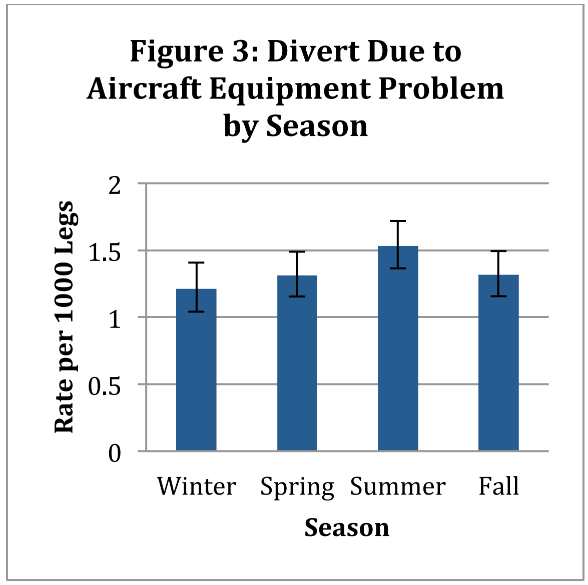
---

[74] Of the 18,368 reports obtained, 61 (0.33%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 180.61 per 1,000 legs compared to 1.61 per 1,000 legs for the other cases. When the suspect observations are included, little change in the trend was observed. The rate of reported events did not change linearly across the observation period (b=.008, Wald $X^2(1)$=0.568, p=.451). However, the overall fit of the model decreased (Model with outliers: Pearson $X^2(18366)$=32984.507, ratio=1.796; Model without outliers: Pearson $X^2(18305)$=23866.328; ratio=1.304). These cases were omitted from the remaining analyses.

Year/Season

No significant differences between years (Wald $X^2(2)=1.394$, p=.498; see Figure 2) or seasons (Wald $X^2(3)=6.312$, p=.097; see Figure 3) were observed. There was no year by season interaction (Wald $X^2(6)=4.795$, p=.570).



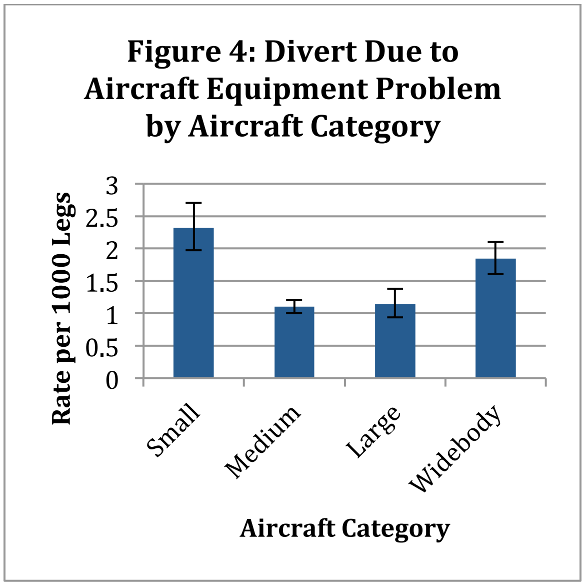


Aircraft Category

There were significant differences between aircraft categories in the rates of these events (Wald $X^2(3)=79.234$, $p<.001$; see Figure 4). Pilots of small aircraft reported the highest rates of these events, significantly higher than all other aircraft (Wald $X^2(1)=22.138$, $p<.001$). Pilots of wide-body aircraft reported the second highest rates of these events, significantly higher than medium (Wald $X^2(1)=26.061$, $p<.001$) and large aircraft (Wald $X^2(1)=15.794$, $p<.001$), which did not differ (Wald $X^2(1)=.066$, $p=.797$).

<u>Interactions</u>

There were no interactions between aircraft category and year/season (Wald $X^2(33)=34.106$, $p=.414$).

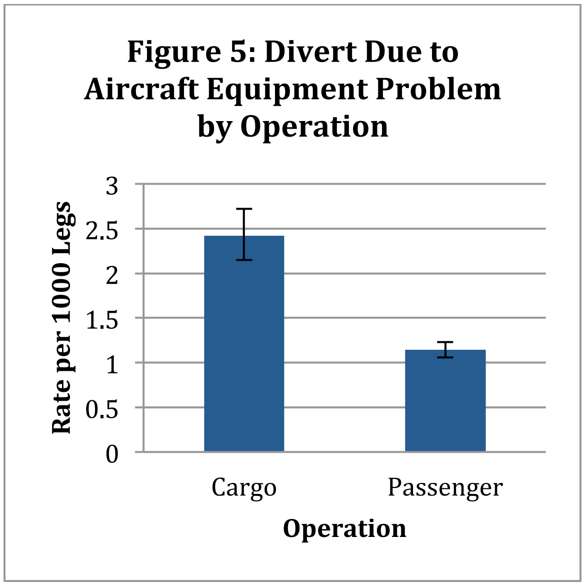


Operation

Pilots of cargo aircraft reported significantly greater rates of diversions due to equipment problems than did pilots of passenger aircraft (Wald $X^2(1)=94.400$, $p<.001$).

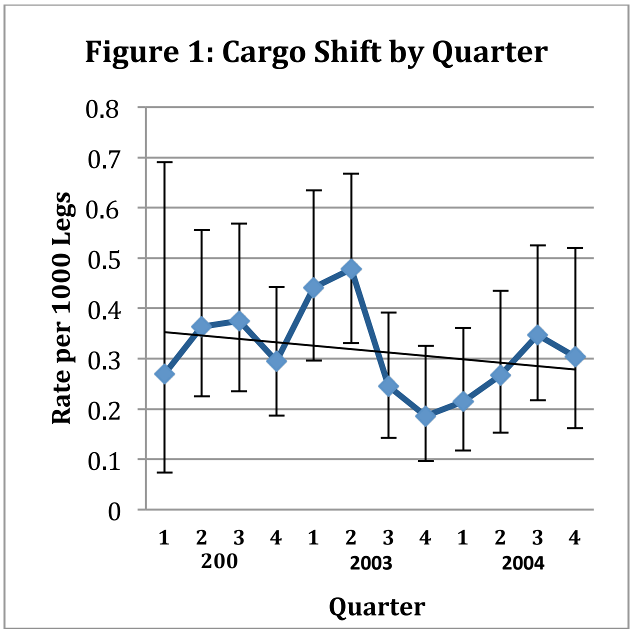
Interactions

There were no interactions between operation and year/season (Wald $X^2(11)=14.652$, $p=.199$).



ER3: Cargo Shift

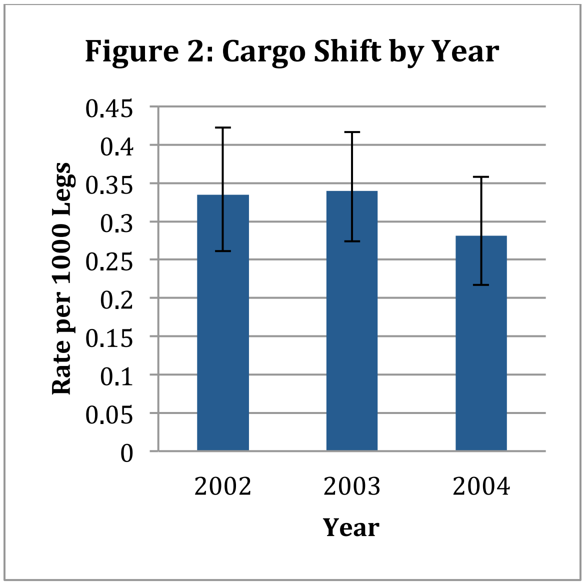
The rate of instances in which pilots reported cargo shifts did not change linearly across the observation period (b=-.034, Wald $X^2(1)$ =2.387, p=.122; see Figure 1).[75] Across the study period, a mean rate of 0.32 events per 1,000 legs was reported. The lowest rate was observed in the fourth quarter of 2003 (0.19) and the highest rate was observed in the second quarter of 2003 (0.48).

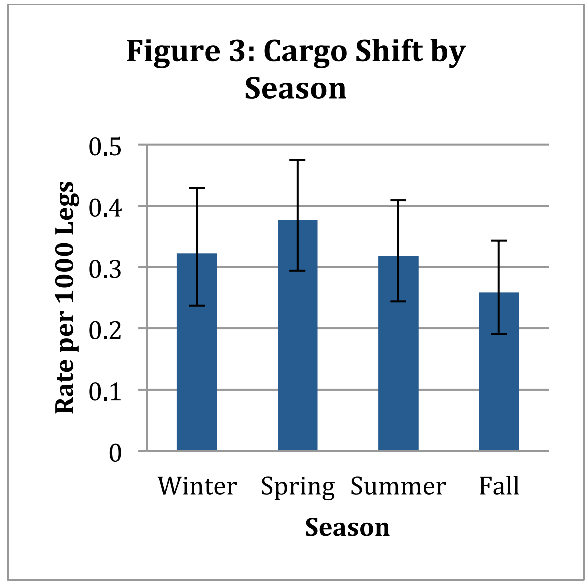


[75] Of the 18,356 reports obtained, 47 (0.26%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 165.46 per 1,000 legs compared to 0.32 per 1,000 legs for the other cases. When the suspect observations are included, the downward linear trend is significant (b=-.085, Wald $X^2(1)$=23.591, p<.001). However, the overall fit of the model decreased substantially (Model with outliers: Pearson $X^2(18354)$=66212.724, ratio=3.608; Model without outliers: Pearson $X^2(18307)$=27357.162; ratio=1.494). These cases were omitted from the remaining analyses.

Year/Season

No significant differences between years (Wald $X^2(2) = 0.718$, p=.698; see Figure 2) or seasons (Wald $X^2(3) = 3.311$, p=.346; see Figure 3) were observed and there was no year by season interaction (Wald $X^2(6) = 10.208$, p=.116).





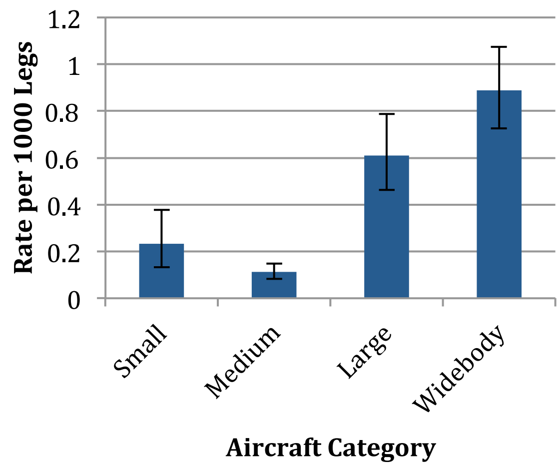
Aircraft Category

There were significant differences in the reported rates of cargo shifts between aircraft categories (Wald $X^2(3)=145.887$, $p<.001$; see Figure 4). Pilots of medium aircraft reported the lowest rates. The rates of these events reported by pilots of small aircraft were marginally higher (Wald $X^2(1)=3.719$, $p=.054$). Pilots of large aircraft reported significantly higher rates of these events than did pilots of small aircraft (Wald $X^2(1)=13.559$, $p<.001$). Pilots of wide-body aircraft reported the highest rates of cargo shifts, significantly higher than the rates reported by pilots of large aircraft (Wald X2(1)=4.795, p=.029).

<u>Interaction</u>

Although the full factorial model could not be estimated, partial models revealed significant interactions between aircraft category and year/season (ΔLLR $X^2(15)=37.820$, $p=.001$). However, there are no significant year/season effects for any aircraft category examined separately.

Figure 4: Cargo Shift by Aircraft Category

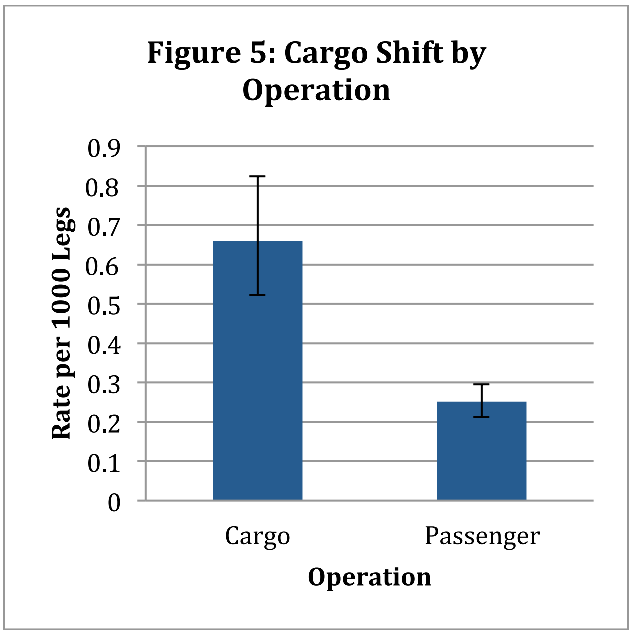


Operation

Pilots of passenger aircraft reported significantly lower rates of these events than did pilots of cargo aircraft (Wald $X^2(1)=42.720$, $p<.001$; see Figure 5).

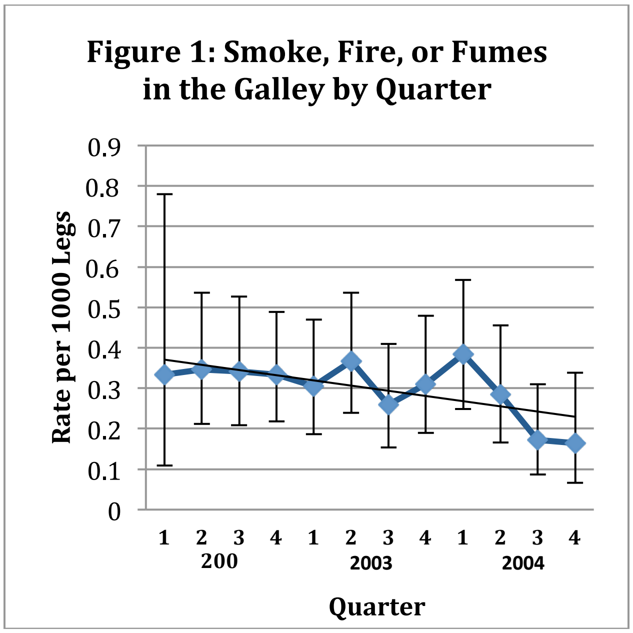
<u>Interactions</u>

Although the full interaction model could not be fit properly, there were significant interactions between operation and season (Wald $X^2(3)=16.273$, $p=.001$) in a smaller model. Cargo operations demonstrated a seasonal pattern (Wald $X^2(3)=13.710$, $p=.003$). For these operations, the reported rate of these events was highest in the spring; significantly higher than the fall (Wald $X^2(1)=7.280$, $p=.007$). The reported rate for the fall was not significantly different from the summer (Wald $X^2(1)=.257$, $p=.612$) but significantly higher than the winter (Wald $X^2(1)=9.584$). Passenger operations did not demonstrate a seasonal pattern (Wald $X^2(3)=5.495$, $p=.139$).



ER5D: Smoke, Fire, or Fumes in the Galley

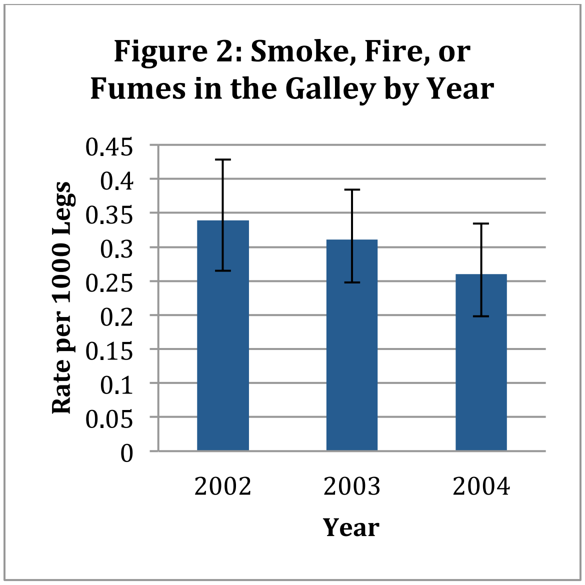
Across the study period, a mean rate of 0.30 events per 1,000 legs of smoke, fire or fumes in the galley was reported. The lowest rate was observed in the fourth quarter of 2004 (0.16) and the highest rate was observed in the first quarter of 2004 (0.38). The rate decreased across the observation period (b=-.044, Wald $X^2(1)$ =3.816, p=.051; see Figure 1)[76].

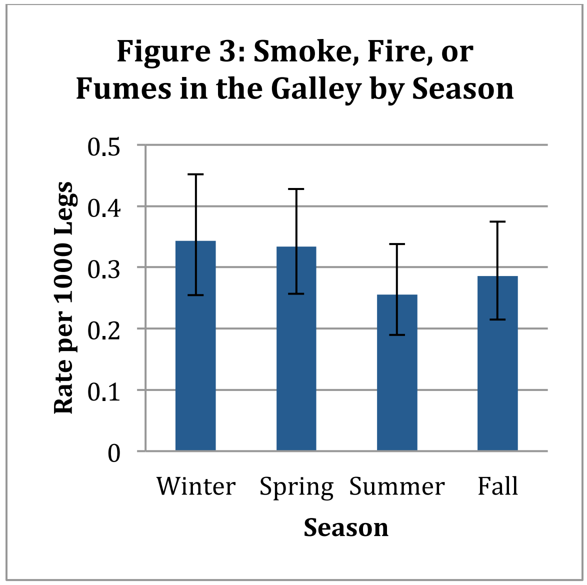


[76] Of the 18,363 reports obtained, 61 (0.33%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. These cases were omitted from the remaining analyses. The mean reported event rate amongst these cases was 98.81 per 1,000 legs compared to 0.27 per 1,000 legs for the other cases. When the suspect observations are included, little change in the trend was observed. The rate of reported events decreased across the observation period (b=-.047, Wald $X^2(1)$=5.948, p=.015). However, the overall fit of the model decreased (Model with outliers: Pearson $X^2(18361)$=34707.545, ratio=1.890; Model without outliers: Pearson $X^2(18300)$=18240.828; ratio=0.997).

Year/Season

There were no significant yearly (Wald $X^2(1)$ =3.313, p=.191; see Figure 2) or seasonal patterns (Wald $X^2(3)$ =3.425, p=.331; see Figure 3) and no year by season interaction (Wald $X^2(6)$=4.571, p=.600).





Aircraft Category

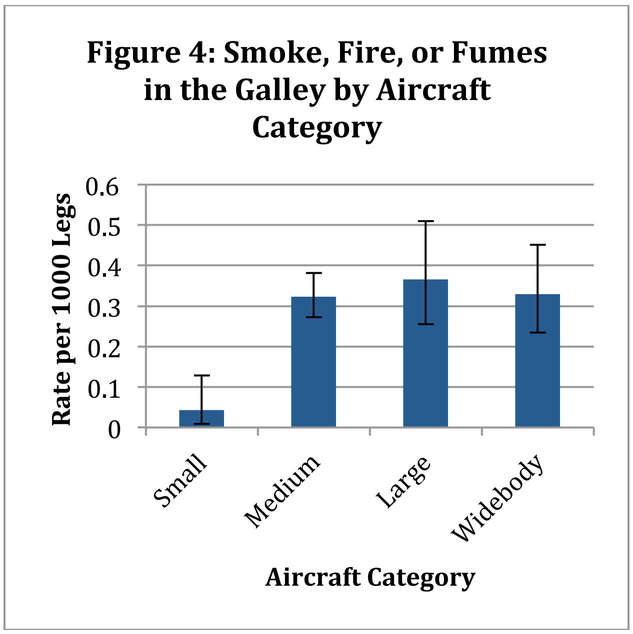
There were significant differences between aircraft categories (Wald $X^2(3)=12.554$, p=.006). However, because many small aircraft do not have galleys, data from small aircraft were eliminated from this analysis. No significant differences between the remaining aircraft categories were observed (Wald $X^2(2)=.398$, p=.819).

Interactions

No significant interactions between aircraft category and year/season were observed (ΔLLR $X^2(9)=11.534$, p=.241).[77]

Operation

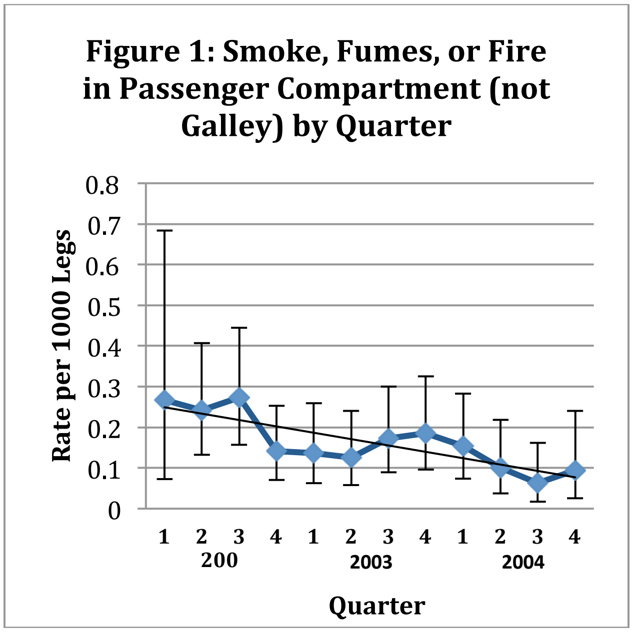
Because few cargo aircraft have galleys, no analysis by operation was performed.



[77] Note: With or without the small aircraft in the analysis, the full model could not be properly estimated given the large number of parameters and the pattern of the data. Smaller models were fit and the same conclusion obtained.

ER5E: Smoke, Fire, or Fumes in the Passenger Compartment (not in Galley)

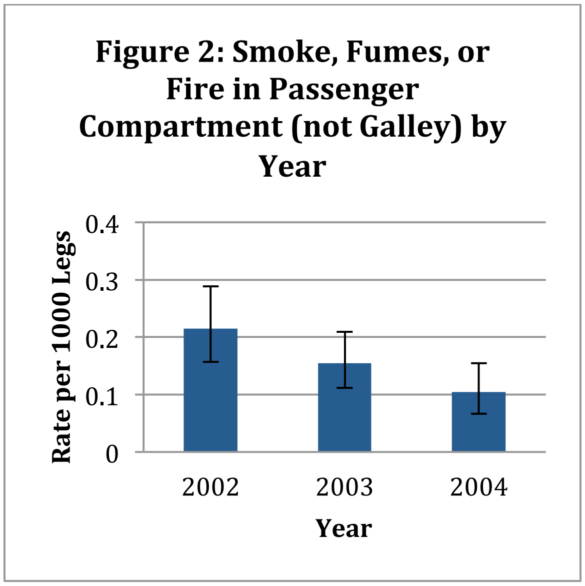
Across the study period, a mean rate of 0.16 events per 1,000 legs of smoke, fire or fumes in the passenger compartment other than in the galley was reported. The lowest rate was observed in the third quarter of 2004 (0.06) and the highest rate was observed in the third quarter of 2002 (0.27). The rate decreased across the observation period (b=-.095, Wald $X^2(1)$ =8.844, p=.003; see Figure 1)[78].

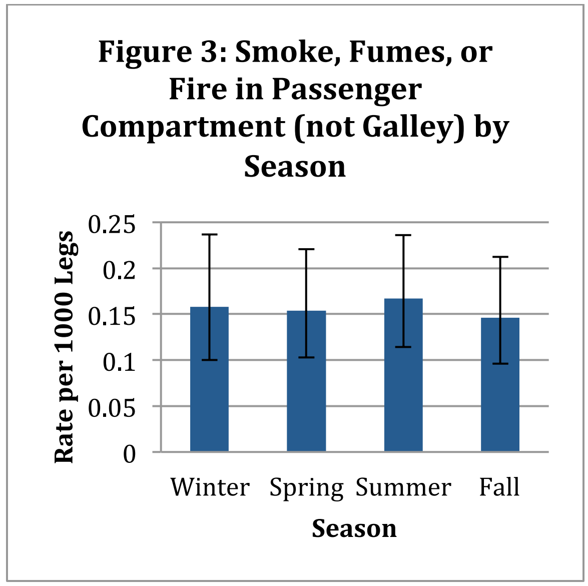


[78] Of the 18,350 reports obtained, 50 (0.27%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. These cases were omitted from the remaining analyses. The mean reported event rate amongst these cases was 90.27 per 1,000 legs compared to 0.11 per 1,000 legs for the other cases. When the suspect observations are included, little change in the trend was observed. The rate of reported events decreased across the observation period (b=-.087, Wald $X^2(1)$=11.801, p=.001). However, the overall fit of the model decreased (Model with outliers: Pearson $X^2(18348)$=34961.191, ratio=1.905; Model without outliers: Pearson $X^2(18298)$=13149.763; ratio=0.719).

Year/Season

Significant differences between years were observed (Wald $X^2(2)$ =8.651, p=.013; see Figure 2). The rate of reported events decreased linearly (Wald $X^2(1)$=7.889, p=.005). No differences between seasons were observed (Wald $X^2(3)$ =0.838, p=.840; see Figure 3) and no year by season interaction was observed (Wald $X^2(6)$=5.605, p=.469).



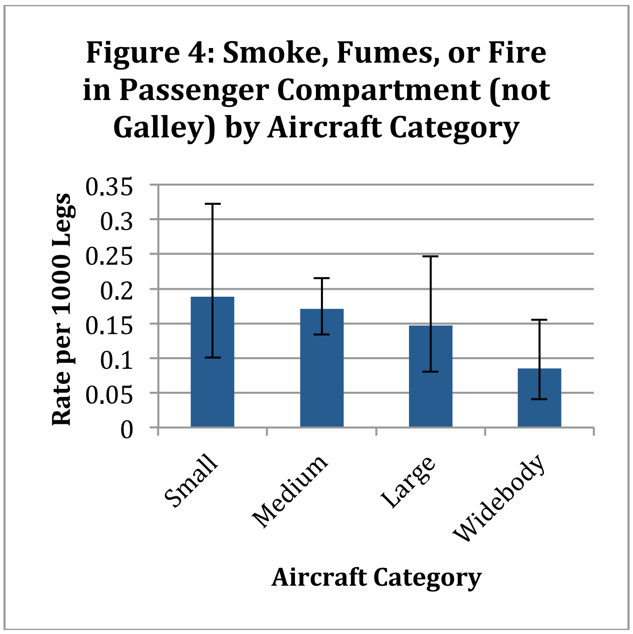


Aircraft Category

There were no significant differences between aircraft types (Wald $X^2(3)$ =4.504, p=.212; see Figure 4).

Interactions

Significant interactions between aircraft category and year/season were observed (ΔLLR $X^2(33)$=48.118, p=.043). Only the medium aircraft demonstrated a consistent linear decrease across the years (Wald $X^2(1)$=12.480, p<.001). Large aircraft demonstrated a generally downward trend across the observed period (Wald $X^2(1)$=4.484, p=.034) but the rates of these events in 2004 was somewhat above the rates in 2003 though lower than in 2002 (Wald $X^2(1)$=4.523, p=.033). No yearly pattern was observed for the other aircraft types.

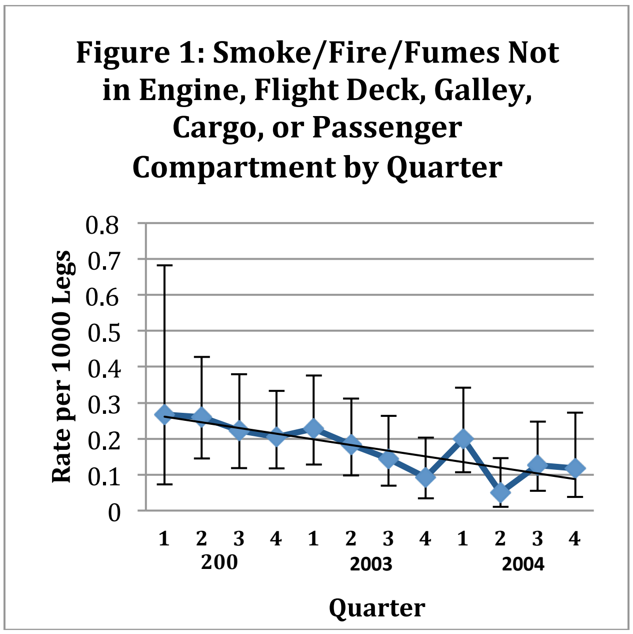


Operation

Because this question asked about the passenger compartment, there were no reports of these events from cargo aircraft.

