

March 2009

AVIATION SAFETY

NASA's National Aviation Operations Monitoring Service Project Was Designed Appropriately, but Sampling and Other Issues Complicate Data Analysis



GAO

Accountability * Integrity * Reliability



Highlights of [GAO-09-112](#), a report to congressional requesters

Why GAO Did This Study

The National Aviation Operations Monitoring Service (NAOMS), begun by the National Aeronautics and Space Administration (NASA) in 1997, aimed to develop a methodology that could be used to survey a wide range of aviation personnel to monitor aviation safety. NASA expected NAOMS surveys to be permanently implemented and to complement existing federal and industry air safety databases by generating ongoing data to track event rates into the future. The project never met these goals and was curtailed in January 2007.

GAO was asked to answer these questions: (1) What were the nature and history of NASA's NAOMS project? (2) Was the survey planned, designed, and implemented in accordance with generally accepted survey principles? (3) What steps would make a new survey similar to NAOMS better and more useful?

To complete this work, GAO reviewed and analyzed material related to the NAOMS project and interviewed officials from NASA, the Federal Aviation Administration, and the National Transportation Safety Board. GAO also compared the development of the NAOMS survey with guidelines issued from the Office of Management and Budget, and asked external experts to review and assess the survey's design and implementation.

View [GAO-09-112](#) or [key components](#). For more information, contact Nancy R. Kingsbury at (202) 512-2700 or kingsburyn@gao.gov, or Gerald L. Dillingham at (202) 512-2834 or dillinghamg@gao.gov.

AVIATION SAFETY

NASA's National Aviation Operations Monitoring Service Project Was Designed Appropriately, but Sampling and Other Issues Complicate Data Analysis

What GAO Found

NAOMS was intended to demonstrate the feasibility of using surveys to identify accident precursors and potential safety issues. The project was conceived and designed to provide broad, long-term measures on trends and to measure the effects of new technologies and aviation safety policies. Researchers planned to interview a range of aviation personnel to collect data in order to generate statistically reliable estimates of risks and trends. After planning and development, a field trial, and eventual implementation of the air carrier pilot survey and the development of a smaller survey of general aviation pilots, the project effectively ended when NASA transmitted a Web-based version of the air carrier pilot survey to the Air Line Pilots Association.

NAOMS's air carrier pilot survey was planned and designed in accordance with generally accepted survey principles, including its research and development, consultation with stakeholders, memory experiments to enhance the questionnaire, and a large-scale field trial. The survey's sample design and selection also met generally accepted research principles, but there were some limitations, and the survey data may not adequately represent the target population. Sample frame and design decisions to maintain program independence and pilot privacy complicate analysis of NAOMS data. Certain implementation decisions, including extended methodological experiments and data entry issues, also complicate analytical strategies. Also, working groups of aviation stakeholders were convened as part of NAOMS to assess the validity and utility of the data, but these groups never had access to the raw data and were disbanded before achieving consensus. To date, NAOMS data have not been fully analyzed or benchmarked against other data sources.

While NAOMS's limitations are not insurmountable, a new survey would require more coherent planning and sampling methods, a cost-benefit analysis, closer collaboration with potential customers, a detailed analysis plan, a reexamination of the sampling strategy, and a detailed project management plan to accommodate concerns inherent in any survey endeavor. As a research and development project, NAOMS was a successful proof of concept with many strong methodological features, but the air carrier pilot survey could not be reinstated without revisions to address some of its methodological limitations. The designers of a new survey would want to supplement NAOMS where it was self-limiting. Alternatively, a newly constituted research team might lead operational, survey, and statistical experts in extensively analyzing existing data to illuminate future projects.

In reviewing a draft of this report, NASA reiterated that NAOMS was a research and development project and provided technical comments, which GAO incorporated as appropriate. NASA also expressed concern about protecting NAOMS respondents' confidentiality, a concern GAO shares. However, GAO noted that other agencies have developed mechanisms for releasing sensitive data to appropriate researchers. The Department of Transportation had no comments.

Contents

Letter		1
	Scope and Methodology	2
	Results in Brief	5
	NAOMS Was Intended to Identify Accident Precursors and Potential Safety Issues	8
	NAOMS's Planning and Design Were Robust, but Implementation Decisions Complicate Data Analysis	23
	A New Survey Would Require Detailed Planning and Revisiting Sampling Strategies	71
	Concluding Observations	77
	Agency Comments and Our Evaluation	78
Appendix I	Technical Issues Relating to NAOMS's Development and Data	81
Appendix II	Comments from the National Aeronautics and Space Administration	90
Appendix III	GAO Contacts and Staff Acknowledgments	92
Bibliography		93
Tables		
	Table 1: NAOMS Briefings, Presentations, Workshops, and Working Group Meetings, 1997–2005	13
	Table 2: Principles We Used to Assess the NAOMS Survey	24
Figures		
	Figure 1: NAOMS's Original Milestones, Fiscal Year 1997 to Implementation as a Permanent Survey	12
	Figure 2: NAOMS's Milestones for Fiscal Years 1997–2007, from Development to Delivery to an Operating Organization	15
	Figure 3: The Rationale for NAOMS's Questionnaire Structure	20

Figure 4: Example of an Air Carrier Pilot Survey Drill-Down Question	34
Figure 5: NAOMS's Preliminary Estimates of Pilot-Reported Flight Hours and Flight Legs, by Aircraft Size, 2002	43
Figure 6: NAOMS Air Carrier Questionnaire Section B, Question ER4 on Uncommanded Movements	61
Figure 7: NAOMS's Preliminary Findings On Pre- and Post-September 11, 2001, Event Rates	64

Abbreviations

ALPA	Air Line Pilots Association
ASMM	Aviation System Monitoring and Modeling
ASRS	Aviation Safety Reporting System
ATO	Air Traffic Organization
BTS	Bureau of Transportation Statistics
CAST	Commercial Aviation Safety Team
CATI	computer-assisted telephone interviewing
FAA	Federal Aviation Administration
FOIA	Freedom of Information Act
GTOW	gross takeoff weight
NAOMS	National Aviation Operations Monitoring Service
NAS	national airspace system
NASA	National Aeronautics and Space Administration
NTSB	National Transportation Safety Board
OIG	Office of Inspector General
OMB	Office of Management and Budget

This is a work of the U.S. government and is not subject to copyright protection in the United States. The published product may be reproduced and distributed in its entirety without further permission from GAO. However, because this work may contain copyrighted images or other material, permission from the copyright holder may be necessary if you wish to reproduce this material separately.



United States Government Accountability Office
Washington, DC 20548

March 13, 2009

Congressional Requesters

The National Aviation Operations Monitoring Service (NAOMS) was a National Aeronautics and Space Administration (NASA) initiative that aimed to develop a methodology to survey a wide range of aviation personnel to monitor safety in the national airspace system (NAS).¹ The foundation for the NAOMS project was President Clinton's August 1996 White House Commission on Aviation Safety and Security, whose principal charge was to develop, domestically and internationally, a strategy to improve aviation safety and security.² By interviewing a probability sample of pilots and other aviation professionals, project staff planned to collect data about the respondents' experiences and thus make possible statistically reliable measurements of rates and rate trends on a wide array of types of safety events in the NAS, from passenger disturbances to engine failures to bird strikes.³ Part of a larger NASA research and development initiative on aviation safety, the NAOMS project was to demonstrate the feasibility of and develop the capacity for using survey research to measure the occurrence of safety events. NASA expected surveys developed under NAOMS to complement existing federal and industry aviation safety databases.⁴ While NASA originally intended for NAOMS to collect data regularly from air carrier and general aviation pilots, air traffic controllers, flight attendants, and mechanics and to hand off the survey data collection to a different entity for permanent implementation, the project never met these goals.

¹The NAS, also known as the national aviation system, comprises the people, procedures, facilities, equipment, and infrastructure that enable air travel in the United States. This includes, but is not limited to, air traffic controllers, safety inspectors and technicians, mechanics, pilots, radar systems, airports, and aircraft.

²Executive Order 13,015; 61 *Federal Register* 43937 (Aug. 27, 1996).

³By "project staff"—and, alternatively, the "NAOMS team" or "NAOMS researchers"—we mean in this report the two researchers experienced in aviation safety that NASA appointed to lead NAOMS, and the contractor staff from the Battelle Memorial Institute (Battelle) who administered the project and worked with experts (Battelle subcontractors) in survey methodology and aviation safety to help with questionnaire construction and project management.

⁴GAO expects to report on its assessment of the Federal Aviation Administration's existing data sources later in 2009.

NAOMS was essentially a survey of air carrier pilots, and it stopped collecting data in 2004.⁵ However, neither project staff nor other aviation safety stakeholders ever fully analyzed its data. NAOMS was curtailed at the end of its first and only decade, when NASA transferred a Web-based version of its data collection system to the Air Line Pilots Association (ALPA) in January 2007. Where hope had been that the NAOMS project would provide a comprehensive, systemwide, statistically sound survey mechanism for monitoring the performance and safety of the overall NAS, ALPA did not plan to permanently implement the air carrier pilot survey as it was designed. The data collection system was never fully implemented, and its future is uncertain.

Our objective in this report is to answer the following three questions:

- What were the nature and history of NASA’s NAOMS project?
- Was the survey planned, designed, and implemented in accordance with generally accepted survey principles?
- What steps would make a new survey similar to NAOMS better and more useful?

Scope and Methodology

To describe the history and nature of the NAOMS project, we researched, reviewed, and analyzed related material posted on several NASA Web sites and provided to us directly by NASA and its contractor for NAOMS. We reviewed relevant documents on the House of Representatives’ Committee on Science and Technology Web site. We examined relevant documents produced by the Battelle Memorial Institute (Battelle), National Academies, and others as well as information produced for the National Research Council. In addition, we reviewed a number of relevant reports, articles, correspondence, and fact sheets on the NAOMS project and air safety. Many of the publicly available materials we reviewed are named in the bibliography at the end of this report.

⁵In this report, we use the term “NAOMS project” to refer to the original project as it was initially conceived, as a monitoring system with multiple surveys of a variety of aviation personnel. However, we primarily use the short form “NAOMS,” and, alternatively, the “NAOMS survey,” to refer to the most extensively developed part of the project, the air carrier pilot survey.

To analyze the NAOMS air carrier pilot survey’s planning, design, and implementation (including pretest, interview, and data collection methods); interviewer training; development of survey questions, including which safety events to include in the survey; and sampling, we interviewed officials from NASA, the Federal Aviation Administration (FAA), and the National Transportation Safety Board (NTSB) and NAOMS project staff. We also reviewed relevant documents. We discussed the survey with NAOMS team members to obtain their recollections of the work, particularly regarding limitations, gaps, and inconsistencies in the documentation. GAO internal experts in survey research reviewed the Office of Management and Budget’s (OMB) *Standards and Guidelines for Statistical Surveys* and derived a number of survey research principles relevant to assessing the NAOMS survey.⁶ We compared the NAOMS survey’s design and implementation with these principles. Although OMB’s standards as they are used today were not final until 2006, the vast majority of OMB’s guidelines represent long-established, generally accepted professional survey practices that preceded the 2006 standards by several decades. We also examined the potential risk for survey error—that is, “errors inherent in the methodology which inhibit the researchers from obtaining their goals in using surveys” or “deviations of obtained survey results from those that are true reflections of the population.”⁷ Survey error could result from issues related to sampling (including noncoverage of the target population and problems with the sampling frame), measurement error, data processing errors, and nonresponse.⁸

We asked three external experts to review and assess the NAOMS air carrier pilot survey’s design and implementation as well as considerations for analysis of collected data. These external reviews and assessments were conducted independently of our own review activities. We selected the experts for their overall knowledge and experience in survey research methodology and, specifically, for their expertise in measurement (particularly the aspects of memory and recall), survey administration and

⁶OMB, Statistical Programs and Standards, *Standards and Guidelines for Statistical Surveys* (Washington, D.C.: September 2006). See www.whitehouse.gov/omb (last accessed Mar. 1, 2009).

⁷Robert M. Groves, *Survey Errors and Survey Costs* (New York, N.Y.: John Wiley and Sons, April 1989), 6.

⁸OMB, Federal Committee on Statistical Methodology, “Measuring and Reporting Sources of Error in Surveys, Statistical Policy [Working Paper 31]” (Washington, D.C.: July 2001).

management, and sampling and estimation. The experts included Robert F. Belli, Professor, Department of Psychology, University of Nebraska, Lincoln, Nebraska; Chester Bowie, Senior Vice President and Director, Economics, Labor, and Population Studies, National Opinion Research Center, Bethesda, Maryland; and Steve Heeringa, Senior Research Scientist at the Survey Research Center and Director of the Statistical Design Group at the Institute for Social Research, University of Michigan, Ann Arbor, Michigan.

To determine what steps or other considerations might improve the quality and usefulness of a survey like NAOMS if one were to be implemented in the future, we identified and described methodological deviations that we found from GAO's guidance and OMB's standards. We also obtained the views of internal and external experts on how limitations caused by such deviations might be overcome. We assessed the potential or known effects of design or implementation limitations we identified.

We focused our review on the most extensively developed part of the NAOMS effort, the air carrier pilot survey. We discuss the general aviation study as it relates to the air carrier survey and overall project evolution, but we do not focus on its development or implementation.⁹ We attempted to identify any problems that might have prevented the NAOMS survey data from producing meaningful results, and that might not materially affect the survey results but could result from accepting the reasonable risk and trade-offs inherent in any survey research project. We note that limitations may not necessarily be weaknesses.

We conducted our work from March 2008 to March 2009 in accordance with generally accepted government auditing standards. Those standards require that we plan and perform the audit to obtain sufficient, appropriate evidence to provide a reasonable basis for our findings and conclusions based on our audit objectives. We believe that the evidence obtained provides a reasonable basis for our findings and conclusions based on our audit objectives.

⁹The general aviation survey adapted the questionnaire and expanded the sample used in NAOMS to survey nonmilitary pilots who were not commercial air carrier pilots.

Results in Brief

The NAOMS project was originally intended to develop a survey methodology to identify accident precursors and potential safety issues. The project was conceived and designed in 1997 to provide broad, long-term measures on trends and to measure the effect of new technologies and policies on aviation safety. NAOMS was to supplement other aviation safety systems by interviewing aviation personnel to collect data that could be used to generate statistically reliable estimates of risks and trends. The project was a developmental effort by NASA that was part of a larger aviation safety initiative, and it aimed to demonstrate the viability of using a survey methodology to monitor trends in aviation safety. It did not have an investigatory mission or aim to provide policy responses or interventions. NASA originally intended that the permanent implementation of surveys developed under NAOMS would generate ongoing data to track event rates into the future. Despite initial plans to administer the survey to pilots, air traffic controllers, flight attendants, and mechanics, NAOMS focused its development efforts primarily on air carrier pilots. After planning and development, a field trial, and eventual implementation of the air carrier pilot survey and a smaller survey of general aviation pilots, the project effectively ended when NASA transmitted a Web-based version of the air carrier pilot data collection system to ALPA in January 2007.

While the NAOMS air carrier pilot survey's planning and design were robust, implementation decisions complicate data analysis. NASA's project team planned and developed NAOMS in accordance with generally accepted survey research principles. The team thoroughly researched the survey's development; consulted with stakeholders in industry, government, and academia during the project's conception and evolution; conducted innovative memory experiments that enhanced the questionnaire; and conducted a large-scale field trial of air carrier pilots to answer key questions about data collection and response rate. The survey's sample design and selection met generally accepted research principles, with some limitations. For example, NAOMS was handicapped by its sampling frame and filter to identify air carrier pilots; while programmatically appropriate, the frame may not have adequately represented the target population. Furthermore, the use of sample selection criteria that potentially biased the data, along with design decisions to protect pilot confidentiality and limited sample sizes, complicate the development of analytical strategies to account for operational differences across aircraft of different sizes and for the potential for multiple pilots to witness the same event. Similarly, implementation decisions met many important survey research principles but also complicate analysis of NAOMS data. The team did not decide on

an optimal recall period for the questionnaire until approximately a quarter of the way into the final survey; additional analysis would be required to determine whether the data from different recall periods are sufficiently similar to be combined. Interviewers were experienced, and the survey attained high completion rates. However, interviewers' skill could not overcome challenges created by problematic questions or data entry issues. Working groups of aviation stakeholders, never having had access to the raw NAOMS data, were disbanded before achieving consensus on the validity and utility of these data; consequently, data validation efforts for NAOMS were limited primarily to preliminary assessments to gauge face validity.¹⁰ Inadequate records preclude the ability to leverage information from the sample when analyzing NAOMS data and hinder evaluation of the project's management and goal attainment.

A new survey would require more coherent planning and sampling methods linked to analytical goals. Sufficient survey methodology literature and documentation on NAOMS's memory experiments are available to conduct another survey of NAOMS's kind with similarly strong survey development techniques. The project's limitations are not insurmountable, and a future effort could successfully go forward from where NAOMS ended. Researchers would benefit from a cost-benefit analysis to ensure that a survey like NAOMS could cost-effectively generate essential safety information. Experimentation and testing that the NAOMS team conducted could provide an effective foundation from which to construct and test a new questionnaire. Closer collaboration with potential customers to formally and specifically codify the expected uses of the data would help ensure the data's utility. Similarly, a detailed analysis plan specifying any likely adjustments or weights, written in concert with the questionnaire, would help ensure that data could be appropriately analyzed. For example, researchers might reconsider the balance between confidentiality and the potential benefits of a questionnaire that allowed pilots to link reported events to particular aircraft and to identify aircraft they flew as air carrier pilots and in other capacities. Researchers should revisit sampling strategies to ensure that the selected frame was the most cost-effective way of sufficiently identifying the target population, and that potential biases could be remedied before or after data analysis. Finally, a detailed project

¹⁰"Face validity," a qualitative measure, refers to whether data look like they measure what is intended, rather than to whether they can be quantified with statistical methods.

management plan would help researchers accommodate the risks and trade-offs inherent in any survey endeavor without jeopardizing eventual analysis of the data.

Overall, we concluded that as a research and development project, NAOMS was a successful proof of concept, with many strong methodological features. For example, in using a probability sample and asking about experiences rather than opinions, by and large, NAOMS satisfied its stakeholders' goal of moving beyond accident-driven safety policy. Despite having successfully demonstrated the feasibility of using a survey to collect safety information from air carrier and general aviation pilots, the NAOMS project never met its goal of collecting data on an ongoing basis from a full range of aviation personnel, including helicopter pilots, air traffic controllers, flight attendants, and mechanics. While NASA eventually conveyed an air carrier pilot survey data collection operation to another entity, the project fell short of attaining permanent implementation of the original survey to track event rates into the future. NAOMS was essentially a survey of air carrier pilots that stopped data collection in 2004, and it could not be reinstated without revisions to address certain methodological limitations. NAOMS data were never fully analyzed, and, depending on the research objective, the existing data would require multiple adjustments for proper analysis. Although potentially useful for historic analysis, these data are limited in their ability to provide insight into the current health of the NAS. While NAOMS's design, data collection methods, and implementation were well-intentioned and strong in many respects, the designers of a new survey would want to supplement NAOMS where it was self-limiting. Alternatively, a newly constituted research team might lead operational, survey, and statistical experts in extensively analyzing existing data to illuminate future projects of the same kind.

We provided NASA and the Department of Transportation with drafts of this report for their review and comment. NASA reiterated that NAOMS was a research and development project and provided technical comments, which we incorporated as appropriate. NASA also expressed concern about protecting NAOMS respondents' confidentiality, a concern we share. However, we noted that other agencies have developed mechanisms for releasing sensitive data to appropriate researchers. The Department of Transportation had no comments. We also provided a draft of this report to Battelle (NASA's contractor for NAOMS) and the survey methodologist for NAOMS for their review. Battelle provided no comments on the draft report. The survey methodologist reported that he found the draft report to be objective and detailed, and that he believed it

will contribute to the public debate on NAOMS. He also provided technical clarifications, which we incorporated into the report as appropriate.

NAOMS Was Intended to Identify Accident Precursors and Potential Safety Issues

The NAOMS project was conceived and designed in 1997 to provide broad, long-term measures on trends and to measure the effect of new technologies and policies on aviation safety. Following the 1996 formation of the White House Commission on Aviation Safety and Security, and the commission's 1997 report to the President committing the government and industry to "establish[ing] a national goal to reduce the aviation fatal accident rate by a factor of five within ten years and conduct[ing] safety research to support that goal," NASA worked with FAA and NTSB to set up the Aviation Safety Investment Strategy Team within NASA.¹¹ This team organized workshops, examined options, and recommended a strategy for improving aviation safety and security. One of its recommendations led to NASA's Aviation System Monitoring and Modeling (ASMM) project, a program to identify existing accident precursors in the aviation system and to forecast and identify potential safety issues to guide the development of safety technology.¹²

ASMM, within NASA's Aviation Safety and Security Program, was to provide systemwide analytic tools for identifying and correcting the predisposing conditions of accidents and to provide methodologies, computational tools, and infrastructure to help experts make the best possible decisions. ASMM was expected to accomplish this by, among other things:

¹¹White House Commission on Aviation Safety and Security, *Final Report to President Clinton*, recommendation 1.1 (Washington, D.C.: The White House, Feb. 12, 1997), 8.

¹²A precursor is "the *symptom* of a systemic problem that is a confluence of causal factors conducive to undesired system behavior (e.g., human fatigue, organizational culture, equipment failure, or procedural discrepancy) that, if left unresolved, has the potential to result in *increased probability* of an accident. A precursor is a measurable deviation from expectations or the norm, and it is important that it not be viewed as being synonymous with causality." See Irving C. Statler, *The Aviation System Monitoring and Modeling (ASMM) Project: A Documentation of Its History and Accomplishments 1999–2005* (Washington, D.C.: NASA, June 2007), 5.

-
- intramural monitoring, providing air carriers and air traffic control facilities with tools for monitoring their own performance and safety within their own organizations, and
 - extramural monitoring, providing a comprehensive, systemwide, statistically sound survey mechanism for monitoring the performance and safety of the overall National Air Transportation System by seeking the perspectives of flight crews, air traffic controllers, cabin crews, mechanics, and other frontline operators (NAOMS was developed as the primary mechanism for collecting this information).

Agencies, airlines, and other private organizations had realized that quantitative and anecdotal information they had been collecting could not be used to calculate statistically reliable risk levels. The project team identified eight major aviation safety data sources that were available when NAOMS was created.¹³ For example, flight operational quality assurance data could have helped in deriving statistically reliable estimates from digital measurements of flight parameters, but these data do not cover all airlines or include information on human cognition or affect. Another dataset was from the Aviation Safety Reporting System (ASRS), which for 30 years had been successfully collecting information from pilots, controllers, mechanics, and other operating personnel about human behavior that resulted in unsafe occurrences or hazardous situations.¹⁴ However, because ASRS reports are submitted voluntarily, the resulting data cannot be used to generate reliable rate estimates. Under ASRS, pilots describe events briefly by mail or on NASA's ASRS Web site. NASA reviews each report and enters detailed information about the events into an anonymous database that it maintains. According to the ASRS Director, the system is subject to volatility in reporting, as in 2006, when the data witnessed a spike in reports of wrong runway use following

¹³The eight data sources are listed in a Battelle document entitled *NAOMS Reference Report: Concepts, Methods, and Development Roadmap*, prepared for the NASA Ames Research Center (Nov. 30, 2007), table 2.1.

¹⁴FAA instituted its voluntary ASRS program in 1975. To enhance the program by increasing the anonymity of reporters and others, FAA delegated reporting, processing, and analysis of raw data from Aviation Safety Reports to NASA as a third party. Under the terms of a memorandum of understanding originally signed in 1975, NASA designed ASRS to receive Aviation Safety Reports, and administers the program independent of FAA. (See U.S. Department of Transportation, FAA, Advisory Circular 00-46D (Feb. 26, 1997).)

a fatal accident in Kentucky, where pilots turned onto a taxiway that was too short for their aircraft to attain lift-off speed.¹⁵

Also, ASRS is not statistically generalizable. Although it does not constrain the types of events that can be reported, ASRS reporting is voluntary and unlikely to cover the universe of safety events, and it cannot be used to calculate trends. To complement this system and other safety databases, the NAOMS project was to interview a statistical sample of professionals participating in the air transportation system, including pilots, about their experiences. Data from the interviews were to enable statistically reliable measurements of rates and rate trends for a wide array of types of safety events, such as the professionals experiencing fire in the cargo or passenger compartment or encountering severe turbulence in clear air, collisions with birds, airframe icing, and total engine failure. As the project evolved, the NAOMS researchers decided to deemphasize NAOMS's potential to calculate rates in isolation, instead highlighting the project's primary capability to identify trends worthy of investigation, thereby complementing other data sources. The premise of the NAOMS project was that aviation personnel were the best source of information on day-to-day, safety-related events. In measuring the occurrence of safety incidents that might increase the risk of an accident, rather than accidents themselves, the project would serve a monitoring role rather than an investigative role. Instead of directly informing policy interventions, NASA expected that trends seen in the NAOMS data would point aviation safety experts toward what to examine in other data systems. However, to date, the accuracy of rate and trend estimates based on NAOMS data has not been established.

NASA appointed two researchers with aviation safety experience to lead a project team in developing surveys for NAOMS as a part of ASMM. The researchers contracted with Battelle to administer the project. Battelle, in turn, subcontracted with experts in survey methodology and aviation safety to help with questionnaire construction and project execution.

¹⁵To encourage operational personnel to report incidents or situations that they believe compromise aviation safety, FAA provides ASRS reporters with limited legal immunity from regulatory enforcement action. The Administrator of FAA is prohibited from using reports submitted to NASA under ASRS (or information derived from them) in any enforcement action, except that it may use information concerning criminal offenses or accidents, which are not covered under the program (14 C.F.R. § 91.25 (2008); see <http://asrs.arc.nasa.gov/overview/immunity.html>, last accessed Mar. 1, 2009).

NASA housed the project within the external monitoring aspect of the ASMM program, which aimed to develop a comprehensive survey methodology for monitoring the overall state of the NAS that could, on implementation, provide aviation decision makers with regular, accurate, and insightful measures of the system's health, performance, and safety. ASMM's plan discussed the importance of developing surveys for NAOMS with methodological rigor, noting that the success of NAOMS depended at least on the

"1) plausibility and understandability of NAOMS statistics (e.g., reasonable and reliable representation of the relative frequencies with which unwanted events occur),

"2) stability and interpretability of NAOMS statistical trends,

"3) sensitivity to industry concerns about data misuse, and

"4) timely and appropriate disclosures of NAOMS findings."¹⁶

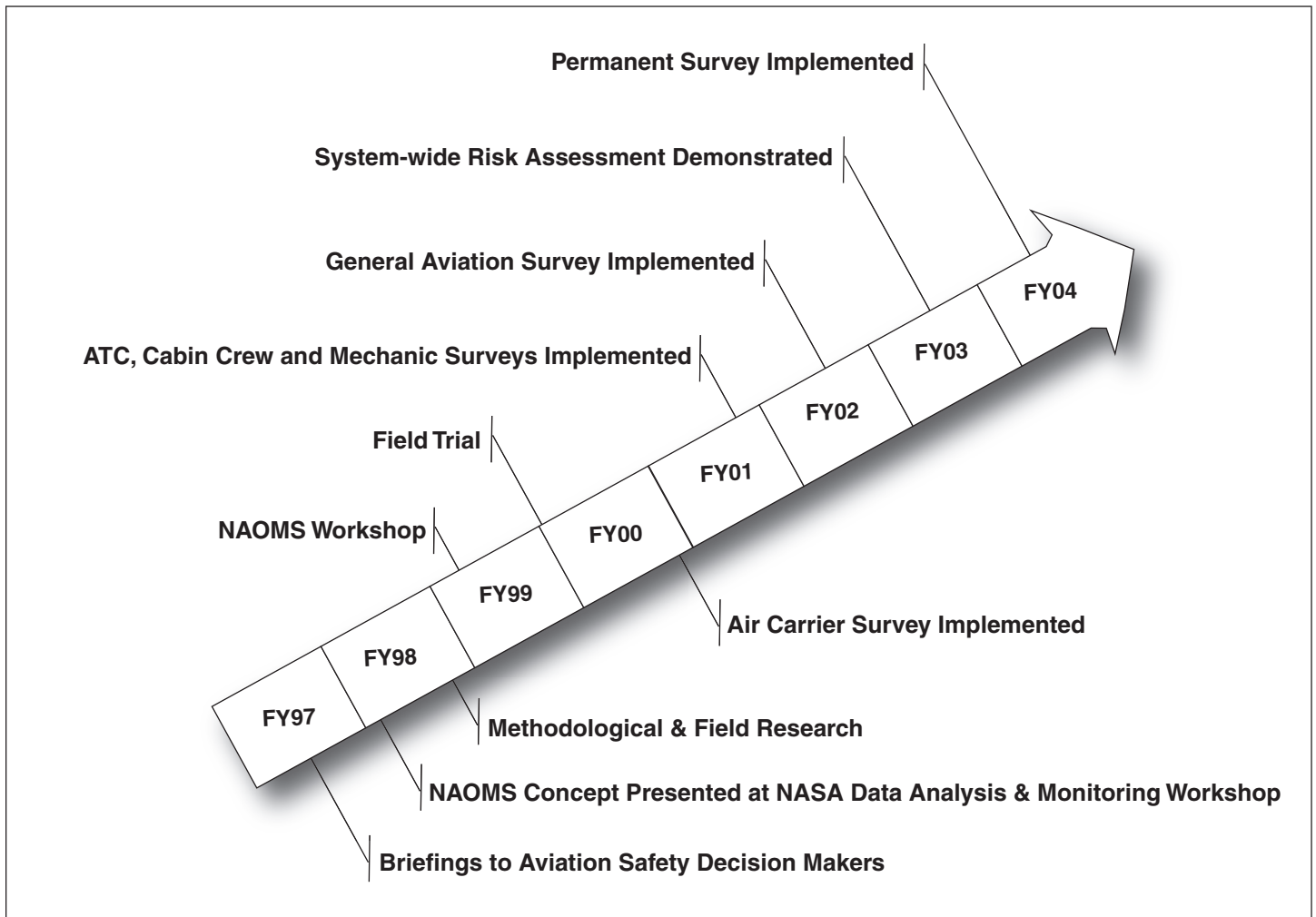
A primary objective of NAOMS was to demonstrate that surveys of personnel from all aspects of the aviation community could be cost-effectively implemented to help develop a full and reliable view of the NAS. NASA also sought to find a permanent "home" for the surveys, having planned to develop "scientific methodologies to maximize the useful information and minimize the cost, but not . . . provide for permanent service" or funding for NAOMS.¹⁷

That is, NASA intended the NAOMS project to collect data continually from air carrier and general aviation pilots, helicopter pilots, air traffic controllers, flight attendants, and mechanics. It sought to design a permanent survey data collection operation that, once implemented, could generate ongoing data to track event rates into the future (see fig. 1). NASA was to conduct the research and development steps necessary to demonstrate a survey methodology that would quantitatively measure aviation safety throughout the NAS, but it expected that a different organization, possibly FAA, would permanently implement the surveys NASA developed.

¹⁶Irving C. Statler, *Aviation Safety and Security Program (AvSSP): 2.1 Aviation System Monitoring and Modeling (ASMM) Sub-Project Plan, Version 4.0* (Washington, D.C.: NASA, February 2004), 40.

¹⁷Statler, *Aviation Safety and Security Program (AvSSP)*, 42.

Figure 1: NAOMS's Original Milestones, Fiscal Year 1997 to Implementation as a Permanent Survey



Source: Linda Connell, *NAOMS Workshop: National Aviation Operations Monitoring Service (NAOMS)* (Washington, D.C.: NASA, Mar. 1, 2000), 113.

NASA's project leaders outlined these objectives in briefings, presentations, workshops, and meetings as they explained the project's concept and progress (see table 1). The NAOMS team briefed officials overseeing the ASRS project, for example, on NAOMS's concept as early as 1997. In 2005, the team showed the Commercial Aviation Safety Team

(CAST) how the NAOMS air carrier pilot survey could help develop metrics to assess the effectiveness of safety interventions.¹⁸

Table 1: NAOMS Briefings, Presentations, Workshops, and Working Group Meetings, 1997–2005

Year	Date	Topic or title	Audience	Place
1997	^a	Concept for Monitoring	NASA Aviation Safety Reporting System Advisory Committee	^a
	^a	Review of Concept for Monitoring	International workshop participants at NASA headquarters	Washington, D.C.
1998	^a	Monitoring Concept Described	Office of System Safety, FAA	^a
	Mar. 5	Creation of NAOMS: Proposed Phase 1, A Monitoring Proposal	Flight Safety Foundation Icarus Committee Working Group on Flight Operational Risk Assessment	Washington, D.C.
	Nov. 13	Development and Proof of Concept	NASA Aviation Safety Reporting System Advisory Subcommittee	Moffett Field, California
1999	May 11	Program Concept and Methodology Workshop	FAA and other government agencies, and aviation industry groups	Alexandria, Virginia
2000	Jan. 26	Program Overview; Partial Field Trial Results	Aviation Specialty Corporation	^a
	Mar. 1	Workshop: Field Trial Results and Methodology	FAA and other government agencies and aviation industry groups	Washington, D.C.
2002	Aug. 28	In-Close Approach Changes Level 2 Milestone Workshop	NASA Ames and ICAC contractors	^a
	Dec. 5	Program Overview: Preliminary Results	Aviation Safety and Security Program Office, NASA Langley Research Center	Hampton, Virginia
2003	Apr. 9	Program Overview and Preliminary Results	FAA	Washington, D.C.
	May 7	Program Review	National Research Council Review Committee	Moffett Field, California
	Aug. 5	Overview and Status	FAA and Joint Implementation Measurement Data Analysis Team of CAST	Newport, Rhode Island
	Dec. 18	Project Overview: Background, Approach, Development, Methodology, and Current Status	NAOMS Working Group 1	Seattle, Washington
2004	^a	Survey Methodology and Design Decisions	NTSB	Washington, D.C.
	May 5	Project Status and Results Review	NAOMS Working Group 2	Washington, D.C.

¹⁸CAST, a government-industry group, identifies top safety areas by analyzing accident and incident data and identifies and implements safety enhancements aimed at reducing fatalities.

Year	Date	Topic or title	Audience	Place
	June 16	Construction of Joint Implementation Measurement Data Analysis Team, Air Carrier Questionnaire Section C	Joint Implementation Measurement Data Analysis Team of CAST	San Francisco, California
	Sept. 1	Program Overview, Air Carrier Questionnaire, Section C, In-Close Approach Changes Results	Air Traffic Organization, FAA	Washington, D.C.
	Sept. 8	Project Overview	FAA La Pointe Technical Center	Mountain View, California
2005	Jan. 26 and 28	Joint Implementation Measurement Data Analysis Team, Air Carrier Questionnaire Section C Results	Joint Implementation Measurement Data Analysis Team of CAST	^a

Source: GAO.

Note: We found no information on briefings, presentations, workshops, or meetings in 2001.

^aWe were unable to determine the missing data in the table.

Another early presentation, in March 1998, demonstrated NAOMS’s concept and goals while spelling out in detail the project’s phase one. Project staff planned to profile and summarize participant demographics in a technical document, develop a preliminary statistical design, identify high-value survey topics, incorporate these topics into a draft survey instrument, and analyze and validate the survey design to refine the survey instrument.¹⁹ The presentation delineated four distinct project phases:

- develop the methodology, while engaging stakeholder support;
- conduct a test survey to prove the concept;
- implement the full nationwide survey incrementally; and
- hand off the instrument to an organization interested in operating it over the long term.²⁰

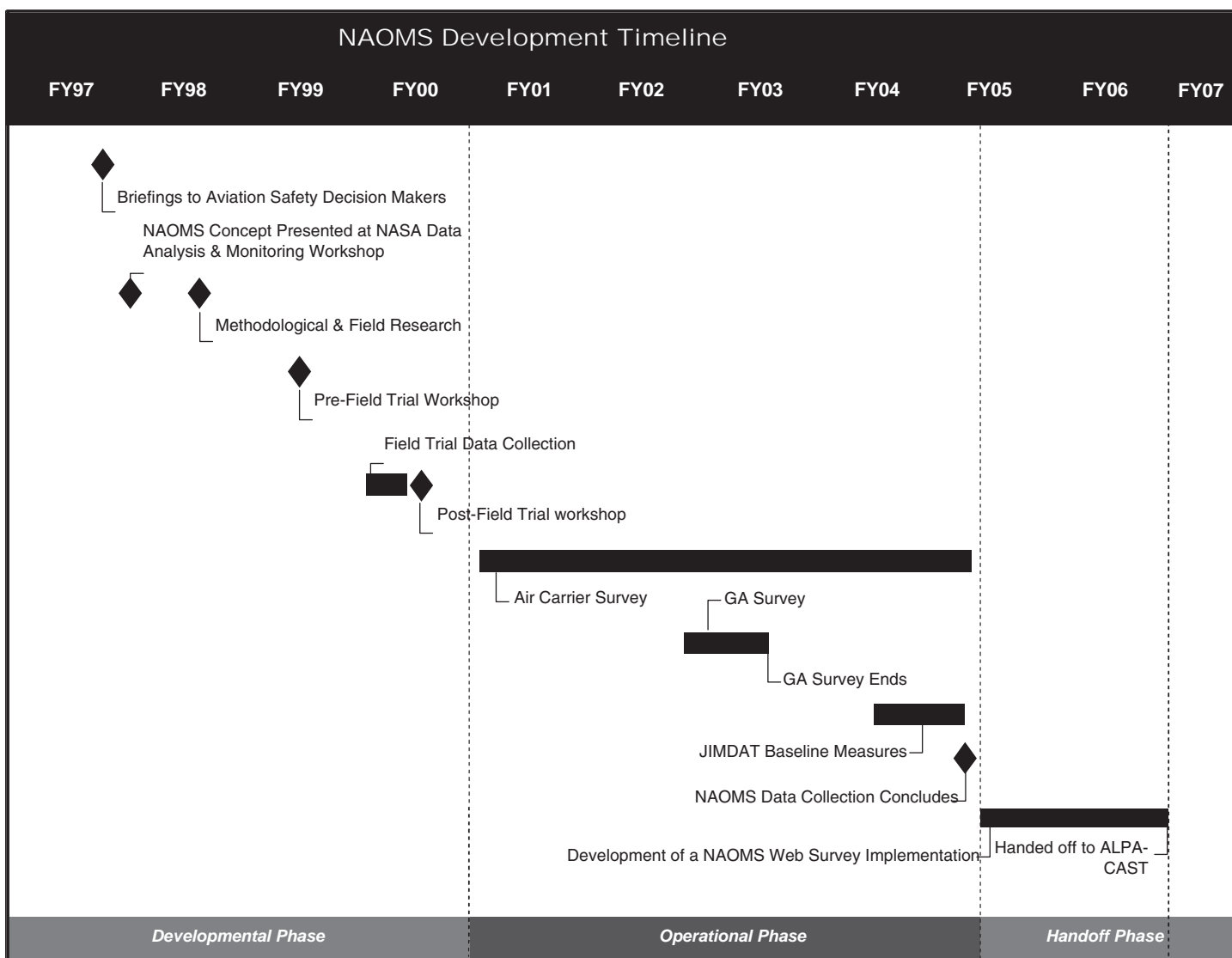
Project staff were later to describe the first two stages as one “methods development” phase. Figure 2 outlines the completion of these phases as expressed first in 1997 briefings to aviation safety decision makers in the development stage to the delivery of NAOMS’s data collection system to ALPA in January 2007. The figure reflects changes in the NAOMS project

¹⁹NASA, “Creation of a National Aviation Operational Monitoring Service (NAOMS): Proposed Phase One Effort” (Washington, D.C.: Mar. 5, 1998), 21.

²⁰NASA, “Creation of a National Aviation Operational Monitoring Service (NAOMS),” 4.

resulting from NASA's decision to halt development of the full array of surveys indicated in figure 1. By 2004, which was the original target date for permanent implementation of surveys, the team had been able to develop and begin only the pilot surveys (both air carrier and general aviation pilots), not those for other personnel as initially was planned.

Figure 2: NAOMS's Milestones for Fiscal Years 1997–2007, from Development to Delivery to an Operating Organization



Source: Robert S. Dodd, "NAOMS Development and Application," presentation to the Aeronautics and Space Engineering Board, National Academies (Washington, D.C.: June 9, 2008), 5.

As shown in figures 1 and 2, NASA originally planned to end funding in 2004 but extended it to 2007 to “properly fund transition of the data” to the larger safety community.²¹ A Web-based version of the air carrier pilot survey and related information were handed off to ALPA in January 2007.