ER5F: Smoke, Fire, or Fumes Other than in the Engine Nacelle, Flight Deck, Cargo Hold, Galley, or Passenger Compartment

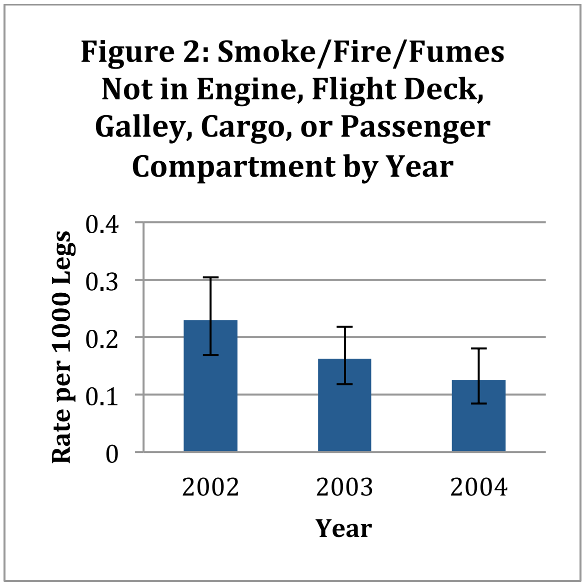
Across the study period, a mean rate of 0.17 events per 1,000 legs of smoke, fire, or fumes other than in the engine nacelle, flight deck, cargo hold, galley, or passenger compartment was reported. The lowest rate was observed in the second quarter of 2004 (0.05) and the highest rate was observed in the first quarter of 2002 (0.27). The rate decreased linearly across the observation period (b=-.096, Wald $X^2(1)$ =9.825, p=.002; see Figure 1)[79].

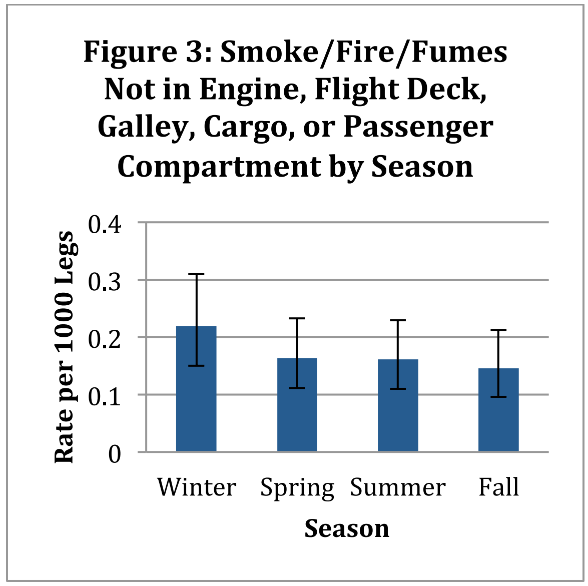


[79] Of the 18,373 reports obtained, 50 (0.27%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. These cases were omitted from the remaining analyses. The mean reported event rate amongst these cases was 88.07 per 1,000 legs compared to 0.14 per 1,000 legs for the other cases. When the suspect observations are included, little change in the trend was observed. The rate of reported events decreased across the observation period (b=-.067, Wald $X^2(1)$=7.809, p=.005). However, the overall fit of the model decreased (Model with outliers: Pearson $X^2(18371)$=36737.145, ratio=2.000; Model without outliers: Pearson $X^2(18321)$=15502.864; ratio=0.846). These cases were omitted from the remaining analyses.

Year/Season

Significant differences between years were observed (Wald $X^2(2)$ =8.210, p=.016; see Figure 2). The rate of reported events decreased linearly across the years (Wald $X^2(1)$=7.297, p=.007). No differences between seasons were observed (Wald $X^2(3)$ =4.516, p=.211; see Figure 3) and no year by season interaction was observed (Wald $X^2(6)$=4.001, p=.676).



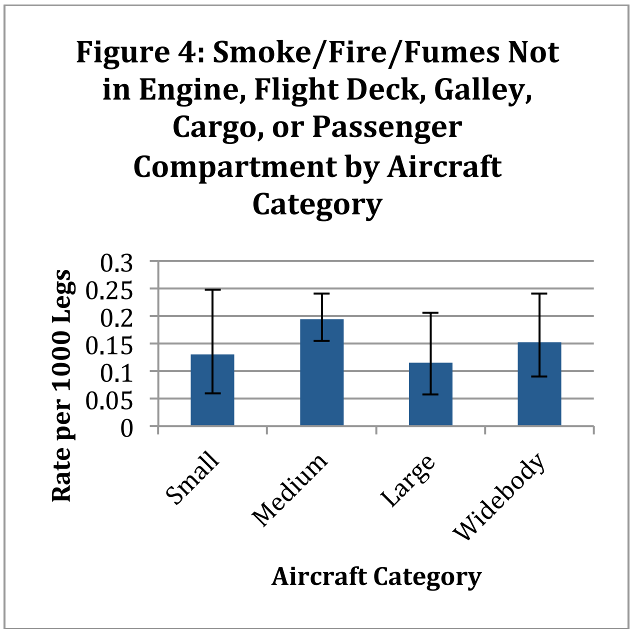


Aircraft Category

There were no significant differences between aircraft categories (Wald $X^2(3)$ =3.784, p=.286; see Figure 4).

Interactions

No interactions between aircraft category and year/season were observed (ΔLLR $X^2(33)$=35.198, p=.364)[80].



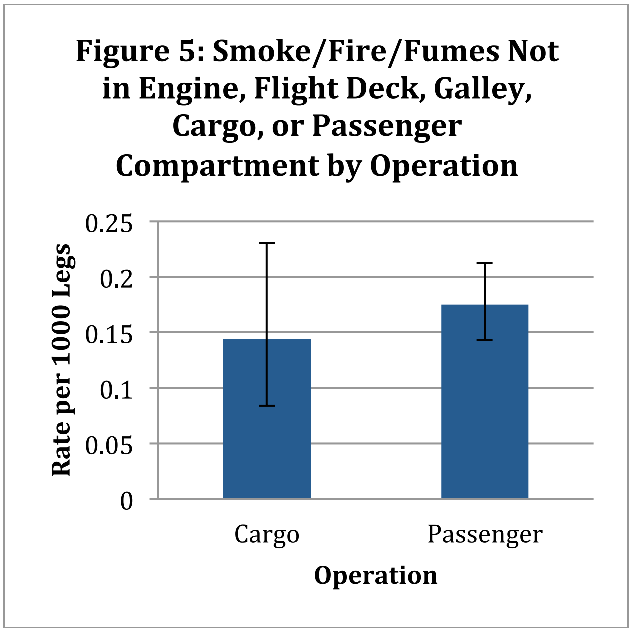
[80] Note: The full model could not be properly estimated given the large number of parameters and the pattern of the data. Smaller models were fit and the same conclusion obtained.

Operation

No effect of type of operation was observed (Wald $X^2(1)=0.530$, p=.466).

Interactions

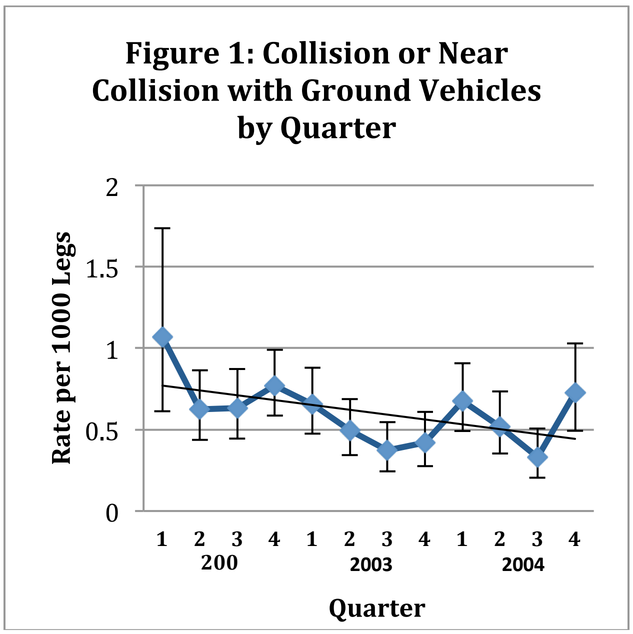
No interactions between operation and year/season were observed (ΔLLR $X^2(11)=7.508$, p=.757).



Questions about ground operations

GE02: Collision or Near Collision with Ground Vehicles

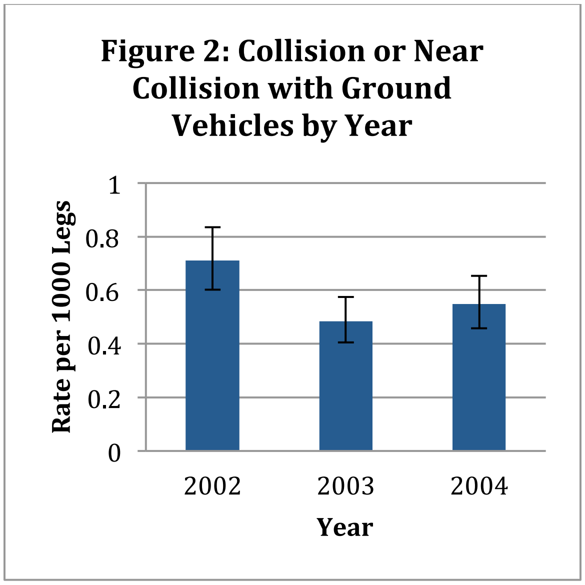
Across the study period, a mean rate of 0.57 events per 1,000 legs of collisions or near collisions with ground vehicles was reported. The lowest rate was observed in the third quarter of 2004 (0.33) and the highest rate was observed in the first quarter of 2002 (1.07). The rate decreased linearly across the observation period (b=-.041, Wald $X^2(1)$ =6.038, p=.014; see Figure 1).[81]

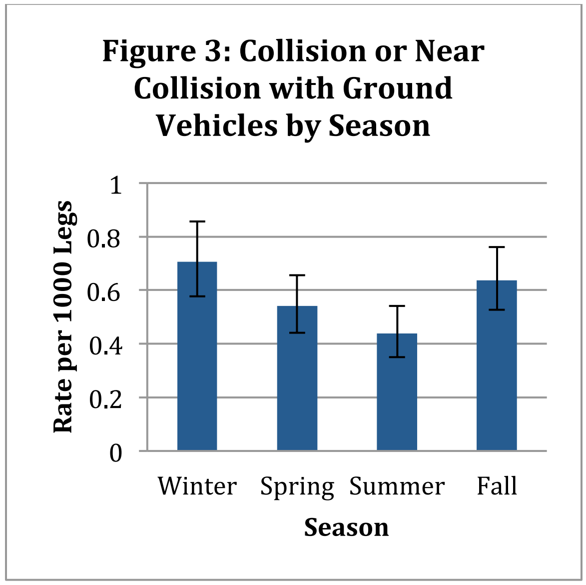


[81] Of the 18,375 reports obtained, 49 (0.27%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. These cases were omitted from the remaining analyses. The mean reported event rate amongst these cases was 136.65 per 1,000 legs compared to 0.50 per 1,000 legs for the other cases. When the suspect observations are included, little change in the results is observed (b=-.032, Wald $X^2(1)$=4.760, p=.029). However, the overall fit of the model decreased (Model with outliers: Pearson $X^2(18373)$=39206.147, ratio=2.134; Model without outliers: Pearson $X^2(18324)$=19644.182; ratio=1.072). These cases were omitted from the remaining analyses. The observed trend does not depend on the first quarter. However, when data from this quarter are eliminated, the trend weakens and becomes marginally significant (b=-.033, Wald $X^2(1)$=3.498, p=.061).

Year/Season

Significant differences between years were observed (Wald $X^2(2)$ =13.204, p=.001; see Figure 2). The rate of these events was highest in 2002 and lower in 2003 and 2004 (Wald $X^2(1)$=9.870, p=.002), which did not differ (Wald $X^2(1)$=.938, p=.333). Significant differences between seasons were also observed (Wald $X^2(3)$ =14.202, p=.003; see Figure 3). The rates of these events were significantly lower in the spring and summer than in the fall and winter (Wald $X^2(1)$=10.432, p=.001). No other differences between the seasons were statistically significant. No year by season interaction was observed (Wald $X^2(6)$=5.089, p=.532).



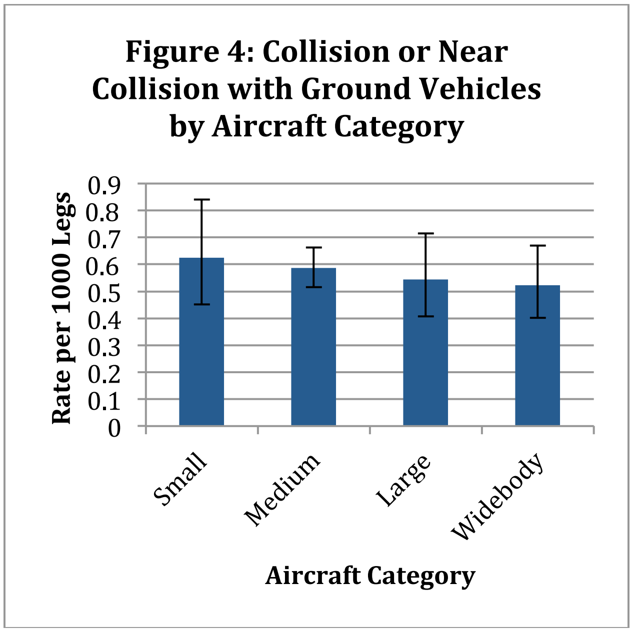


Aircraft Category

There were no significant differences between aircraft categories (Wald $X^2(3)=1.061$, p=.786).

Interactions

No significant interactions between aircraft category and year/season were observed (ΔLLR $X^2(33)=44.82$, p=.082).[82]



---

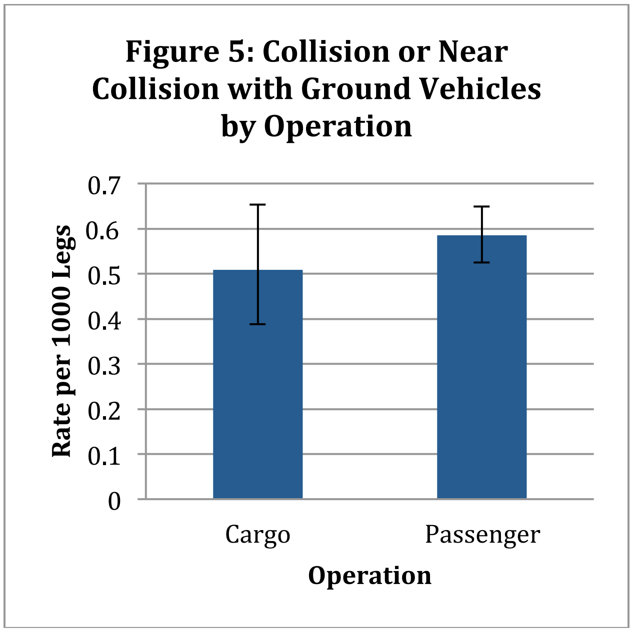
[82] Note: The full model could not be properly estimated given the large number of parameters and the pattern of the data. Smaller models were fit and the same conclusion obtained.

Operation

There was no significant difference between passenger and cargo operations in the rates of these events (Wald $X^2(1)=0.939$, $p=.333$).

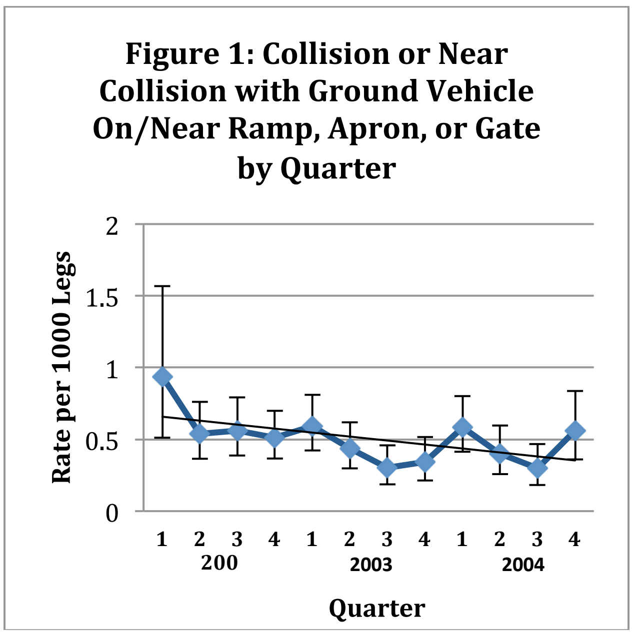
<u>Interactions</u>

No significant interactions between operation and year/season were observed (ΔLLR $X^2(11)=7.77$, $p=.734$).



GE2A: Collide or Nearly Collide with a Ground Vehicle On/Near the Ramp, Apron, or Gate

The rate of instances in which pilots reported collisions or near collisions with ground vehicles in the vicinity of the ramp, apron, or gate decreased linearly across the observation period (b=-.042, Wald $X^2(1)$ =5.380, p=.020; see Figure 1).[83] Across the study period, a mean rate of 0.47 events per 1,000 legs was reported. The lowest rate was observed in the third quarter of 2004 (0.30) and the highest rate was observed in the first quarter of 2002 (0.94). The unusually high rate for the first quarter contributes substantially to the observed downward trend. When the data from this quarter is removed from the analysis, the downward linear trend weakens and becomes marginally significant (b=-.033, Wald $X^2(1)$ =2.896, p=.089).

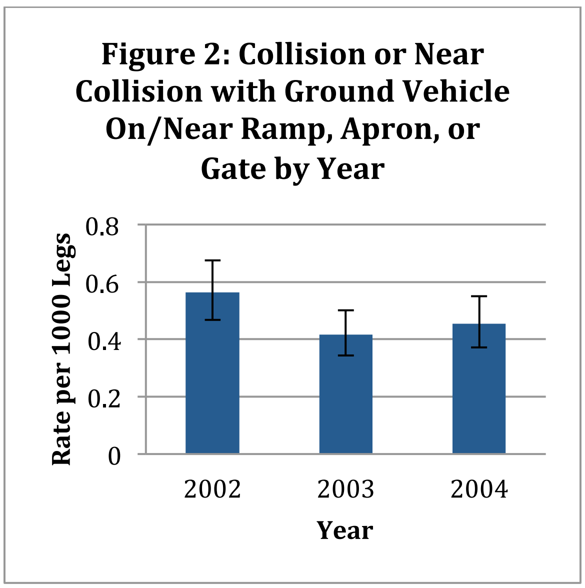


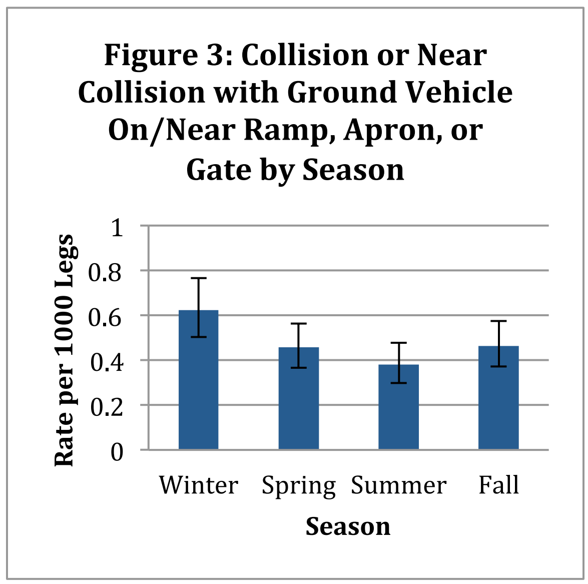
[83] Of the 18,375 reports obtained, 47 (0.26%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 127.56 per 1,000 legs compared to 0.40 per 1,000 legs for the other cases. When the suspect observations are included, little change in the trend was observed. The rate of reported events decreased linearly across the observation period (b=-.030, Wald $X^2(1)$=3.866, p=.049). However, the overall fit of the model decreased (Model with outliers: Pearson $X^2(18373)$=39579.254, ratio=2.154; Model without outliers: Pearson $X^2(18326)$=18368.131; ratio=1.002). These cases were omitted from the remaining analyses.

Year/Season

Significant differences between years were observed (Wald $X^2(2)$ =9.646, p=.008; see Figure 2). The rate of collisions or near collisions in the ramp/apron/gate area in 2002 was significantly higher than in 2003 (Wald $X^2(1)$ =8.137, p=.004). The rate of these events in 2003 did not differ from the rates in 2004 (Wald $X^2(1)$ =0.463, p=.496). Significant differences between seasons also were observed (Wald $X^2(3)$ =13.147, p=.004; see Figure 3). The reported rates of these events were higher in the winter than in the other seasons (Wald $X^2(1)$=8.315, p=.004). No other differences between seasons were found. No year by season interaction was observed (Wald $X^2(6)$ =5.110, p=.530).

When the first quarter is eliminated, the general pattern of results remains the same. There are significant differences between the years (Wald $X^2(2)$=7.175, p=.028) and the seasons (Wald $X^2(3)$=10.085, p=.018) and no year by season interaction (Wald $X^2(5)$=4.875, p=.431). However, the observed effects are weaker. The reported rates of events in 2002 are significantly higher than in 2003 (Wald $X^2(1)$=3.918, p=.048) and there is no difference between 2003 and 2004 (Wald $X^2(1)$=0.463, p=.496). The rates of events reported during winter are significantly higher than those reported in the other seasons (Wald $X^2(1)$=4.211, p=.040) no other differences are statistically significant.



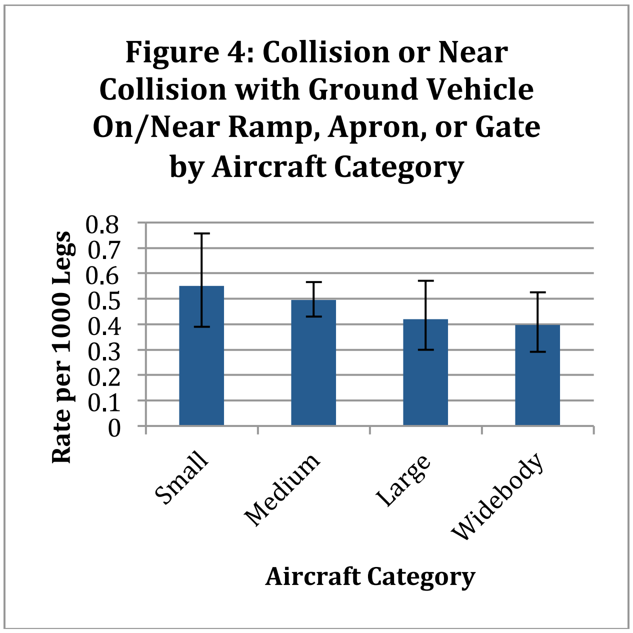


Aircraft Category

There were no significant differences between aircraft categories in the rates of these events (Wald $X^2(3)=3.245$, p=.355).

Interactions

Although full factorial analyses could not be run on these data, analysis of partial models did not reveal any interactions between aircraft category and year or season.

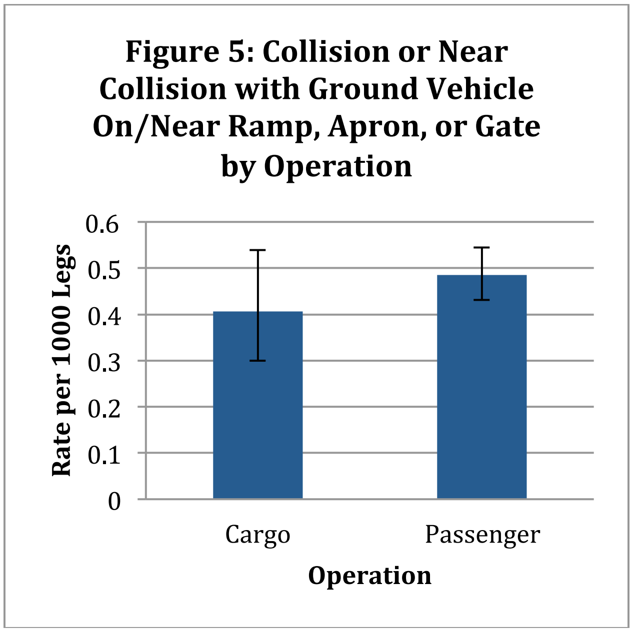


Operation

There were no significant differences in the rates of these events between cargo and passenger operations (Wald $X^2(1)=1.213$, p=.271).

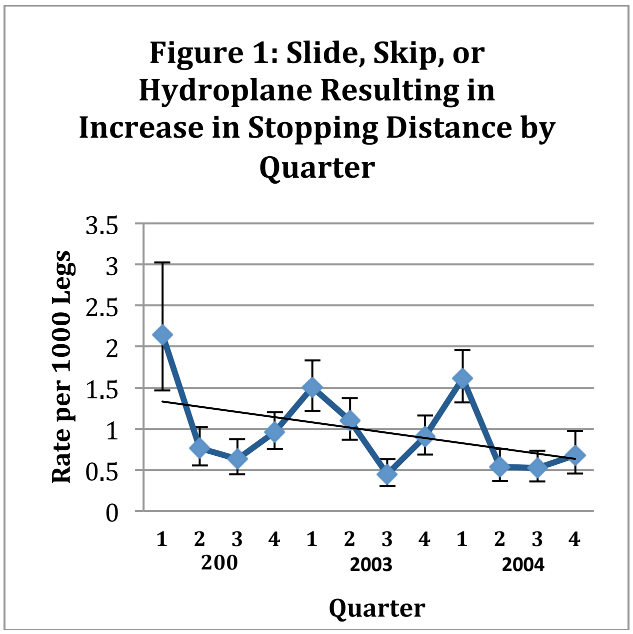
Interactions

There were no interactions between operation and year/season (ΔLLR $X^2(11)=7.808$, p=.730).



GE3: Slide, Skip, or Hydroplane Resulting in Increase in Stopping Distance

The rate of instances in which pilots reported sliding, skipping, or hydroplaning sufficient to cause an increase in stopping distance decreased linearly across the observation period (b=-.036, Wald $X^2(1)$ =7.069, p=.008; see Figure 1).[84] However, this trend is dependent on an unusually high rate of events in the first quarter of 2002. If that quarter is eliminated from the analysis, the downward trend is no longer statistically significant (b=-.021, Wald $X^2(1)$=2.177, p=.140). Across the study period, a mean rate of 0.92 events per 1,000 legs was reported. The highest rate was observed in the first quarter of 2002 (2.14) and the lowest rate was observed in the third quarter of 2003 (0.45).



[84] Of the 18,371 reports obtained, 53 (0.29%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 147.76 per 1,000 legs compared to 0.92 per 1,000 legs for the other cases. When the suspect observations are included, little change in the trend was observed. The rate of reported events decreased linearly across the observation period (b=-.039, Wald $X^2(1)$=10.425, p=.001) – but see text above. However, the overall fit of the model decreased (Model with outliers: Pearson $X^2(18369)$=39511.418, ratio=2.151; Model without outliers: Pearson $X^2(18316)$=21573.377; ratio=1.178). These cases were omitted from the remaining analyses.

Year/Season

When the data from all quarters are included, marginally significant differences between years (Wald $X^2(2)$ =5.830, p=.054; see Figure 2) and significant differences between seasons were observed (Wald $X^2(3)$ =93.271, p<.001; see Figure 3). However, when the data from the first quarter are excluded, the differences between years are not statistically significant (Wald $X^2(2)$=4.236, p=.120) but the differences between seasons remain (Wald $X^2(3)$=11.862). There were significantly higher rates of these events in the winter than during the other seasons (Wald $X^2(1)$=47.483, p<.001). The rates in summer were lower than in the spring (Wald $X^2(1)$=7.198, p=.007) or fall (Wald $X^2(1)$=11.294, p=.001). The rates in spring did not differ from those in fall (Wald $X^2(1)$=0.562, p=.453). Marginally significant interactions between year and season were also observed (Wald $X^2(6)$ =12.420, p=.053). However, these reflect minor variations across years in the seasonal pattern described above.