The Survey’s Development: Feasibility, Methodology, and Field Testing

In 1998, members of the NAOMS team—NASA managers, survey methodologists, experts in survey implementation, aviation safety analysts, and statisticians working with support service contractors from Battelle—began to study long-term surveys that had helped support government policymaking since at least 1948. The team intended for NAOMS to employ the best practices of surveys used in other policy areas providing comparable benefits. The team members reviewed an extensive variety of surveys used for national estimates and for risk monitoring. These surveys included the Centers for Disease Control and Prevention’s Behavioral Risk Factor Surveillance System, which provides information on, among others, rates of smoking, exercise, and seat-belt use, and the Bureau of Labor Statistics’ Consumer Expenditure Survey, which provides data to construct the consumer price index. The team’s aim was to learn how the NAOMS survey could measure actual experiences.

The NAOMS team came to the conviction that the survey should collect the information they needed from the people

“who were watching the operation of the aviation system first-hand and who knew what was happening in the field . . . [and that] this use of the survey method was in keeping with many other long-term federally funded survey projects that provide valuable information to monitor public risk, identify sources of risk that could be minimized, identify upward or downward trends in specific risk areas, to call attention to successes, identify areas needing improvement, and thereby save lives”²²

²¹Michael D. Griffin, Administrator, and Bryan D. O’Connor, Chief, Safety and Mission Assurance, NASA, “Release of Aviation Safety Data,” media briefing (Washington, D.C.: Dec. 31, 2007), 17 (Michael D. Griffin statement).

²²Jon A. Krosnick, statement on the National Aviation Operations Monitoring Service before the Committee on Science and Technology, House of Representatives, U.S. Congress (Washington, D.C.: Oct. 31, 2007), 2–3.

The team decided that in a well-designed and implemented survey process,

“only the aviation systems operators—its pilots, air traffic controllers, mechanics, flight attendants, and others—[had] the situational awareness and breadth of understanding to measure and track the frequency of unwanted safety events and to provide insights on the dynamics of the safety events they observe. The challenge was to collect these data in a systematic and objective manner.”²³

In 1999, the team established a plan of action that included a feasibility assessment, with a literature review, to study methodological issues, estimate sample size requirements, and enlist the support of the aviation community. The assessment also planned for research that included a series of focus groups to help determine likely responses to a survey and a study of how pilots recall experiences and events. It also outlined a field trial to begin in fiscal year 1999 and, finally, a staged implementation, beginning with air carrier pilots, progressing to a regular series of surveys, and moving on to other aviation constituencies.²⁴

For the feasibility assessment, NAOMS researchers consulted with industry and government safety groups, including members of CAST and FAA and analysts with ASRS. They reviewed aviation event databases such as ASRS, the National Airspace Information Monitoring System, and Bureau of Transportation Statistics (BTS) data on air carrier traffic. The team drew on information from this research, as well as team members’ own expertise, to construct and revise a preliminary questionnaire for air carrier pilots.

After the feasibility assessment, the team conducted a large-scale field trial from November 1999 to February 2000 to help resolve the following issues about the air carrier pilot questionnaire:

“• What risk-elevating events should we ask the pilots to count?”

“• How shall we gather the information from pilots—written questionnaires, telephone interviews, or face-to-face interviews?”

²³Battelle, *NAOMS Reference Report*, 6.

²⁴Linda Connell, *Workshop on the Concept of the National Aviation Operational Monitoring Service (NAOMS)* (Alexandria, Va.: May 11, 1999), 24. See also Robert S. Dodd, “NAOMS Development and Application,” presentation to the Aeronautics and Space Engineering Board, National Academies (Washington, D.C.: June 9, 2008), 5.

“• How far back in the past can we ask pilots to remember without reducing the accuracy of their recollections?”

“• In what order should the events be asked about in the questionnaire?”²⁵

As a result of the 600 air carrier pilot interviews conducted for the field trial, the researchers decided that telephone interviewing was sufficiently cost-effective and had a high enough response rate to use in the final survey. The field trial had tested question content that derived from previous research and had experimented with the order of different sections of the survey. The field trial gave the team confidence that the NAOMS survey was a viable means of monitoring safety information. However, the field trial did not fully resolve questions about the period of time that would best accommodate pilots’ ability to recall their experiences or about the best data collection strategy.

Getting the Survey Under Way

The team had decided before the field trial that the NAOMS questionnaire content and structure were to be governed by (1) measures of respondent risk exposure, such as the numbers of flight hours and flight legs flown; (2) estimates of the numbers of safety incidents and related unwanted events respondents experienced during the recall period; (3) answers to questions on special focus topics stakeholders requested; and (4) feedback on the quality of the questions and the overall survey process.²⁶

After the team analyzed the data from the field trial and conducted further extensive research, it decided that the NAOMS survey should address as many safety events identified during its preliminary research as practical, that its questions should be ordered to match clusters from the field trial based on causes and phases of flight, and that a sample size of approximately 8,000 to 9,000 interviews per year would provide sufficient sensitivity to detect changes in rates. The team structured the survey in four sections in accordance with their original expectations of what the

²⁵Krosnick, statement before the Committee on Science and Technology, 7–8.

²⁶Battelle, *NAOMS Reference Report*, 14–15, and Connell, *Workshop on the Concept of the National Aviation Operational Monitoring Service*, 51–58. Flight hours are used to calculate risk exposure for events that can occur any time during flight; flight legs are used for events that occur mainly during terminal operations. See Mary Connors and Linda Connell, “The National Aviation Operations Monitoring Service: A Project Overview of Background, Approach, Development, and Current Status,” presentation to the NAOMS Working Group 1 (Seattle, Wash.: Dec. 18, 2003), 16.

survey should cover. NAOMS's project managers explained the rationale for this structure, shown in figure 3, in a 2004 presentation to FAA's Air Traffic Organization (ATO).²⁷

²⁷ ATO employs approximately 35,000 air traffic controllers, technicians, engineers, and support personnel who provide air traffic services to the nation to facilitate the safe and efficient movement of aircraft throughout the NAS. See www.faa.gov/about/office_org/headquarters_offices/ato (last accessed Mar. 1, 2009).

Figure 3: The Rationale for NAOMS's Questionnaire Structure

Questionnaire Structure



- Section A: Operational Exposure
 - Measures operational activity levels (risk exposure)
- Section B: Safety Event Experiences (Core Questions)
 - Counts standard event frequencies with long-term trends in mind
- Section C: Focus Topics
 - Provides a moving “searchlight” that can be redirected as needed to topics of interest
- Section D: Participant Feedback
 - Seeks continuing feedback on the validity of the NAOMS survey process and survey questions

Source: Mary Connors and Linda Connell, “National Aviation Operations Monitoring Service Project Overview: Background, Development, Approach, and Current Status,” presentation to the Air Traffic Organization (Washington, D.C.: NASA, Sept. 1, 2004), 12.

NASA’s contractors began computer-assisted telephone interviewing (CATI) data collection for the full air carrier pilot survey in March 2001. Using a sample that was drawn quarterly from a subset of a publicly available FAA database, interviewers surveyed pilots regularly over approximately 45 months of data collection. The survey methodology changed during the first few months of the survey: that is, researchers settled on which recall period to use and a cross-sectional data collection

strategy approximately 1 year after the operational survey began. Interviewing ended in December 2004, by which time more than 25,000 air carrier pilot interviews had been completed.

In addition to the air carrier pilot survey, NAOMS researchers explored elements of the original action plan for the project. They conducted focus groups with air traffic controllers and drafted preliminary survey questions. Building on research done for the main air carrier survey, NAOMS staff also developed and implemented a survey for general aviation pilots that ran for approximately 9 months in late 2002 and early 2003. However, by the end of 2002, NASA realized that it would not be feasible to expand the project to other aviation personnel under its initial plan to hand off the surveys for permanent service at the end of fiscal year 2004. NAOMS staff focused their attention on establishing the NAOMS air carrier pilot survey as a permanent service, noting that the system was still under development and that its benefits had not been fully demonstrated. They suggested that it would be difficult to find an organization that would be willing to commit to the financial and developmental resources necessary to manage an uncompleted project.

The Survey's Handoff and Results, and the NASA Inspector General's Review

NASA's documentation had repeatedly shown that the NAOMS project's purpose was "the development of methodologies for collecting aviation safety data," with their eventual transition "to the larger safety community" for permanent implementation. NAOMS had met its key objectives of demonstrating a survey methodology to quantitatively measure aviation safety and track trends in event rates by the end of 2004, when original funding for the project had been scheduled to end. Seeking to ensure the future of the survey while streamlining the project, project staff tested whether Web-based data collection was a cost-effective measure.

NASA established an agreement with ALPA, which planned to initiate a Web-based version of the air carrier pilot survey on behalf of CAST and its Joint Implementation Measurement Data Analysis Team.²⁸ NASA extended NAOMS original funding into 2007 to accommodate the transition to

²⁸The Joint Implementation Measurement Data Analysis Team is a CAST working group that assesses proposed safety enhancements and prepares safety plans to track progress in implementing them.

ALPA.²⁹ NASA conducted training sessions for ALPA staff on the NAOMS Web application in early fiscal year 2007 and conveyed the operational data collection system to ALPA in January 2007. However, ALPA never fully implemented the Web survey. According to an ALPA official in late 2007, the organization was exploring how to modify the survey before implementing it.³⁰ Although ALPA never had access to existing NAOMS data, this official also expressed uncertainty about what should be done with the existing data. The project effectively ended at the point of transfer.

In October 2007, following NASA's rejection of an Associated Press reporter's request for NAOMS data under the Freedom of Information Act (FOIA), the House Committee on Science and Technology held hearings about the development and execution of NAOMS. NASA's Office of Inspector General (OIG) subsequently initiated an investigation into NAOMS's project management. The OIG's March 2008 report, summarizing the history and status of the NAOMS project, found that NAOMS had achieved many of its objectives. Specifically, NAOMS had

“demonstrated a survey methodology to quantitatively measure aviation safety, tracked trends in event rates over time, identified effects of new procedures introduced into the operating environment, and generated interest and acceptance of NAOMS by some of the aviation community as described in the Project Plans.”³¹

The OIG report identified several shortcomings of the project, including that (1) the “contracting officers did not adequately specify project requirements” or “hold Battelle responsible for completing the NAOMS Project as designed or proposed”; (2) the “contractor underestimated the level of effort required to design and implement the NAOMS survey”; (3) “NASA had no formal agreement in place for the transfer and permanent service of NAOMS”; and (4) “NAOMS working groups failed to achieve their objectives of validating the survey data and gaining

²⁹Griffin and O'Connor, “Release of Aviation Safety Data,” 17 (Michael D. Griffin statement).

³⁰Terry McVenes, Executive Air Safety Chairman, ALPA International, statement on the National Aviation Operations Monitoring Service before the Committee on Science and Technology, House of Representatives, U.S. Congress (Washington, D.C.: Oct. 31, 2007).

³¹NASA, Assistant Inspector General for Auditing, Office of Inspector General, “Final Memorandum on the Review of the National Aviation Operations Monitoring Service (Report No. IG-08-014; Assignment No. S-08-004-00),” to the Associate Administrator for Aeronautics Research, NASA (Washington, D.C.: Mar. 31, 2008), 9.

consensus among aviation safety stakeholders about what NAOMS survey data should be released.”³² An additional deficiency, according to the OIG, was that, as of February 2008, “NASA had not published an analysis of the NAOMS data nor adequately publicized the details of the NAOMS Project and its primary purpose as a contributor to the ASMM Project.”³³

NAOMS’s Planning and Design Were Robust, but Implementation Decisions Complicate Data Analysis

We found that, overall, the NAOMS project followed generally accepted survey design and implementation principles, but decisions made in developing and executing the air carrier pilot survey complicate data analysis. We discuss in this report each of the three major stages of survey development—planning and design, sample design and selection, and implementation—in turn. While we document the many strengths of the NAOMS survey and its evolution, we also discuss limitations that raise the risk of potential errors in various aspects of the survey’s results. We also note where design, sampling, and implementation decisions directly or potentially affect the analysis and interpretation of NAOMS’s data.

Table 2 outlines the generally accepted survey research principles, derived in part from OMB guidelines, that we used in our assessment. The table is a guide primarily to how we answered our second question on the strengths and limitations of the design, sampling, and implementation of the NAOMS survey. However, we caution that survey development is not a linear process; steps appearing in one section of table 2 may also apply to other aspects of the project. Direct fulfillment of each step, while good practice, is not sufficient to ensure quality. Additional related practices, and the interaction of various steps throughout the course of project development and implementation, are essential to a successful survey effort. Table 2 should be viewed not as a simple checklist of survey requirements, but as guiding principles that underlie the narrative of our report and our overall evaluation of the NAOMS survey.

³²NASA, “Final Memorandum,” 10.

³³NASA, “Final Memorandum,” 10. To respond to the Inspector General’s recommendation that NASA lead aviation stakeholder efforts to assess the utility of NAOMS data, NASA contracted with the National Research Council of the National Academies to provide an independent assessment of NAOMS’s methodology. NASA estimated that the council would complete such an assessment in June 2009.

Table 2: Principles We Used to Assess the NAOMS Survey

Survey element	Principle
Planning and design	The survey had a clear rationale?
	A review of existing studies, surveys, reports, or other literature informed the survey?
	Potential users were consulted to identify their requirements and expectations?
	The scope of survey data items was defined and justified?
	A management plan preserved the survey data and documentation of survey records?
	The design identified the frequency and timing of data collection?
	The design identified survey data collection methods?
	The questionnaire design minimized respondent burden and maximized data quality?
	The questionnaire was pretested and all components of the final survey system were field tested?
	The design planned for the highest practical rates of response before data collection?
Sample design and selection	The proposed target population was clearly identified?
	The sample frame and design were appropriate?
	Sample design coverage issues were described and handled appropriately?
	Sample size calculations were appropriate?
	Potential nonsampling errors were estimated?
Implementation	Sample administration and disposition monitoring were appropriate?
	Appropriate steps were taken to communicate confidentiality to respondents and to preserve the confidentiality of their data?
	The respondents were provided with appropriate informational materials?
	Response maximization efforts, including period of data collection and interviewer training, were appropriate?
	Steps to ensure the quality of the data were appropriate?
	Appropriate checks and edits on the data collection system mitigated errors?
	Actions taken during data editing or other changes to the data were documented?
	Survey response rates were calculated using standard formulas?
	Nonresponse analysis was conducted appropriately?
	The survey system documentation included all information necessary to analyze the data appropriately?
The survey system documentation was sufficient to evaluate the overall survey?	

Sources: GAO and OMB, *Standards and Guidelines for Statistical Surveys* (Washington, D.C.: September 2006).

The Survey's Planning and Design

Early documentation of the NAOMS project shows that the project was planned and developed in accordance with generally accepted principles of survey planning and design. As we have previously discussed, the project team established a clear rationale for the air carrier pilot survey and its use for ongoing data collection at its conception. Team members considered the survey's scope and role in light of other sources of available data, basing the questionnaire on a solid foundation of available data, literature, and information from aviation stakeholders. They devised mechanisms to protect respondent confidentiality. Researchers collected preliminary information from focus groups and interviews that they used in conducting confirmatory memory experiments and in developing the questionnaire to reduce respondent burden and increase data quality. The team was also concerned with validating the concept of NAOMS and achieving buy-in from members of industry and others to help ensure the relevance and usefulness of the NAOMS data to potential users, although they were not able to fully resolve questions some stakeholders had in the utility of the data. The team's field trial of air carrier pilots allowed them to answer key questions about data collection and response rate. The field trial was followed with supplemental steps to revise the questionnaire before the full air carrier pilot survey.

Notwithstanding the survey design's strengths, it exhibited some limitations, such as a failure to use the field trial to fully test questionnaire content and order and fragmented management plans.³⁴ We found potential risk for survey errors involving measurement, with low implications for risk of error in the survey's data.

Preliminary Research Supported the Survey's Development

In its planning, the NAOMS team extensively researched survey methodology, existing safety databases, and literature on aviation safety and personnel. The team also conducted interviews and focus groups with pilots. To generate publicity and support from aviation stakeholders, the NAOMS team made multiple presentations to and conducted workshops with government officials and aviation stakeholders (see table 1). The preliminary research and feedback from stakeholders helped the team define the scope of data collection.

³⁴According to NASA officials, issues of content and order were addressed before the full air carrier pilot survey was implemented.

Literature Reviews and Planning

Initial literature reviews focused primarily on the data collection methods that would be most likely to ensure response accuracy, on question wording and ordering that would maximize recall validity, and on preventing respondents from underreporting for fear of being held accountable for mistakes. A document summarizing several early team memorandums addressed theories and literature on “satisficing”—or the notion that survey respondents seek strategies to minimize respondent burden and cognitive engagement—and the relationship between the data collection method and respondent motivation. This document, which was reprinted, in part, in the contractor’s reference report on NAOMS, also examined literature on social desirability, particularly how confidentiality affects response accuracy. It included reviews of academic literature on how interviewing methods can dampen or enhance tendencies toward socially desirable responses.

The summary document discussed the importance of the questionnaire’s accounting for memory organization as a way to minimize response burden and maximize respondent recall using specific cues to take full advantage of how pilots organize events in memory, thus maximizing their ability to recall and report events in the reference period. It outlined specific strategies that have been used to assess memory organization. The document proposed steps the NAOMS researchers could take to assess memory organization; identify optimal recall periods; and construct, validate, pretest, and refine the survey questionnaire. It also outlined a way to implement and evaluate different data collection methods and included initial sample size calculations to compare response rates and potential sampling frames.

Another planning document enumerated in detail the populations of interest in addition to pilots, including air traffic controllers, mechanics, dispatchers, and flight attendants. The project team compiled an annotated list of sources on aviation safety and their limitations to indicate how the survey might play a role within an overall system to monitor national airspace safety.³⁵ The project team supplemented its research with focus groups and one-on-one interviews with pilots to help in deciding which safety events the questionnaire should cover. These focus groups and interviews are discussed in more detail in appendix I.

³⁵Battelle, *NAOMS Reference Report*, app. 2.

Workshops and Consultations with Stakeholders and Potential Users

After presentations on the NAOMS concept and its relevance to aviation safety in March and November 1998, NAOMS staff held the project's first major workshop on May 11, 1999. A wide range of FAA and NASA officials; representatives from private industry, academia, and labor unions; and methodologists discussed

- the need for NAOMS as a way to fill gaps in safety knowledge and move beyond accident-driven safety policy (often called the “accident du jour” syndrome);
- government's and others' use of survey research, citing specific surveys that are used to measure rates, trends, risks, and safety information in other fields;
- the intent to focus NAOMS questions on individuals' experiences, rather than on their opinions; and
- the need to involve industry and labor stakeholders to ensure high participation rates and relevant safety content.³⁶

In addition to introducing the concept of NAOMS and its likely form, the team expressly sought labor and industry participation in developing NAOMS and to ensure high response rates; the relevance of specific questions; and the survey's output application to decision making on policies, procedures, and technology.

Several aviation stakeholders participating in the workshop offered feedback on the survey in general and on individual questions raised in focus groups and the early field research. For example, a summary of comments from FAA staff raised questions about response rate, the scope of questions, and strategies for data validation.³⁷ We found that NAOMS staff clearly thought through many of these issues, including matters of response rate and questionnaire consistency, and worked to address them as the project developed. However, as we discuss in the following text,

³⁶Connell, *Workshop on the Concept of the National Aviation Operational Monitoring Service*, especially 6, 14, 32–34, 62, and 64.

³⁷NASA, “FAA NAOMS Workshop FAA Attendees Interviews, Summary” (Washington, D.C.: September 1999).

while NASA initially expected that FAA would be a primary customer of NAOMS data, it failed to attain consensus with the agency on the project's merits and on whether NAOMS's goal of establishing statistically reliable rates, in addition to trends, was possible.

Defining the Scope of the Data NAOMS Would Collect

The NAOMS team determined that the NAOMS survey would usefully supplement other safety resources whose goals were investigative or were to identify causation. Unlike those resources, NAOMS was to capture not just incidents but also precursors to accidents and "more subtle associations that may precede safety events."³⁸ The 2007 ASMM summary report noted that one must know where to look in order to investigate precursors.³⁹ NAOMS was designed to point toward such research. The project team expected that trends seen in the NAOMS data would point aviation safety experts toward what to examine in other data systems. Researchers and FAA officials told us that many data, such as radar track data and traffic collision avoidance data, do not cover the entire NAS and were not regularly analyzed at the time that NAOMS was being developed.

Following the 1999 workshop on the concept of NAOMS and the preliminary air carrier pilot questionnaire, a summary of comments from FAA showed some support for NAOMS. However, the summary expressed concern that much of the data being gathered were too broad to permit the development of appropriate intervention strategies. An FAA memorandum later, following meetings with NAOMS staff in 2003, requested extensive questionnaire revisions and suggested that certain questions were irrelevant, should be dropped, or were covered by other safety systems. FAA also sought more detailed investigatory questions to assess the causes of some events, such as engine shutdowns, and revisions to questions that it saw as too subjective and too broad to provide real safety insight. To ensure that question consistency over time would enable trend calculations, NASA researchers did not make most of the revisions. Instead, they responded that to the extent that NAOMS might provide "a broad base of understanding about the safety performance of the aviation

³⁸NASA, "NAOMS Response to FAA Questions and Concerns" (Washington, D.C.: August 2003), especially 1.

³⁹Statler, *The Aviation System Monitoring and Modeling (ASMM) Project*, 10.

system” and allow for the computation of general trends over time, its questions could help supplement other safety systems.⁴⁰

The project team’s concerns about respondent confidentiality influenced the questionnaire’s design. For example, they expressed some fear that questions that attributed blame to respondents reporting safety events would lead to underreporting. These concerns motivated decisions to exclude from the questionnaire most of the information that could have identified respondents. Pilots were not asked to give dates or identify aircraft associated with events they reported. Additionally, the database that tracked sampling and contact information for individual pilots recorded only the weeks in which interviews took place, not their specific dates.

Project Management Plans Were Not Comprehensive

The NAOMS team’s project management plans were not comprehensive. From 1998 to 2001, the activities of Battelle and its subcontractors were covered by statements of work to plan and track the survey’s development. These documents enumerated tasks, deliverables, and projected timelines. Similar documents do not exist for the 2002 to 2003 data collection period, when NASA changed priorities for NAOMS. Battelle developed a new implementation plan to address changes in NASA’s priorities in 2004, but plans from 2002 onward were largely subsumed in a series of contract modifications and were not centralized. Twenty-four base contracts and modifications contained information to track overall progress, but, according to NASA, the overall ASMM project plan (while in accordance with NASA policy) did not contain sufficient detail to correlate the plan with contract task modifications such as those used for NAOMS. The lack of a central plan makes it difficult to evaluate specific aspects of NAOMS against preestablished benchmarks. Furthermore, the failure to maintain management or work plans during data collection or to adapt the initial work plans to accommodate project changes may have contributed to the gaps in record-keeping regarding sampling, as discussed later in this report.

Innovative Memory Experiments Enhanced the Questionnaire

Research demonstrates that designing a survey to accommodate the population’s predominant memory structure can reduce respondents’ cognitive burden and increase the likelihood of collecting high-quality data. The NAOMS team conducted innovative experiments to help in developing a survey that would reduce respondent burden and

⁴⁰NASA, “NAOMS Response to FAA,” 5.

accommodate the air carrier pilots' memory organization and their ability to recall events, thus increasing the likelihood of accuracy. While researching and testing hypotheses about memory organization to enhance questionnaire design are excellent survey research practices, few researchers have the time or resources to conduct extensive experiments on their target population. The NAOMS survey methodologist ran experiments from 1998 through 1999 to generate and test hypotheses that could be incorporated into the design of the air carrier pilot survey.

Several of the project's experiments to determine pilots' recall and memory structures were based on relatively few pilots. These were supplemented with other experiments and additional data analysis to validate the researchers' hypotheses. However, these experiments were limited to the core questions on safety in the air carrier pilot survey and did not extend to other sections of the survey or other populations, whether general aviation pilots, mechanics, or flight crew. The memory experiments led researchers to design the core safety events section of the survey according to a hybrid scheme of memory organization—that is, it used groupings and cues related to causes of events as well as phases of flight, such as ground operations and cruising.⁴¹

After the memory experiments, the NAOMS survey methodologist recommended that project staff undertake cognitive interviews to ensure that the questionnaire to be used in a planned field trial could be understood and was complete, recommending also that a final version of the questionnaire be tested with a separate group of pilots. A memorandum indicated that at least five cognitive interviews were held before the field trial, but we could not identify documentation on their effect on the questionnaire's structure or content.

A Large-Scale Field Trial Resolved Many Issues, but Not Others

In 1999, following more than 1 year of research, experiments, and questionnaire development, NAOMS researchers conducted a large-scale field trial. It was to help decide the appropriate recall period for the survey questions; major issues of order and content for the questionnaire; and the appropriate method of survey administration to minimize cost, while maximizing response rate and data quality. The field trial also allowed the NAOMS team to assess whether the survey methodology was a viable means of measuring safety events. Although largely in accordance with

⁴¹More details about these experiments are in appendix I of this report as well as in Battelle, *NAOMS Reference Report*, app. 4-1.

generally accepted survey principles, the field trial had some limitations and did not resolve important questions about the survey's methodology.

To administer the trial, team members randomly assigned pilots to various experimental conditions: three different interviewing methods (self-administered questionnaires, and CATI and in-person interviews), six different recall periods, and the presentation of the main questions of the core safety questions first or following the topical focus section. Interviewers for the CATI and in-person interviews received group and individual training, and the researchers used widely accepted practices to enhance response rates for the self-administered questionnaire, with notifications and reminder letters to maximize response rate. Their analysis of the data appeared to show that experimental assignments were sufficiently random and different in data quality to allow some decisions about response mode and recall period—showing, for example, that different modes resulted in different completion rates, and that longer recall periods produced higher event counts.

Recall Period Research and Testing

The NAOMS researchers hoped to reliably measure highly infrequent events—the severest of which pilots were likely to recall quite well—without jeopardizing the measurement of more frequent, less memorable events that had safety implications. Literature on survey research did not point to one specific reference period for events such as those in the NAOMS survey. To evaluate the effect of recall period on a pilot's ability to accurately remember events, the project's survey expert asked five pilots to fill out, from memory, a calendar of the dates and places of each of their takeoffs and landings in the past 4 weeks. Then they were asked to fill out an identical calendar at home, using information they had recorded in their logbooks.

The survey methodologist used these data to support his recommendation that NAOMS use a 1-week recall period, noting that this would require a substantial increase in sample size to measure events with the precision NAOMS originally intended. However, because the experiment was designed to measure only takeoffs and landings—routine activities that were unlikely to carry the weight in memory of more severe or infrequent safety events at the heart of the NAOMS project—the survey methodologist added the caveat that the final decision about recall interval would have to be informed by the particular list of events in the final NAOMS questionnaire and the rates at which pilots witnessed them.

Following the logbook experiment, NAOMS researchers tested several potential recall periods in the field trial, including 1 and 2 weeks and 1, 2, 4, and 6 months. Data from the field trial show an increase in the number of hours flown and event reporting commensurate with extensions of the recall period and possible overreporting for the 1-week period relative to the others. Aside from the logbook experiment, however, no efforts were made to validate the accuracy of field trial reports of safety events or flight hours and legs flown in survey data collected within different recall periods.⁴²

The project team also obtained feedback from the pilots participating in the field trial. This feedback indicated that most who commented on recall periods said they were too short; the pilots wanted to report incidents that happened recently, but not within the recall period. The researchers noted that the pilots' discomfort with a short recall period did not necessarily mean the data collected within that period were inaccurate; it meant only that it was possible that they wanted to report events outside the recall period to avoid giving the impression that certain events never occurred. Researchers also studied pilots' reported confidence in their responses as an indication of data quality obtained with different recall periods. However, the information from the field trial tests and respondent feedback did not resolve the question of which recall period to use. Researchers decided to use approximately the first 9 months of NAOMS data collection as an experimental period to resolve questions the field trial could not answer, and they settled on a 60-day recall period several quarters after full data collection began.⁴³

Data Collection Methods

The contractor administering the field trial randomly assigned pilots to mail questionnaires, face-to-face interviewing, or CATI. Face-to-face data collection was stopped after it proved to be too costly and complicated. The project team then compared the costs and response rates of the two other methods as well as the completeness of responses as a measure of

⁴²NASA and its contractors attempted to validate flight hour and leg reports in the full air carrier pilot survey by comparing it with existing BTS data.

⁴³Battelle's final reference report on NAOMS suggests that the experiments on data collection method and recall period persisted throughout first year of the full air carrier pilot survey's operation. (See Battelle, *NAOMS Reference Report*, 26.) NASA staff have clarified that the NAOMS team decided to discard panel-based data collection in favor of the cross-sectional approach approximately 9 months into the survey's operation.

data quality. Completed mail questionnaires cost \$67 each and had a response rate of 70 percent, and 4.8 percent of the questions went unanswered. Telephone interviews cost \$85 and attained a response rate of 81 percent, and all of the questions were answered.⁴⁴

The project team decided that the CATI collection method was preferable, given the response rate, the cost, and a tighter relationship between the numbers of hours flown and aggregated events reported. We found ample information to support this data collection method. In contrast, the field trial did not provide the researchers with an opportunity to validate the sample strategy for data collection—either cross-sectional (drawing each sample anew over time) or panel (surveying the same set of respondents over time). As with the recall period, researchers used the early part of the full survey to experiment with both panel and cross-sectional approaches. They decided on a final data collection approach approximately 9 months after the full survey began.

Questionnaire Order and Content

Team members developed different versions of the field trial questionnaire to test whether to survey pilots first about main events—the core safety issues in section B—or about focus events—the issues on specific topics in section C (see fig. 3). The researchers’ quantitative analysis of the field trial data suggested that different section orders did not affect data quality. However, we found it unusual that the field trial questionnaire did not fully incorporate the specific question order suggested by experiments or literature in the main events section. While questionnaires contained content areas from the memory experiment that combined the causes of events and the phases of flights, individual topics within the core safety events section of the field trial survey were not ordered from least to most severe as the survey methodologist recommended. NASA later clarified that the NAOMS team incorporated the results of the field trial into the final survey instrument.

Additionally, the field trial questionnaire did not contain the “drill-down” questions that appeared in the final questionnaire—that is, questions asking for multiple response levels (see fig. 4). The failure to include these questions appears to violate the generally accepted survey practice of

⁴⁴Final cost numbers presented to the National Academies in 2008 differ from the estimates of fully operational costs presented in the Battelle, *NAOMS Reference Report*, 31.

using a field trial to test a questionnaire that has been made as similar as possible to the final questionnaire. While questionnaires almost inevitably change between a field trial and their final form, the results of the experiments, cognitive interviews, and full set of questions should have been incorporated into the test questionnaire before the development of the final survey.

Figure 4: Example of an Air Carrier Pilot Survey Drill-Down Question

<p>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?</p>	<p># HAZMAT <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/></p> <p style="text-align: right;">IF 0, SKIP TO ER3.</p>
<p>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?</p>	<p># IN CARGO COMPARTMENT <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/></p> <div style="border: 1px solid black; padding: 5px; text-align: center;"> <p>THE AMOUNT IN ER2A CANNOT BE GREATER THAN THE AMOUNT IN ER2.</p> </div>
<p>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?</p>	<p># IN PASSENGER COMPARTMENT..... <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/></p> <div style="border: 1px solid black; padding: 5px; text-align: center;"> <p>THE AMOUNT IN ER2A AND ER2B COMBINED CANNOT BE GREATER THAN THE AMOUNT IN ER2.</p> </div>
<p>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?</p>	<p># OUT OF COMPLIANCE WITH REGULATIONS <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/></p> <div style="border: 1px solid black; padding: 5px; text-align: center;"> <p>THE AMOUNT IN ER2C CANNOT BE GREATER THAN THE AMOUNT IN ER2.</p> </div>

Source: Battelle Memorial Institute, *NAOMS Reference Report: Concepts, Methods, and Development Roadmap*, prepared for the NASA Ames Research Center (Nov. 30, 2007), app. 11-5.

Supplementary Steps Led to Questionnaire Revisions before the Main Survey

In addition to subject matter and survey methodology research, experiments, and field testing, NAOMS staff used other commonly used survey research techniques to develop and revise the air carrier pilot survey questionnaire. For example, we found that at least five cognitive interviews were conducted before the field trial, but we found no

documentation that described these interviews or their effect.⁴⁵ Additional cognitive interviews were conducted after the field trial on nearly final versions of the questionnaire before the survey's full implementation, resulting in changes to the questionnaire (see app. I). The project team did not record field trial interviews; doing so would have allowed verbal behavioral coding, which is a supplemental means of assessing problems with survey questions for both respondents and interviewers.

Besides the changes the team made to the questionnaire from the results of the cognitive interviews, team members reviewed the survey instrument in great detail, adding and deleting questions to make it easier for the interviewers to manage and for the respondents to understand. However, as we have previously mentioned, the questionnaire used in the field trial did not fully incorporate the order of events suggested by the memory experiments. This order appears to have been addressed after the cognitive interviewing that took place just before the final survey began.

We found evidence that the NAOMS team made some changes to the questionnaire as a result of respondent comments on the field trial, such as discarding a planned section on minimum equipment lists, seen by many respondents as ambiguous and unclear, in favor of a different set of questions. However, there is no documentation of additional question revisions in response to empirical information from the field trial. Additionally, except for CATI testing involving Battelle managers and interviewers, we could not find evidence of a pretest of the final questionnaire incorporating all order and wording changes before the main survey was implemented. NASA recently told us that the results of the field trial, as well as inputs from other research, were fully incorporated into the final survey instrument.

The Survey's Sample Design and Selection

We found that for its time, NAOMS's practices regarding sample frame design and sample selection met generally accepted survey research principles, with some limitations. The project team clearly identified a target population and potential sample sources. To maintain program independence, the team constructed the sampling frame from a publicly available database that was known to exclude a sizable proportion of air carrier pilots, and applied filtering criteria to the frame to increase the

⁴⁵Cognitive interviews are individual pretests of the survey in which the survey developers solicit feedback on the language and comprehensibility of specific questions.

likelihood that the pilots NAOMS contacted would be air carrier pilots, rather than general aviation pilots. It is not known for certain whether the approximately 36,000 pilots NAOMS identified for its sample frame were representative of the roughly 100,000 believed to exist.⁴⁶ The implications for the risk of error were high; the most significant sources of potential survey error stem from coverage and sampling.

In addition to increasing the risk of error, sampling decisions potentially affect the analysis and interpretation of NAOMS data. Sample size calculations may not be sufficient to generate reliable trend estimates because of the infrequency of events that have great safety significance and concerns about operational characteristics and potential bias resulting from the sample filter. Additionally, developing estimates of event counts for air carrier operations in the NAS (which was not a primary objective of NAOMS) from a sample of pilots is complicated by the fact that rates from NAOMS are based on individuals' reports, rather than on direct measures of safety events.⁴⁷ Also, the survey has the potential for multiple individuals to observe the same event.

Potential Problems Related to the Sampling Strategy Require Additional Assessment

While NAOMS researchers designed and selected a sample in accordance with generally accepted survey research principles, sampling decisions they made to address complications influenced the nature of the data collected. NAOMS's sampling strategy for the air carrier pilot survey was complicated by the needs to (1) link a target population to specific analytical goals; (2) identify an appropriate frame from which to draw a sample; and (3) locate commercial pilots, rather than general aviation pilots. Eventually, the team constructed a frame from a publicly available pilot registration database that excluded some pilots and lacked information on where pilots worked, compelling the team to use a filter to increase the likelihood of sampling air carrier pilots. The contractor drew a simple random sample each quarter from the freshly updated, filtered, and cleaned database and divided the sample into random replicates that

⁴⁶Preliminary analysis suggests that relative to BTS data and data from air carrier pilots in the general aviation study, NAOMS air carrier pilot survey data overrepresent pilots flying widebody aircraft with long flight times and pilots flying as captains, rather than as first officers or in some other capacity.

⁴⁷That a survey (or safety monitoring system) relies on individuals' reports is not a flaw, but a design feature that must be accounted for when analyzing data.

were released weekly for interviewing.⁴⁸ After the first year of the air carrier pilot survey, which adapted sampling to accommodate experiments on recall period and panel approach to data collection, the survey sampled approximately 3,600 air carrier pilots for most quarters of data collection. This sampling strategy resulted in 25,720 completed interviews by the end of the air carrier interviewing.

Identifying a Target Population

To develop NAOMS's sampling strategy, the team first needed to identify a target population. Although an ideal target population corresponds directly with a specific unit of analysis of interest, researchers often rely on proxies when they cannot directly sample the unit. With NAOMS's goal of estimating trends of safety events per air carrier flight hour or flight leg in the NAS, a target population might have been all air carrier flights in the NAS. Theoretically, one could draw a sample of all air carrier flights in the NAS, locate the pilots on these flights, and interview them about events specific to a particular flight.

Given that such a sample would be prohibitively resource-intensive, the NAOMS team identified an alternative target population—namely, air carrier pilots. Surveying air carrier pilots would provide information on safety events as well as on how many flight hours or flight legs that pilots flew. If the frame fully covered the population of air carrier pilots, the team's planned simple random sample from the frame would allow an estimation of individual air carrier pilots' rates of events experienced per hour or leg flown. In isolation, these individual-based estimates would fall short of cleanly characterizing the NAS, which involves other pilots besides air carrier pilots and other personnel, including other crew members on each flight. However, the estimates could address NAOMS's goal of estimating rates (for individual air carrier pilots) on the basis of risk exposure and trends in safety events over time, to supplement other systems of information about safety.

⁴⁸To generate a replicate, survey researchers take smaller samples from the full sample, using the same sampling design. By releasing small replicates on a regular basis, instead of the entire sample at once, researchers can begin generating estimates for the entire sample as each replicate is released and help ensure that systematic differences between those who respond to the survey rapidly and those who take longer to interview do not compound over the time that the survey is administered.

One potential difficulty with this target population was that the number of pilots actively employed as air carrier pilots was not known when the project began. Although the NAOMS team extensively reviewed the size of the pilot population, we found multiple estimates of the target population from the NAOMS documentation. NAOMS's preliminary research suggested that approximately 90,000 pilots were flying for major national and regional air carriers and air cargo carriers.⁴⁹ Other information suggested that the population could have been as large as 120,000 pilots. For example, the 60,000 air carrier pilots in ALPA's membership represented "roughly one-half to two-thirds" of all air carrier pilots, or, alternatively, up to 80 percent of the target population.⁵⁰ In light of these different estimates, we assume for purposes of discussion a target population of about 100,000 air carrier pilots.

Constructing a Sampling Frame

NAOMS researchers next needed to identify a source of information on its target population to provide a sampling frame from which it could sample air carrier pilots. As we have previously mentioned, because there was no central list of air carrier pilots that would ensure coverage of the target population, researchers had to choose an alternative frame. Initially, they considered using ALPA's membership list of air carrier pilots. However, to maintain the project's independence and to be as inclusive of pilots as possible, regardless of their employer or union status, they decided against using this or any other industry list, such as personnel information from airlines.

The project team also considered using FAA's Airmen Registration Database.⁵¹ Its information on pilots included certification type and number, ratings, medical certification, and other personal data. When the

⁴⁹See Battelle, *NAOMS Reference Report*, app. 2.

⁵⁰Robert Dodd, "Airline Pilot Self Selection Bias in the FAA Airmen Certification Database and Methods to Evaluate Its Effect: Questions on Airline Size," memorandum to Mary Connors and Linda Connell, NASA (May 31, 2002), 1.

⁵¹FAA's Airmen Registration Database is a searchable set of files that is updated monthly (see www.faa.gov/licenses_certificates/airmen_certification/releasable_airmen_download, last accessed Mar. 1, 2009) and maintained by FAA's Airmen Certification Branch in Oklahoma City. Since 2000, airmen may opt to restrict public access to information in the database, including their name, address, and ratings, in accordance with the Wendell H. Ford Aviation Investment and Reform Act for the 21st Century, Public Law 106-181 (Apr. 5, 2000).

survey was first being developed, limited information for all pilots in the Airmen Registration Database was publicly available as the Airmen Directory Releasable File. In 2000, after the field trial but before the full air carrier pilot survey was about to be implemented, FAA began allowing pilots to opt out of the publicly releasable database. NASA officials told us that the team had considered asking FAA for the full database but decided against formally pursuing access to it for several reasons. These included ensuring continuing access to a public, updated database; ensuring access to a database that contained contact information for pilots; and maintaining independence from FAA as an aviation regulatory agency. Also, NASA was concerned about using the full data, because it wanted to maintain the privacy of pilots who had removed their names from the list explicitly to avoid contacts from solicitors, purveyors, or the like.

NAOMS staff had access to the full database when it was still publicly available in 2000 for the air carrier pilot survey's field trial sample. However, NASA officials believed that they could not use it for the full-scale survey from 2001 to 2004 because the nature of the frame—in terms of how well it represented the current air carrier pilot population—would change over time. Instead, the team decided to use as the frame for the full-scale air carrier pilot survey the Airmen Directory Releasable File that excluded pilots who had opted out; this file was regularly updated over the course of the air carrier pilot survey.⁵² The choice of frame may have been appropriate, given programmatic constraints, but posed several challenges. First, pilots in the publicly available Airmen Directory Releasable File were not necessarily representative of pilots in FAA's full Airmen Registration Database. Second, the database lacked information on whether airmen actively flew for a commercial airline. Lastly, only a relatively small portion of the 688,000 pilots in the database at the time of the field trial were air carrier pilots.