Figure 2: Slide, Skip, or Hydroplane Resulting in Increase in Stopping Distance by Year

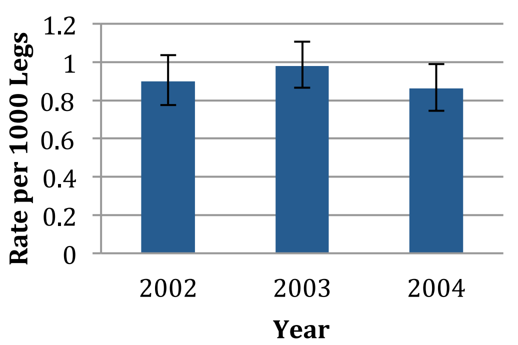
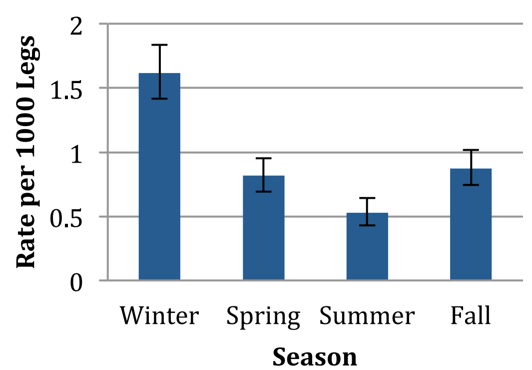


Figure 3: Slide, Skip, or Hydroplane Resulting in Increase in Stopping Distance by Season

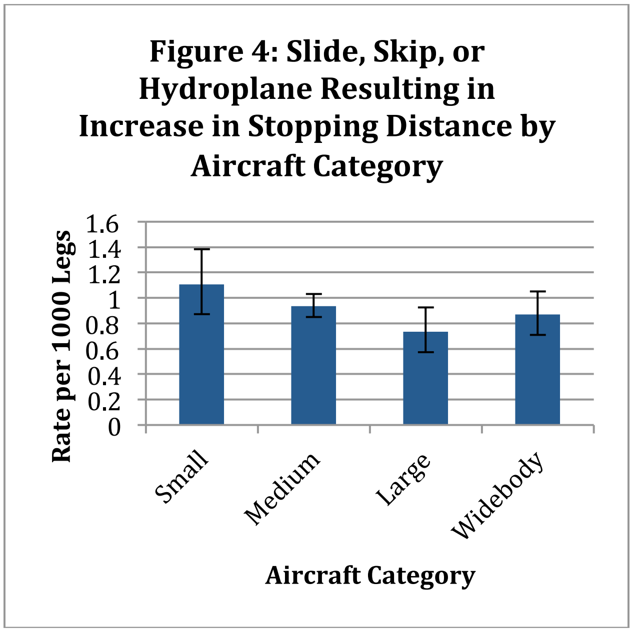


Aircraft Category

There were significant differences in reported rates of these events between aircraft categories (Wald $X^2(3)=8.243$, p=.041; see Figure 4). Pilots of large aircraft reported the lowest rates, significantly lower than the rates reported by pilots of small (Wald $X^2(1)=6.834$, p=.009) and medium aircraft (Wald $X^2(1)=4.253$, p=.039). No other differences between aircraft categories were statistically significant.

Interaction

No interactions between aircraft category and year/season were observed (ΔLLR $X^2(33)=41.834$, p=.139).

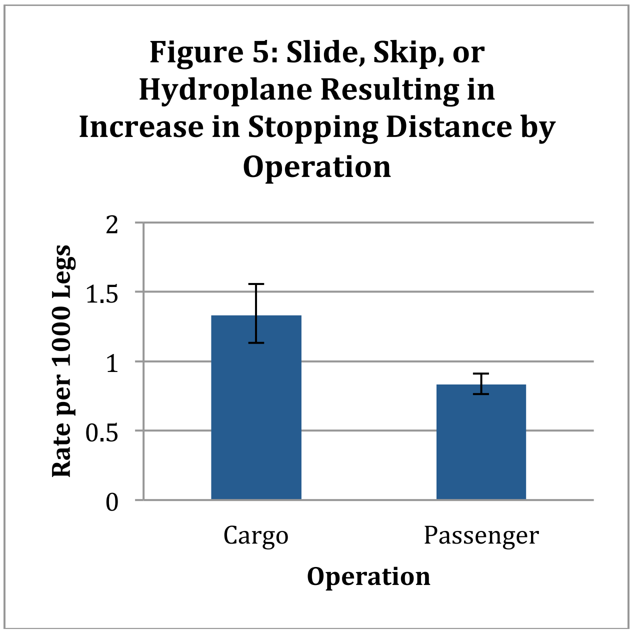


Operation

Pilots of cargo aircraft reported a higher rate of these events than did pilots of passenger aircraft (Wald $X^2(1)=24.370$, $p<.001$; see Figure 5).

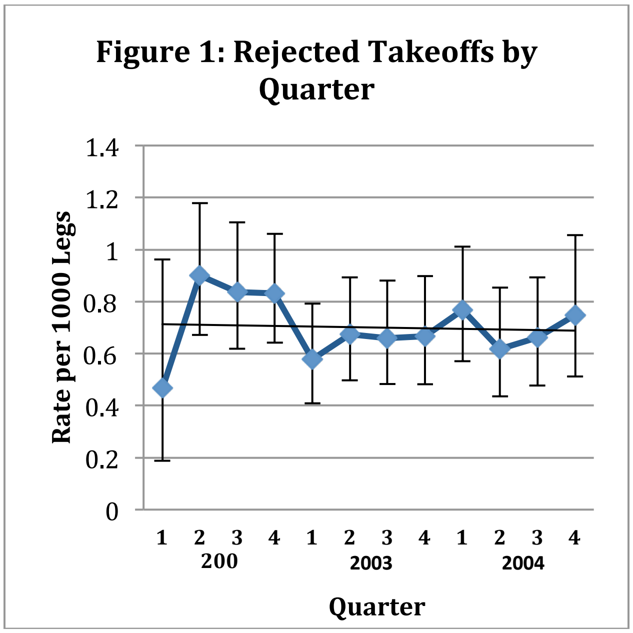
Interaction

There were no significant interactions between operation and year/season (Δ LLR $X^2(11)=13.448$, $p=.265$).



GE4: Rejected Takeoffs

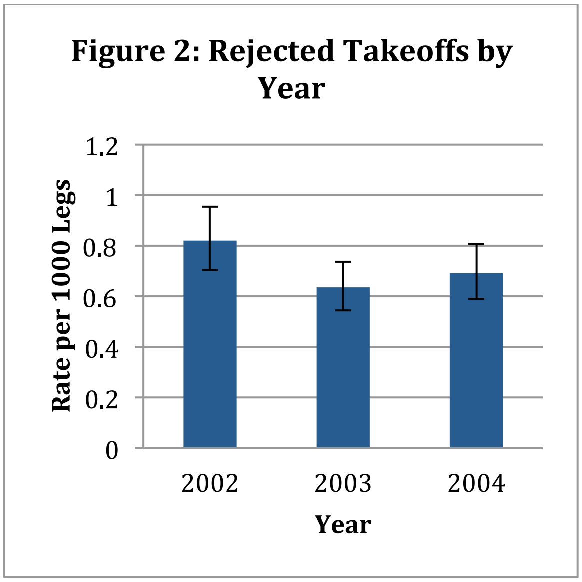
Across the study period, a mean rate of 0.71 rejected takeoffs per 1,000 legs was reported. The lowest rate was observed in the first quarter of 2002 (0.47) and the highest rate was observed in the second quarter of 2002 (0.90). The rate did not change linearly across the observation period (b=-.018, Wald $X^2(1)$ =1.514, p=.219; see Figure 1).[85]

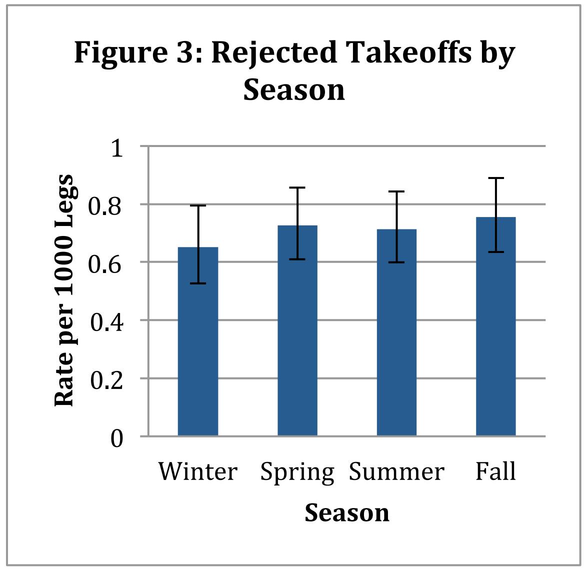


[85] Of the 18,374 reports obtained, 52 (0.28%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. These cases were omitted from the remaining analyses. The mean reported event rate amongst these cases was 140.70 per 1,000 legs compared to 0.72 per 1,000 legs for the other cases. When the suspect observations are included, little change in the trend was observed. The rate of reported events decreased across the observation period (b=-.019, Wald $X^2(1)$=1.807, p=.179). However, the overall fit of the model decreased (Model with outliers: Pearson $X^2(18372)$=26127.617, ratio=1.422; Model without outliers: Pearson $X^2(18320)$=18414.623; ratio=1.005).

Year/Season

No significant differences between years (Wald $X^2(2)$ =1.236, p=.539; see Figure 2) or seasons (Wald $X^2(3)$ =1.864, p=.601; see Figure 3) were observed and no year by season interaction was observed (Wald $X^2(6)$=4.025, p=.673).



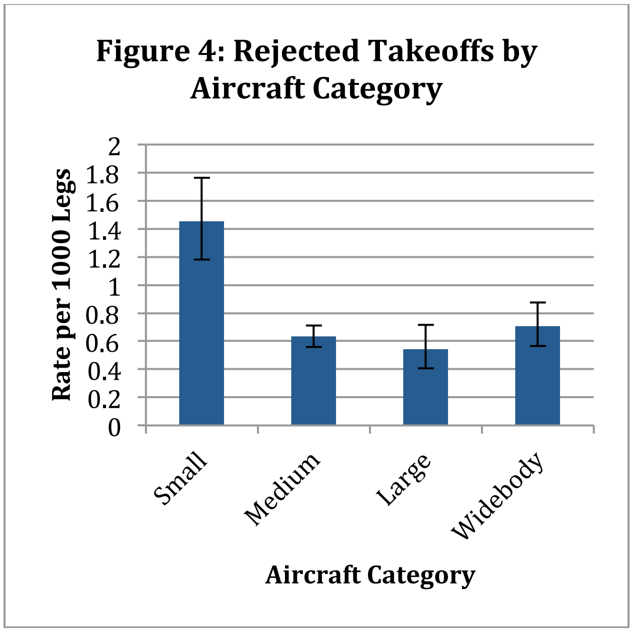


Aircraft Category

There were significant differences between aircraft categories (Wald $X^2(3) = 50.653$, $p < .001$; see Figure 4). Pilots of small aircraft reported higher rates of rejected takeoffs than did pilots of larger aircraft (Wald $X^2(1) = 25.075$, $p < .001$). No other differences between aircraft categories were observed.

Interactions

No significant interactions between aircraft category and year/season were observed (ΔLLR $X^2(33) = 37.532$, $p = .269$).[86]



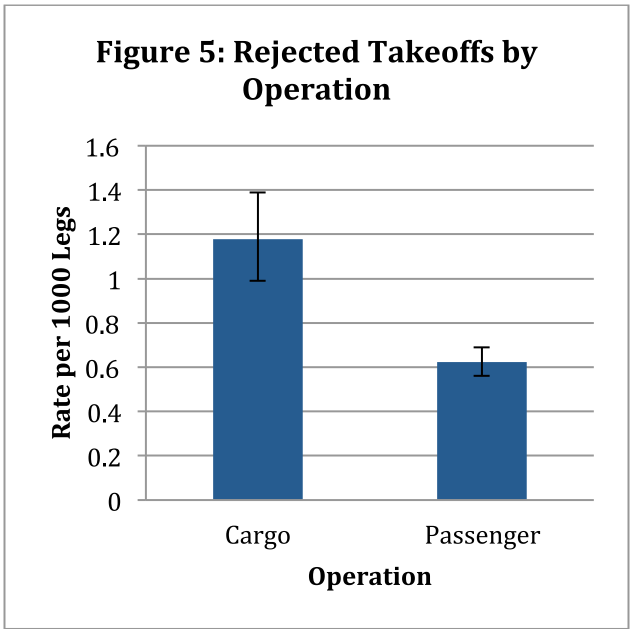
[86] Note: The full model could not be properly estimated given the large number of parameters and the pattern of the data. Smaller models were fit and the same conclusion obtained.

Operation

Pilots of cargo aircraft reported a higher rate of rejected takeoffs than did pilots of passenger aircraft (Wald $X^2(1)=37.922$, $p<.001$).

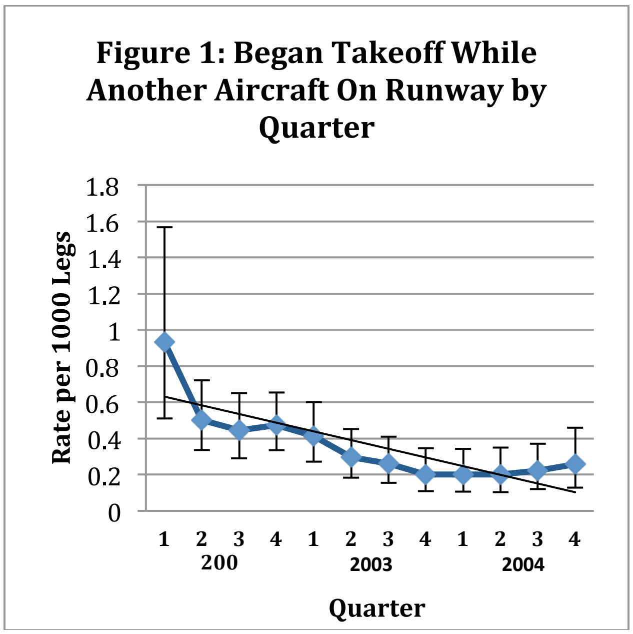
Interactions

There were no significant interactions between operation and year/season (ΔLLR $X^2(11)=16.85$, $p=.112$).



GE8: Began Takeoff While Another Aircraft was on the Runway

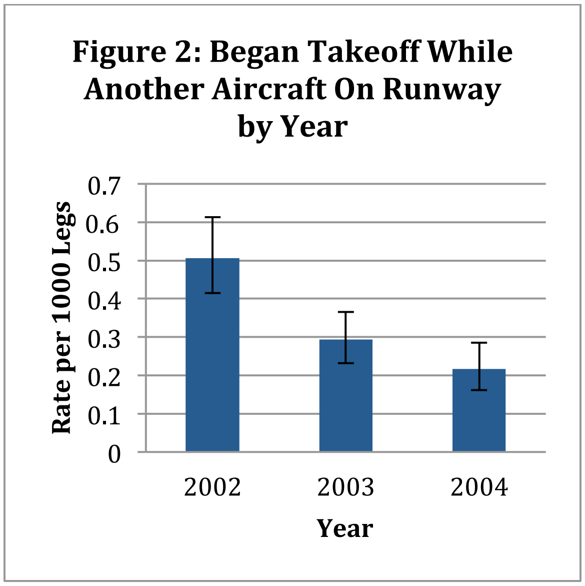
The rate of instances in which pilots reported beginning takeoffs while another aircraft was on the runway decreased linearly across the observation period (b=-.119, Wald $X^2(1)$ =28.316, p<.001; see Figure 1).[87] This trend remained statistically significant when the data from the first quarter of 2002 is omitted from the analysis (b=-.106, Wald $X^2(1)$=20.279, p<.001). Across the study period, a mean rate of 0.33 events per 1,000 legs was reported. The lowest rate was observed in the first quarter of 2004 (0.20) and the highest rate was observed in the first quarter of 2002 (0.93).

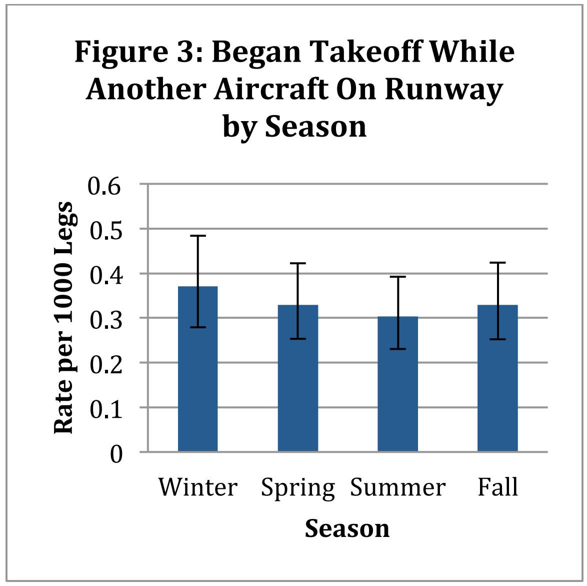


[87] Of the 18,351 reports obtained, 60 (0.33%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 147.72 per 1,000 legs compared to 0.26 per 1,000 legs for the other cases. When the suspect observations are included, the downward trend remains (b=-.041, Wald $X^2(1)$=7.204, p=.007). However, the overall fit of the model decreased (Model with outliers: Pearson $X^2(18349)$=66878.386, ratio=3.645; Model without outliers: Pearson $X^2(18289)$=26397.657; ratio=1.443). These cases were omitted from the remaining analyses.

Year/Season

Significant differences between years were observed (Wald $X^2(2)$ =32.560, p<.001; see Figure 2). The reported rates of these events decreased linearly across the years in the sample (Wald $X^2(1)$ =25.174, p<.001). No differences between seasons were observed (Wald $X^2(3)$ =4.686, p=.196; see Figure 3). No year by season interaction was observed (Wald $X^2(6)$ =5.260, p=.511).



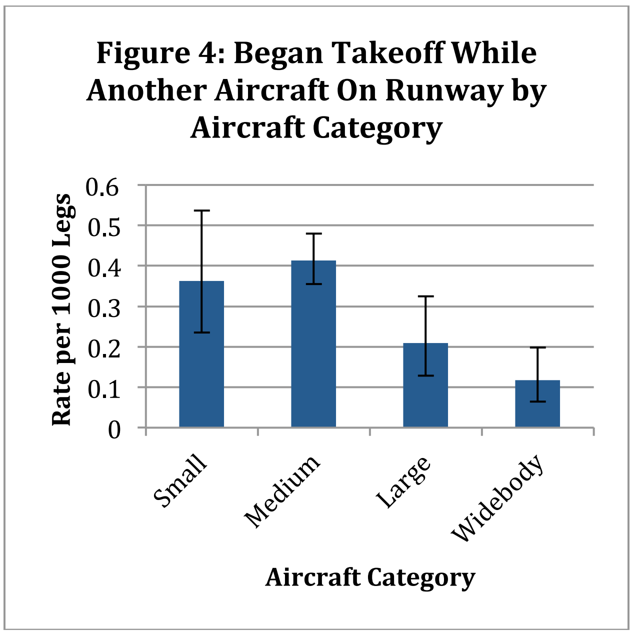


Aircraft Category

There were significant differences in the reported rates of these events between aircraft categories (Wald $X^2(3)=25.614$, $p<.001$; see Figure 4). Pilots of medium aircraft reported the highest rates, significantly higher than those reported by pilots of large (Wald $X^2(1)=12.138$, $p<.001$) or wide-body (Wald $X^2(1)=40.607$, $p<.001$) aircraft. The difference in reported rates between small and wide-body aircraft was also statistically significant (Wald $X^2(1)=8.651$, $p=.003$). No other differences between aircraft categories were significant.

Interaction

No interactions between aircraft category and year/season were observed (ΔLLR $X^2(15)=17.942$, $p=.266$).[88]



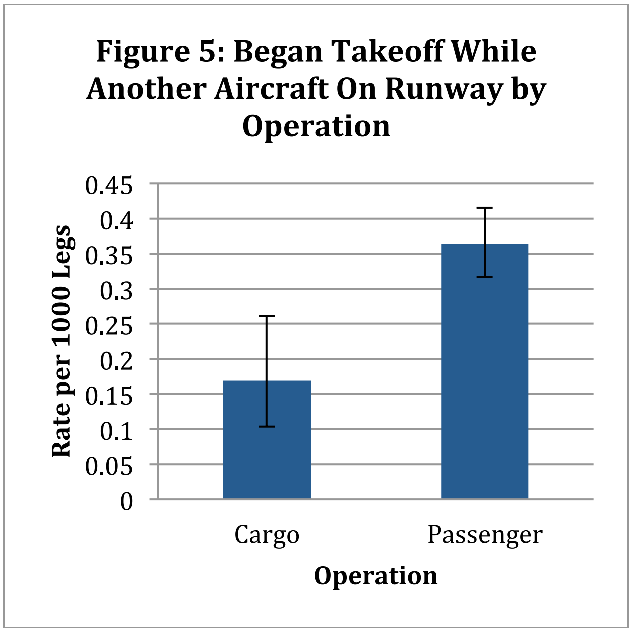
[88] The full factorial model did not fulfill convergence criteria. However, smaller models that did fulfill criteria were fit to the data as well and these also did not demonstrate any interaction effects.

Operation

Pilots of cargo aircraft reported significantly lower rates of beginning takeoffs with another aircraft on the runway than did pilots of passenger aircraft (Wald $X^2(1)=10.090$, p=.001; see Figure 5).

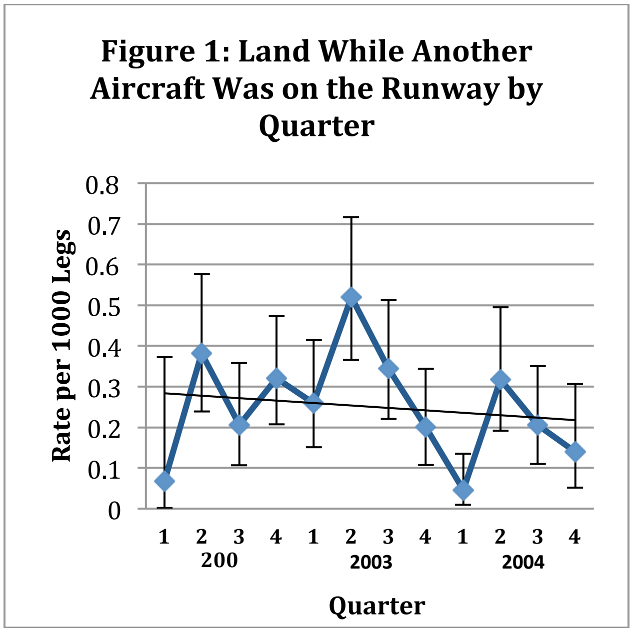
Interaction

There were no significant interactions between operation and years/seasons (Δ LLR $X^2(5)=3.702$, p<.593).



GE9: Landing While Another Aircraft is on the Runway

The rate of instances in which pilots reported landing while another aircraft was on or crossing the runway decreased linearly across the observation period (b=-.053, Wald $X^2(1)$ =4.891, p=.027; see Figure 1).[89] Across the study period, a mean rate of 0.27 events per 1,000 legs was reported. The lowest rate was observed in the first quarter of 2004 (0.05) and the highest rate was observed in the second quarter of 2003 (0.52).

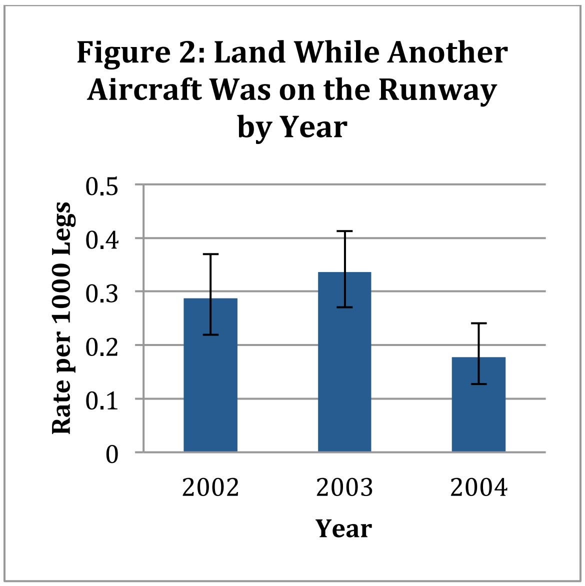


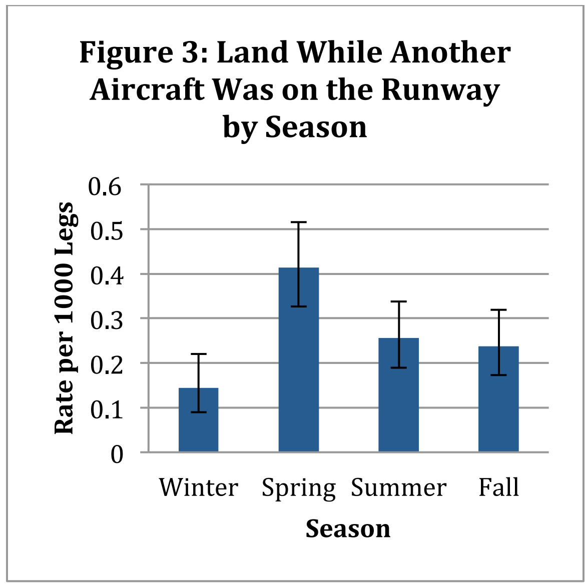
[89] Of the 18,367 reports obtained, 37 (0.20%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 181.35 per 1,000 legs compared to 0.24 per 1,000 legs for the other cases. The outlier cases were omitted from the remaining analyses. When the suspect observations are included, little change in the trend was observed. The rate of reported events decreased linearly across the observation period (b=-.103, Wald $X^2(1)$=32.782, p<.001). However, the overall fit of the model was extremely poor (Model with outliers: Pearson $X^2(18365)$=72236.226, ratio=3.933). The fit of the model without outliers was substantially better (Pearson $X^2(18328)$=32360.873; ratio=1.766), but there were still substantial departures. The reported results were confirmed using a variety of alternate analyses.

Year/Season

Significant differences between years were observed (Wald $X^2(2) = 11.980$, p=.003; see Figure 2). The reported rate of landings while other aircraft were on the runway in 2004 was significantly lower than in 2003 (Wald $X^2(1) = 13.404$, p<.001). No other difference between the years was significant.

There were also significant differences between the seasons (Wald $X^2(3) = 19.221$, p<.001). The lowest rates of these events occurred in the winter. The rate during the winter was significantly lower than in the spring (Wald X2(1)=25.314, p<.001), summer (Wald $X^2(1)=8.616$, p=.003), and fall (Wald $X^2(1)=5.011$, p=.025). The summer and fall did not differ (Wald X2(1)=0.456, p=.499). In the spring, the reported rate of these events was higher than in the summer (Wald $X^2(1)=6.036$, p=.014) and fall (Wald $X^2(1)=9.229$, p=.002). No year by season interaction was observed (Wald $X^2(6) = 9.148$, p=.165).





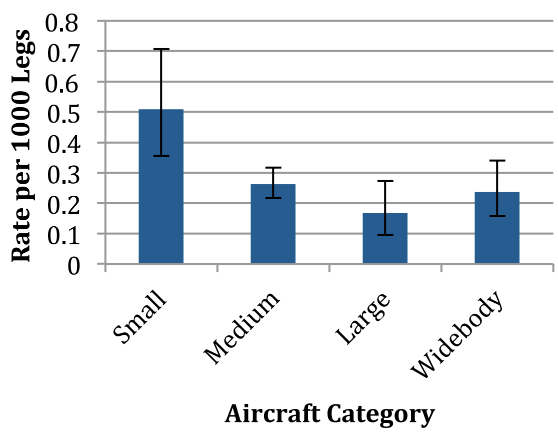
Aircraft Category

There were significant differences in the rates of these events between aircraft categories (Wald $X^2(3)=14.754$, p=.002). Pilots of small aircraft reported significantly more landings while another aircraft was on or crossing the runway than did pilots of aircraft in other categories (Wald $X^2(1)=7.948$, p=.005; see Figure 4). However, this result is due entirely to an unusually high rate of these events during the spring of 2003. When data from that quarter are removed, there are no significant differences between aircraft categories (Wald X2(3)=1.479, p=.687). There were no other significant differences between aircraft categories.

Interactions

Due to the small number of events and the unequal distribution of events across quarters, neither a full interaction model nor a model containing only the 2-way interactions could be properly estimated. When separate models are fit for each aircraft category, no significant year or season effects were observed for any aircraft category.