Potential Effect of the Opt-out Policy

NAOMS staff, realizing the potential limitations of using the publicly available data, were concerned about whether the frame provided adequate coverage of the target population or introduced bias into the data—that is, whether pilots in the public, opt-out database were sufficiently representative of air carrier pilots overall. For example, ALPA

⁵²The NAOMS sample was drawn from the version of the Airmen Directory Releasable File that was posted at www.landings.com (last accessed Mar. 1, 2009).

had provided its membership (which comprises approximately two-thirds of air carrier pilots) with information about the opt-out policy and with a form letter to pilots to facilitate their removal from the list. It is, therefore, possible that ALPA pilots removed their names from public access at a higher rate than non-ALPA pilots.

NAOMS researchers' analysis suggests that air carrier pilots may have removed their names from the public database at a disproportionately greater rate than did general aviation pilots.⁵³ One Battelle statistician expressed concern to other NAOMS team members that the sample, therefore, might not represent the population of interest. To help assess potential bias as a result of the opt-out policy (and the filter, discussed in the following text), researchers added a question to the survey—part way through the data collection phase—asking pilots to identify the size category of the aircraft fleet of the air carrier for which they flew. This information would allow for a comparison with air carrier fleet sizes known to exist in the NAS.⁵⁴

Identifying Air Carrier Pilots from the Sampling Frame

The database from which the project drew its sample of pilots lacked information on where the pilots worked and, therefore, could not be used to identify pilots flying commercial aircraft. The incidence of air carrier pilots in the full Airmen Registration Database was fairly low—approximately one in seven pilots would have been an air carrier pilot. (We could not find documentation on the number or proportion of air carrier pilots in the opt-out database, but we believe it to have had a similarly low incidence.) Therefore, the NAOMS researchers decided to use a filter to increase the likelihood that those contacted for the survey would be air carrier pilots.

⁵³The NAOMS team compared the pilots in its field trial air carrier sampling frame (based on the full FAA directory) with the publicly available opt-out directory for the next year, and found that 39 percent of its sample pilots were not available in the new directory. This contrasted with the team's understanding that roughly 8 to 10 percent of all pilots had opted out of the public list by 2002. The team reported that an unknown portion of the greater attrition was likely to have resulted from air carrier pilot retirements or withdrawals for medical reasons.

⁵⁴Preliminary analysis of the question was eventually used to illustrate that NAOMS data overrepresented large air carriers and underrepresented small carriers.

The filter required that pilots be U.S. residents certified for air transport, with flight engineer certification and a multiengine rating—a rating that sets specific standards for pilot experience and skill in operating a multiengine aircraft. By construction, all pilots in the public (opt-out) Airmen Directory Releasable File who did not fulfill these filtering requirements fell into the sampling frame to be used for the general aviation survey. After the filter was applied, the final frame for air carrier sampling had approximately 37,000 pilots in the first several quarters; records on the size of the frame’s later quarters were not maintained.

With these filtering criteria, approximately 70 percent to 80 percent of those contacted for the air carrier sample were, in fact, air carrier pilots who had flown within the recall period specified on the questionnaire. Although the contractor collected some information on pilots who were contacted but deemed ineligible for the survey, the data were not analyzed specifically to establish how effective the filter was at identifying air carrier pilots, even if they did not qualify for the survey. Without data on which people were excluded because they were general aviation, rather than air carrier pilots, these pilots would be wrongly omitted from the sampling frame for the general aviation survey.

As data collection progressed, the NAOMS team realized that the data were biased toward more experienced pilots, pilots flying primarily as captains, and pilots flying widebody aircraft over longer flight times.⁵⁵ After extensive analysis of the observed bias, the team attributed the bias primarily to two of the four filtering criteria—that is, that pilots were required to have both air transport and flight engineer certifications. Team researchers explored various strategies for addressing the observed bias and made several recommendations for data collection and analysis. The team considered whether using stratification to select samples according to alternative or additional characteristics would help reduce the observed bias toward more experienced pilots flying larger aircraft, but it eventually decided against changing the sampling strategy midsurvey.⁵⁶

⁵⁵The NAOMS data appeared to be biased in comparison with BTS benchmark data and air carrier pilots in the general aviation sample.

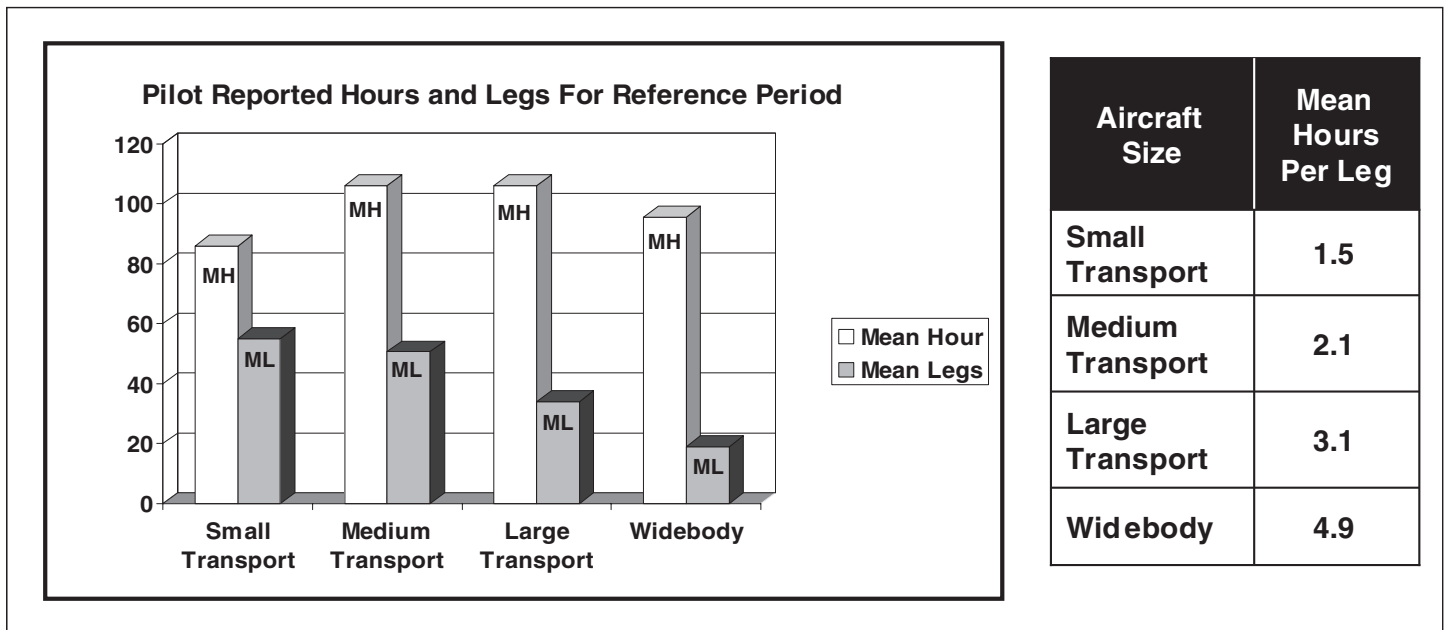
⁵⁶A stratified sample involves dividing the sampling frame into mutually exclusive subgroups thought to be similar on the basis of available information on each case; simple random or systematic random samples are then selected from within each subgroup. Researchers often use stratification to help ensure that they obtain reasonably precise estimates for each subgroup of interest in a population.

To determine whether the filter systematically excluded certain types of respondents—for example, air carrier pilots flying smaller aircraft or pilots with less experience—the NAOMS team recommended capitalizing on the implementation of NAOMS’s general aviation portion. The sampling frame for the general aviation survey included all pilots not filtered into the air carrier sample. Accordingly, project staff could examine the characteristics of air carrier pilots who fell into the general aviation sample because they did not meet filtering requirements, to establish whether they differed notably from those surveyed using the filtered sample. Preliminary analysis confirmed that pilots surveyed from the filtered sample exhibited systematic differences from air carrier pilots in the general aviation survey. Specifically, pilots surveyed with the air carrier sampling filters overrepresented captains and international flights, underrepresented smaller aircraft and airlines, and overrepresented the largest aircraft and airlines.

Following these analyses, the NAOMS team advocated incorporating operating characteristics into all analyses to mitigate potential bias. For the most part, the team recommended using operational size categories—that is, small transport aircraft and medium, large, and widebody aircraft—to stratify and possibly weight analyses, since different types of aircraft face different event risks and since safety issues may be more or less serious, depending on operating characteristics or aircraft make and model.⁵⁷ The team’s presentations of preliminary results frequently incorporated such analyses, as shown in figure 5. While other operational stratifications were suggested, such as specific aircraft make and model, it was acknowledged that this kind of analysis would dramatically reduce the effective sample size available for analysis in each category. A smaller effective sample size would decrease the precision of estimates from the survey, making it more difficult to detect changes in rates over time, especially for infrequent events.

⁵⁷The team categorized aircraft in its preliminary analyses into four operational size categories (see fig. 5): small transport = less than 100,000 pounds per gross takeoff weight (GTOW); medium transport = 100,000 to 200,000 pounds GTOW; large transport = more than 200,000 pounds GTOW with a single aisle; and widebody = more than 300,000 pounds GTOW with two aisles.

Figure 5: NAOMS's Preliminary Estimates of Pilot-Reported Flight Hours and Flight Legs, by Aircraft Size, 2002



Source: Linda Connell and Mary Connors, "National Aviation Operation Monitoring Service (NAOMS)," presentation to the Aviation Safety and Security Program Office (Hampton, Va.: NASA, Dec. 5, 2002), 44.

Additionally, to the extent that the data were to be analyzed as rates per flight leg or flight hour, an analysis segregated by operational characteristics would represent a fair description of these rates if it were assumed that the data adequately represented aircraft and pilots experiencing safety events within those operational categories—for example, if the widebodies and their pilots in the sample were fairly representative of air carrier widebody aircraft and pilots in the NAS.

Sample Size Calculations May Have Curtailed Statistically Reliable Trend Estimates for All Questions

NAOMS aimed to generate statistically reliable rates and trends that would allow analysts to identify a 20 percent yearly change with 95 percent confidence. However, the ability to detect such trends depended not only on the sample size, but also on the frequency of events. One statistician who had worked with the project team reported recently that detecting changes in trends of very rare events, such as complete engine failure, would require a prohibitively large sample of approximately 40,000 pilots. NAOMS's sample sizes were insufficient to allow analysis of all questions on the air carrier pilot survey or to accommodate analytical strategies that researchers eventually deemed necessary after data collection had begun, such as analysis by aircraft size category.

During the field trial, sample sizes were calculated to distinguish response rates between the three data collection methods (face-to-face and telephone interviews and mail questionnaires) to answer questions such as the following: Did an 81 percent completion rate for telephone interviews differ significantly from a 70 percent response rate for mail questionnaires? Later sample calculations for the full survey focused more directly on establishing the ability to detect a 20 percent change in event rates over time. Data from the field trial were analyzed to estimate how frequently an air carrier pilot experienced each specific event, enabling the team to assess how reliably different sample sizes could detect increases or decreases of 20 percent. From the field trial data, the contractor estimated that 8,000 interviews would allow detection of changes in rates with 95 percent confidence for approximately one-half of the core safety event questions.

The team eventually settled on a sample size of approximately 8,000 cases a year, declaring in its application to OMB that this would be the minimum size required to reliably detect a 20 percent change.⁵⁸ The application clarifies that just 5,000 unique pilots would be interviewed in the first year to gather 8,000 completed surveys (4,000 in cross-sectional samples, and 1,000 in four waves of the panel), but sample size calculations submitted to OMB do not expressly consider the impact of the panel's smaller sample size on the ability of NAOMS data to detect trends.⁵⁹ In the 3 years after data collection experiments in recall and method were discontinued, the survey interviewed approximately 7,000 cases a year.

At the time the NAOMS OMB application was submitted, project staff did not have adequate data to know for certain how frequently individual safety events would be reported, or to know an exact number of interviews that could actually be attained in a year. The NAOMS OMB application reported that pilots experience certain events quite

⁵⁸NASA, *National Aviation Operations Monitoring Service Application for OMB Clearance* (Moffett Field, Calif.: Ames Research Center, June 12, 2000), 15. Under the Paperwork Reduction Act of 1995, OMB requires federal agencies seeking to conduct new surveys to submit an application that establishes the necessity of new data collection in light of other data systems; estimates cost and respondent burden; and provides specific details about the survey and its sampling, implementation, and likely use. See John D. Graham, Administrator, Office of Management and Budget, *Memorandum for the President's Management Council: Guidance on Agency Survey and Statistical Information Collections* (Washington, D.C.: Executive Office of the President, Jan. 20, 2006).

⁵⁹NASA, *National Aviation Operations Monitoring Service Application*, 15.

infrequently, without expressly calculating how well a sample size of 8,000 could generate reliable estimates for such events. The sample size calculations in the application also assumed that the first-year data could be aggregated across recall periods and both the panel and cross-sectional data collection approaches that were used. NAOMS project staff later told us that further analysis would be essential to establish whether rates and trends generated from different recall periods and data collection approaches were sufficiently similar to allow combining the data. NASA believes that, even without data from the experimental period, the subsequent 3 years of air carrier pilot data were sufficient to demonstrate the survey's capability of detecting trends reliably.

Partway through data collection for the full air carrier pilot survey, NASA's contractor conducted simulations using early NAOMS data to better establish sample sizes at which 20 percent changes in rates for individual questions could be detected. These data confirmed that a sample of 8,000 cases a year would be sufficient to detect a 20 percent change for roughly one-half the core safety event questions, assuming all cases were analyzed simultaneously. By this point, however, the project team had already established the importance of breaking out NAOMS's estimates according to the size category of the aircraft flown to compensate for operational differences and the effects of the sampling procedures that we have previously described. Thus, sample size calculations may have overstated the ability of the NAOMS data to reliably detect trends at given significance levels, if segregating answers by operational characteristics is critical. Additional simulations that accounted for likely analytical considerations would be essential to determine whether the NAOMS project could attain its goal of measuring 20 percent changes in rates of different safety events with statistical confidence.

Sampling and Design Decisions
Bear on NAOMS's Rate
Calculations and
Characterization of the
National Air Space

When analyzing NAOMS's data, researchers must consider the effect of several design and sampling decisions that the project team made to accommodate pilots' confidentiality and the infeasibility of directly sampling all flights in the NAS. For example, the likelihood that a particular event would be reported by a pilot responding to the NAOMS survey increased with the number of crew witnessing the event and the number of aircraft involved. However, in designing a questionnaire to lessen the likelihood of respondent identification, the NAOMS team decided not to link pilots' reports of specific events to particular aircraft flown during those events or on the dates on which those events happened. Furthermore, the team's choice of sampling frame and filter resulted in a disproportionate selection of captains relative to other crew members. While sampling and design choices were rational in light of

concerns about confidentiality and program independence, such decisions have had implications on how to calculate and interpret rates from NAOMS and on whether analysts can extrapolate the data to characterize the national air space. NAOMS staff failed to identify specific analytical strategies to accommodate these issues in advance of data collection.

Using NAOMS Data to Calculate Rates and Trends

Survey design and sampling decisions affect how rates from NAOMS data can be calculated. For example, the NAOMS survey has the potential to collect multiple reports of safety events if more than one crew member on an aircraft or crew members on different aircraft observed the same safety event.⁶⁰ Safety events happening on aircraft with more crew members would also have had a greater likelihood of being reported, since more individuals who experienced the same event could have been subject to selection into the sample. These issues are not a problem, unless researchers fail to address them appropriately in an analysis.

Analytic goals must determine whether one adjusts for the potential that an event is observed by multiple crew members in the sampled population. Given that one of NAOMS's goals was to characterize the rate at which individual air carrier crew members experienced events per flight hour or flight leg, and assuming all crew members in an aircraft were equally likely to be sampled, multiple crew members observing an event involving one aircraft would not pose a problem. However, other considerations bear on whether and how to make adjustments. For example, bias resulting from the sampling frame and filter suggests that captains were more likely to have been selected into the air carrier sample than first officers or other crew members; additionally, many pilots flew in more than one crew capacity during the recall period. Events involving multiple aircraft also complicate estimates, partly because individuals not qualified for the air carrier pilot survey might have flown many of these aircraft. Extrapolating from individually derived rate estimates to system counts would also require making substantial assumptions and adjustments (see the following text).

One potential strategy to address the possibility of multiple observations of the same event would be to allocate events according to the number of

⁶⁰NASA officials and project staff have used the phrase "double-counting" as shorthand to denote the potential for an event to be reported by more than one pilot.

crew members who might have witnessed them (more details on alternative strategies are in app. D). For example, a report of a bird strike from a pilot flying a widebody aircraft with two additional crew members could be counted as one-third of a bird strike. Appropriate allocation presumes, however, that the analyst can identify the number of crew members present for any given report of a safety event. In general, the NAOMS recall period extended over 60 days, during which some pilots flew two or more types of aircraft of different size categories, implying different numbers of crew.⁶¹ Additionally, the questionnaire did not allow a pilot who flew more than one aircraft to identify which aircraft a reported safety event was associated with or in which role he or she served as crew. Analysts seeking to address the potential effect of multiple reports of the same event would have to develop allocation strategies that account for these design issues.

Researchers must also develop allocation strategies for other aspects and types of analysis using NAOMS data, such as trends or rate estimates for different aircraft types. We have previously mentioned that the NAOMS team recommended analyzing data by operational size category because of sampling considerations and because the effect and exposure to certain risks varied by class of aircraft. They also noted the importance of seasonal variations in relation to safety events—for example, icing is less likely to be a problem in summer than winter.

In its preliminary analysis, the NAOMS team attempted to resolve the issue of seasonal assignment by using nonproportional allocation strategies. The team used a midpoint date of the recall period—for example, October 1 if an interview recall period ran from September 1 to October 30—to determine a seasonal assignment for each interview in the analysis. For pilots flying different aircraft during the recall period, team members assigned an operational size class, based on the aircraft predominantly flown. For pilots who reported flying different operational sizes of aircraft equally over the recall period, project staff used a random number generator to determine the size class for preliminary analysis.

Extrapolating to the National Airspace System

The NAOMS team disagreed on the survey's ability to provide information on systemwide event counts versus rates and on trends based on

⁶¹The first-year data included 30-day, 60-day, and 90-day recall periods.

individuals' risk exposure. In preliminary analysis, the contractors often used BTS data to weight NAOMS data to generate systemwide event counts for air carrier operations in the NAS, and to provide baseline measures to assess potential bias resulting from sampling and filtering procedures.⁶² Since BTS's data collection processes changed during the NAOMS data collection period, however, the contractor stopped using these data to weight its estimates.

Because of the distinction between the NAOMS's unit of analysis and the sampling frame, as well as other sampling issues we found, it may not be possible to establish systemwide event counts for air carrier flights from the NAOMS data without using an external benchmarking dataset. However, extrapolating to systemwide event counts was not an explicit goal of the project. To the extent that analysts seek to use an external dataset to weight the NAOMS data in estimates of systemwide counts, that dataset's collection procedures and reliability would require assessment. Additionally, caution should be exercised, since changes in data collection or editing procedures over time could confound actual trends with changes resulting from variations in any external weighting dataset.

The Survey's Implementation

We found that NAOMS researchers followed generally accepted survey principles for many aspects of the survey's implementation, with some limitations. Sample administration, information systems, and confidentiality provisions appear to have been adequate, and telephone interviewers were successful in administering technical questions and attaining high completion rates. However, despite adequate records of data editing and checks, analysis and interpretation of NAOMS data are complicated by first-year experiments in recall period and data collection approaches and CATI programming choices, along with sampling and design decisions. Researchers did not conduct full data validation or nonresponse bias assessments to ensure the quality of the data. We found deficiencies in record-keeping and moderate implications for the risk of survey error; the potential survey errors involved processing, sampling, and nonresponse.

⁶²These baseline measures include flight hours and legs flown by commercial aircraft.

Information Systems and
Sample Management
Maintained Confidentiality, but
Data Checks and Record-
Keeping Were Limited

We found several issues with NAOMS information systems. Sample administration and management, including notification of and informational materials for pilots and release of sample for interviewing, met generally accepted survey principles. Pilot confidentiality seriously concerned project staff, and steps to protect confidentiality appear to have been adequate. In contrast, CATI programming and data checks, along with record-keeping, had greater limitations.

Sample Administration and Management

Taking its sample from the Airmen Directory Releasable File, NAOMS sampled using pilots' certificate numbers, with a filter designed to target air carrier pilots. After adjusting for duplicate certificate numbers that had entered the sample some time in the previous year (regardless of whether an interview was completed), the team obtained pilots' updated addresses from the U.S. Postal Service's change-of-address file and submitted them to Telematch to obtain telephone numbers for each address.⁶³ This process resulted in an approximately 60 percent match of addresses to telephone numbers, which researchers saw as sufficient because they believed the Airmen Directory included some records for individuals who had retired or were deceased. Each quarterly sample was then divided randomly into 13 parts to be released weekly. On the Friday before each week's release, project staff sent pilots a notification on NASA letterhead that described the study and its confidentiality provisions and informed them that an interviewer would be calling. To pilots for whom Telematch could not provide a valid telephone number, or who had "bad" numbers from the field trial, project staff sent postcards asking them to call NAOMS interviewers directly or to send in an updated telephone number.

The project team monitored the disposition of the sample on a weekly or quarterly basis, including the proportion of respondents who were ineligible, refused, or could not be located. While between 17 and 29 percent of pilots in each quarterly sample could not be located, and consequently were not interviewed, approximately 5 percent of the completed interviews resulted from cases that had not been matched to a telephone number through Telematch. The NAOMS team aimed initially

⁶³Telematch, in Springfield, Virginia, is a national database that consists of some 170 million directory assistance consumer and business listing records sourced directly from telephone companies and updated daily. The NAOMS team used Telematch to find telephone numbers based on each pilot's address in the Airmen Directory. See Telematch at www.telematch.com (last accessed Mar. 1, 2009).

for a 6-week fielding period, or “call window,” to allow interviewers sufficient time to call back each nonresponding pilot in the sample before assigning the case a final disposition (such as “no-locate” or “refusal”) and removing the pilot from the sample. However, researchers found that a 3-month call window was necessary to attain a sufficient response rate. The team did not indicate having compared the answer patterns of pilots they reached early in the sample with the answer patterns of pilots who were hard to track down, to ensure the patterns were comparable across the full sample field period.

Information Systems and Pilot Confidentiality

The survey’s management techniques and documentation for interviewers indicate that the NAOMS project team was particularly attentive to confidentiality. The questionnaire did not ask pilots to link safety events to specific flights, airlines, or times. Interviewers were informed that “Battelle [can] not link data items with individual pilots. All reports will be presented using aggregate information.”⁶⁴ Battelle used separate systems to track the sampling and to store the interview data, which ensured that pilots’ answers could not be linked to any identifying information. In the system with sampling information, the specific date of each interview was not recorded, only the week in which it happened. The *NAOMS Reference Report* described NAOMS’s responses as “functionally anonymous” and suggested that the promise of confidentiality enhanced the respondents’ rapport with the interviewers.⁶⁵

The NAOMS team never sought to release unedited individual-level data from the survey. The project’s OMB application describes plans for ensuring the confidentiality of respondents, including provisions for confidentiality statements on behalf of interviewers and staff, and separate computer systems for sampling and interviewing so that respondents’ answers could not be linked back to identifying information. The application also states that:

⁶⁴NASA, *Interviewer Training Manual from Year 1* (2001), I-10.

⁶⁵Battelle, *NAOMS Reference Report*, 8. See also NASA, *Interviewer Training Manual from Year 1*, I-10.

-
- “• 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.”⁶⁶

Among analytical products for the aviation community, researchers planned to release summary reports and “structured, fully de-identified datasets.” According to a presentation at the first NAOMS workshop, NAOMS products would be subject to FOIA after they were in “a finished state.”⁶⁷ NASA officials told us that they agreed that there would be little risk of violating pilots’ confidentiality if data were released in aggregate as initially was planned.

In meetings with NASA, as well as in the agency’s written comments responding to our draft report, officials expressed serious concern about the importance of protecting pilots’ identity, a concern we share. The officials offered several specific examples of how they felt NAOMS data could be used to identify individual pilots. However, many government agencies that collect sensitive information, such as the Institute for Education Sciences, the Census Bureau, and the National Center for Health Statistics, have successfully allowed individual researchers access to extremely sensitive raw data on individuals. These agencies have effectively addressed the issue of individual privacy by, for example, requiring researchers to attain clearance to use data that could reveal sensitive information, to sign nondisclosure agreements, and to submit to stiff penalties for noncompliance. Additionally, agencies may restrict the types of analyses that can be performed with the data, where data can be analyzed, and how the data are reported. For example, the National Center for Health Statistics may prevent researchers from accessing table cells that contain fewer than five observations to lessen the likelihood that an individual respondent can be identified.

We realize that given the evolution of data mining techniques, one could conceive of a full, raw NAOMS dataset being linked to proprietary information from airlines or a host of other safety systems in ways that

⁶⁶NASA, *National Aviation Operations Monitoring Service Application*, 9, sec. I-J.

⁶⁷Connell, *Workshop on the Concept of the National Aviation Operational Monitoring Service*, 66.

might enable a dedicated data analyst to identify a particular pilot from the air carrier survey.⁶⁸ This breach seems unlikely to happen, however, given the relative absence of identifiable information in the survey data and the lack of connection between the tracking database and the CATI data. If the survey were to be implemented as it was planned and the data released publicly only in aggregate, the confidentiality provisions of the air carrier pilot survey appear to have been adequate. The risk that individual pilots might be identified from the raw data would be greater for the general aviation survey, which involved a wider range of aircraft types, several of which might be linked to very small populations of pilots.

NASA officials also expressed concern that pilots might have understood NAOMS's promises of confidentiality as conferring the kind of legal protection that voluntary reporting to a system like ASRS provides. We found no evidence substantiating or refuting this understanding. To the extent that confidentiality protections in NAOMS were adequate, any fear that pilots would invoke legal protections that did not exist are unfounded.

CATI Programming and Data Checks

Partly because NASA emphasized the importance of not second-guessing pilots, and partly because project staff wanted to avoid truncating answers unnecessarily, the contractor built only limited edit checks into the CATI data collection system, despite initial plans to the contrary. The questionnaire used in training interviewers identified one structured prompt for the number of hours a pilot reported having flown during the recall period.⁶⁹ It did not include any other instructions to recheck values

⁶⁸NASA's concerns about pilot confidentiality underlie the agency's recent efforts to develop a redacted version of the NAOMS data for public release.

⁶⁹A structured prompt is a scripted instruction available for an interviewer to clarify a respondent's question or response. When an interviewer enters a respondent's answer into the computer system and the entry does not meet certain criteria that have been established by the survey's designers, the system may be programmed to provide interviewers with a structured prompt that is to be read verbatim to clarify a respondent's answer. For example, if the data system recorded a value of 1,000 for the number of hours a pilot flew in a week the interviewer would be instructed to ask the respondent whether that value was correct. Structured prompts ensure that the interaction between interviewers and respondents is consistent and help to mitigate the effects of misunderstandings and data entry problems, in that an unusually high or low value that persists after a structured prompt can be treated as an outlier, rather than as a typing error or miscommunication.

reported for specific questions if they seemed unreasonable (perhaps indicating mistyping or an interviewer-respondent misunderstanding).

Although the contractor documented edits and quality checks that it performed on the collected data, the CATI system may not have included all initially planned edit checks. The final questionnaire for interviewer training suggests that additional edit checks were built into the CATI system, but the contractor's data editing protocols suggest that the edit checks were not consistently integrated into the program. For example, when pilots were asked to break the time that they flew different aircraft into percentages—such as 50 percent of the time flying a Boeing 737, 25 percent flying a McDonnell Douglas MD-80, and 25 percent flying a Boeing 727—the CATI system was supposed to have forced interviewers to reenter information if the responses did not add to 100 percent. Therefore, if, for example, the interviewer had mistakenly entered 25 percent for each of the three separate aircraft categories, the total percentage (75 percent) should have triggered the CATI system to force the interviewer to reenter information until it added to 100 percent, but the system did not in a handful of cases.⁷⁰ Although such anomalies were extremely rare in the air carrier pilot data, multiple managerial reviews and tests of the CATI programming before the survey was implemented failed to identify the anomalies in advance of survey fielding.

For many of the questions that pilots were asked, the concern that answers not be truncated unnecessarily by imposing predetermined edit checks seems reasonable, given that the goal was to generate statistically reliable information on aviation safety that was otherwise unavailable. For other questions, such as those on total engine failure and other rare events, input from aviation experts and operational staff would have helped in constructing thresholds for the checks in the CATI system. The additional data would have helped analysts distinguish between true outliers and data entry errors and between interviewer and respondent misunderstandings.

Survey completion rates were relatively high, and the NAOMS team reported exceptionally few break-offs partway through the interviews. It is

⁷⁰The raw data suggest that the CATI programming appropriately prevented interviewers from entering responses in cases where the sum of the aircraft flown questions would exceed 100 percent, but not in cases where the sum was less than 100 percent. Similarly, another set of questions included cases for which the sum of events from individual question subparts erroneously exceeded the top-level question value.

impossible to know for certain whether the high completion rates were because interviewers did not second-guess pilots by asking them to repeat answers that researchers had deemed unlikely. To the extent that interviewer rapport with pilots was enhanced because the pilots were not second-guessed, the decision to limit the number of built-in CATI edit checks may have enhanced the completion rates, at the expense of complicating data cleaning and outlier identification.

Record-Keeping

NAOMS record-keeping was fairly decentralized. While many of the individual steps of the NAOMS project appear to have been documented in some form, the project staff and contractors did not assemble a coordinated, clear history detailing the project's management that would facilitate evaluation of the overall air carrier pilot survey. Information on the project's steps is largely dispersed across a series of contracts and modifications between NASA and Battelle and internal NAOMS team documents on individual pieces of the project. The lack of summary documentation for various aspects of the project makes it difficult to (1) distinguish between what was planned at the beginning of the project and what phases were accomplished in later years, following NASA priority changes for NAOMS's resources, and (2) assess whether aspects of project and budget management raised the potential risk of survey error.⁷¹

Regarding the sample, the contractor kept limited information on the size of the frame before and after filtering to identify air carrier pilots. The size information the contractor maintained was not enough to reconstruct the sampling fraction—the percentage of pilots sampled each quarter from the filtered frame—for all quarters of the air carrier pilot survey. Additionally, Battelle's procedures for maintaining pilot confidentiality aimed to make it extraordinarily difficult to identify which pilots were in the sample frame at any given time. At the time of sampling, Battelle maintained enough information to remove pilots who had already been sampled from future samples for the next four quarters. Battelle did this partly because the population was relatively small, and because they did not want to interview the same pilot more than once a year. Although the contractor

⁷¹Chester Bowie, National Opinion Research Center, "Review and Evaluation of the Survey Management Component of the Air Carrier Survey in NASA's National Aviation Operation's Monitoring Service (NAOMS)," paper prepared for GAO (Bethesda, Md: Aug. 29, 2008), 3–4.

Experiments in Data Collection and Recall Period Length May Have Restricted the Utility of the First-Year Data

lacked formal records, it estimated that the procedure led to the exclusion of approximately 20 percent of the filtered sampling frame in any given year.

Regarding NAOMS data, the lack of sampling records prevents analysts from leveraging sampling information when producing estimates or calculating sampling errors. Furthermore, the lack of these data hinders the kinds of nonresponse bias analysis that the project team originally planned. Without reliable information on the proportion of cases that were removed from the sample in any given quarter, analysts must rely on more conservative variance estimates than might have been necessary, making the detection of changes over time more difficult.

Two main experiments that NAOMS researchers conducted in the initial year of interviewing may have restricted the utility of first-year data. Because the field trial had not resolved the optimal length of time the survey's questions should cover, researchers used the final survey to test first two and then three different recall periods for several months. Subject matter experts on the team also advocated a second experiment to determine the relative merits of a panel or cross-sectional data collection approach. NASA officials told us that they viewed the first months of the survey as part of a development phase, rather than full implementation of the survey. Nevertheless, NAOMS project staff have noted that adequate research on the feasibility of combining data from the experimentation has not yet been done. Depending on the results of such research, it may be imprudent to evaluate NAOMS's first-year responses as if they were similar to the trend data collected in subsequent years.⁷² Approximately one-quarter of NAOMS air carrier pilot survey interviews were collected under experimental conditions; the subsequent 3 years of the survey used a cross-sectional data collection approach with a 60-day recall period.

Panel or Cross-Sectional Approach

As interviewing for the full survey began, project staff had not reached consensus on whether to use a panel or a cross-sectional approach for data collection.⁷³ Panel data are observations collected on the same sample

⁷²Additionally, the events of September 11, 2001, affected the airline industry and may have had an impact on the nature of subsequently collected data, according to NASA officials.

⁷³A November 2000 team document showed 2 arguments in favor of a panel approach and 10 against.

of respondents over a period of time. Cross-sectional data are observations collected on respondents at a single point in time.⁷⁴ While some team members opposed the panel approach because of potential respondent attrition, others thought that it might “encourage participants to become even more acute observers of aviation system safety” and “produce a higher response rate and higher response quality.”⁷⁵ However, confidentiality procedures that removed the link between the sample tracking system and respondent’s answers meant that panel data would not necessarily provide repeated observations for analysis in the NAOMS data. According to the interviewers’ manual,

“We will be asking panel members to give us a code word that we can use to link interviews, but this code word will not be kept in our tracking system. Pilots forgetting the word will not have their data linked.”⁷⁶

The NAOMS team decided to begin its first full year of air carrier data collection using both panel and cross-sectional approaches.

After analyzing the first half-year of data, the team noted that, among other things, the panel approach may have heightened pilots’ awareness of the timing of safety events but not the number of events recalled.⁷⁷ The project team decided, for the following four reasons, to abandon the panel design in favor of cross-sectional data collection: (1) the panel design resulted in fewer independent observations; (2) the panel design was logistically difficult to administer; (3) NAOMS’s confidentiality procedures made analyzing repeated observations over time impossible (the proportion of pilots who remembered the password and thus could have data linked was not reported); and (4) the cross-sectional design had yielded a sufficiently high response rate to allay worries that pilots would be unwilling to respond unless enlisted as panel members.

⁷⁴See, for example, GAO, *Designing Evaluations*, [GAO/PEMD-10.1.4](#) (Washington, D.C.: March 1991).

⁷⁵Battelle, *NAOMS Reference Report*, 35, app. 9-6.

⁷⁶NASA, *Interviewer Training Manual from Year 1*, I-10.

⁷⁷A contractor document about potential bounding effects (whereby the time from one panel survey interview to the next provides a mental benchmark for the respondent) and a 7-day recall period demonstrate how collection method and recall period can affect the nature of the data collected. To the extent that different approaches resulted in substantive differences in these data, results from interviews collected under different methodologies should not be combined.

Recall Period

As we have previously discussed, the lack of literature on pilots' recall, in particular, and the wide variation in the literature's recommended recall periods, more generally, made it difficult for the team to decide on the most appropriate recall period. Team members had extensively analyzed data from the field trial to determine any differences among the recall periods tested in that survey. Researchers' analysis showed that, as expected, respondents with longer recall periods reported having flown more hours and legs than those with shorter recall periods. Researchers' regression analysis also confirmed a positive relationship between recall period and the total number of events that pilots reported; the magnitude and statistical significance of this relationship was strongest between 2 weeks (14 days) and 2 months (60 days). Additionally, the team examined pilots' comments on whether their particular recall period had been appropriate.

Despite these analyses, the team decided to delay the decision on recall period until they had collected more data in the initial months of the full air carrier survey. After reviewing the field trial results and pilots' comments, the team was firm only in the belief that a 7-day period was too short, despite a small-scale experiment suggesting this period was optimal for pilots' memory of *routine* events. (However, a 7-day period would have been too short to capture *infrequent* risk events.) The team explored various tolerances for error, event periodicity, and cost before testing 30-day and 90-day recall periods in the survey's first two quarters of sampling.

After the first two waves of data collection, team members explored data on the length of the recall period. Then they tested a three-way split design, collecting an additional 2 months of cross-sectional data to assess whether 60 days would be the best compromise between the 30-day and 90-day periods. Using these data, the project team compared the mean event rate over time across all core safety event questions—noting that longer recall periods should result in pilots reporting more events—and the standard deviation associated with these rates, which declined as the recall period increased. However, the team did not analyze the relationship between recall periods and specific events or the correlation of exposure units (flight hours and flight legs) to safety events for the

different periods.⁷⁸ Eventually, staff chose 60 days as providing a reasonable balance between the recall of events and avoidance of error. According to NASA officials, the selected recall period was seen as a compromise between cost and reliability. Despite the theoretical merits of the analyses justifying this decision, researchers cannot independently confirm the accuracy of reporting under different recall periods without separate data validation efforts as part of the field trial or full survey. However, the practicality of efforts to validate respondent accuracy depends on the nature of the data being collected, the existence of alternative data sources, and the design of the questionnaire. As NAOMS's survey methodologist has observed, surveys would be unnecessary if a true population value were known.⁷⁹

Because NASA's objective in designing and implementing the NAOMS survey was to develop a data collection methodology, the team was warranted in deciding to use the first year of data analysis to resolve questions that had not been fully answered by the field trial. This is particularly true for their decision to test various recall periods that would help them find an appropriate balance between recall period and budget and sampling constraints. As we have previously mentioned, further analysis would be required to establish whether data collected during the experimentation can be combined with later data using only the 60-day recall period and cross-sectional approach. However, NASA officials told us that the subsequent 3 years of cross-sectional data collection with a 60-day recall period was sufficient to demonstrate the capability of the air carrier pilot survey to measure trends.

Experienced Professional Interviewers Administered Technical Questions

Training materials, questionnaire copies and revisions, specificity in interviewers' scripts, and cooperation among staff demonstrate that the team selected appropriate interviewers and was sensitive to key issues throughout the questionnaire's development. The NAOMS project team decided not to use aviation experts as interviewers in the belief that the "lack of expert knowledge can be a benefit since the interviewers are only recording what they hear rather than interpreting it through the lens of

⁷⁸Robert F. Belli, "NAOMS Survey Review," paper prepared for GAO (Lincoln, Neb.: Aug. 27, 2008), 7. Belli has noted that there were no individual-level validation data on which to make a determination of the efficacy of these different recall periods.

⁷⁹The survey methodologist also suggested that many federal surveys do not conduct independent data validation.

their own experiences.”⁸⁰ To mitigate issues that might have resulted from using interviewers unfamiliar with the subject matter, the team emphasized the importance of the clarity of the questions and consistency in how the interviewers read them and responded to the respondents’ questions.

The project staff emphasized the importance of using professional and experienced interviewers and giving them adequate training to administer the survey. NAOMS’s principal investigator told us that the interviewers Battelle used for the NAOMS survey were exceptionally professional and were accustomed to conducting interviews on sensitive topics.⁸¹ Interviewers received a training manual for the project’s first year, which included the following: a background on the rationale for the NAOMS survey, a description of how the survey could shed light on safety systems, the survey’s confidentiality protections, and information on the survey’s sampling and tracking information.⁸² They also received a paper copy of the questionnaire with interviewer notes, pronunciation information, and a glossary of aviation terms.

The NAOMS team conducted a series of cognitive interviews with pilots to learn whether they would understand the questions and whether the incidents they reported were those that the team sought to measure. These interviews led to questionnaire revisions to address potential ambiguities for both respondents and interviewers. Regardless of efforts to develop clear questions that interviewers could read directly and respondents could easily interpret and answer, the team acknowledged that certain questions turned out to be less reliable than others. For example, in considering a question series on the uncommanded movements of rudders, ailerons, spoilers, and other such equipment (see fig. 6), the team’s concern was that pilots might be unaware of these events or might interpret uncommanded movements as including autopilot adjustments.⁸³ The survey instrument did not include instructions to interviewers to clarify the intended meaning of this set of questions, and question

⁸⁰Battelle, *NAOMS Reference Report*, 29.

⁸¹Battelle’s 2007 report describes the interviewers’ training and certification for the field trial. See Battelle, *NAOMS Reference Report*, 7–8 and 29.

⁸²NASA, *Interviewer Training Manual from Year 1*, sec. I.

⁸³Ailerons and spoilers are parts of aircraft wings that provide (or decrease) lift, and that help to control the airplane’s stability in flight.

standardization alone could not overcome the questions' potential ambiguity, despite interviewers' skill.