Figure 4: Land While Another Aircraft Was on the Runway by Aircraft Category

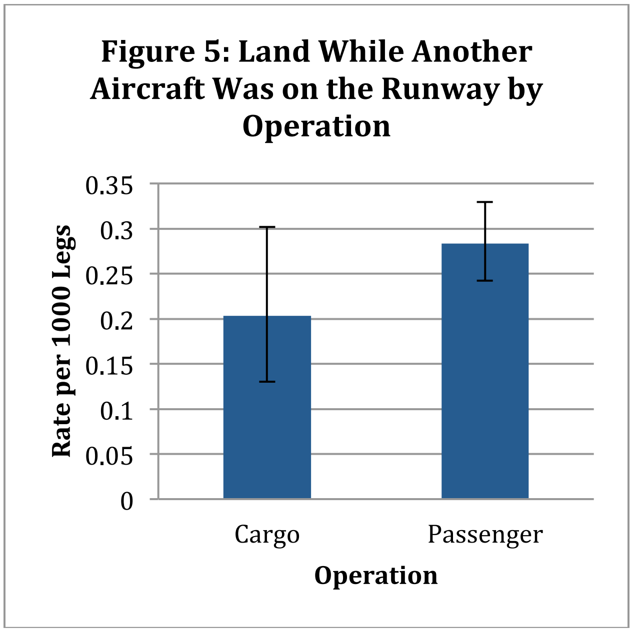


Operation

There were no significant differences between passenger and cargo operations in the rates of these events (Wald $X^2(1)=2.216$, p=.137).

Interactions

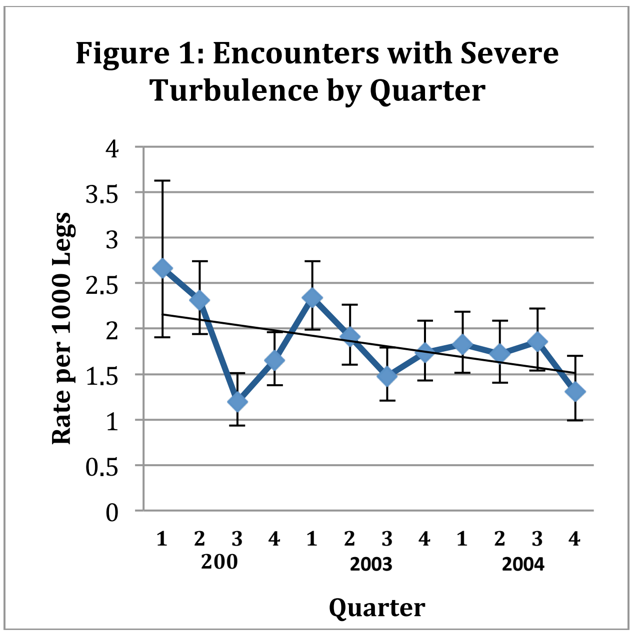
A full factorial model could not be fit; a smaller model containing only 2-way interaction terms revealed no significant interactions between type of operation and year/season (ΔLLR $X^2(5)=2.296$, p=.807).



Questions about turbulence

TU1: Encounters with Severe Turbulence

Across the study period, a mean rate of 1.79 encounters with severe turbulence per 1,000 legs was reported. The lowest rate was observed in the third quarter of 2002 (1.19) and the highest rate was observed in the first quarter of 2002 (2.67). The rate of reported instances in which pilots encountered severe turbulence that caused large abrupt changes in altitude, airspeed, or attitude did not change linearly across the observation period (b=-.016, Wald $X^2(1)$ =2.563, p=.109; see Figure 1).[90]



[90] Of the 18,376 reports obtained, 59 (0.32%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 259.50 per 1,000 legs compared to 2.38 per 1,000 legs for the other cases. When the suspect observations are included, little change in the trend was observed (b=-.007, Wald $X^2(1)$=0.600, p=.438). However, the overall fit of the model decreased (Model with outliers: Pearson $X^2(18374)$=46101.899, ratio=2.509; Model without outliers: Pearson $X^2(18315)$=30567.792; ratio=1.669). The outlier cases were omitted from the remaining analyses.

Year/Season

There were no differences across the years (Wald $X^2(2)$ =2.058, p=.357; see Figure 2), but there was a significant seasonal pattern (Wald $X^2(3)$ =25.747, p<.001; see Figure 3). The rates of reported encounters with severe turbulence were higher in the winter and spring compared to the summer and fall (Wald $X^2(1)$=22.898, p<.001). No other differences between seasons were significant. There was a statistically significant year by season interaction (Wald $X^2(6)$ =15.811, p=.015). This was due to an unusually high rate of these events during the summer of 2004.

Figure 2: Encounters with Severe Turbulence by Year

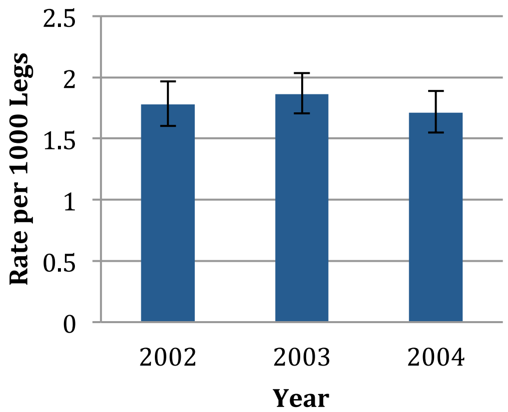
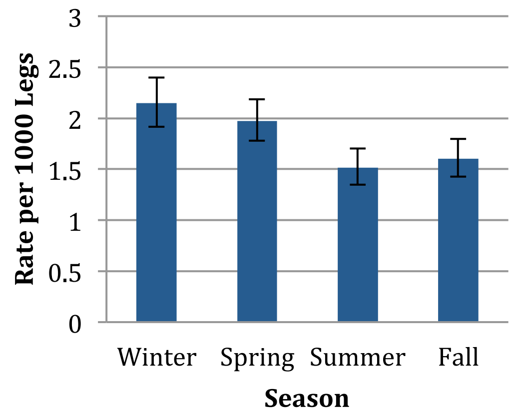


Figure 3: Encounters with Severe Turbulence by Season

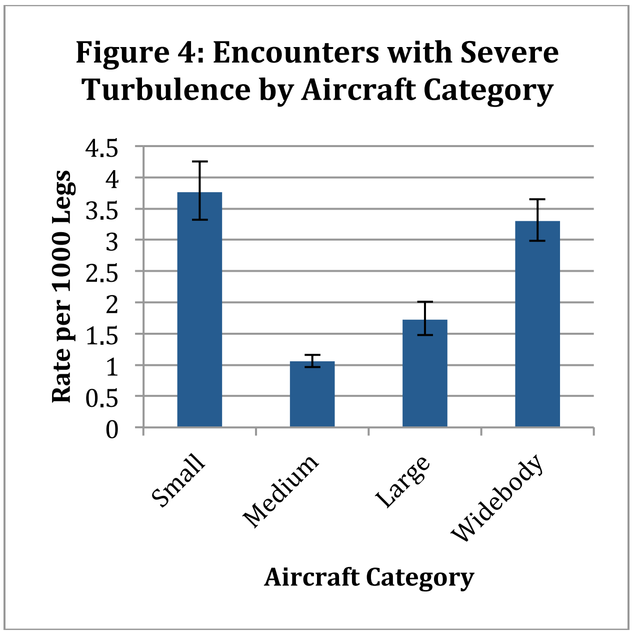


Aircraft Category

The rates of reported encounters with severe turbulence varied by aircraft category (Wald $X^2(3)=306.834$, $p<.001$; see Figure 4). Pilots of small and wide-body aircraft reported higher rates of encounters with severe turbulence than did pilots of medium or large aircraft (Wald $X^2(1)=124.151$, $p<.001$). In addition, pilots of large aircraft reported significantly higher rates than did pilots of medium aircraft (Wald $X^2(1)=17.957$, $p<.001$). No other differences between aircraft categories were statistically significant.

<u>Interactions</u>

There were significant interactions between aircraft category and year/season (Δ LLR $X^2(33)=51.236$, $p=.022$). This was due largely to greater variability in the rates reported by pilots of small aircraft. When only the data from the larger aircraft are considered, the interaction between aircraft category and year/season is not statistically significant (Δ LLR $X^2(22)=32.392$, $p=.071$).

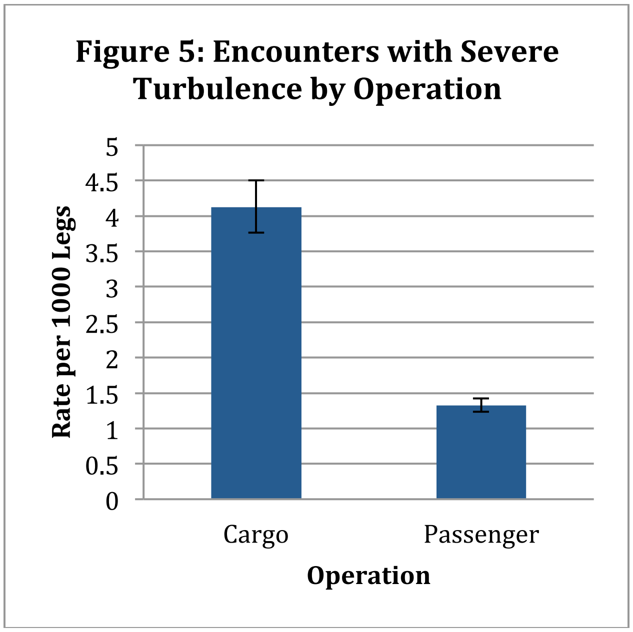


Operation

Pilots of cargo aircraft reported higher rates of encounters with severe turbulence than did pilots of passenger aircraft (Wald $X^2(1)$=316.612, p<.001).

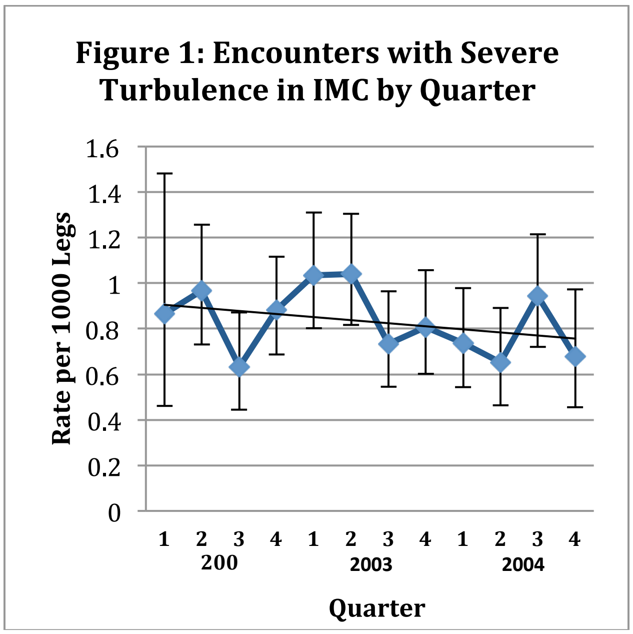
Interactions

There was no interaction between operation and year/season (ΔLLR $X^2(11)$ =9.274, p=.597).



TU1A: Encounters with Severe Turbulence in IMC

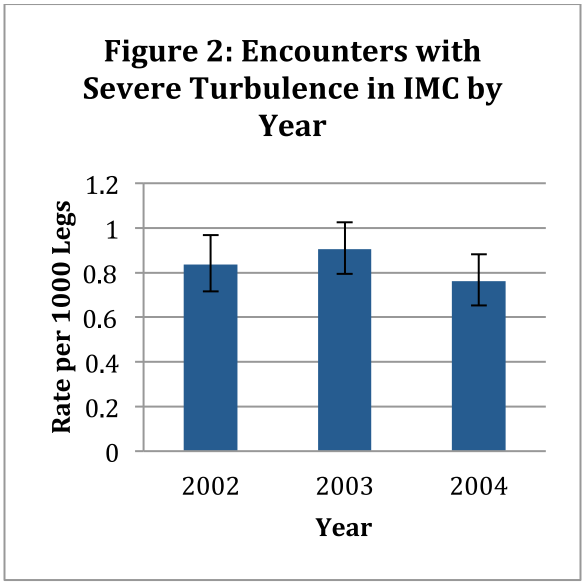
Across the study period, a mean rate of 1.04 encounters with severe turbulence in IMC per 1,000 legs was reported. The lowest rate was observed in the third quarter of 2002 (0.63) and the highest rate was observed in the second quarter of 2002 (1.04). The rate of reported instances in which pilots encountered severe turbulence in IMC that caused large abrupt changes in altitude, airspeed, or attitude did not change linearly across the observation period (b=-.018, Wald $X^2(1)$ =1.654, p=.198; see Figure 1).[91]

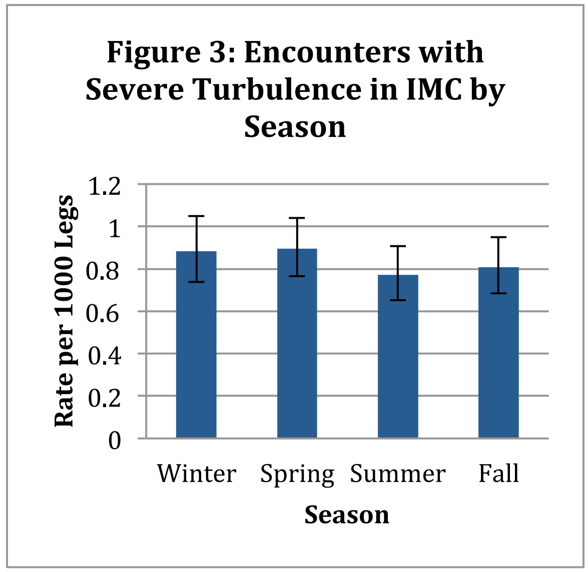


[91] Of the 18,372 reports obtained, 46 (0.25%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 213.85 per 1,000 legs compared to 1.06 per 1,000 legs for the other cases. When the suspect observations are included, little change in the trend was observed (b=-.011, Wald $X^2(1)$=0.713, p=.398). However, the overall fit of the model decreased (Model with outliers: Pearson $X^2(18370)$=44769.294, ratio=2.437; Model without outliers: Pearson $X^2(18324)$=27383.360; ratio=1.494). The outlier cases were omitted from the remaining analyses.

Year/Season

There were no differences across the years (Wald $X^2(2)$ =3.381, p=.184; see Figure 2) or seasons (Wald $X^2(3)$ =1.890, p=.596; see Figure 3) and there was no year by season interaction (Wald $X^2(6)$ =10.761, p=.096).



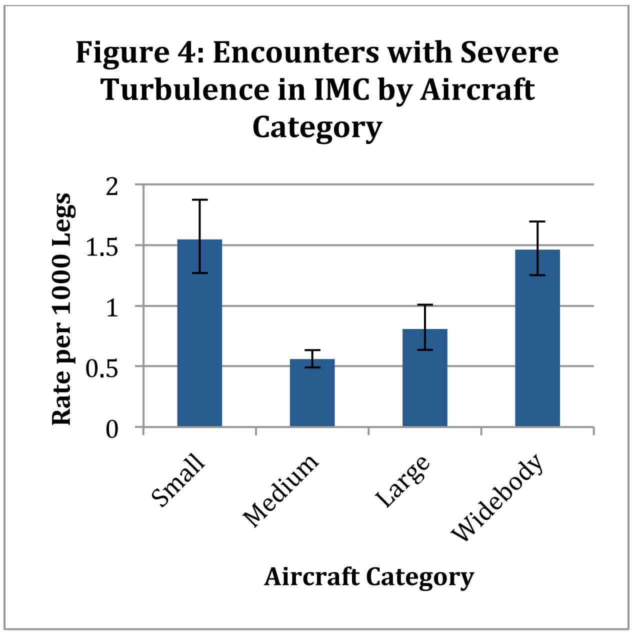


Aircraft Category

The rates of reported encounters with severe turbulence in IMC varied by aircraft category (Wald $X^2(3)=116.477$, $p<.001$; see Figure 4). Pilots of small and wide-body aircraft reported higher rates of encounters with severe turbulence in IMC than did pilots of medium or large aircraft (Wald $X^2(1)=51.992$, $p<.001$). In particular, pilots of medium aircraft reported the lowest rates, significantly lower than pilots of large aircraft (Wald $X^2(1)=5.811$, $p=.016$). In turn, pilots of large aircraft reported significantly lower rates of these events than did pilots of wide-body aircraft (Wald $X^2(1)=17.950$, $p<.001$) or small aircraft (Wald $X^2(1)=16.954$, $p<.001$). No difference between small and wide-body aircraft categories was observed (Wald $X^2(1)=0.939$, $p=.333$).

Interactions

There were no interactions between aircraft category and year/season (Δ LLR $X^2(33)=38.358$, $p=.239$).

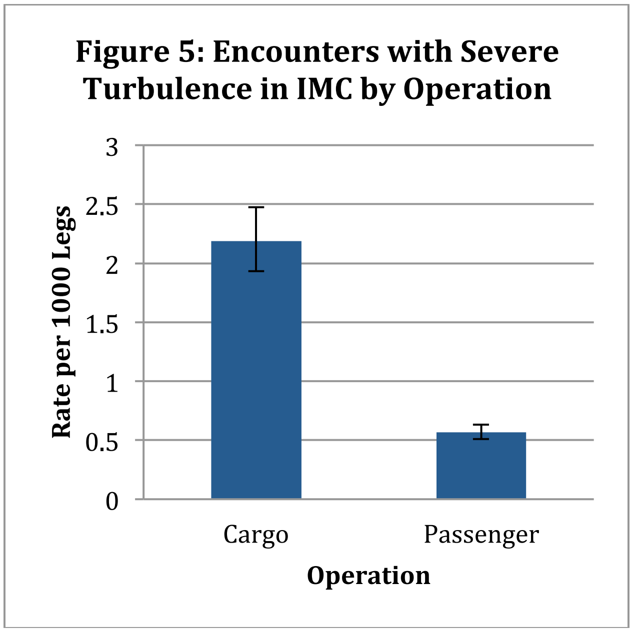


Operation

Pilots of cargo aircraft reported higher rates of encounters with severe turbulence in IMC than did pilots of passenger aircraft (Wald $X^2(1)=237.442$, $p<.001$).

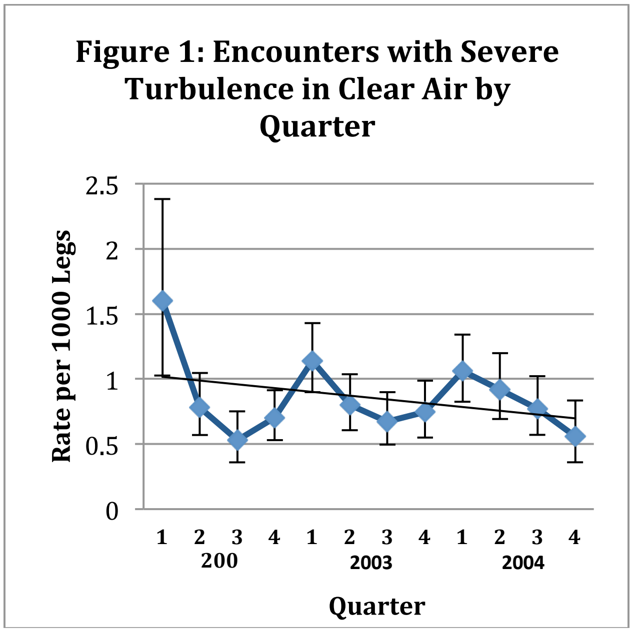
<u>Interactions</u>

There was no interaction between operation and year/season (ΔLLR $X^2(11)=8.246$, $p=.691$).



TU1B: Encounters with Severe Turbulence in Clear Air

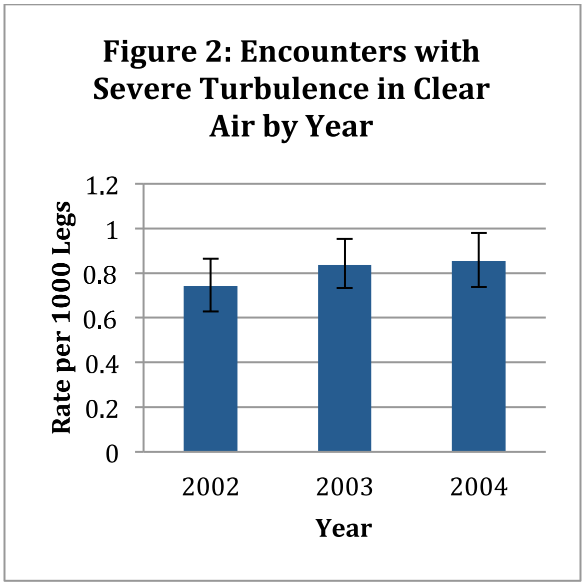
Across the study period, a mean rate of 0.81 encounters with severe turbulence in clear air per 1,000 legs was reported. The lowest rate was observed in the third quarter of 2002 (0.53) and the highest rate was observed in the first quarter of 2002 (1.60). The rate of reported instances in which pilots encountered severe turbulence in clear air that caused large abrupt changes in altitude, airspeed, or attitude did not change linearly across the observation period (b=-.006, Wald $X^2(1)$ =0.158, p=.691; see Figure 1).[92]

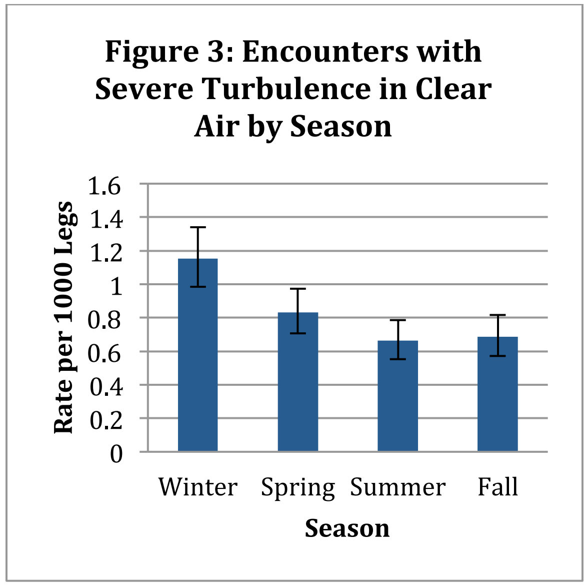


[92] Of the 18,370 reports obtained, 49 (0.27%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 171.77 per 1,000 legs compared to 1.07 per 1,000 legs for the other cases. When the suspect observations are included, little change in the trend was observed (b=.013, Wald $X^2(1)$=0.972, p=.324). However, the overall fit of the model decreased (Model with outliers: Pearson $X^2(18368)$=44109.608, ratio=2.401; Model without outliers: Pearson $X^2(18319)$=28338.229; ratio=1.547). The outlier cases were omitted from the remaining analyses.

Year/Season

There were no differences across the years (Wald $X^2(2)$ =0.033, p=.984; see Figure 2). However, there was a significant seasonal pattern (Wald $X^2(3)$ =33.109, p<.001; see Figure 3). The reported rate of these events was higher in the winter than during the other seasons (Wald $X^2(1)$=19.685, p<.001). The mean rate during the spring was lower than in the winter (Wald $X^2(1)$=9.857, p=.002) but higher than in the summer (Wald $X^2(1)$=4.332, p=.037) and marginally higher than in the fall (Wald $X^2(1)$=3.225, p=.073). The rates reported in the summer did not differ from those reported in the fall (Wald $X^2(1)$=.055, p=.815). There was no year by season interaction (Wald $X^2(6)$ =6.988, p=.322).



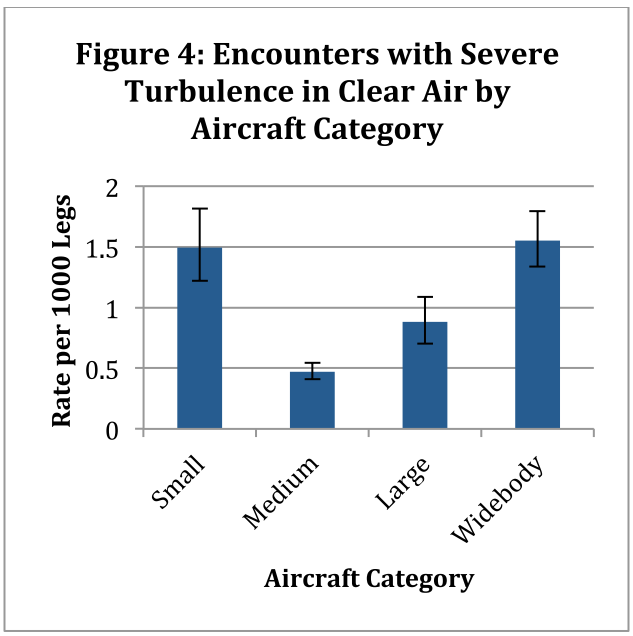


Aircraft Category

The rates of reported encounters with severe turbulence in clear air varied by aircraft category (Wald $X^2(3)=152.322$, $p<.001$; see Figure 4). Pilots of small and wide-body aircraft reported higher rates of encounters with severe turbulence in clear air than did pilots of medium or large aircraft (Wald $X^2(1)=56.464$, $p<.001$). In particular, pilots of medium aircraft reported the lowest rates, significantly lower than pilots of large aircraft (Wald $X^2(1)=14.087$, $p<.001$). In turn, pilots of large aircraft reported significantly lower rates of these events than did pilots of wide-body aircraft (Wald $X^2(1)=18.978$, $p<.001$) or small aircraft (Wald $X^2(1)=12.636$, $p<.001$). No difference between small and wide-body aircraft categories was observed (Wald $X^2(1)=0.036$, $p=.850$).

<u>Interactions</u>

There were significant interactions between aircraft category and year/season (Δ LLR $X^2(33)=62.078$, $p<.002$). The seasonal effect was significant for the wide-body aircraft (Wald $X^2(3)=24.906$, $p<.001$), marginally significant for the large aircraft (Wald $X^2(3)=7.012$, $p=.072$) and not significant for the medium aircraft (Wald $X^2(3)=5.089$, $p=.165$) and small aircraft (Wald $X^2(3)=4.389$, $p=.109$).

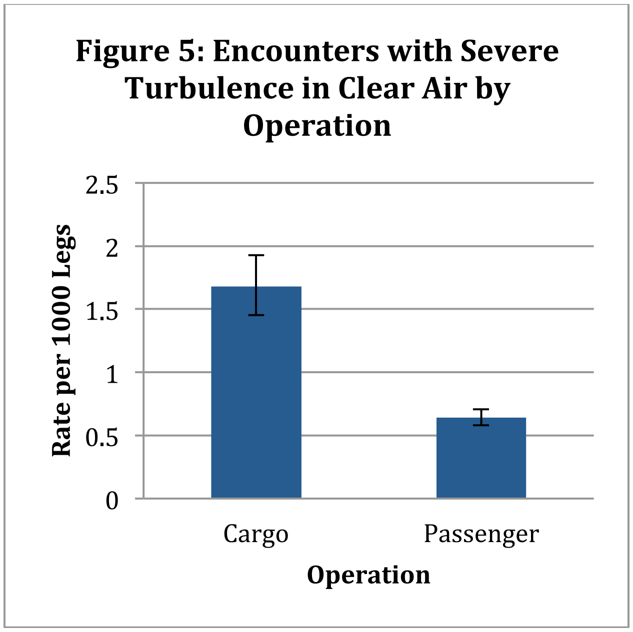


Operation

Pilots of cargo aircraft reported higher rates of encounters with severe turbulence in clear air than did pilots of passenger aircraft (Wald $X^2(1)=105.276$, $p<.001$).

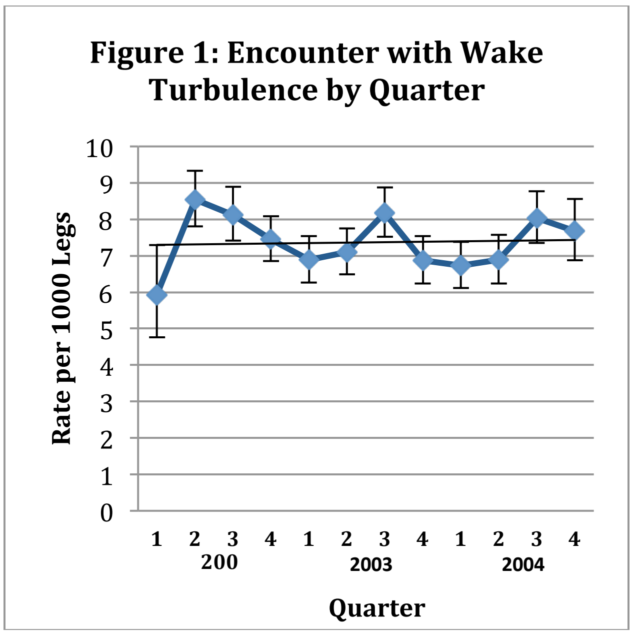
<u>Interactions</u>

There was a significant interaction between operation and year/season (ΔLLR $X^2(11)=24.808$, $p=.009$). Pilots of cargo aircraft reported unusually high rates of severe turbulence in clear air in the winter of 2003. Pilots of passenger aircraft reported unusually high rates of severe turbulence in clear air during the winter of 2002.