Figure 6: NAOMS Air Carrier Questionnaire Section B, Question ER4 on Uncommanded Movements

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	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
b. Uncommanded movements of the rudder?	# RUDDER.....	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
c. Uncommanded movements of the ailerons?	# AILERONS.....	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
d. Uncommanded movements of the spoilers?	# SPOILERS.....	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
e. Uncommanded movements of the speedbrakes? .	# SPEEDBRAKERS.....	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
f. Uncommanded movements of the trim tabs?	# TRIM TABS.....	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
g. Uncommanded movements of the flaps?	# FLAPS.....	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
h. Uncommanded movements of the slats?	# SLATS.....	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
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.....	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
------------------------------	---

Source: Battelle Memorial Institute, *NAOMS Reference Report: Concepts, Methods, and Development Roadmap*, prepared for the NASA Ames Research Center (Nov. 30, 2007), app. 11-6.

In its quality assurance procedures, Battelle monitored and documented approximately 10 percent of the interviews. However, it did not record audio of the interviews. Battelle’s documentation states that the monitoring procedure took the form of live supervisory monitoring of interviews in progress, as well as callbacks to respondents to ask about their interviewing experience and to administer key questionnaire items again to see whether answers were reliable. However, NASA officials told us that the callbacks were never performed, in keeping with the project’s concerns about pilot confidentiality.

Telephone Interviews Attained High Completion Rates, but Validation Efforts Focused Primarily on Face Validity

While interviewers for NAOMS attained high completion rates from pilots in the sample, limited validation efforts hinder confirmation of data quality. Roughly 80 percent of sampled pilots thought to be eligible for the NAOMS air carrier pilot survey completed telephone interviews, and a notable portion of those who were contacted were found to be ineligible. The project team decided against conducting nonresponse bias analysis and did not pursue other formal data validation, focusing instead on the face validity of preliminary NAOMS rates and trends.

Completion Versus Response Rates

In public presentations and documents of air carrier pilot survey results, NAOMS staff often discussed the rate of sample cases that were located and the proportion of interviews completed. The completion rate, distinct from a response rate, surpassed 80 percent by the end of the air carrier survey. Throughout the air carrier survey, approximately 23 percent of those contacted were deemed ineligible because they were not commercial air carrier pilots or had not flown in the recall period. Additionally, approximately 24 percent of cases drawn for the air carrier sample were never located and, thus, their eligibility for the sample could not be determined.

A survey's response rate, defined, in general, as the number of completed interviews divided by the eligible number of reporting units in the sample, is often used as an indicator of data quality and as a factor in deciding to pursue nonresponse bias analyses or additional survey follow-up.⁸⁴ OMB's guidelines, although not yet formal when the NAOMS survey was implemented, call for a nonresponse bias analysis when survey response rates fall below 80 percent. OMB guidelines cite survey industry standards for response rate calculations; these calculations generally include either unknown sample cases or an estimate of likely eligibles among unknown cases, in the denominator of the calculations.⁸⁵ A calculation of response rates that excludes unknown cases rests on the assumption that all of

⁸⁴See, for example, American Association for Public Opinion Research, *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*, 5th ed. (Lenexa, Kans.: 2008), 34.

⁸⁵The first calculation would be the American Association for Public Opinion Research's response rate 1; the second would be response rate 3. In a calculation of rates, the numerator was data NAOMS collected on events, the denominator was data it collected on exposures per flight hour or flight leg. See Connors and Connell, "The National Aviation Operations Monitoring Service," 16.

those cases would have proven ineligible. For NAOMS data, a response rate calculation that included cases of indeterminate eligibility in the denominator (because the pilots could not be located) would be closer to 64 percent. If the cases not located fell out of scope at approximately the same rate as the cases that were located and contacted, the NAOMS response rate would be approximately 67 percent.

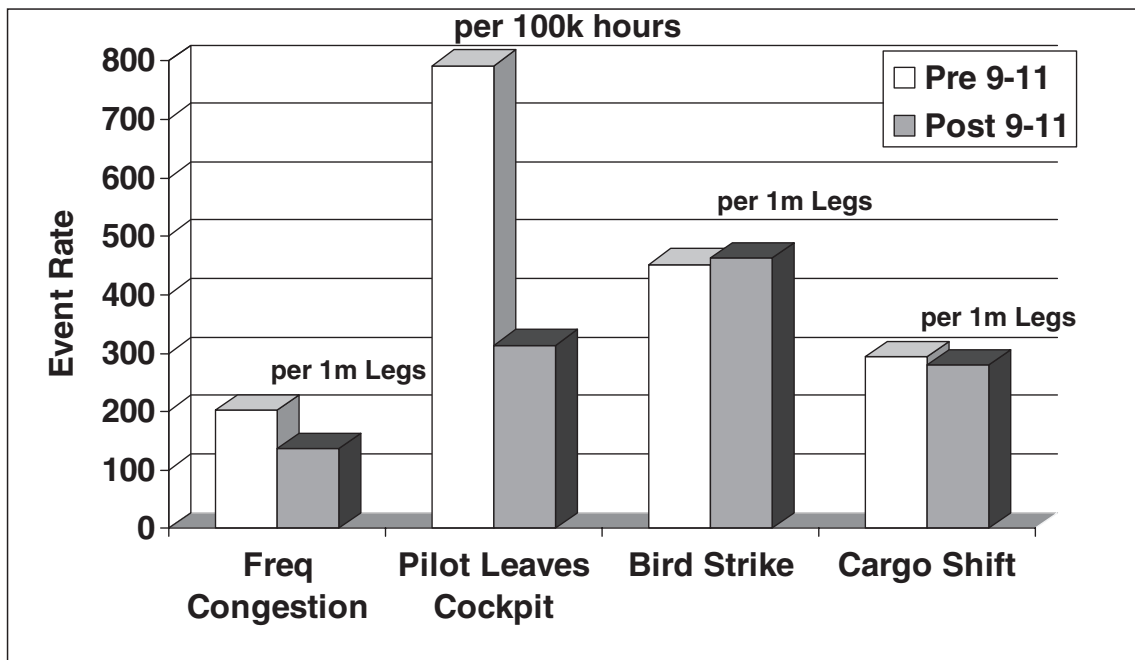
NAOMS staff told us that they decided against pursuing nonresponse bias analyses as initially planned because they thought that air carrier completion rates were quite high for pilots who were located and contacted and because NASA's priorities had changed, resulting in fewer resources for staff to complete such activities. However, more conservative calculations of response rates might have merited further scrutiny, such as a nonresponse bias analysis or other research into reasons for the sample rate of unlocated pilots. Comparing information from the sample frame respondents' and unlocated pilots' characteristics might have provided insight into any systematic differences between the two groups.

Establishing Validity

NAOMS project staff attempted to validate the data in a variety of limited ways. Besides the interview monitoring, they made preliminary calculations, such as a comparison of the hourly rate at which pilots left the cockpit to deal with passenger disturbances. They found that, unlike some other events, the rate dropped dramatically after September 11, 2001 (see fig. 7), which demonstrated the importance of enforcing existing rules requiring the cockpit door to be closed during flight. Other validation attempts included checking on the seasonality of events—for example, on whether reports of icing problems increased in winter.

Figure 7: NAOMS's Preliminary Findings On Pre- and Post-September 11, 2001, Event Rates

Pre and Post 9-11 Evaluation of Sample Events



Source: NASA, "National Aviation Operations Monitoring Service (NAOMS)," presentation to FAA (Washington, D.C.: Apr. 9, 2003).

The NAOMS staff recommended more formal validation efforts, suggesting the examination of questions that had been included in the survey specifically because they could be benchmarked against other FAA data systems, such as ASRS and the Wildlife Strike Database. Such work would have been complicated, however, by the decision to use NAOMS data to fill in data gaps from other safety systems and not to ask questions that directly overlapped them, even for items included for benchmarking. For example, NAOMS asked pilots about all bird strikes without establishing a

threshold for their severity. FAA does not, however, require pilots to report all bird strikes to its Wildlife Strike Database, only those bird strikes that cause “significant” damage. Additionally, aviation researchers have estimated that up to 80 percent of bird strikes with civil aircraft are not reported to FAA’s Wildlife Strike Database.⁸⁶ Therefore, it is not surprising that NAOMS data imply a much higher incidence of bird strikes than other systems.

In addition to considering examples such as pre- and post-September 11, 2001, rates, NAOMS staff had also examined other issues that had intuitive appeal, such as seasonal fluctuations in reported bird strikes.⁸⁷ Project staff also suggested that the data corresponded well with other data systems, citing as an example both runway incursions—a decline in which the NAOMS team attributed to an FAA policy change—and reserve fuel tank use—an increase in which had reportedly been seen in ASRS.⁸⁸ Additionally, for field trial data, project staff examined the strength of the relationship between the number of events reported and the hours flown or the length of the recall period, because pilots flying more hours or recalling events over longer recall periods should report more events than those with fewer hours flown or shorter recall periods. In addition to having face validity, the survey methodologist noted that the relationship between events reported and flight hours and legs is also a measure of construct validity, in that it demonstrated that NAOMS’s measures corresponded well with theoretical expectations. However, the relationship does not confirm whether the events that pilots reported actually happened. No other data validation efforts were undertaken on the full survey.⁸⁹ NAOMS project staff reported that several questions in

⁸⁶Sandra E. Wright and Richard A. Dolbeer, *The National Wildlife Strike Database for the U.S.A.: 1990 to 2002 and Beyond*, Bird Strike Committee Proceedings 2003, Bird Strike Committee U.S.A. and Canada, 5th Joint Annual Meeting, Toronto (Lincoln, Neb.: University of Nebraska, 2003), 1.

⁸⁷The survey methodologist for NAOMS reported that the pattern of birdstrikes followed the expected seasonal pattern, which helped to give the researchers additional confidence in the validity of the data.

⁸⁸In 2002, FAA completed a national runway safety plan with 39 safety objectives, such as enhancing runway markings and lighting. For more details on FAA policy on runway incursions, see GAO, *Aviation Runway and Ramp Safety: Sustained Efforts to Address Leadership, Technology, and Other Challenges Needed to Reduce Accidents and Incidents*, [GAO-08-29](#) (Washington, D.C.: Nov. 20, 2007).

⁸⁹NAOMS staff provided us with a list of potential benchmarking questions and corresponding data sources to facilitate future analysis.

Stakeholders Disagreed on the Utility and Value of the NAOMS Data

the NAOMS data had face validity, but the data still had to be benchmarked. While such benchmarking is critical for validating NAOMS data, it may not be sufficient to confirm the accuracy of pilot recall for most NAOMS questions or to estimate the potential effect of nonresponse bias.

The effectiveness of NAOMS as a monitoring tool depended on its ability to provide reliable and valid estimates to address customers' concerns. NAOMS team members promoted the survey's potential for generating rates and trends but also debated whether the data could be used to establish baseline counts of events for the NAS. NAOMS working groups were started but disbanded before resolving this issue or benchmarking the data against what was known from other safety data.

NAOMS Data and Systemwide Event Counts

NAOMS team members agreed that the survey was designed to measure the occurrence of events, rather than their causes. They did not clearly agree on the survey's ability to provide systemwide counts of events, rather than rates per flight hour or flight leg, or rate trends over time. According to the project's leaders, NAOMS was never intended to generate an absolute picture of the NAS (i.e., total counts of the number of events in the NAS each year). They told us that its utility was understood to lie in its ability to measure *relative* frequencies that could be used to generate trends over time. However, NASA's OIG found "a disparity between the stated goals of NAOMS and the manner in which NAOMS project management initially presented the data to FAA," a point that FAA also raised.⁹⁰ Senior FAA officials told us that NAOMS staff repeatedly indicated that the project would provide "true" estimates of rates of safety events in the NAS at the project's beginning, a capability that FAA disputed. NAOMS's emphasis on relative trends, which FAA believed NAOMS could depict, happened only in later stages of the project.

Regardless of whether NAOMS data were presented as counts or rates, the data were never designed to serve as a stand-alone system. The survey's methodologist told us that he believed that NASA staff were always clear about the goal of establishing rates and trends, but that in the absence of a baseline count of how frequently safety events occurred, these rates were insufficient to specifically quantify change from the survey's beginning.

⁹⁰NASA, "Final Memorandum," 13.

However, in theory, such data could be used to generate trends if the nature of any sampling and nonsampling error in data collection remained constant over time.

Additionally, the NAOMS survey methodologist described issues that might jeopardize inferences about trends based on hourly rates. For example, because rates per-exposure unit are a per-pilot measure, rather than a system or aircraft measure, one could incorrectly attribute a change in rates to a systemwide shift that might instead have resulted from a change in technology that affected the number of individuals in the cockpit crew. As we have previously mentioned, the sampling frame, the filter, and potential noncoverage and nonresponse issues would make further analysis necessary before one could conclude that NAOMS's measures of rates per-exposure unit could be generalized to the full population of air carrier pilots.

According to NASA's researchers, when the NAOMS contractors began to work closely with the data, they began to extrapolate and generate systemwide count estimates. NASA reported that one contractor believed it was essential to report system counts: that is, counts were necessary to convey the meaning of the data from a policymaker's perspective and rates did not convey the significance of a given result. Battelle staff used BTS data to weight NAOMS data according to systemwide numbers of flight hours or flight legs and used these estimates in several presentations of NAOMS preliminary results. The staff reported to us later that they had decided against weighting up to the full population of aircraft types because they did not think that it made sense to combine operational size categories of aircraft.

The early presentations of the NAOMS data raised concerns for FAA, because the numbers presented as systemwide estimates did not match FAA's other information sources. Several FAA and NASA officials with whom we spoke asserted that data from several specific survey items did not correspond with the content of other reporting systems. However, the items cited were not intended to overlap directly with data FAA had already collected. NASA officials conceded that how NAOMS defined the question wording might have contributed to one cited discrepancy. In addition, FAA officials thought NAOMS was unable to accurately measure systemwide rates of safety events and asked for extensive, specific revisions to the survey to address specific questions. Among other things, these officials wanted NAOMS to ask questions that were more investigatory in nature than the broad monitoring concept that NASA had envisioned. NASA did not make the changes that FAA recommended part

way through the survey. In correspondence with FAA, NAOMS researchers emphasized that the survey's ability to measure trends required consistent question wording. FAA officials were also concerned about the quality of NAOMS data because the survey's questions were based solely on pilots' perceptions.

NAOMS's Working Groups

NASA's project leaders reported that the working groups were to play a critical role in evaluating the validity of the NAOMS data and in establishing whether the survey's information seemed reasonable, given what was known about safety from other data sources.⁹¹ The two working groups, established in 2003 and 2004, were distinct from the two workshops conducted in 1999 and 2000, although the groups and workshops were similar in that they both aimed to introduce the NAOMS project to a wide range of stakeholders, including FAA and industry members, and that they solicited input on the survey's goals and questionnaires.⁹²

NASA envisioned a wide range of participants in the working groups, including pilots; flight attendants; people familiar with alternative data systems; and other aviation stakeholders, such as academic researchers and industry. Project leaders told us that they did not expect that participants would necessarily attain consensus, except to the extent that the groups thought the NAOMS data appeared to be valid and could publicly present the data in a way that would not be automatically translated into systemwide extrapolation of event counts. According to a presentation at the first working group meeting, in December 2003, "the

⁹¹See, for example, NASA, "Final Memorandum," 15–16.

⁹²The workshop agendas, participants, and feedback discussions are detailed in Battelle, *NAOMS Reference Report*, apps. 8 and 10. The 1999 workshop agenda (app. 8-1) included work group summaries and discussions, but these work groups are not to be confused with the working groups established later, during NAOMS's implementation. For the two workshops, see also Connell, *Workshop on the Concept of the National Aviation Operational Monitoring Service*, and Linda Connell, *NAOMS Workshop: National Aviation Operations Monitoring Service (NAOMS)* (Washington, D.C.: Mar. 1, 2000). An earlier workshop, in 1998, also conducted during development, was NASA, "Creation of a National Aviation Operational Monitoring Service (NAOMS)." All three development workshops, as well as the implementation's working groups, are found at NASA, National Aviation Operational Monitoring Service (NAOMS) Information Release, NAOMS Project Presentations and Associated Documents, www.nasa.gov/news/reports/NAOMS_pres.html (last accessed Mar. 1, 2009). See also table 1 of this report.

release of NAOMS data, and its future directions, will be guided by the Working Group [sic].”⁹³ NASA and FAA representatives had agreed earlier that year not to release any survey results before the working groups reviewed them and came to a consensus on the timing, content, and level of the release of NAOMS data.

Discussing the fate of the 2003 and 2004 working groups, NASA’s OIG concluded in March 2008 that “the NAOMS working groups failed to achieve their objectives of validating the survey data and gaining consensus among aviation safety stakeholders about what NAOMS survey data should be released.”⁹⁴ The working groups’ limited effect may have stemmed partly from disagreement over their composition. NASA project leaders suggested that FAA had wanted an existing advisory group to oversee efforts to validate the data, whereas NASA wanted a different combination of academicians—specifically, FAA staff, subject matter experts, and industry stakeholders.⁹⁵ FAA officials told us that they had serious concerns about some of NASA’s proposed experts, because these experts cited preliminary estimates from NAOMS data that FAA found not to be credible.

Additionally, portions of the working group agendas were dedicated to discussing the importance of survey research for reliably measuring trends. These discussions might indicate that some working group members doubted the core foundations of the NAOMS project or the survey’s ability to supplement aviation safety systems.⁹⁶ According to an official in NASA’s OIG, he believed that the presentations at the working groups were, in a sense, an attempt to get the working group participants on board with the NAOMS project.

NASA’s project team suggested that the two working group meetings took place necessarily late in the NAOMS project to allow for the collection of enough preliminary data and to work through nondisclosure issues. The

⁹³Connors and Connell, “The National Aviation Operations Monitoring Service,” agenda item 8 (“Future Directions”), 2.

⁹⁴NASA, “Final Memorandum,” 10.

⁹⁵Senior FAA officials recently told us that the existing advisory group they proposed was CAST’s Joint Implementation Measurement Data Analysis Team.

⁹⁶The December 18, 2003, and May 5, 2004, working groups’ agendas and presentations are available at www.nasa.gov/news/reports/NAOMS_pres.html (last accessed Mar. 1, 2009).

team also suggested that the meetings “were largely dedicated to organizational, procedural, and membership issues.”⁹⁷ Moreover, presentations at the two working group meetings showed only the contractor’s preliminary aggregate analysis. Because the working group members never had the raw data, they had no opportunity to achieve consensus on the validity of NAOMS data or appropriate uses of these data. NASA’s project leaders have asserted, moreover, that the “Working Group approach” was “terminated prematurely because the NAOMS resources were re-directed to another approach.”⁹⁸ According to the project leaders, policy changes resulted in the disbanding of all advisory groups before a more formalized NAOMS group could be assembled after the first two groups failed to reach their objectives. Reestablishing any sort of advisory group would be difficult, because NASA procedures would require prospective participants to undergo a strict nondisclosure procedure.

Given that the working group members did not have access to the raw data and did not agree on the groups’ goals or composition, it is not surprising that they were unable to productively pursue consensus on the validity and utility of NAOMS data. Additionally, to the extent that some participants rejected NAOMS’s premise that a survey is a valid and reliable way to generate safety-related data, they are not likely to have believed that the data the project collected could be validated. For example, while acknowledging that NAOMS had the potential to allow reliable estimates of relative trends, FAA officials told us that they disagreed that NAOMS could generate statistically reliable rate estimates because of the subjectivity of NAOMS questions. These officials questioned the ability of NAOMS’s information to generate rates or its capacity for validation by existing databases.⁹⁹ Additionally, FAA officials noted that they did not believe any potential customers would have confidence in aggregate NAOMS results unless the source data were released to the customers

⁹⁷Irving Statler, “Comments on the OIG Review of the National Aviation Operations Monitoring Service,” NASA (Mountain View, Calif.: Mar. 5, 2008), 3–4.

⁹⁸Statler, “Comments on the OIG Review of the National Aviation Operations Monitoring Service,” 3.

⁹⁹In addition, FAA maintained that it had long believed that the survey would be “overtaken by events,” such as the collection of digital flight data, also known as flight operational quality assurance data. Such data could provide precise rates of occurrence on multiple parameters and, thus, in FAA’s view, could obviate NAOMS’s potential benefits. As of October 2008, 21 air carriers had FAA- and airline-approved digital flight data programs. When NAOMS began in 1997, only 3 air carriers participated in such programs.

directly, rather than to a working group. FAA also expressed concern that pilots would lack causal knowledge to answer the survey's questions. However, we have noted in this report that the questionnaire was not designed to collect causal information. Additionally, we believe that knowledge of why an event occurred should not be needed to report whether a pilot witnessed or experienced a specific event.