TU2: Encounter Wake Turbulence

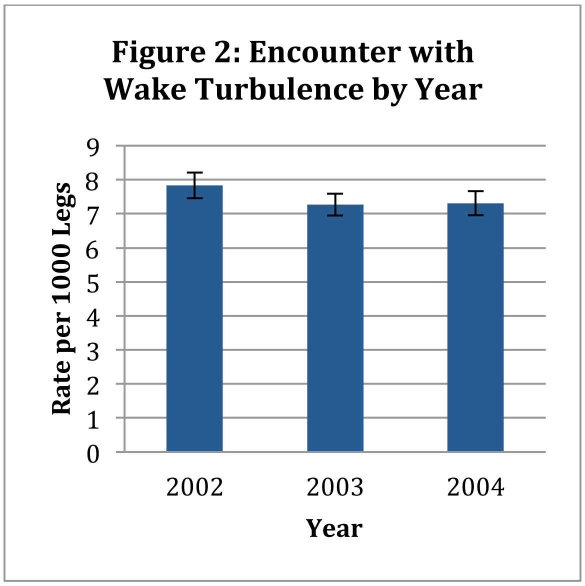
Across the study period, a mean rate of 7.44 encounters with wake turbulence per 1,000 legs was reported. The lowest rate was observed in the first quarter of 2002 (5.93) and the highest rate was observed in the second quarter of 2002 (8.54). The rate of reported encounters with wake turbulence did not change linearly across the observation period (b=-.006, Wald $X^2(1)$ =0.894, p=.344; see Figure 1).[93]

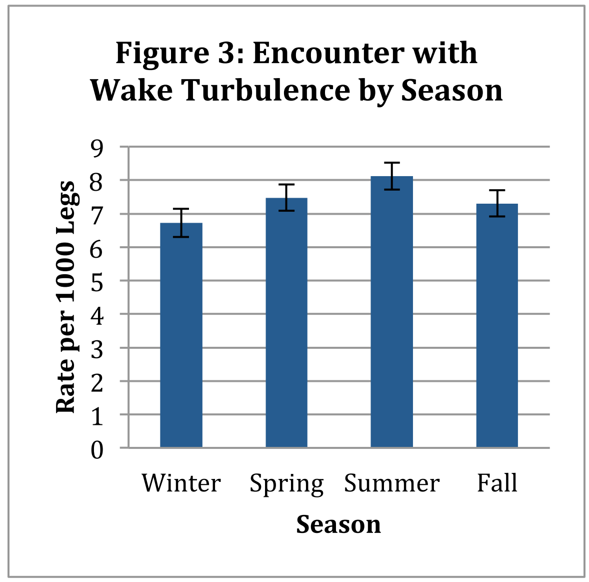


[93] Of the 18,365 reports obtained, 45 (0.25%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 343.79 per 1,000 legs compared to 7.87 per 1,000 legs for the other cases. When the suspect observations are included, little change in the trend was observed (b=-.003, Wald $X^2(1)$=0.355, p=.551). However, the overall fit of the model decreased (Model with outliers: Pearson $X^2(18363)$=28207.934, ratio=1.536; Model without outliers: Pearson $X^2(18318)$=21510.751; ratio=1.174). The outlier cases were omitted from the remaining analyses.

Year/Season

There were no differences across the years (Wald $X^2(2)$ =0.151, p=.927; see Figure 2), but there was a significant seasonal pattern (Wald $X^2(3)$ =13.540, p=.004; see Figure 3). The rate of reported events was lower in the winter compared to all of the other seasons (Wald $X^2(1)$ =9.901, p=.002). The rate of these events was higher in the summer than in the fall (Wald $X^2(1)$=4.209, p=.040) but did not differ between the spring and the summer (Wald $X^2(1)$=2.273, p=.132). There was no year by season interaction (Wald $X^2(6)$ =6.293, p=.391).



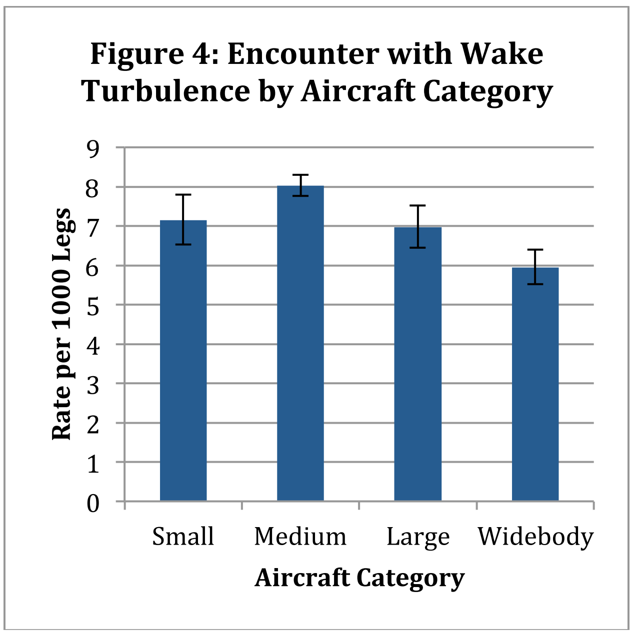


Aircraft Category

The reported rates of encounters with wake turbulence differed by aircraft category (Wald $X^2(3)=48.911$, $p<.001$; see Figure 4). Pilots of wide-body aircraft reported lower rates of these events than did pilots of the other aircraft (Wald $X^2(1)=25.591$, $p<.001$). Pilots of medium aircraft reported the highest rates, significantly greater than the rates reported by pilots of large (Wald $X^2(1)=11.493$, $p=.001$) and wide-body aircraft (Wald $X^2(1)=52.105$, $p<.001$), but not significantly different from the rate reported by pilots of small aircraft (Wald $X^2(1)=1.739$, $p=.187$).

<u>Interactions</u>

There was a significant interaction between aircraft category and year/season (ΔLLR $X^2(33)=53.856$, $p=.012$). This was due to an anomalously high rate of events reported by pilots of small aircraft during the final quarter of 2004. When this quarter is eliminated from the analysis, no interaction between aircraft category and year or season is observed (ΔLLR $X^2(30)=36.682$, $p=.187$).

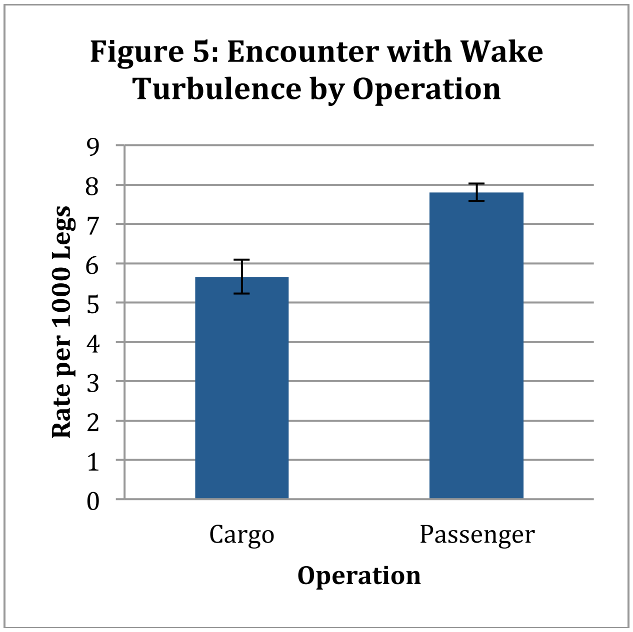


Operation

Pilots of cargo aircraft reported significantly lower rates of encounters with wake turbulence than did pilots of passenger aircraft (Wald $X^2(1)=42.173$, $p<.001$).

Interactions

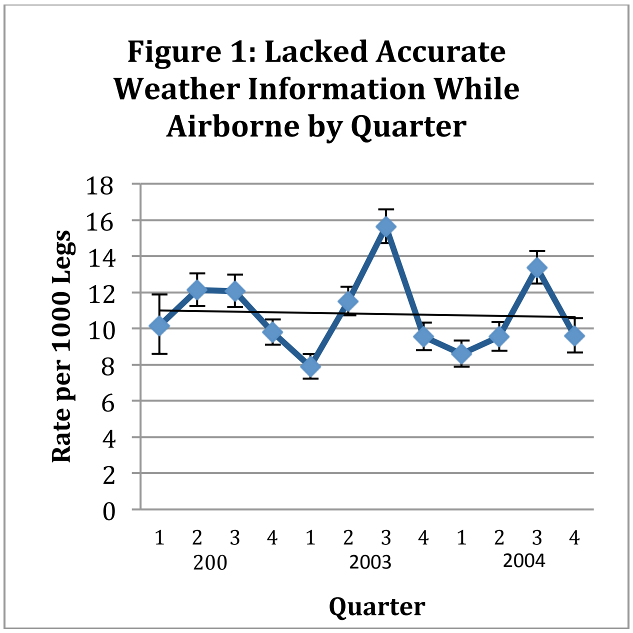
There was no interaction between operation and year/season (ΔLLR $X^2(11)=16.042$, $p=.140$).



Questions about weather-related events while airborne

WE1: Lacked Accurate Weather Information While Airborne

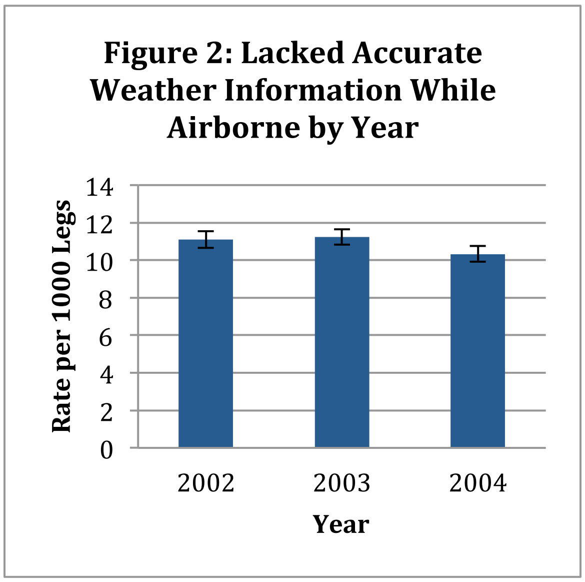
Across the study period, a mean rate of 10.9 events in which pilots reported lacking accurate weather information while airborne per 1,000 legs was reported. The lowest rate was observed in the first quarter of 2003 (7.89) and the highest rate was observed in the third quarter of 2003 (15.63). The rate of reported instances in which aircrews lacked accurate weather information while airborne did not increase or decrease linearly across the observed quarters (b=-.002, Wald $X^2(1)=.170$, p=.680).[94]

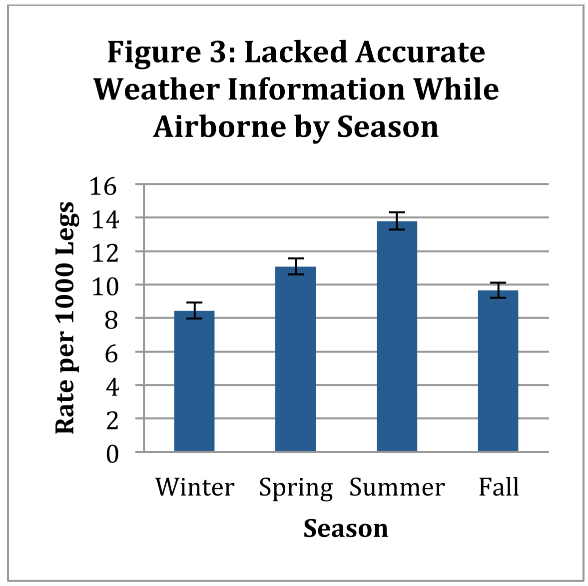


[94] Of the 18,350 reports obtained, 54 (0.29%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 1049.41 per 1,000 legs compared to 14.40 per 1,000 legs for the other cases. When the suspect observations are included, a statistically significant decrease in the rate of these events cross the observation period was observed (b=-.015, Wald $X^2(1)=8.598$, p=.003). However, the overall fit of the model decreased (Model with outliers: Pearson $X^2(18348)=96337.885$, ratio=5.251; Model without outliers: Pearson $X^2(18294)=63964.113$; ratio=3.496). The high X^2/df ratios indicate that these data depart from the negative binomial (2) distribution. The outlier cases were omitted from the remaining analyses. These cases were omitted from the remaining analyses.

Year/Season

The rate of reported events did not change over the observed years (Wald $X^2(2)=2.182$, $p=.336$; see Figure 2). However, there was a significant seasonal pattern (Wald $X^2(3)=69.910$, $p<.001$; see Figure 3). All seasons were significantly different from each other. There was also a year by season interaction (Wald $X^2(6)=18.730$, $p=.005$). In 2003 and 2004, the rates of these events were highest in the summer. However, in 2002 the rates of these events were highest in the spring and summer, which did not differ.



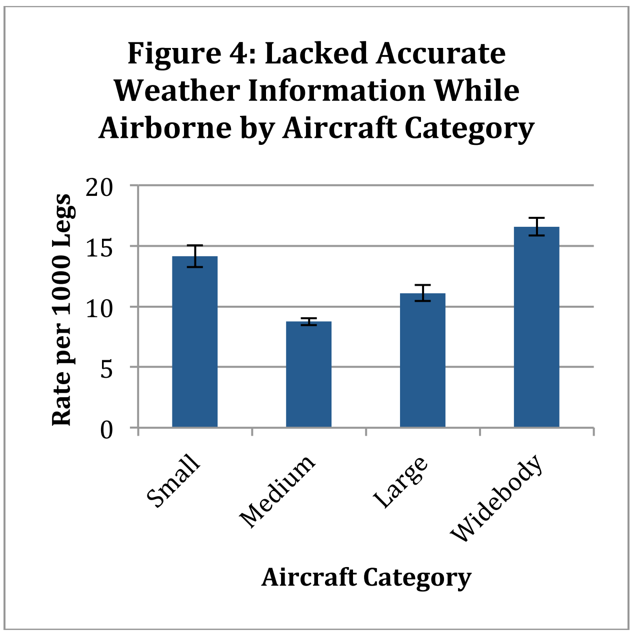


Aircraft Category

The rates of reported events differed across aircraft categories (Wald $X^2(3)=238.414$, $p<.001$; see Figure 4). Pilots of small and wide-body aircraft reported a higher rate of these events than did pilots of medium and large aircraft (Wald $X^2(1)=100.125$, $p<.001$). There was no difference between the rates reported by pilots of small and wide-body aircraft (Wald X2(1)=0.195, p=.658). Pilots of large aircraft reported higher rates than did pilots of medium aircraft (Wald X2(1)=14.700, p<.001).

Interactions.

Statistically significant interactions between aircraft category and year/season were observed (ΔLLR $X^2(33)=85.998$, $p<.001$). The data obtained from pilots of medium, large, and wide-body aircraft demonstrated significant seasonal and year by season effects. The data obtained from pilots of small aircraft did not.

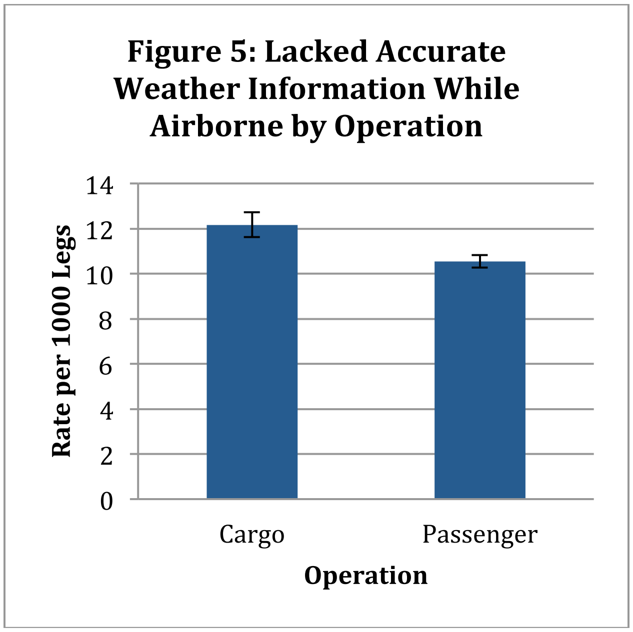


Operation

Pilots of cargo aircraft reported higher rates of these events than did pilots of passenger aircraft (Wald $X^2(1)=120.708$, $p<.001$; see Figure 5).

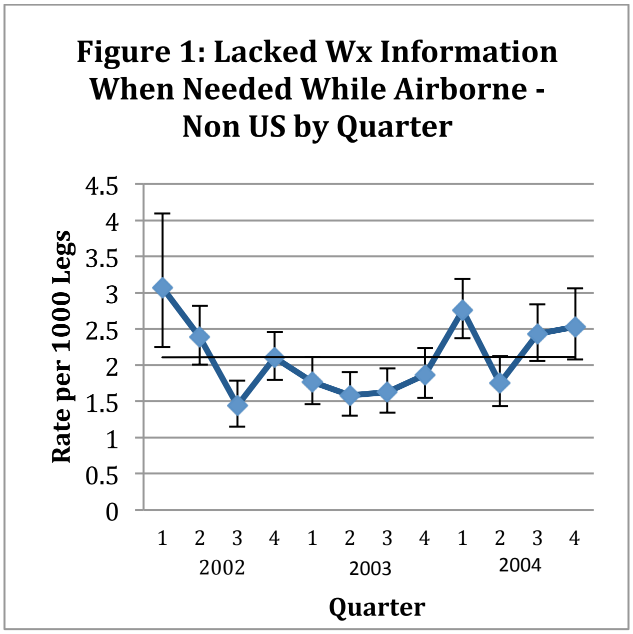
Interactions.

Statistically significant interactions between operation and year/season were observed (ΔLLR $X^2(11)=37.742$, $p<.001$). The rates reported for cargo aircraft demonstrated somewhat more variability than did those reported for passenger aircraft.



WE1A: Lacked Wx Information When Needed While Airborne—Non-US

Across the study period, a mean rate of 2.01 events in which crews reported that they lacked weather information while airborne outside the United States per 1,000 legs was reported. The lowest rate was observed in the third quarter of 2002 (1.44) and the highest rate was observed in the first quarter of 2002 (3.07). The rate of events in which crews reported that they lacked weather information while airborne outside the United States did not change linearly across the observation period (b=.018, Wald $X^2(1)$ =3.520, p=.061; see Figure 1).[95]



[95] Of the 18,353 reports obtained, 65 (0.35%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 528.69 per 1,000 legs compared to 3.23 per 1,000 legs for the other cases. When the suspect observations are included, little change in the trend was observed (but see main text). There was no change in the rate of reported events across quarters during the observation period (b=.002, Wald $X^2(1)$=0.038, p=.845). However, the overall fit of the model decreased (Model with outliers: Pearson $X^2(18351)$=132127.453, ratio=7.200; Model without outliers: Pearson $X^2(18286)$=66636.062; ratio=3.644). These cases were omitted from the remaining analyses.

Year/Season

There were significant differences in the rates of these events across years (Wald $X^2(2)=19.821$, $p<.001$; see Figure 2). The reported rates of these events were lower in 2003 than in 2002 and 2004 (Wald X2(1)=18.412, $p<.001$). There were significant differences in the rates of these events across seasons (Wald $X^2(3)=13.035$, $p<.005$). The reported rates of these events were lower in the spring and summer than in the winter and fall (Wald X2(1)=12.161, $p<.001$). However, there was a significant year by season interaction, indicating that the seasonal pattern differed across years (Wald $X^2(6)=19.104$, $p<.004$; see Figure 3). In 2002, the rates reported during the summer were significantly lower than during the other quarters, which did not differ. In 2003, there were no significant differences across seasons. In 2004, the rates reported during the spring were significantly lower than during the other quarters, which did not differ.

Figure 2: Lacked Wx Information When Needed While Airborne - Non US by Year

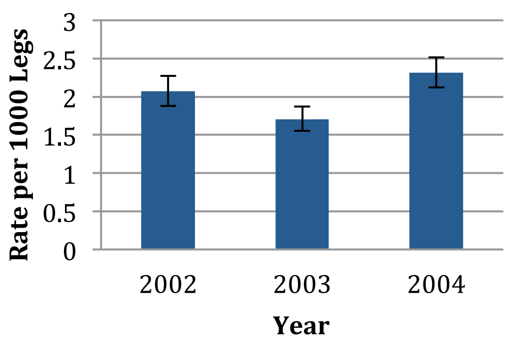
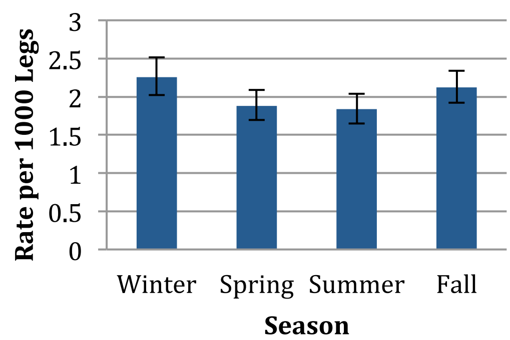


Figure 3: Lacked Wx Information When Needed While Airborne - Non US by Season

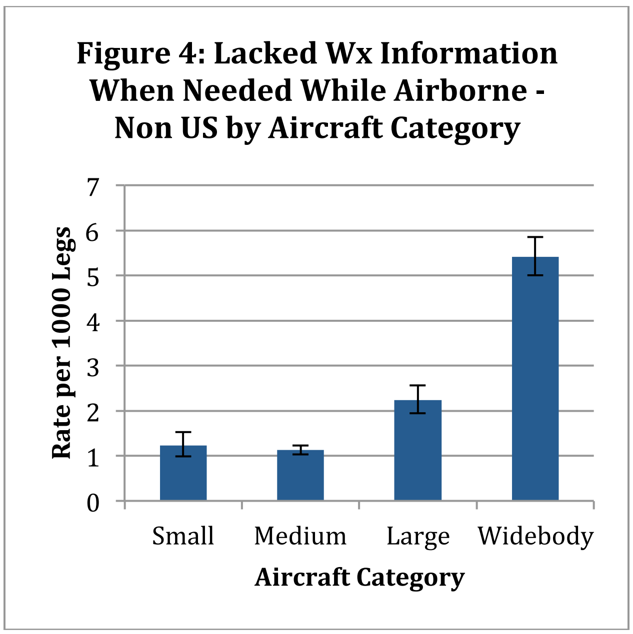


Aircraft Category

The rates of these events differed by aircraft category (Wald $X^2(3)=606.397$, $p<.001$). Not surprisingly, pilots of wide-body aircraft reported the highest rate of these events by far (see Figure 4). These aircraft are frequently used for long haul operations. This rate was significantly higher than that reported by pilots of smaller aircraft (Wald $X^2(1)=226.080$, $p<.001$). Pilots of large aircraft reported significantly higher rates of these events than did pilots of smaller aircraft (Wald $X^2(1)=30.592$, $p<.001$). No difference in the rates reported by pilots of small and medium aircraft was observed (Wald $X^2(1)=0.885$, $p=.347$).

Interactions

There was a significant interaction between aircraft category and year/season (ΔLLR $X^2(33)=140.906$, $p<.001$). Each category of aircraft demonstrated somewhat different patterns across seasons and years.

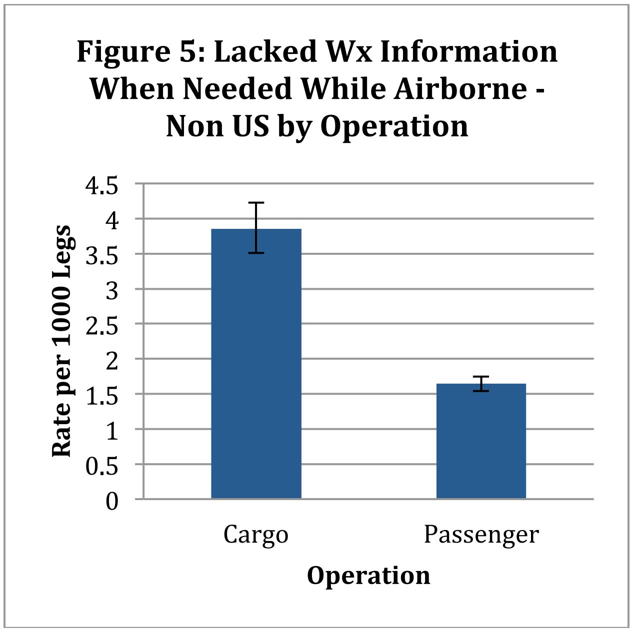


Operation

Pilots of cargo aircraft reported significantly higher rates of these events than did pilots of passenger aircraft (Wald $X^2(1)=176.780$, $p<.001$).

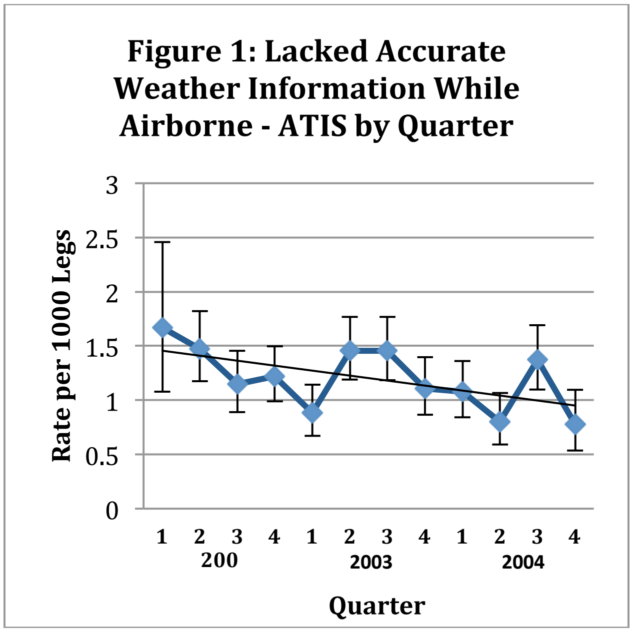
Interactions

There were significant interactions between operation and year/season (ΔLLR $X^2(11)=25.018$, $p=.009$). The rates reported by pilots of cargo aircraft demonstrated considerably greater variability than did those reported by pilots of passenger aircraft.



WE1B: Lacked Accurate Weather Information While Airborne - ATIS

Across the study period, a mean rate of 1.19 instances per 1,000 legs of pilots lacking accurate weather information while airborne due to ATIS problems was reported. The lowest rate was observed in the fourth quarter of 2004 (0.78) and the highest rate was observed in the first quarter of 2002 (1.67). The rate of reported instances of these events decreased linearly across the observation period (b=-.030, Wald $X^2(1)$ =6.198, p=.013; see Figure 1).[96]



[96] Of the 18,327 reports obtained, 54 (0.29%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 275.22 per 1,000 legs compared to 1.24 per 1,000 legs for the other cases. When the suspect observations are included, little change in the trend was observed (b=-.023, Wald $X^2(1)$=4.725, p=.030). However, the overall fit of the model decreased (Model with outliers: Pearson $X^2(18325)$=66810.909, ratio=3,646; Model without outliers: Pearson $X^2(18271)$=34064.788; ratio=1.864). The outlier cases were omitted from the remaining analyses.

Year/Season

There were significant differences across the years in the reported rates of these events (Wald $X^2(2) = 10.489$, $p=.005$; see Figure 2). There was a significant linear decrease across years (Wald $X^2(1)=9.810$, $p=.002$). There was no seasonal pattern (Wald $X^2(3) = 4.527$, $p=.210$; see Figure 3) but there was a significant season by year interaction (Wald $X^2(6)=18.987$, $p=.004$). There were no significant differences across seasons in 2002 (Wald $X^2(3)=3.822$, $p=.281$). There were significant differences across seasons in 2003 (Wald $X^2(3)=9.604$, $p=.022$) and 2004 Wald $X^2(3)=11.159$, $p=.011$), but the patterns were different.

Figure 2: Lacked Accurate Weather Information While Airborne - ATIS by Year

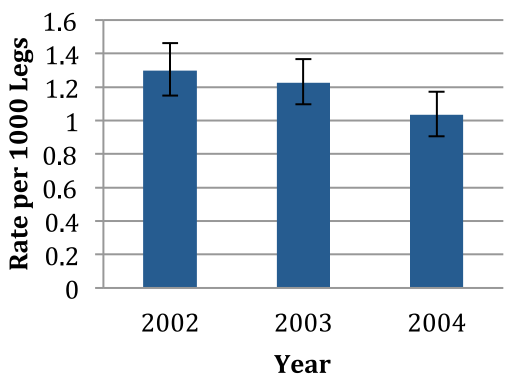
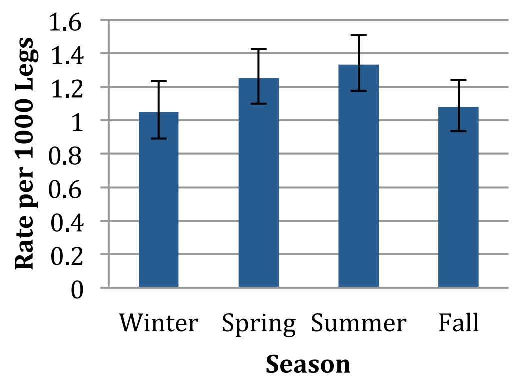


Figure 3: Lacked Accurate Weather Information While Airborne - ATIS by Season

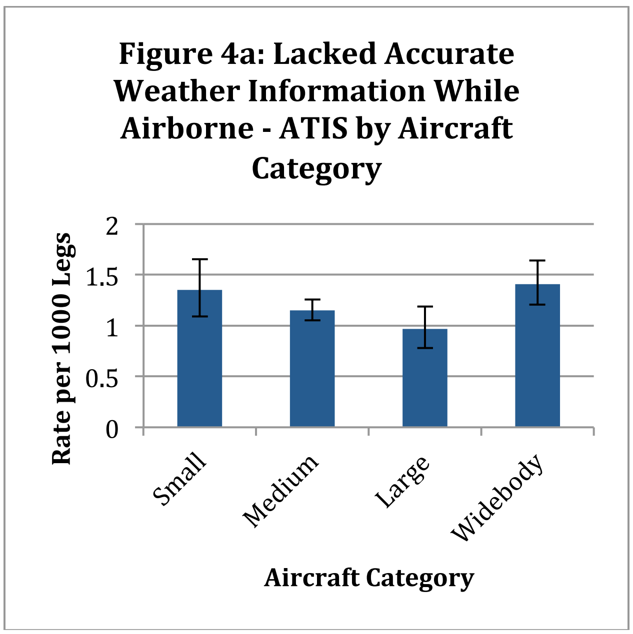


Aircraft Category

The rates of these events varied by aircraft category (Wald $X^2(3)=7.851$, p<.049; see Figure 4). Pilots of large aircraft reported the lowest rates of lacking accurate weather information while airborne due to problems with ATIS, significantly lower than the rates of these events reported by pilots of small (Wald $X^2(1)=4.058$, p=.044) or wide-body aircraft (Wald $X^2(1)=6.611$, p=.010) but not significantly different from the rates reported by pilots of medium aircraft (Wald $X^2(1)=2.448$, p=.118). No other differences between aircraft categories were statistically significant.

<u>Interactions</u>

There was a significant interaction between aircraft category and year/season (ΔLLR $X^2(33)=56.966$, p=.006). Small aircraft demonstrated a significant decrease across years (Wald X2(1)=4.844, p=.028). The decrease across years was nearly significant for large aircraft (Wald X2(1)=3.671, p=.055). There were no significant patterns across years or seasons for medium or wide-body aircraft.

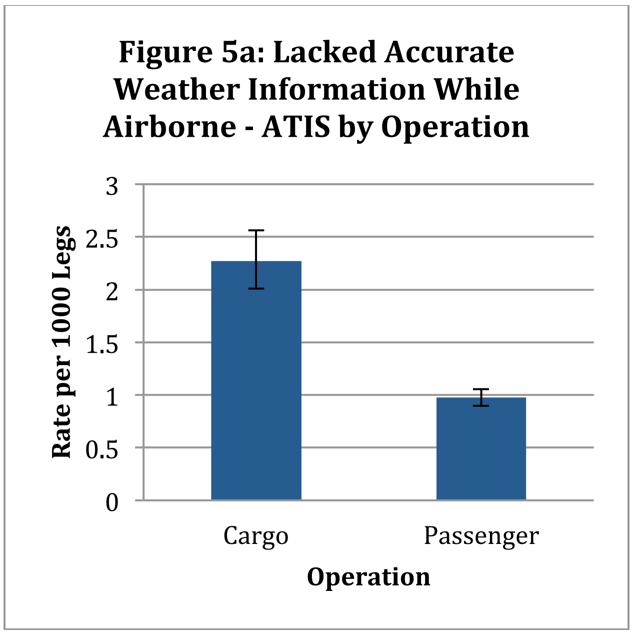


Operation

Pilots of cargo aircraft reported higher rates of these events than did pilots of passenger aircraft (Wald $X^2(1)$=101.670, p<.001).

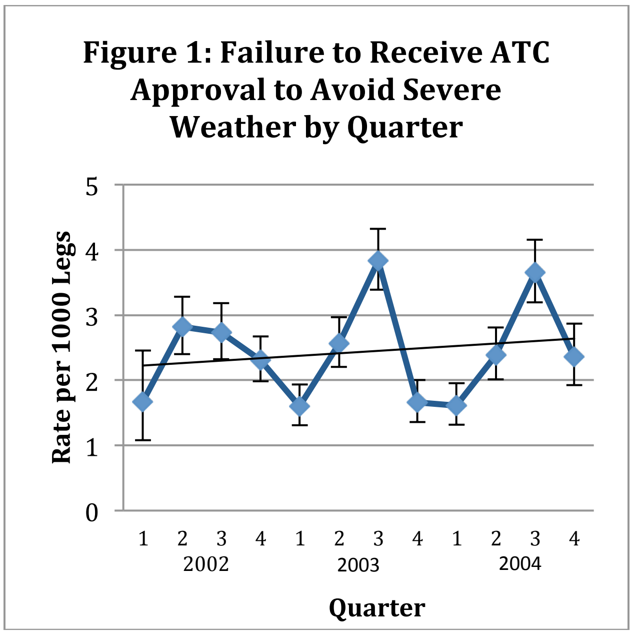
Interactions

However, there were significant interactions between operation and year/season (ΔLLR $X^2(11)$ =27.844, p=.003). The linear decrease in the rate of these events across years was more pronounced and less variable across seasons for cargo aircraft (Wald $X^2(1)$=4.321, p=.038) than for passenger aircraft (Wald $X^2(1)$=3.703, p=.054).



WE2: Failure to Receive ATC Approval for a Request to Avoid Severe Weather

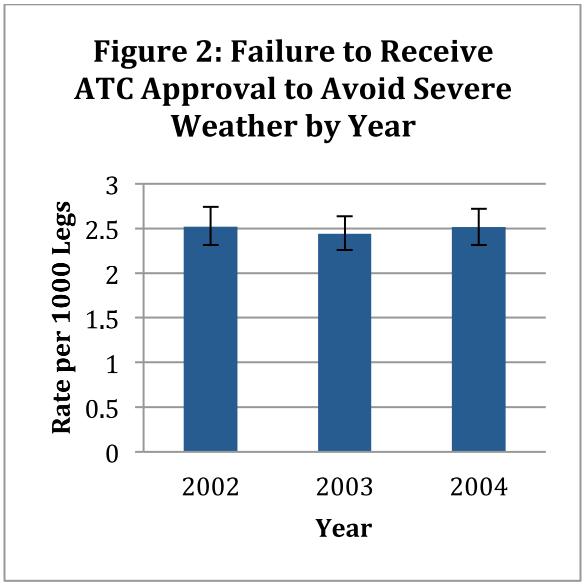
Across the study period, a mean rate of 2.49 failures to receive ATC approval for a request to avoid severe weather per 1,000 legs was reported. The lowest rate was observed in the first quarter of 2003 (1.60) and the highest rate was observed in the third quarter of 2003 (3.83). The rate of reported failures to receive ATC approval for a request to avoid severe weather did not increase or decrease linearly across the observed quarters (b=.007, Wald $X^2(1)=.722$, p=.395).[97]

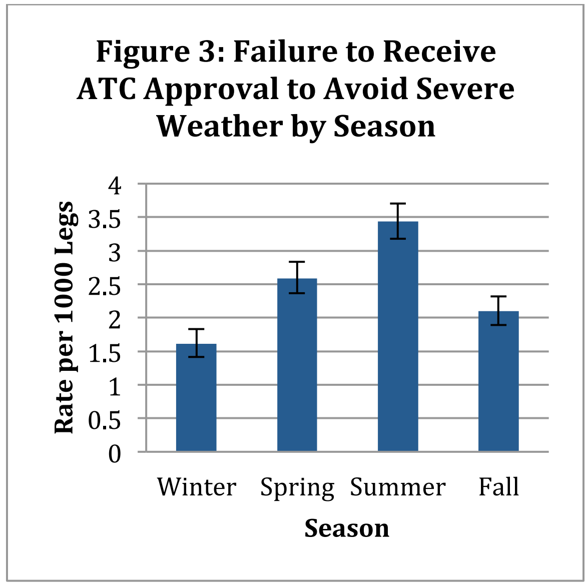


[97] Of the 18,367 reports obtained, 47 (0.26%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 666.25 per 1,000 legs compared to 3.35 per 1,000 legs for the other cases. When the suspect observations are included, the trend becomes statistically significant (b=.031, Wald $X^2(1)=16.371$, p<.001). However, the overall fit of the model decreased substantially (Model with outliers: Pearson $X^2(18365)=142660.521$, ratio=7.768; Model without outliers: Pearson $X^2(18318)=56978.881$; ratio=3.111). The outlier cases were omitted from the remaining analyses.

Year/Season

The rate of reported events did not change over the observed years (Wald $X^2(2)=0.617$, p=.735; see Figure 2). However, there was a significant seasonal pattern (Wald $X^2(3)=78.713$, p<.001; see Figure 3). The rate of reported events was highest in the summer, significantly lower in the spring (Wald $X^2(1)=14.933$, p<.001), and still lower in the fall than in the spring (Wald $X^2(1)=10.144$, p=.001). The rate of reported events in the fall was significantly higher than the rate in the winter (Wald $X^2(1)=4.300$, p=.038). There was a year by season interaction (Wald $X^2(6)=18.963$, p=.004). There was an unusually low rate of these events (not significantly different from the spring) during the summer of 2002.



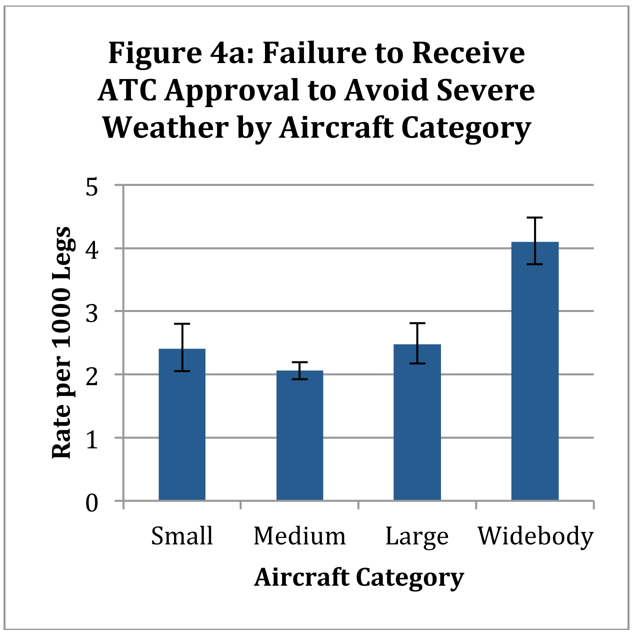


Aircraft Category

There were significant differences in these event rates between aircraft categories (Wald $X^2(3)=109.907$, $p<.001$). Pilots of wide-body aircraft reported higher rates of events that did pilots of other aircraft (Wald $X^2(1)=52.378$, $p<.001$; see Figure 4). Pilots of medium aircraft reported lower rates than did pilots of large aircraft (Wald $X^2(1)=4.766$, $p=.029$) but only marginally lower than pilots of small aircraft (Wald $X^2(1)=3.365$, $p=.067$).

<u>Interactions</u>

There were significant differences between aircraft categories in the year/seasonal pattern described previously (Δ LLR $X^2(33)=108.716$, $p<.001$). Pilots of small aircraft reported an unusually high rate of these events during the spring of 2002. Pilots of wide-body aircraft generally reported higher and more variable rates of these events.

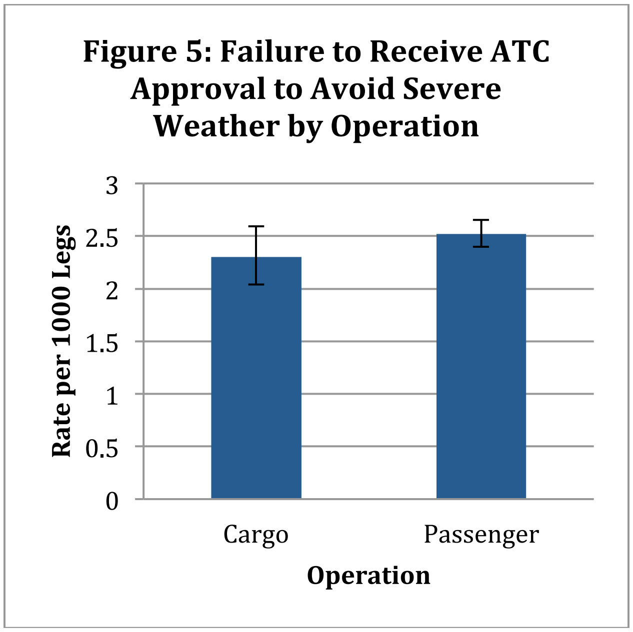


Operation

No differences in the rate of events by type of operation were observed (Wald $X^2(1)=2.178$, $p=.140$).

Interactions

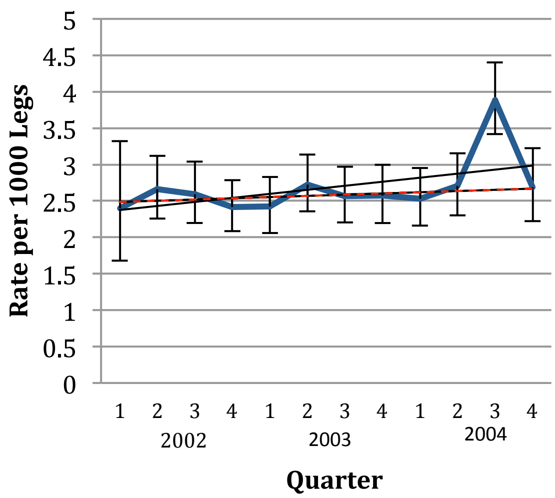
A full factorial model could not be fit to these data. A smaller model containing only 2-way interactions revealed no differences in the observed year/season patterns by type of operation (Δ LLR $X^2(5)=9.966$, $p=.076$).



WE3: Diversion to Alternate Due to Weather

Across the study period, a mean rate of 2.69 diversions due to weather per 1,000 legs was reported. The lowest rate was observed in the first quarter of 2002 (2.40) and the highest rate was observed in the third quarter of 2004 (3.89). The rate of reported diversions due to weather appears to increase across the observed quarters (b=.023, Wald $X^2(1)=7.639$, p=.006; see Figure 1).[98] However, this trend is due entirely to an anomalously high rate of reported diversions in the summer of 2004. When this quarter is omitted from the analysis, no significant trend is observed (b=.006, Wald $X^2(1)=.435$, p=.514).

Figure 1: Divert to an Alternate Due to Weather by Quarter

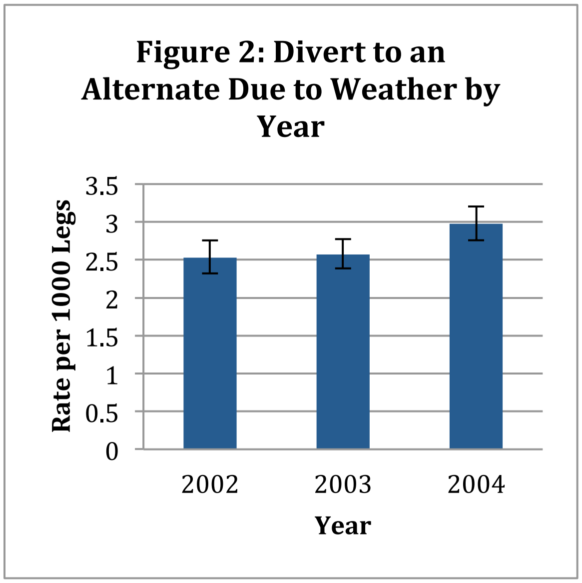


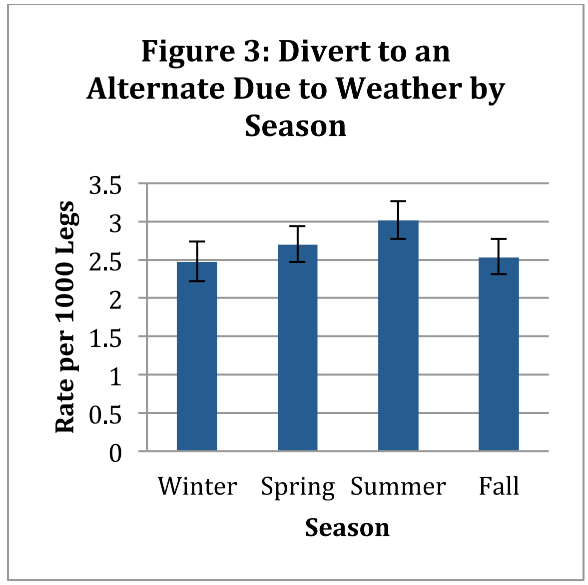
---

[98] Of the 18,371 reports obtained, 34 (0.19%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 285.49 per 1,000 legs compared to 2.95 per 1,000 legs for the other cases. When the suspect observations are included, little change in the trend was observed (b=.024, Wald $X^2(1)=8.421$, p=.004). However, the overall fit of the model decreased somewhat (Model with outliers: Pearson $X^2(18369)=25169.677$, ratio=1.370; Model without outliers: Pearson $X^2(18335)=20324.300$; ratio=1.108). However, as noted in the text, the apparent trend is due entirely to one anomalous quarter. The outlier cases were omitted from the remaining analyses.

Year/Season

The rate of reported events did not change over the observed years (Wald $X^2(2)=5.429$, p=.066; see Figure 2) and there was no significant seasonal pattern (Wald $X^2(3)=4.797$, p=.187; see Figure 3) or year by season interaction (Wald $X^2(6)=8.552$, p=.200).



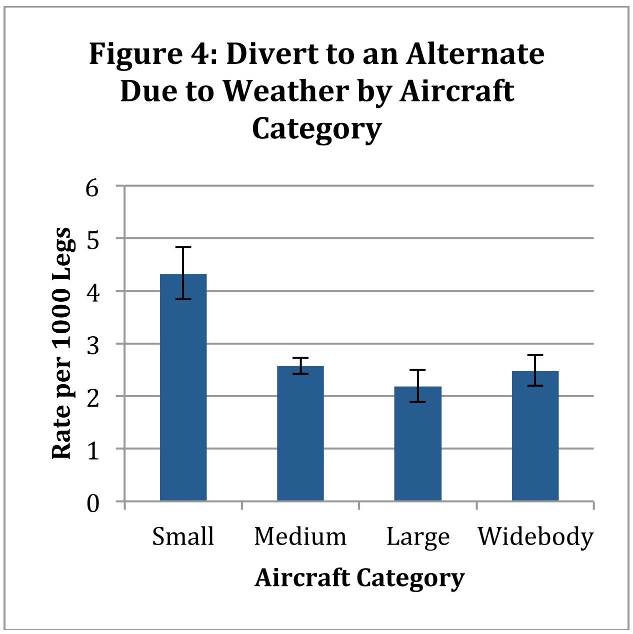


Aircraft Category

There were significant differences between aircraft categories in the rates of these events (Wald $X^2(3)=71.332$, $p<.001$). Pilots of small aircraft reported significantly higher rates of diversions due to weather than did pilots of larger aircraft (Wald $X^2(1)=42.318$, $p<.001$; see Figure 4). Pilots of large aircraft reported slightly lower rates of these events than did pilots of medium (Wald $X^2(1)=5.453$, $p=.020$) but not wide-body (Wald $X^2(1)=1.594$, $p=.207$) aircraft. No difference in the rate of these events between medium and wide-body aircraft was observed. This pattern of results is essentially unchanged if the data from summer 2004 is omitted from the analysis.

Interactions

There were significant interactions between aircraft categories and years/seasons (Δ LLR $X^2(33)=53.794$; $p=.006$). However, this effect is due entirely to the anomalously high rate reported for summer 2004. When this quarter is omitted from the analysis, no interaction with aircraft category is observed (Δ LLR $X^2(33)=37.706$, $p=.263$).



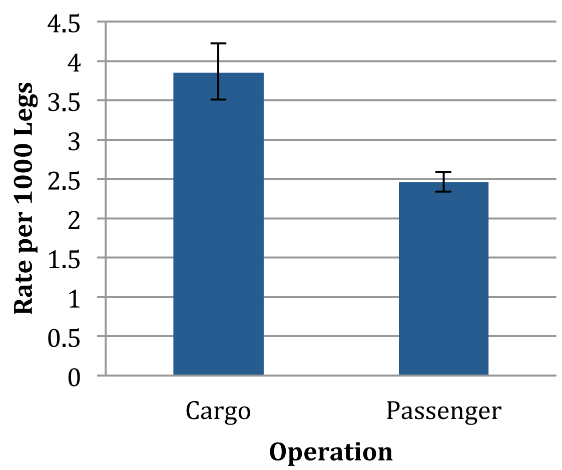
Operation

Pilots of cargo aircraft reported significantly higher rates of these events than did pilots of passenger aircraft (Wald $X^2(1)=51.851$, p<.001).

Interactions

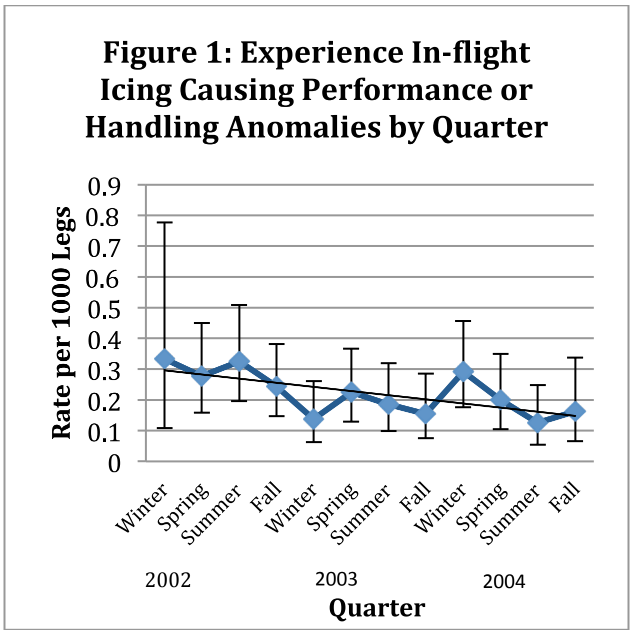
A significant interaction between year/season and type of operation was observed (Δ LLR $X^2(11)=38.100$, p<.001). When the data from summer 2004 are omitted, the interaction remains though it is considerably weaker (Wald $X^2(1)=19.924$, p=.046). Cargo and passenger aircraft demonstrated different seasonal patterns. However, these patterns were not significant when data from the summer of 2004 is omitted.

Figure 5: Divert to an Alternate Due to Weather by Operation



WE4: Experience In-flight Icing Causing Performance or Handling Anomalies

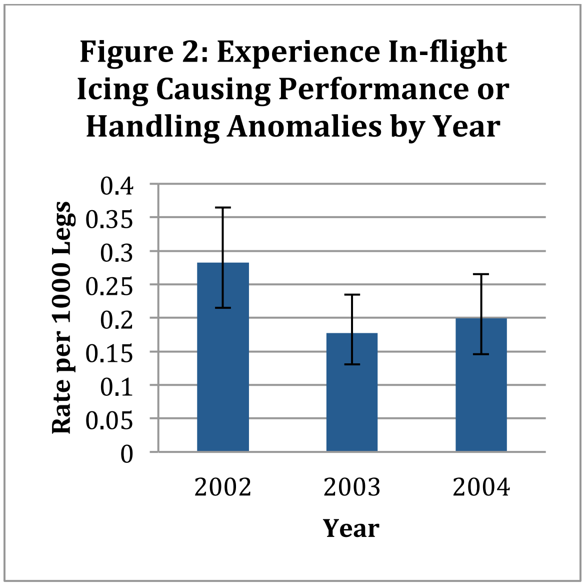
Across the study period, a mean rate of 0.22 icing events per 1,000 legs was reported. The highest rate was observed in the first quarter of 2002 (0.33) and the lowest rate was observed in the third quarter of 2004 (0.13). The reported rate of icing events causing handling or performance anomalies decreased across the observation period (b=-.054, Wald $X^2(1)$ =4.143, p=.042; see Figure 1).[99]

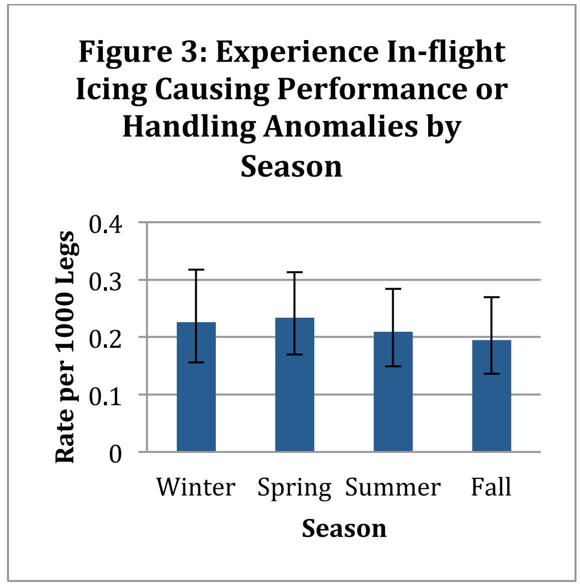


[99] Of the 18,364 reports obtained, 43 (0.23%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 111.52 per 1,000 legs compared to 0.14 per 1,000 legs for the other cases. When the suspect observations are included, little change in the trend was observed. There was a significant decrease in the rate of reported bird strikes across the observation period (b=-.091, Wald $X^2(1)$=19.863, p<.001). However, the overall fit of the model decreased (Model with outliers: Pearson $X^2(18362)$=48184.286, ratio=2.624; Model without outliers: Pearson $X^2(18319)$=17505.501; ratio=0.956). These cases were omitted from the remaining analyses.

Year/Season

The change across years was statistically significant (Wald $X^2(2)$ =6.589, p=.037; see Figure 2). The reported rates of these events were significantly lower in 2003 (Wald $X^2(1)$=5.138, p=.023) and marginally lower in 2004 (Wald $X^2(1)$=3.646, p=.056) than in 2002. There were no differences across seasons (Wald $X^2(3)$ =1.357, p=.716; see Figure 3). No year by season interaction was observed (Wald $X^2(6)$ =5.291, p=.507).



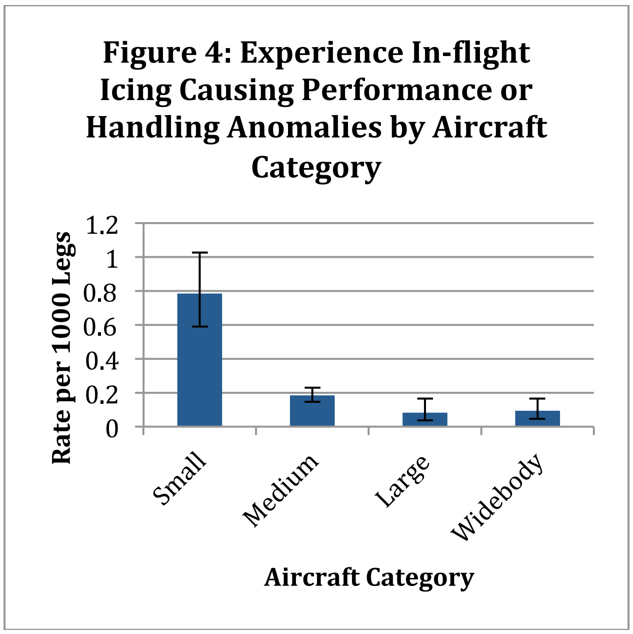


Aircraft Category

There were significant differences in the reported rates of these events between aircraft categories (Wald $X^2(3)=81.823$, $p<.001$; see Figure 4). The reported rates of these events were significantly higher for small aircraft than for larger aircraft (Wald $X^2(1)=30.475$, $p<.001$). Pilots of medium aircraft reported significantly higher rates than did pilots of larger aircraft (Wald $X^2(1)=10.937$, $p=.001$).

<u>Interactions</u>

There was no year by category interaction (Wald $X^2(6)=1.520$, $p=.958$). No other interaction models could be estimated.

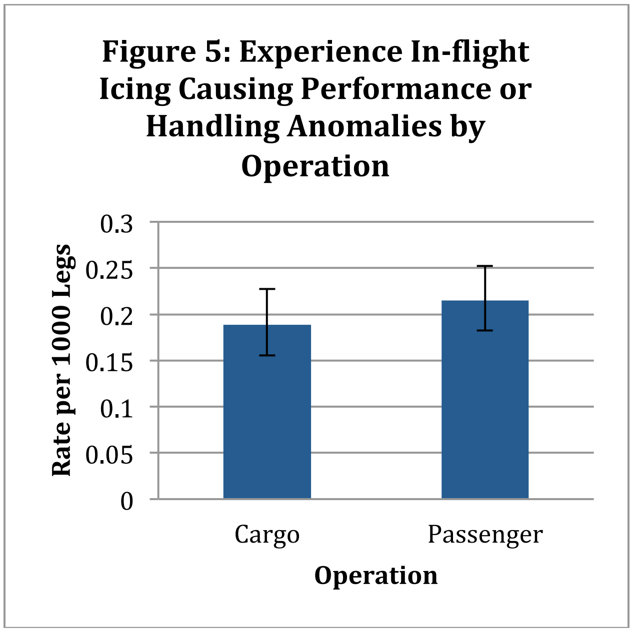


Operation

Pilots of aircraft engaged in passenger operations reported significantly higher rates of these events than did pilots of cargo aircraft (Wald $X^2(1)=10.927$, p=.001).

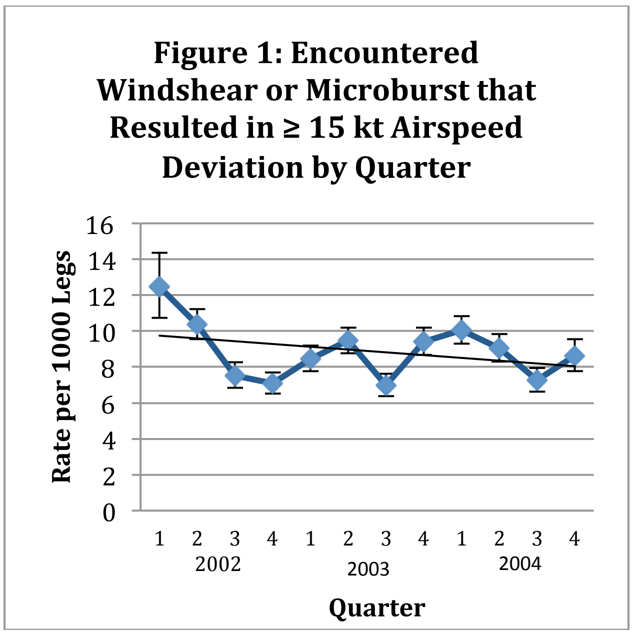
<u>Interactions</u>

The full interaction model could not be estimated. There were no interactions between operation and year/season in a smaller model containing only 2-way interaction terms (ΔLLR $X^2(11)=9.76$, p=.082).



WE5: Encountered Wind shear or Microburst that Resulted in ≥ 15 kt Airspeed Deviation

Across the study period, a mean rate of 8.61 encounters with wind shear or microbursts that resulted in airspeed deviations of 15 knots or more per 1,000 legs was reported. The lowest rate was observed in the third quarter of 2003 (6.98) and the highest rate was observed in the first quarter of 2002 (12.46). The rate of reported encounters with wind shear or microbursts did not show a significant linear trend across the observed quarters (b=-.006, Wald $X^2(1)=1.112$, p=.292; see Figure 1).[100]



[100] Of the 18,369 reports obtained, 28 (0.15%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 447.44 per 1,000 legs compared to 11.03 per 1,000 legs for the other cases. When the suspect observations are included, little change in the trend was observed (b=-.008, Wald $X^2(1)=2.195$, p=.138). However, the overall fit of the model decreased somewhat (Model with outliers: Pearson $X^2(18367)=25936.583$, ratio=1.412; Model without outliers: Pearson $X^2(18339)=22329.398$; ratio=1.218).

Year/Season

The rate of reported events did not change over the observed years (Wald $X^2(2)=1.863$, p=.394; see Figure 2). The reported rate did change across seasons (Wald $X^2(3)=50.022$, p<.001; see Figure 3). The lowest reported rates of these events were in the summer, significantly lower than in the fall (Wald $X^2(1)=11.859$, p=.001). The reported rate was lower in the fall than in the winter (Wald X2(1)=8.637, p=.003) and spring (Wald $X^2(1)=5.358$, p=.021), which did not differ (Wald $X^2(1)=1.102$, p=.294). However, the seasonal pattern varied across the years (Wald $X^2(6)=27.118$, p<.001). In particular, 2002 demonstrated a significantly different pattern. During the winter, the rate of events in 2002 was unusually high and significantly higher (p<.02) than the rate observed during the winter of 2004 (the rate in the winter of 2003 did not differ from the winter rate in either of the other years). During the fall, the rate of events in 2002 was unusually low and significantly (p<.04) lower than the rates in the fall of 2003 and 2004 (which did not differ from each other).

Figure 2: Encountered Windshear or Microburst that Resulted in ≥ 15 kt Airspeed Deviation by Year

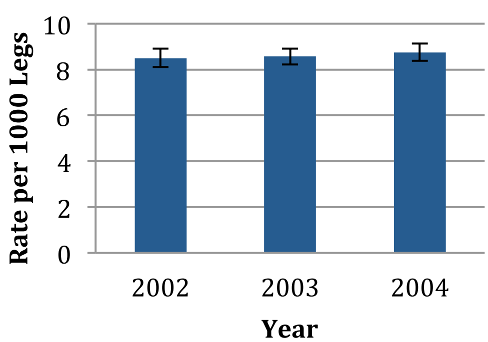
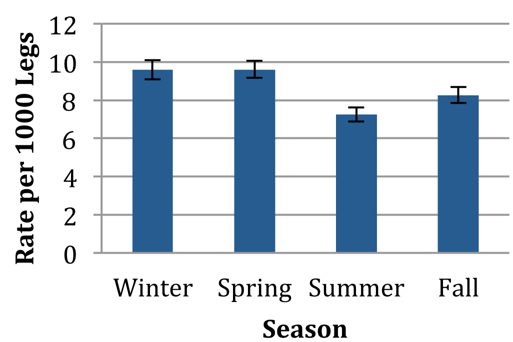


Figure 3: Encountered Windshear or Microburst that Resulted in ≥ 15 kt Airspeed Deviation by Season

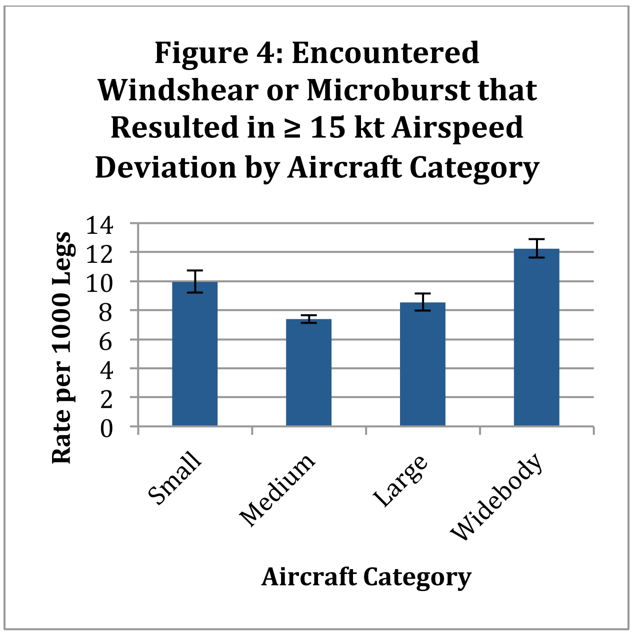


Aircraft Category

The reported rates of encounters with wind shear or microbursts that caused airspeed deviations of 15 knots or greater varied by aircraft category (Wald $X^2(3)=141.308$, $p<.001$; see Figure 4). Pilots of wide-body aircraft reported the highest rate of these events. This rate was significantly greater than the rate reported by pilots of small aircraft (Wald $X^2(1)=5.580$, $p=.018$). Pilots of large aircraft reported slightly lower rates than pilots of small aircraft. (Wald $X^2(1)=6.431$, $p=.011$). Pilots of medium aircraft reported rates lower than those of large aircraft (Wald $X^2(1)=5.705$, $p=.017$).

<u>Interactions</u>

There were no significant interactions between aircraft categories and years/seasons (Δ LLR $X^2(33)= 38.854$; $p=.223$).

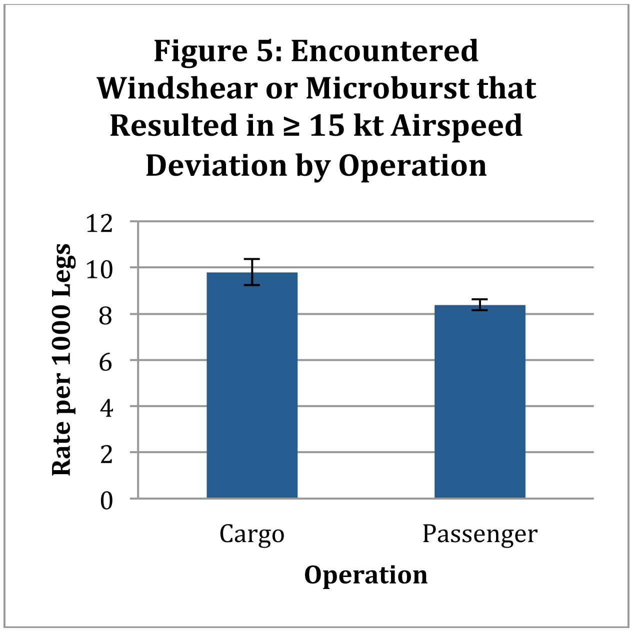


Operation

Pilots of passenger aircraft reported a significantly lower rate of these events than did pilots of cargo aircraft (Wald $X^2(1)=13.281$, $p<.001$; see Figure 5).

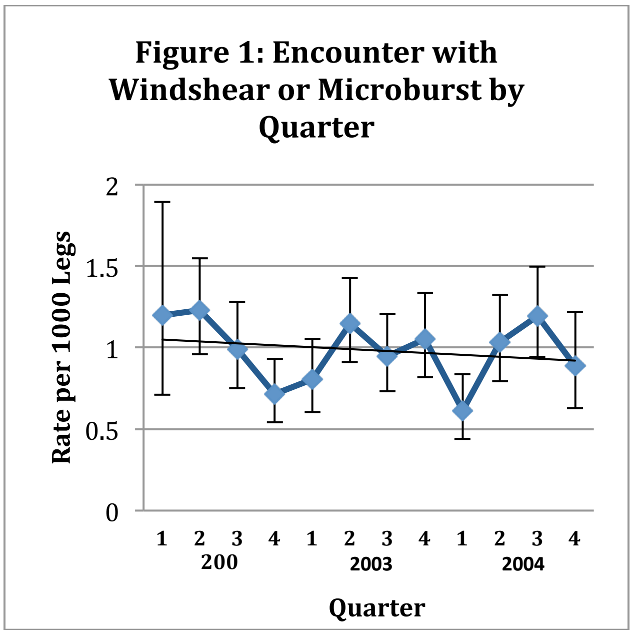
Interactions

There was a statistically significant interaction between operation and year/season (Δ LLR $X^2(11)=22.006$, $p=.024$). Pilots of cargo aircraft reported higher rates of these events than did pilots of passenger aircraft on 8/12 quarters. However, the seasonal patterns varied across years and by type of operation.



WE6: Encounter with Wind shear or Microburst that Required an Avoidance Maneuver

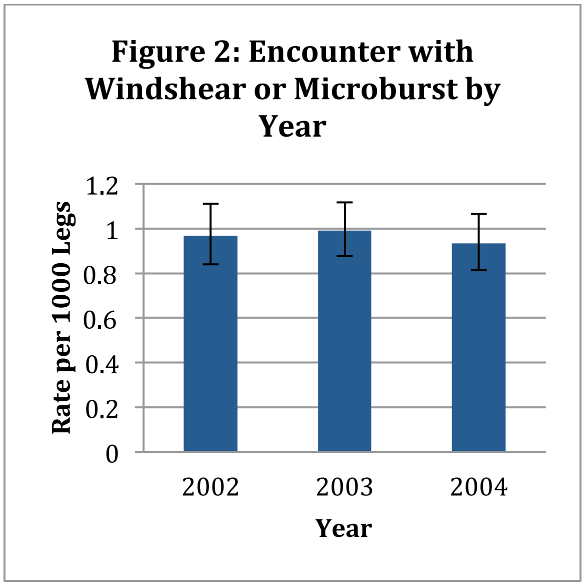
Across the study period, a mean rate of 0.97 encounters with wind shear or microbursts that required an avoidance maneuver per 1,000 legs was reported. The lowest rate was observed in the first quarter of 2004 (0.61) and the highest rate was observed in the second quarter of 2002 (1.23). The rate of reported encounters with wind shear or microbursts did not vary linearly across the observation period (b=-.005, Wald $X^2(1)=0.129$, p=.720; see Figure 1).[101]

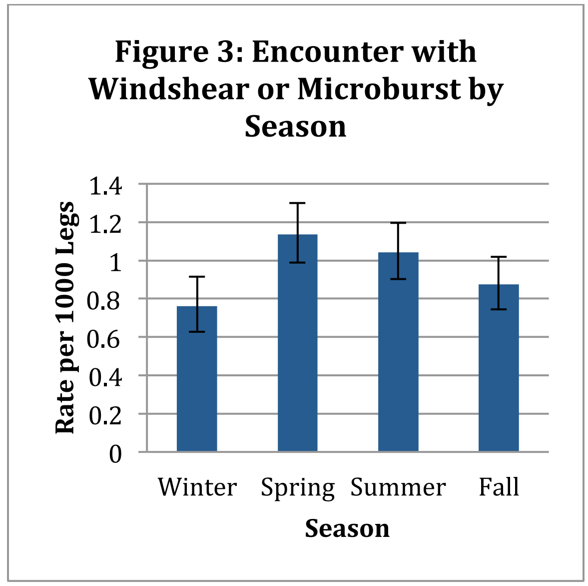


[101] Of the 18,369 reports obtained, 37 (0.20%) reported event rates that are suspect because they departed substantially (> 8 sd) from the overall distribution of events. The mean reported event rate amongst these cases was 196.69 per 1,000 legs compared to 1.10 per 1,000 legs for the other cases. When the suspect observations are included, little change in the trend was observed. The rate of reported encounters with wind shear or microbursts that required an avoidance maneuver did not show a linear trend across the observation period (b=-.008, Wald $X^2(1)=0.450$, p=.520). However, the overall fit of the model decreased (Model with outliers: Pearson $X^2(18367)=35370.228$, ratio=1.926; Model without outliers: Pearson $X^2(18330)=23725.521$; ratio=1.294). These cases were omitted from the remaining analyses.

Year/Season

No differences between years were observed (Wald $X^2(2) = 1.256$, p=.534; see Figure 2). The seasonal pattern was significant (Wald $X^2(3) = 8.426$, p=.038; see Figure 3). The rate of wind shear or microburst encounters that required an avoidance maneuver was higher in the spring and summer than in the fall and winter (Wald $X^2(1) = 7.846$, p=.005). There was no year by season interaction (Wald $X^2(6)=11.662$, p=.070).



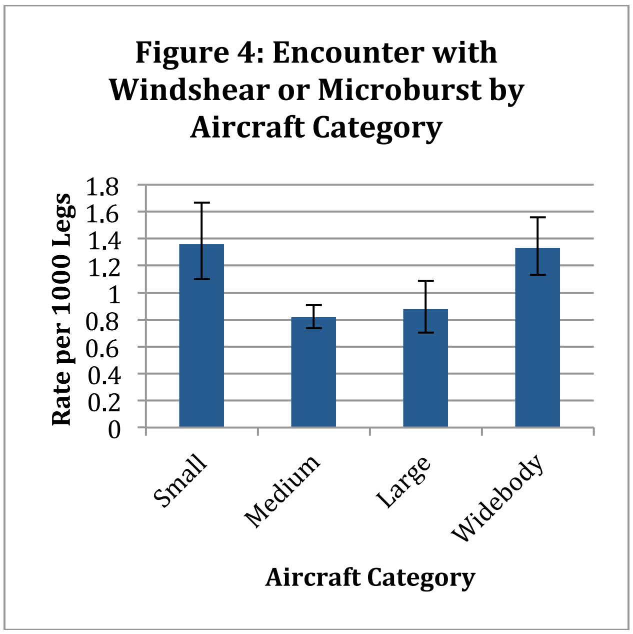


Aircraft Category

There were significant differences in the rates of these events between aircraft categories (Wald $X^2(3)=32.204$, $p<.001$). Pilots of small and wide-body aircraft reported significantly higher rates of these events than did pilots of medium and large aircraft (Wald $X^2(1) = 20.064$, $p<.001$; see Figure 4). No other differences between aircraft categories were observed.

<u>Interactions</u>

The full factorial model could not be estimated. A smaller model containing only 2-way interactions was estimated. No interactions between aircraft category and year/season were observed (ΔLLR $X^2(15)=19.96$, $p=.173$).

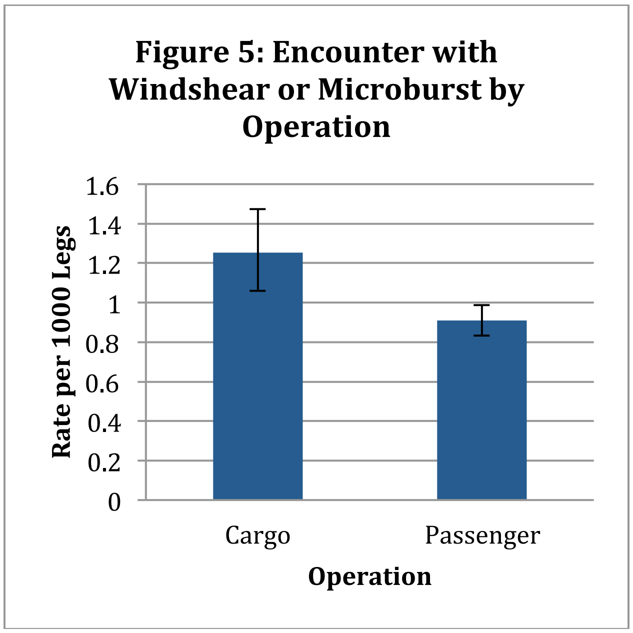


Operation

Pilots of cargo aircraft reported higher rates of these events than did pilots of passenger aircraft (Wald $X^2(1)=10.476$, $p=.001$).

Interactions

No interactions between operation and year/season were observed (ΔLLR $X^2(11)=14.526$, $p=.205$).



Appendix N. Results of the Study with the CAST/JIMDAT

The Civil Aviation Safety Team (CAST) had recommended many Safety Enhancements (SEs) for implementation by the industry to mitigate the occurrences of problems they found were common in aviation accidents. The CAST organized the Joint Implementation Measurement Data Analyses Team (JIMDAT) to monitor and measure the effectiveness of these SEs and they expressed interest in the possibility of using NAOMS for this purpose.

A working group was formed composed of members of the JIMDAT and the NAOMS Team to evaluate the use of NAOMS to assess the efficacies and the extent of implementation of SEs. This group decided that they should focus their experiment on the SEs that were concerned specifically with training and procedures and that these could be appropriately addressed in Section C of the pilot-survey questionnaire, the section that addresses special topics. The CAST and the JIMDAT specified the topics they wanted addressed by the questions in the new Section C of the survey instrument. They related to the following:

- Avionics and Navaids
- SOPs and Current Practices
- Training
- Flight Deck Innovations
- Air Carrier Safety Culture

The joint working group designed a new survey instrument that incorporated a revised set of questions for Section C on these issues. The specific questions used in this small survey for the JIMDAT are provided in Appendix E as part of the air-carrier pilot questionnaire and replaced the Section C concerned with ICAC that had been part of the survey of the 25,000 air-carrier pilots conducted from 2001 to 2004. Telephone interviews were conducted with 1,194 additional air-carrier pilots over a three-month period ending in December 2004 to collect the data for this special study. Of these, 18 interviews were deleted due, primarily, to outlier exceedances; 26 interviews were excluded because they were business- aircraft pilots, and 2 interviewees stopped midway through their interviews. The responses from the remaining 1,145 pilots to the questions in Section A were used for normalization of their responses to the questions in Section C.

This Appendix documents the complete set of the results for all of the questions in Section C used for the JIMDAT study just as they were presented to the full JIMDAT and to the CAST.

PROCEDURES AND TRAINING

Section A Results: Basic Activity Measures

JIMDAT SECTION A RESULTS: BASIC ACTIVITY MEASURES				
Question	Question	Response	Response Value	Response %
A1	During the last 60 days, how many hours did you fly as a crewmember on a commercial aircraft?	Low Value	2	na
		Mean Value	108	na
		High Value	320	na
A2	During the last 60 days, how many legs did you fly as a crewmember on a commercial aircraft?	Low Value	1	na
		Mean Value	41	na
		High Value	496	na
A2.1	During the last 60 days, how many of the legs you flew involved taking off or landing at an airport outside the United States?	Low Value	1	na
		Mean Value	6	na
		High Value	80	na
A4-A6	During the last 60 days, what percent of the hours did you fly as a crewmember on flights with ….?	Revenue Passengers	na	73%
		Cargo w/o passengers	na	25%
		No passenger or cargo	na	2%

JIMDAT SECTION A RESULTS: BASIC ACTIVITY MEASURES				
Question	**Question**	**Response**	**Response Value**	**Response %**
A7.1	Which of the following three categories best describes the number of airplanes currently operated by your airline? Please do not include airplanes operated by code-share partners?	350 or more	769	67%
		150 to 349	200	17%
		149 or less	174	15%
		Don't Know	7	1%
		Total	1150	100%
A7a	During the last 60 days, did you fly a commercial aircraft as a: *	Captain?	631	55%
		First Officer?	505	44%
		Flight Engineer?	56	5%
		Relief Pilot?	95	8%
	* (Total % exceeds 100% since the responses are not mutually exclusive)	Other Position?	29	2%
Constructed Variable	Aircraft Size Category	Wide body	402	35%
		Large	214	19%
		Medium	501	44%
		Small	33	3%
		Total	1150	100%

SECTION C RESULTS:
GPWS Questions

JIMDAT SECTION C RESULTS: GPWS QUESTIONS				
Question Number	**Question**	**Response**	**Response Value**	**Respons e %**
JD1	Is the aircraft you flew (most) during the last 60 days equipped with GPWS?	Yes	1097	95%
		No	52	5%
		Refuse	0	0%
		Don't know	1	0%
		Total	1150	100%
If No, Refused or Don't Know, Skip to JD2				
→» JD1a	Is it equipped with a terrain display, such as you find in an enhanced GPWS, or Terrain Avoidance Warning System, also known as TAWS?	Yes	1011	92%
		No	83	8%
		Refuse	0	0%
		Don't know	3	0%
		Total	1097	100%
If No, Refused or Don't Know, Skip to JD2				
→» JD1b	Does your airline require the terrain display to be selected during takeoff at specific airports?	Yes	687	68%
		No	304	30%
		Refuse	0	0%
		Don't know	20	2%
		Total	1011	100%
If No, Refused or Don't Know, Skip to JD2				

JIMDAT SECTION C RESULTS: GPWS QUESTIONS				
Question Number	**Question**	**Response**	**Response Value**	**Respons e %**
→» **JD1c**	Does your airline require the terrain display to be selected during descent and landing?	Yes	584	85%
		No	99	14%
		Refuse	0	0.0%
		Don't know	4	1%
		Total	687	100%
→» **JD1d**	For times that terrain display is not required, do you usually use it during takeoff?	Yes	535	78%
		No	150	22%
		Refuse	0	0%
		Don't know	2	0%
		Total	687	100%
→» **JD1e**	For times that terrain display is not required, do you usually use it during descent and landing?	Yes	540	79%
		No	145	21%
		Refuse	1	0%
		Don't know	1	0%
		Total	687	100%

JIMDAT SECTION C RESULTS: GPWS QUESTIONS				
Question Number	**Question**	**Response**	**Response Value**	**Respons e %**
→» JD1f	Has the terrain display experienced a map shift on any aircraft on which you were a crew member?	Yes	103	15%
		No	582	85%
		Refuse	0	0%
		Don't know	2	0%
		Total	687	100%
JD2	During the last 60 days, how many times did an aircraft on which you were a crewmember experience a ground proximity warning?	0	1091	95%
		1	43	4%
		2 or more	15	1%
		Refuse	0	0%
		Don't know	1	0%
		Total	1150	100%
If zero, Refused or Don't Know skip to JD3				
→» JD2a	Was (this warning/ the most recent of these warnings) valid?	Yes	32	55%
		No	25	43%
		Refuse	0	0%
		Don't know	1	2%
		Total	58	100%
If No, Refused or Don't Know, Skip to JD3				

JIMDAT SECTION C RESULTS: GPWS QUESTIONS				
Question Number	**Question**	**Response**	**Response Value**	**Respons e %**
→» JD2b	During this (most recent) warning, did you see the approaching terrain on the terrain display before you heard the aural warning?	Yes	22	69%
		No	10	31%
		Refuse	0	0%
		Don't know	0	0%
		Total	32	100%
If No, Refused or Don't Know, Skip to JD3				
JD3	During the last 60 days, how many times did an aircraft on which you were a crewmember receive a Minimum Safe Altitude Warning Alert, also known as an MSAW or an altitude awareness call from an A.T.C controller?	0	1138	99%
		1	12	1%
		Refuse	0	0%
		Don't know	0	0%
		Total	1150	100%
If zero, refused or don't know skip to JD4				
→» JD3a	(During the most recent of these events,) What did your aircraft do in response to the warning?	12 comments were received		

JIMDAT SECTION C RESULTS: GPWS QUESTIONS				
Question Number	**Question**	**Response**	**Response Value**	**Respons e %**
→» JD3b	(During this most recent A.T.C. warning event,) Did the aircraft have an enhanced G.P.W.S. or T.A.W.S. (taws) installed?	Yes	8	67%
		No	4	33%
		Refuse	0	0%
		Don't know	0	0%
		Total	12	100%
If No, Refused or Don't Know, Skip to JD4				
→» JD3b1	Did your aircraft also receive a ground proximity warning from this system?	Yes	0	0%
		No	8	100%
		Refuse	0	0%
		Don't know	0	0%
		Total	8	100%

SECTION C RESULTS:
Approach Related Questions

JIMDAT SECTION C RESULTS: APPROACH RELATED QUESTIONS				
Question Number	**Question**	**Response**	**Response Value**	**Response %**
JD4	How many times in the last 60 days, did an aircraft on which you were a crewmember fly a non-precision approach?	0	448	39%
		1	164	14%
		2	177	15%
		3	95	8%
		4 to 9	181	16%
		10 or more	83	7%
		Refuse	0	0%
		Don't know	2	0%
		Total	1150	100%
If zero, Refused or Don't Know skip to JD8				
→»JD4a	Was this non-precision approach flown in I.M.C? / How many of these non-precision approaches were flown in I.M.C?)	0	242	35%
		1	185	26%
		2	127	18%
		3 or more	146	21%
		Refuse	0	0%
		Don't know	0	0%
		Total	700	100%

JIMDAT SECTION C RESULTS: APPROACH RELATED QUESTIONS				
Question Number	**Question**	**Response**	**Response Value**	**Response %**
JD5	How many times in the last 60 days did an aircraft on which you were a crewmember fly an un-stabilized non-precision approach where the aircraft was not in landing configuration, on airspeed, or on glide-slope by 1,000 feet I.M.C or 500 feet V.M.C?	0	658	94%
		1	27	4%
		2 or more	14	2%
		Refuse	0	0%
		Don't know	1	0%
		Total	700	100%
If zero, Refused or Don't Know skip to JD6				
→»JD5a	(During the most recent un-stabilized non precision approach,) What factors contributed to the inability to conduct a stabilized approach?	Fourty three comments were provided		

JIMDAT SECTION C RESULTS: APPROACH RELATED QUESTIONS				
Question Number	**Question**	**Response**	**Response Value**	**Response %**
JD6	During the last 60 days, did an aircraft on which you were a crewmember have the choice between flying a constant angle approach or step-down non-precision approach?	Yes	260	37%
		No	431	62%
		Refuse	0	0%
		Don't know	9	1%
		Total	700	100%
If No, Refused or Don't Know, Skip to JD7				
→»JD6a	Which did you choose most often, the constant angle approach or the step-down non-precision approach?	Constant Angle	227	87%
		Step Down	24	9%
		Roughly Equal	9	3%
		Refuse	0	0%
		Don't know	0	0%
		Total	260	100%

JIMDAT SECTION C RESULTS: APPROACH RELATED QUESTIONS				
Question Number	Question	Response	Response Value	Response %
JD7	During the last 60 days, how many times did an aircraft on which you were a crewmeber fly a non-precision approach to a runway when glide-slope information was available to you?	0	547	78%
		1	43	6%
		2	24	3%
		3	17	2%
		4 or more	63	9%
		Refuse	2	0%
		Don't know	4	1%
		Total	700	100%
If zero, Refused or Don't Know skip to JD8				
→»JD7a	During (this/the most recent) non-precision approach, did you use the glide-slope information?	Yes	103	70%
		No	44	30%
		Refuse	0	0%
		Don't know	0	0%
		Total	147	100%

JIMDAT SECTION C RESULTS: APPROACH RELATED QUESTIONS				
Question Number	**Question**	**Response**	**Response Value**	**Response %**
JD8	(Is the aircraft you fly/Are any of the aircraft you fly) LNAV / VNAV (L-nav/ V-nav) capable?	Yes	927	80%
		No	215	19%
		Refuse	0	0%
		Don't know	8	1%
		Total	1150	100%
If No, Refused or Don't Know, Skip to JD9				
→»JD8a	Does your airline ever require pilots to use LNAV / VNAV (L-nav/V-nav) to fly constant angle approaches?	Yes	483	52%
		No	434	47%
		Refuse	0	0%
		Don't know	10	1%
		Total	927	100%

JIMDAT SECTION C RESULTS: APPROACH RELATED QUESTIONS				
Question Number	**Question**	**Response**	**Response Value**	**Response %**
→»JD8a1	In the last 60 days, how many times did an aircraft on which you were a crewmember use LNAV / VNAV (L-nav/V-nav) to fly constant angle approaches?	0	160	33%
		1	82	17%
		2	73	15%
		3	43	9%
		4	21	4%
		5	22	5%
		6 or more	80	17%
		Refuse	0	0%
		Don't know	2	0%
		Total	483	100%
→»JD8b	During the last 60 days, how many times did an aircraft on which you were a crewmember not fly an LNAV / VNAV (L-nav/V-nav) approach when that option was available?	0	387	80%
		1	11	2%
		2	8	2%
		3 or more	61	13%
		Refuse	1	0%
		Don't know	15	3%
		Total	483	100%
If zero,Refused or Don't Know skip to JD9				

JIMDAT SECTION C RESULTS: APPROACH RELATED QUESTIONS				
Question Number	Question	Response	Response Value	Response %
→»JD8b1	Please explain why the LNAV / VNAV (L-nav/V-nav) approach wasn' t flown (during the most recent time that it was available).	Eighty one comments were provided		
JD9	During the last 60 days, was an aircraft on which you were a crewmember equipped to meet Required Navigation Performance standards, sometimes called R.N.P?	Yes	804	70%
		No	287	25%
		Refuse	0	0%
		Don't know	59	5%
		Total	1150	100%
If No, Refused or Don't Know, Skip to JD10				
→»JD9a	Does your airline choose to use R.N.P?	Yes	724	90%
		No	60	7%
		Refuse	0	0%
		Don't know	20	3%
		Total	804	100%
If No, Refused or Don't Know, Skip to JD10				

JIMDAT SECTION C RESULTS: APPROACH RELATED QUESTIONS				
Question Number	Question	Response	Response Value	Response %
→»JD9b	How many times in the last 60 days did an aircraft on which you were a crewmember fly an R.N.P approach?	0	511	71%
		1	53	7%
		2	48	7%
		3 or more	88	12%
		Refuse	1	0%
		Don't know	23	3%
		Total	724	100%
→»JD9c	During the last 60 days, how many times did any aircraft on which you were a crewmember not fly an R.N.P approach when that option was available?	0	606	84%
		1	14	2%
		2 or more	63	9%
		Refuse	0	0%
		Don't know	41	6%
		Total	724	100%
If zero, Refused or Don't Know skip to JD10				

JIMDAT SECTION C RESULTS: APPROACH RELATED QUESTIONS				
Question Number	**Question**	**Response**	**Response Value**	**Response %**
→»JD9c1	Please explain why the R.N.P. approach was not flown (most recent time that it was available).	Seventy-five comments were provided		
If JD4 = 0, Refused or Don't Know skip to JH11				
JD10	During the last 60 days, how many times did an aircraft on which you were a crewmember fly a non-precision approach into an airport without D.M.E.?	0	761	81%
		1	64	7%
		2	42	4%
		3 or more	62	7%
		Refuse	0	0%
		Don't know	12	1%
		Total	941	100%
If zero, Refused or Don't Know skip to JD11				
→»JD10a	During (this event/the most recent of these events), would D.M.E have improved your ability to land safely?	Yes	107	64%
		No	61	36%
		Refuse	0	0%
		Don't know	0	0%
		Total	168	100%

JIMDAT SECTION C RESULTS: APPROACH RELATED QUESTIONS				
Question Number	**Question**	**Response**	**Response Value**	**Response %**
JD11	During the last 60 days, how many times did an aircraft on which you were a crewmember fly an instrument approach into an airport where glide-slope or other ground based vertical angle guidance information was unavailable?	0	550	48%
		1	171	15%
		2	158	14%
		3 or more	270	23%
		Refuse	0	0%
		Don't know	1	0%
		Total	1150	100%
If zero, Refused or Don't Know kip to JD 12				
→»JD11a	During (this approach/the most recent of these approaches), was D.M.E used to calculate the rate of descent for landing?	Yes	328	55%
		No	267	45%
		Refuse	0	0%
		Don't know	4	0%
		Total	599	100%

JIMDAT SECTION C RESULTS: APPROACH RELATED QUESTIONS				
Question Number	**Question**	**Response**	**Response Value**	**Response %**
JD12	During the last 60 days, how many times did an aircraft on which you were a crewmember land on a runway without VASI (vasi) or PAPI (papi)?	0	459	40%
		1	89	8%
		2	124	11%
		3 or more	456	40%
		Refuse	0	0%
		Don't know	22	2%
		Total	1150	100%
If zero, Refused or Don't Know skip to JD13				
→»JD12a	During the most recent of these events) would VASI (vasi) or PAPI (papi) have improved the aircraft' s ability to land safely?	Yes	518	77%
		No	147	22%
		Refuse	1	0%
		Don't know	3	1%
		Total	669	100%

SECTION C RESULTS:
SOP Related Questions

JIMDAT SECTION C RESULTS: SOP RELATED QUESTIONS				
Question Number	**Question**	**Response**	**Response Value**	**Response %**
JD13	Do your airline's written S.O.Ps include Controlled Flight into Terrain prevention, sometimes called C-FIT?	Yes	1101	96%
		No	34	3%
		Refuse	0	0%
		Don't know	15	1%
		Total	1150	100%
JD14	Do your airline's written S.O.Ps talk about how to avoid circumstances that could lead to an in-flight loss of control?	Yes	1089	95%
		No	46	4%
		Refuse	0	0%
		Don't know	15	1%
		Total	1150	100%
JD15	Do your airline's written S.O.P.s talk about how to perform recovery from unusual attitudes and departure from controlled flight?	Yes	1109	97%
		No	37	3%
		Refuse	0	0%
		Don't know	4	0%
		Total	1150	100%

JIMDAT SECTION C RESULTS: SOP RELATED QUESTIONS				
Question Number	**Question**	**Response**	**Response Value**	**Response %**
JD16	Do your airline' s written S.O.Ps talk about how to avoid approach and landing accidents?	Yes	1107	96%
		No	36	3%
		Refuse	1	0%
		Don't know	6	1%
		Total	1150	100%
JD17	Do your airline's written S.O.Ps talk about how to fly non-precision approaches?	Yes	1142	99%
		No	6	1%
		Refuse	0	0%
		Don't know	2	0%
		Total	1150	100%
JD18	Do your airline's written S.O.Ps require the use of constant angle non-precision approaches when that option is available?	Yes	755	66%
		No	367	32%
		Refuse	1	0%
		Don't know	27	2%
		Total	1150	100%

JIMDAT SECTION C RESULTS: SOP RELATED QUESTIONS				
Question Number	**Question**	**Response**	**Response Value**	**Response %**
JD19	Do your airline's written S.O.Ps talk about how to respond to E.G.P.W.S warnings?	Yes	1131	98%
		No	17	2%
		Refuse	0	0%
		Don't know	2	0%
		Total	1150	100%

SECTION C RESULTS:
Recurrent Training

JIMDAT SECTION C RESULTS: Recurrent Training				
Question Number	**Question**	**Response**	**Response Value**	**Response %**
JD20M	In what month did you receive your most recent recurrent training?	Quarter 1 (1-3)	207	18%
		Quarter 2 (4-6)	356	31%
		Quarter 3 (7-9)	328	29%
		Quarter 4 (10-12)	254	22%
		Refuse	0	0%
		Don't know	5	0%
		Total	1150	100%
JD20Y	In what year did you receive your most recent recurrent training?	2004	1051	91%
		2003	84	7%
		2002	7	1%
		2001 or older	6	1%
		Refuse	0	0%
		Don't know	2	0%
		Total	1150	100%

JIMDAT SECTION C RESULTS: Recurrent Training				
Question Number	**Question**	**Response**	**Response Value**	**Response %**
JD21	Did your most recent recurrent training talk about basic airmanship?	Yes	891	77%
		No	239	21%
		Refuse	0	0%
		Don't know	19	2%
		Total *	1149	100%
(* At this point in the interview, a pilot stopped reducing the total to 1149 from 1150 pilot responses)				
JD21a	Did your most recent recurrent training talk about normal approach procedures?	Yes	1092	95%
		No	54	5%
		Refuse	0	0%
		Don't know	3	0%
		Total	1149	100%
JD21b	Did your most recent recurrent training talk about approach briefings?	Yes	1095	95%
		No	53	5%
		Refuse	0	0%
		Don't know	1	0%
		Total	1149	100%

JIMDAT SECTION C RESULTS: Recurrent Training				
Question Number	Question	Response	Response Value	Response %
JD21c	Did your most recent recurrent training talk about criteria for initiating go-around and missed approaches?	Yes	1109	97%
		No	38	3%
		Refuse	0	0%
		Don't know	2	0%
		Total	1149	100%
JD21d	Did your most recent recurrent training talk about go-around and missed approach execution?	Yes	1114	97%
		Refuse	33	3%
		Refuse	0	0%
		Don't know	2	0%
		Total	1149	100%
JD21e	Did your most recent recurrent training talk about emergency or abnormal conditions procedures?	Yes	1134	99%
		No	14	1%
		Refuse	0	0%
		Don't know	1	0%
		Total	1149	100%

JIMDAT SECTION C RESULTS: Recurrent Training				
Question Number	**Question**	**Response**	**Response Value**	**Response %**
JD22	Have you received C-FIT (C-fit) prevention training from your airline?	Yes	1102	96%
		No	45	4%
		Refuse	0	0%
		Don't know	1	0%
		Total *	1148	100%
(* At this point in the interview, another pilot stopped reducing the total from 1149 to 1148 pilot responses)				
JD22aM	In what month did you receive your most recent C-FIT (C-fit) prevention training?	Quarter 1 (1-3)	217	20%
		Quarter 2 (4-6)	344	31%
		Quarter 3 (7-9)	281	25%
		Quarter 4 (10-12)	244	22%
		Refuse	0	0%
		Don't know	16	1%
		Total	1102	100%

JIMDAT SECTION C RESULTS: Recurrent Training				
Question Number	**Question**	**Response**	**Response Value**	**Response %**
JD22aY	In what year did you receive your most recent C-FIT (C-fit) prevention training?	2004	909	82%
		2003	139	13%
		2002	26	2%
		2001 or older	20	2%
		Refuse	0	0%
		Don't know	8	1%
		Total	1102	100%
JD22b	Did your most recent C-FIT (C-fit) prevention training talk about minimum obstruction clearance altitudes or MOCA?	Yes	869	79%
		No	216	20%
		Refuse	0	0%
		Don't know	17	1%
		Total	1102	100%
JD22c	Did your most recent C-FIT (C-fit) prevention training talk about minimum enroute altitudes or M.E.A?	Yes	869	79%
		No	224	20%
		Refuse	0	0%
		Don't know	9	1%
		Total	1102	100%

JIMDAT SECTION C RESULTS: Recurrent Training				
Question Number	**Question**	**Response**	**Response Value**	**Response %**
JD22d	Did your most recent C-FIT (C-fit) prevention training talk about grid MORAs?	Yes	666	60%
		No	417	38%
		Refuse	0	0%
		Don't know	19	2%
		Total	1102	100%
JD22e	Did your most recent C-FIT (C-fit) prevention training talk about G.P.W.S or E.G.P.W.S?	Yes	1083	98%
		No	18	2%
		Refuse	0	0%
		Don't know	1	0%
		Total	1102	100%
JD22f	Did your most recent C-FIT prevention training talk about escape maneuvers in response to G.P.W.S or E.G.P.W.S warnings?	Yes	1086	98%
		No	16	2%
		Refuse	0	0%
		Don't know	0	0%
		Total	1102	100%

JIMDAT SECTION C RESULTS: Recurrent Training				
Question Number	**Question**	**Response**	**Response Value**	**Response %**
JD22g	Did your most recent C-FIT (C-fit) prevention training talk about drift down procedures after engine failure?	Yes	758	69%
		No	334	30%
		Refuse	0	0%
		Don't know	10	1%
		Total	1102	100%
JD22h	Did your most recent C-FIT (C-fit) prevention training talk about maintaining situational awareness?	Yes	1078	98%
		No	22	2%
		Refuse	0	0%
		Don't know	2	0%
		Total	1102	100%
JD22i	Did your most recent C-FIT (C-fit) prevention training talk about cockpit resource management, or C.R.M as it relates to C-FIT (C-fit) recovery?	Yes	1031	94%
		No	67	6%
		Refuse	0	0%
		Don't know	4	0%
		Total	1102	100%

JIMDAT SECTION C RESULTS: Recurrent Training				
Question Number	**Question**	**Response**	**Response Value**	**Response %**
JD22j	How would you rate the quality of the most recent C-FIT (C-fit) prevention training you received from your airline?	Excellent	601	55%
		Good	436	40%
		Fair	57	5%
		Poor	8	1%
		Very Poor	0	0%
		Refuse	0	0%
		Don't know	0	0%
		Total	1102	100%
JD23	Did you receive training specifically in upset recovery from your airline?	Yes	993	87%
		No	152	13%
		Refuse	0	0%
		Don't know	3	0%
		Total	1148	100%
If No, Refused or Don't Know, Skip to JD24				

JIMDAT SECTION C RESULTS: Recurrent Training				
Question Number	**Question**	**Response**	**Response Value**	**Response %**
JD23aM	In what month did you receive your most recent training in upset recovery?	Quarter 1 (1-3)	200	20%
		Quarter 2 (4-6)	294	30%
		Quarter 3 (7-9)	256	26%
		Quarter 4 (10-12)	219	22%
		Refuse	0	0%
		Don't know	24	2%
		Total	993	100%
JD23aY	In what year did you receive your most recent training in upset recovery?	2004	714	72%
		2003	191	19%
		2002	46	5%
		2001 or older	34	3%
		Refuse	2	0%
		Don't know	6	1%
		Total	993	100%

JIMDAT SECTION C RESULTS: Recurrent Training				
Question Number	**Question**	**Response**	**Response Value**	**Response %**
JD23b	Was this training received in a simulator, in a ground school, or both?	Simulator	507	51%
		Ground School	23	2%
		Both	462	46%
		Refuse	0	0%
		Don't know	1	1%
		Total	993	100%
JD23c	How would you rate the quality of the upset recovery training you received? Would you say it was (READ CATEGORIES)?	Excellent	527	53%
		Good	384	39%
		Fair	75	8%
		Poor	5	1%
		Very Poor	1	0%
		Refuse	0	0%
		Don't know	1	0%
		Total	993	100%

JIMDAT SECTION C RESULTS: Recurrent Training				
Question Number	**Question**	**Response**	**Response Value**	**Response %**
JD24	Does your airline provide training in Cockpit or Crew Resource Management, sometimes called C.R.M?	Yes	1133	98%
		No	14	2%
		Refuse	0	0%
		Don't know	1	0%
		Total	1148	100%
If No, Refused or Don't Know, Skip to JD25				
→»JD24a	Have you received this C.R.M training?	Yes	1132	100%
		No	1	0%
		Refuse	0	0%
		Don't know	0	0%
		Total	1133	100%
If No, Refused or Don't Know, Skip to JD25				
→»JD24b	Did this C.R.M. training change how you manage the flight deck?	Yes	699	62%
		No	428	38%
		Refuse	0	0%
		Don't know	5	0%
		Total	1132	100%

JIMDAT SECTION C RESULTS: Recurrent Training				
Question Number	**Question**	**Response**	**Response Value**	**Response %**
→»JD24c	Do you have suggestions for how the C.R.M training might be improved?	Yes	186	17%
		No	943	83%
		Refuse	2	0%
		Don't know	1	0%
		Total	1132	100%
If No, Refused or Don't Know, Skip to JD25				
→»JD24d	What suggestions do you have?	On hundred and eighty six comments were pvided		

SECTION C RESULTS:
Safety Reporting

JIMDAT SECTION C RESULTS: Safety Reporting				
Question Number	**Question**	**Response**	**Response Value**	**Response %**
JD25	Does your airline have a no-fault missed approach or go-around policy?	Yes	1076	94%
		No	55	5%
		Refused	0	0%
		Don't know	17	1%
		Total	1148	100%
If Yes, Refused or Don't Know, Skip to JD26				
→»JD25a	Would you favor the institution of such a policy, oppose it, or neither favor nor oppose it?	Favor	35	64%
		Oppose	6	11%
		Neither	14	25%
		Refused	0	0%
		Don't know	0	0%
		Total	55	100%
JD26	During the last 60 days did you perform a missed approach or go around?	Yes	193	17%
		No	955	83%
		Refused	0	0%
		Don't know	0	0%
		Total	1148	100%
If No, Refused or Don't Know, Skip to JD27				

Question Number	Question	Response	Response Value	Response %
→»JD26a	Did you receive any feedback from your airline regarding this missed approach or go around?	Yes	8	4%
		No	185	96%
		Refused	0	0%
		Don't know	0	0%
		Total	193	100%
If No, Refused or Don't Know, Skip to JD27				
→»JD26b	Was that feedback positive, negative, or both positive and negative?	Positive	7	87%
		Negative	0	0%
		Neither	1	13%
		Refused	0	0%
		Don't know	0	0%
		Total	8	100%
JD27	Does your airline participate in the safety reporting program called ASAP also known as the Aviation Safety Action Program?	Yes	992	86%
		No	96	9%
		Refused	0	0%
		Don't know	60	5%
		Total	1148	100%
If No, Refused or Don't Know, Skip to JD28				

JIMDAT SECTION C RESULTS: Safety Reporting				
Question Number	**Question**	**Response**	**Response Value**	**Response %**
→»JD27a	Have you been briefed on this ASAP program?	Yes	921	92%
		No	63	6%
		Refused	0	0%
		Don't know	8	1%
		Total	992	100%
→»JD27b	Were you told about the general purpose of the ASAP program?	Yes	954	96%
		No	31	3%
		Refused	0	0%
		Don't know	7	1%
		Total	992	100%
→»JD27c	Were you told how to submit an ASAP report?	Yes	917	92%
		No	64	7%
		Refused	0	0%
		Don't't know	11	1%
		Total	992	100%

JIMDAT SECTION C RESULTS: Safety Reporting				
Question Number	**Question**	**Response**	**Response Value**	**Respons e %**
→»JD27d	If the situation arises in the future, would you submit an ASAP report?	Yes	972	98%
		No	7	1%
		Refused	0	0%
		Don't know	13	1%
		Total	992	100%
If Yes, Refused or Don't Know, Skip to JD27e				
→»JD27d1	Why not?	Seven comments were provided		
→»JD27e	Do you believe that the confidentiality of ASAP data is adequately protected?	Yes	850	86%
		No	56	6%
		Refused	1	0%
		Don't know	85	8%
		Total	992	100%
If Yes, Refused or Don't Know, Skip to JD27f				
→»JD27e1	Why not?	Fifty seven comments were provided		
→»JD27f	Are you aware of any positive changes that have resulted from the ASAP program?	Yes	533	86%
		No	442	6%
		Refused	0	0%
		Don't know	17	8%
		Total	992	100%

JIMDAT SECTION C RESULTS: Safety Reporting				
Question Number	**Question**	**Response**	**Response Value**	**Respons e %**
JD28	Does your airline have a procedure or program other than ASAP for receiving safety reports from pilots?	Yes	137	88%
		No	12	8%
		Refused	0	0%
		Don't know	7	4%
		Total	156	100%
If Refused or Don't Know, Skip to JD29, If No Skip to 28B				
→»JD28a	Are you aware of any positive changes that have resulted from this pilot reporting program?	Yes	89	65%
		No	46	34%
		Refused	0	0%
		Don't know	2	1%
		Total	137	100%
→»JD28b	Would you favor the establishment of an ASAP program, oppose it, or neither favor nor oppose it?	Favor	105	71%
		Oppose	2	2%
		Neither	33	21%
		Refused	0	0%
		Don't know	9	6%
		Total	149	100%

JIMDAT SECTION C RESULTS: Safety Reporting				
Question Number	**Question**	**Response**	**Response Value**	**Respons e %**
JD29	Does your airline have a Flight Operations Quality Assurance Program, sometimes called FOQA?	Yes	835	73%
		No	176	15%
		Refused	0	0%
		Don't know	137	12%
		Total	1148	100%
[If NO, ask JD29A] [If YES, skip to JD29b] [If refused or don't know skip to JD30]				
→»JD29a	Would you favor the establishment of a FOQA program at your airline, oppose it, or neither favor nor oppose?	Favor	69	39%
		Oppose	33	19%
		Neither	59	34%
		Refused	0	0%
		Don't know	15	8%
		Total	176	100%
Skip to JD30				
→»JD29b	Have you been briefed on the program?	Yes	738	88%
		No	92	11%
		Refused	0	0%
		Don't know	5	1%
		Total	835	100%

JIMDAT SECTION C RESULTS: Safety Reporting				
Question Number	**Question**	**Response**	**Response Value**	**Respons e %**
→»JD29c	Do you believe that the confidentiality of FOQA data is adequately protected?	Yes	640	77%
		No	95	11%
		Refused	0	0%
		Don't know	100	12%
		Total	835	100%
→»JD29d	Are you aware of any safety improvements that have resulted from the FOQA program?	Yes	552	66%
		No	269	32%
		Refused	0	0%
		Don't know	14	2%
		Total	835	100%

SECTION C RESULTS:
Corporate Safety

JIMDAT SECTION C RESULTS: Corporate Safety				
Question Number	**Question**	**Response**	**Respons e Value**	**Respons e %**
JD30	Does your airline have a C.E.O. mission statement on safety?	Yes	935	82%
		No	72	6%
		Refuse	0	0%
		Don't know	141	12%
		Total	1148	100%
JD31	Does your airline have a Director of Safety?	Yes	1070	93%
		No	31	3%
		Refuse	0	0%
		Don't know	47	4%
		Total	1148	100%
JD32	Does your airline have a V.P. of Safety?	Yes	535	46%
		No	218	19%
		Refuse	0	0%
		Don't know	395	35%
		Total	1148	100%

JIMDAT SECTION C RESULTS: Corporate Safety				
Question Number	**Question**	**Response**	**Respons e Value**	**Respons e %**
JD33	Have you observed a strong commitment to safety among senior management?	Yes	971	85%
		No	145	12%
		Refuse	1	0%
		Don't know	31	3%
		Total	1148	100%
If NO or Refused or Don't Know, skip to JD34				
JD33a	Is this senior management commitment to safety reflected throughout the organization?	Yes	917	94%
		No	48	5%
		Refuse	0	0%
		Don't know	6	1%
		Total	971	100%
JD34	If you have a safety concern, do you have a mechanism for bringing that concern to the attention of senior management?	Yes	1084	94%
		No	50	5%
		Refuse	1	0%
		Don't know	13	1%
		Total	1148	100%
If NO or Refused or Don't Know, skip to Section D				

JIMDAT SECTION C RESULTS: Corporate Safety				
Question Number	**Question**	**Response**	**Respons e Value**	**Respons e %**
JD34a	How effective is this mechanism in reaching senior management?	Extremely Effective	210	19%
		Very Effective	507	47%
		Somewhat Effective	252	23%
		Not Very Effective	47	5%
		Not At All Effective	11	1%
		Refuse	0	0%
		Don't know	57	5%
		Total	1084	100%

Report Documentation Page	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE (DD-MM-YY)	**2. REPORT TYPE**	**3. DATES COVERED** ***(From – To)***
15-03-2012	Technical Publication	

4. TITLE AND SUBTITLE

The National Aviation Operational Monitory Service (NAOMS): Results of a Survey of Air-Carrier Pilots

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Connors, Mary. M.; Mauro, Robert; Statler, Irving C.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

609866

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESSES(ES)

NASA Ames Research Center
Moffett Field, California 94035-1000

8. PERFORMING ORGANIZATION REPORT NUMBER

TH-089

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

National Aeronautics and Space Administration
Washington, DC 20546-0001

10. SPONSORING/MONITOR'S ACRONYM(S)

NASA

11. SPONSORING/MONITORING REPORT NUMBER

NASA/TP—2012–216013

12. DISTRIBUTION/AVAILABILITY STATEMENT

Unclassified—Unlimited
Subject Category: 6 Availability: NASA CASI (301) 621-0390 Distribution: Nonstandard

13. SUPPLEMENTARY NOTES

Point of Contact: Mary Connors, NASA Ames Research Center, MS 262-2, Moffett Field, CA 94037; (650) 604-6114.

14. ABSTRACT

The National Aviation Operational Monitoring Service (NAOMS) was a research project under NASA's Aviation Safety Program during the years from 2000 to 2005. The purpose of this project was to develop a methodology for gaining reliable information on changes over time in the rates-of-occurrence of safety-related events as a means of assessing the safety of the national airspace. The approach was a scientifically designed survey of the operators of the aviation system concerning their safety-related experiences. This report presents the results of the methodology developed and a demonstration of the NAOMS concept through a survey of nearly 20,000 randomly selected air-carrier pilots. Results give evidence that the NAOMS methodology can provide a statistically sound basis for evaluating trends of incidents that could compromise safety. The approach and results are summarized in the report and supporting documentation and complete analyses of results are presented in 14 appendices.

15. SUBJECT TERMS

safety trends methodology, NASA Aviation Safety Program, National Aviation Operational Monitoring Service

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	*19a. NAME OF RESPONSIBLE PERSON*
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	UU	458	STI Help Desk at email: help@sti.nasa.gov
U	U	U			*19b. TELEPHONE NUMBER (Include area code)* (443) 757-5802

Standard Form 298 (Rev. 8-98)