A New Survey Would Require Detailed Planning and Revisiting Sampling Strategies

A new survey similar to NAOMS would require more coherent planning and sampling methods linked to specific analytic goals. In addition, the NAOMS survey exhibited some limitations that others might want to avoid. Sufficient survey methodology literature and documentation on NAOMS's memory experiments are available to conduct another survey of its kind with similarly strong survey development techniques, built on a similarly strong foundation.¹⁰⁰ The sections that follow suggest some elements of a new survey like NAOMS.

Conduct a Cost-Benefit Analysis

Before undertaking a similar survey, researchers should review developments in aviation safety and also the costs of and potential for the NAOMS data to enhance policymakers' ability to measure trends and effects on safety interventions. As NAOMS's application to OMB observed, managers seek rational and data-driven approaches to aviation safety, which "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."¹⁰¹

NAOMS air carrier data demonstrate that surveys can be used to generate trend data measuring aspects of aviation safety, and some of the team's researchers believe that the data's utility for monitoring the effect of policy interventions has already been demonstrated. A survey like NAOMS could supplement other safety information, but additional analysis must

¹⁰⁰NASA officials noted that they have not found sufficient publicly released overall technical documentation for the NAOMS project. Many of the documents we reviewed for our report were internal team memorandums and analyses that we obtained directly from NASA contractors and subcontractors for the project, and they have not been publicly posted. However, we believe the *NAOMS Reference Report* and the project's OMB paperwork contain sufficient information on the nature of the memory experiments to inform future research.

¹⁰¹NASA, *National Aviation Operations Monitoring Service Application*, 3.

determine whether NAOMS can be sufficiently useful and cost-effective, given more recent events and technological developments. For example, digital flight data could potentially provide monitoring information, but they are not yet comprehensive or regularly and thoroughly analyzed. Additionally, many data sources, such as digital measurements of flight parameters, cannot illuminate behavioral or perceptual information from operators that might bear on aviation safety. Until such capacity exists, a survey like NAOMS may nonetheless cost-effectively supplement other safety information and identify where to look for other sources of safety information.

A thorough cost-benefit analysis should include the cost of additional steps to develop the survey, such as further experiments, questionnaire revisions, and pretesting.¹⁰² Such an analysis should also address the potential costs and benefits of the survey in light of resources required to analyze other sources of safety information. For example, the cost of collecting and analyzing NAOMS-like data may be small relative to the cost of thoroughly analyzing digital flight data, but, depending on the questionnaire design, such analysis may not identify causation.

Capitalize on Experimentation and Testing

A future survey should build on the insights gained from NAOMS's extensive developmental research on pilots' memory organization and ability to recall events. The survey might undertake additional experiments and testing to accommodate survey revisions resulting from stakeholder interests and lessons learned from the NAOMS air carrier pilot survey. A survey might supplement experiments with additional cognitive interviews, behavioral coding, and reviews. Researchers should consider the resources needed for wide-scale testing during the survey's development. Whereas research demonstrates the benefits of adapting a survey's content to the subject matter and population of interest, researchers would want to consider the availability of resources and time to conduct the experiments necessary to reduce respondent burden and increase accuracy. Additionally, researchers should engage in data validation efforts beyond establishing face validity when making important design decisions, such as which recall period to use.

¹⁰²See, for example, GAO, *GAO Cost Estimating and Assessment Guide: Best Practices for Developing and Managing Capital Program Costs*, [GAO-09-3SP](#) (Washington, D.C.: Mar. 2, 2009).

Generally accepted survey practice is to use a field trial to test a questionnaire that is as similar as possible to the final questionnaire. Accordingly, a future survey might attempt to incorporate the results of the experiments, cognitive interviews, and full set of questions into a field trial questionnaire. A future survey should also run a monitored CATI pretest on the final version of the questionnaire, to test the automated programming and ensure that interviewers and respondents appear to interpret questions correctly.

Collaborate with Customers in the Survey's Development

Beyond soliciting and incorporating feedback from aviation safety stakeholders, staff promoting a new survey like NAOMS should work directly with the survey's presumed customers to specify the uses of the data. While it is not essential that these data inform policy interventions, policymakers should agree on their potential utility. A customer's rejection of the premises of a data collection system—as happened with FAA's rejection of the idea that NAOMS would provide a reliable safety monitoring system—should be resolved before full data collection begins, and consensus on the survey's goals and uses should be formally documented. Otherwise, alternative customers should be identified or the survey's design and goals should be revisited. Consulting with potential customers on the wording and likely use of specific questions would enhance the utility of the survey's data. An analysis of the existing NAOMS data by both scientists and customers' representatives could help demonstrate how specific analytic products might directly or indirectly serve organizational missions.

Assess Whether Questionnaire Content Facilitates Planned Analyses

In the NAOMS air carrier pilot survey, there is the potential for more than one crew member on the same aircraft or on separate aircrafts to have reported the same incident. Proportional allocation or segregated analysis of different types of crew might help address the potential for multiple reports of the same event but can be difficult to implement. Nevertheless, survey designers should consider their analytic goals when designing the questionnaire—that is, are they looking for per-crew member risk estimates or system counts? Certain goals may require researchers to adjust the data, while others may not. Overall, survey designers should be prepared to compare the sensitivity of their estimates with different strategies and under different assumptions.

Future efforts to collect safety information from pilots in a survey might also reconsider the potential effect of sampling pilots who fly more than one type of aircraft during the recall period or in more than one crew

capacity. The survey designers might want to consider whether NAOMS's confidentiality considerations outweigh the potential benefits of allowing pilots to link reported events to particular aircraft, given the perceived link between operational size class and risk exposure. To facilitate estimates, the designers of a future survey should also explore the feasibility of modifying the questionnaire to allow pilots to identify specific aircraft and crew capacities associated with each report of a safety event. They would benefit from establishing an analysis plan in conjunction with the questionnaire. Doing so would help determine the utility of adding and deleting questions and would clarify, at the analysis stage, the effect that doing so would have on data collection.

Detail Analytical Goals and Strategies in Advance of Fielding

To ensure consensus on the usefulness of the data, a detailed analysis plan should be developed. The plan should include basic information on likely estimating the strategies and uses of the data, as well as detailed information on likely adjustments or weights needed to take account of questionnaire design and sampling and of the potential uses of the data. Any adjustments to the analysis plan for operational considerations, preliminary results, policy changes, or unforeseen circumstances should be formalized as data collection progresses.

NAOMS was intended to capture precursors to accidents and nonsignificant risks and to supplement other aviation safety information. It was expected that rate trends seen in the NAOMS data would point aviation safety experts toward what to examine in data systems. Therefore, aviation safety experts and stakeholders would have to conduct more extensive analysis than was conducted in the NAOMS project to establish whether rates and trends could be used for this purpose. Additionally, for a similar survey, analysis would have to establish whether data generated from different recall periods, interview methods, or operational size categories were sufficiently similar to allow data to be combined, and whether making adjustments to sampling strategies or question wording is necessary to accommodate analytic goals.

The NAOMS survey was intended to provide a better understanding of the safety performance of the aviation system, and to allow for the computation of general trends over time, in order to supplement safety systems. A survey with a different goal—one that was investigative or intended to understand the causes of events—would seek information different from those asked for in the NAOMS questions. Depending on the customers' intended use of the data, developers of a future survey might consider writing questions that asked about, for example, the causes of

engine failures or details about air crews' experience of engine shutdowns. Whereas questions such as the latter would be consonant with NAOMS's goal of describing precursors to safety events, the former would be more investigative. Developing a detailed analysis plan in conjunction with the questionnaire would help ensure that the survey included questions relevant for specific analyses.

Revisit Sampling Strategy

Given the proportion of out-of-scope cases drawn into NAOMS's filtered sample, and the cost of finding and contacting them, the designers of a future survey should reevaluate the merits of using a database like the Airmen Registration Database as a sampling frame relative to potential alternatives, to ensure that the database is still the most cost-effective or programmatically viable means of identifying the target population.¹⁰³ Other frames, such as industry or union lists, might be considered or alternative stratification and filtering strategies might be used to identify air carrier pilots. Sampling strategies must also consider whether the proliferation of cell phones will require adjusting contact methods to target a population as mobile as pilots.

Analysis of data such as the NAOMS data might compare different approaches to calculating trends and exposure rates to see if substantive conclusions were similar. Analysts might also want to determine how their estimates relate to the overall NAS. For example, if estimates can address only crew-based risk exposure, they probably do not characterize the NAS, although they may provide other important information for aviation safety monitoring. To the extent that characterizing event levels for the NAS is a goal, a survey like NAOMS might require a different sampling strategy than for a survey designed primarily to monitor trends. Sampling records, including sources used to construct a sample frame and the frame itself, should be maintained for potential use in estimates and nonresponse bias analyses.

¹⁰³ A sample frame based on the full version of the FAA Airmen Registration Database would ensure that all potentially eligible pilots were available on the sampling frame, including those who opted out of the publicly available directory. However, even the full database lacks information on where pilots work and, thus, precludes direct identification of air carrier pilots.

Write a Detailed Implementation Plan

A detailed implementation plan would help ensure the continuity of management and record-keeping for the project and would help ensure that steps like data validation and bias analyses are carried through on a schedule. Given the risks and trade-offs inherent in any survey endeavor, such a plan would also help to ensure that future analysis of the data can accommodate decisions made in the face of changing conditions or for practical considerations.

While benchmarking and face validity checks are important aspects of data validation, they may not be sufficient to confirm the accuracy of pilot recall or estimate the potential effect of nonresponse bias. Even so, besides conducting quality checks on the interview process, future survey developers should undertake formal data validation efforts during data collection and questionnaire development. Nonresponse bias analyses should be planned and completed. The survey's sponsors should allocate resources to fully benchmark the data.

NAOMS's confidentiality provisions appear to have been adequate. Nevertheless, researchers interested in implementing a similar survey might find it useful to further delineate the kinds of data that might be released and the techniques that might be used to remove identifiers from datasets before implementing the survey. In light of other agencies' mechanisms for releasing individual-level data to screened researchers in a controlled fashion, survey documentation should also clarify the conditions under which data could be released to outside researchers, as appropriate.

While the NAOMS extended survey sample fielding period may have been necessary to attain a high response rate from a population as mobile as pilots, future researchers should compare the nature of the answers from pilots who were contacted with relative ease with the answers from pilots who it took greater effort to contact. These researchers should also consider an extended field period's implications for how quarterly statistics are generated in light of potential changes to the sampling frame over time.

There is some merit to NASA's assertion that the working groups could not conduct any data validation, without access to the data. In a future survey, such groups might be constituted earlier, so that data are available for discussions on data validation. A future effort might use such working groups in parallel with data collection, thus soliciting and formalizing the participation of stakeholders. This parallel effort might help the new effort

begin validation as soon as sufficient data are collected. It might also help circumvent disputes over the potential uses of the survey data.

Finally, researchers pursuing efforts similar to the NAOMS project might usefully delineate in advance exactly how rates will be calculated, how potential issues will be clarified, and how the data will be interpreted. A future survey might benefit from tighter coordination between its designers and contractors to ensure that public presentations of preliminary results, when there is still significant debate about the validity of the results, show only the numbers agreed to by project staff.

Concluding Observations

As a monitoring tool, NAOMS was intended to point air safety experts toward trends, to help show FAA and others where to look for causes or extremely rare safety events in other datasets. As a research and development project, NAOMS was a successful proof of concept. However, the data that NASA collected under NAOMS have not been fully analyzed or validated by project staff or aviation safety stakeholders. Depending on the research objective, proper analysis of NAOMS data would require multiple adjustments. Additionally, because of their age, existing NAOMS data would most likely not be useful as indicators of the current status of the NAS.

A similar project, adequately funded and appropriately planned, could accomplish what NAOMS intended to do. According to a 2008 FAA presentation to the National Research Council:

“The NAOMS survey could be very useful in sampling flight crew perceptions of safety, and complementing other databases such as ASRS. The survey data, when properly analyzed, could be used to call attention to low-risk events that could serve as potential indicators for further investigation in conjunction with other data sources.”¹⁰⁴

In this report, we have both described NAOMS’s limitations sufficiently to enable others to look at redesigning them and suggested ways in which a newly undertaken project might successfully go forward. The planners and designers of a new survey might want to supplement it where NAOMS was self-limiting, by incorporating research into investigatory questions of the

¹⁰⁴Mike Baseshore, Office of Aviation Safety Analytical Services, FAA, “Discussion on the NASA National Aviation Operational Monitoring Service (NAOMS) Project,” presentation to Aeronautics and Space Engineering Board, National Academies (Washington, D.C.: June 10, 2008).

type that interested FAA, or to more specifically detail its monitoring capacity in conjunction with existing aviation safety systems. Alternatively, a newly constituted research team might lead operational, survey, and statistical experts in extensively analyzing existing data to validate a new survey's utility for various purposes or to illuminate future projects of the same type.

Agency Comments and Our Evaluation

We provided a draft of this report to the National Aeronautics and Space Administration and to the Department of Transportation for their review. Transportation had no comments on the draft report. NASA provided written comments, and appendix II contains a reprint of the agency's letter. NASA also provided technical clarifications, which we incorporated into the report as appropriate.

In response to the draft report's characterization of NAOMS, NASA emphasized that NAOMS was a research and development initiative. We revised the report to more clearly reflect this aspect of NAOMS. NASA also stated that the draft report inappropriately asserted that NAOMS's goals changed over time, and noted that the principal goal of the project was always to develop a methodology to assess trends or changes over time. While we recognize that this was a primary goal of the project and have revised the report to clarify this issue, we believe that the project staff were not consistent in how they presented NAOMS's likely capabilities to other aviation stakeholders over the life of the project. NASA was also concerned about the draft report's discussion about maintaining pilot confidentiality, citing its own research on the risk of pilot disclosure in the NAOMS data and the inability to determine individuals' motivation for trying to identify a specific pilot. We agree with NASA's concern about pilot identification and have revised the report to highlight NASA's concern; however, we also note that other government agencies have developed mechanisms for releasing, in a controlled manner, extremely sensitive raw data with high risk for the identification of individuals to appropriate researchers.

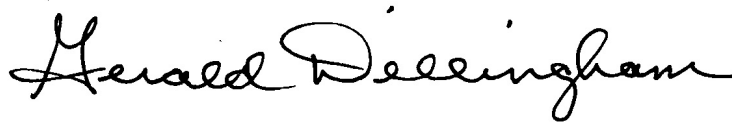
We also provided a draft of this report to Battelle (NASA's contractor for NAOMS) and Jon A. Krosnick, Professor, Stanford University (the survey methodologist for NAOMS) for their review. Battelle provided no comments on the draft report. Dr. Krosnick reported that he found the draft report to be objective and detailed, and that he believed it will contribute to the public debate on NAOMS. He also provided technical clarifications, which we incorporated into the report as appropriate.

As agreed with your offices, unless you publicly announce its contents earlier, we plan no further distribution of this report until 30 days after its issuance date. At that time, we will send copies of this report to relevant congressional committees, the Administrator of the National Aeronautics and Space Administration, the Secretary of the Department of Transportation, and the Administrator of the Federal Aviation Administration, and other interested parties. The report also will be available at no charge on the GAO Web site at <http://www.gao.gov>.

If you or your staffs have questions concerning this report, please contact Nancy Kingsbury at (202) 512-2700, kingsburyn@gao.gov, or Gerald Dillingham at (202) 512-2834, dillingham@gao.gov. Contact points for our Offices of Congressional Relations and Public Affairs are on the last page of the report. GAO staff who made key contributions to this report are acknowledged in appendix III.



Nancy R. Kingsbury, Ph.D.
Managing Director, Applied Research and Methods



Gerald L. Dillingham, Ph.D.
Director, Physical Infrastructure Issues

List of Requesters

The Honorable Bart Gordon
Chairman
Committee on Science and Technology
House of Representatives

The Honorable Gabrielle Giffords
Chair
Subcommittee on Space and Aeronautics
Committee on Science and Technology
House of Representatives

The Honorable Brad Miller
Chairman
Subcommittee on Investigations and Oversight
Committee on Science and Technology
House of Representatives

The Honorable Mark Udall
United States Senate

The Honorable Jerry Costello
House of Representatives

The Honorable Daniel Lipinski
House of Representatives

Appendix I: Technical Issues Relating to NAOMS's Development and Data

In this appendix, we present in more detail a few topics we discuss in the report. They are the (1) National Aviation Operations Monitoring Services' (NAOMS) memory experiments; (2) NAOMS's cognitive interviews with pilots; (3) estimating the effect of the sampling frame, filter, and operational considerations; (4) outlier detection and mitigation; and (5) allocation strategies.

Memory Experiments

The recall and memory experiments for the core safety event section began with three focus groups conducted in August and September 1998, consisting of 37 pilots, and one-on-one “autobiography” interviews of 9 pilots. The autobiographies gave the team insight into pilots' experiences and how they thought about events, enabling the team to develop potential event clusters that matched general categories suggested by the pilots' responses. The focus groups and autobiographies helped in generating questions about different types of events that would link to the major hypothesized memory structures—flight phases, causes, and severity—and eventually a hybrid type that contained causes and flight phases.¹

The NAOMS team and its subject matter experts then listed 96 events—some based on actual experiences, some purely hypothetical—that covered different permutations of these events. For example, they differentiated between minor, moderate, and major problems during takeoff, cruise, and other phases of flight, involving specific causes and resulting in specific events. Examples were “major, approach, weather, spatial deviation” and “minor, landing, people-problem with a conflict or in-flight encounter.” A sorting experiment used the list derived from this process. Researchers gave 14 pilots 96 randomly sorted cards, each containing an individual event, and asked them to sort these cards into stacks containing events that were similar to one another, and to label the stacks descriptively. This sorting task further confirmed potential clusters in the pilots' memory structures. A quantitative analysis of four competing hypotheses of organizational schemes (cause, flight phase, combined cause and flight phase, and severity) showed that the scheme that contained both causes and flight phases best explained the results of the sorting experiment.

¹Battelle, *NAOMS Reference Report: Concepts, Methods, and Development Roadmap*, prepared for the NASA Ames Research Center (Nov. 30, 2007), app. 7.

The project team also assessed the order in which pilots recalled events. The team transcribed the 96 events onto individual sheets of paper and randomly sorted them before presenting them to 9 pilots to read. The pilots then were asked to solve a set of anagrams completely unrelated to aviation—a “distraction” activity to clear their minds—before recalling specific events from the list of 96 events. The researchers tape-recorded what the pilots said, transcribed the responses, and analyzed the resulting data, using an index called “adjusted ratio of clustering” for each of the four hypothesized schemes. Data again indicated that a scheme combining causes and phases of flight best represented pilots’ prevalent memory structures.

For a final confirmatory test of the best organizational approach to pilots’ memory structures, the project team randomly assigned 36 pilots to 1 of 4 experimental conditions. This test was similar to the recall study, except that pilots in 3 of the experimental conditions were offered cues to prompt event recall (cause, phase, or a combination of the two). The cues that combined cause and phase appeared to optimize the number of specific events that a pilot could recall.

A memorandum summarizing these results added a final caveat on question order: that is, events were to be ordered from the weakest in memory to the strongest in memory. This ordering would accord with literature that showed that strong memories can obscure lesser ones in the same memory cluster. The memorandum’s author recommended further research with pilots to develop a ranking of weak to strong memories. It does not appear that formal analysis was conducted, although it is likely that some NAOMS researchers tapped into their own flying and other aviation experience to help sort events on the final questionnaire.

Cognitive Interviews

For the full air carrier pilot survey, researchers interviewed four Aviation Safety Reporting System (ASRS) analysts, all of retired pilots, plus seven active pilots recruited from personal friends of NAOMS staff. At least six of the seven active pilots were air carrier pilots who would have been within NAOMS’s target population.

The questionnaire was revised between the three separate sets of cognitive interviews, but not between participants within a set of interviews—the four ASRS analysts, the six air carrier pilots, and the 7th pilot. The revisions included changes the survey methodologist recommended to more appropriately match the memory structure that the earlier experiments had revealed, as well as changes to accommodate issues

raised in the cognitive interviews. We do not have evidence to suggest whether the questionnaire's final version was cognitively tested before the survey's implementation. Interviewers and Battelle Memorial Institute (Battelle) managers did conduct a series of interviews to test the flow of the computer-assisted telephone interview (CATI) programming before the survey was implemented.

Estimating the Effect of the Sampling Frame, Filter, and Operational Considerations

The decisions that decreased the likelihood of identifying the NAOMS survey respondents made it necessary for analysts to adjust their estimates. In making adjustments, analysts generally look to their analytical goals and to the likely effect of an adjustment on the substantive interpretation of an estimate compared with an alternative. The analysts also try to explore whether adjustments made to address specific problems affect adjustments to address other issues. For example, a series of adjustments to address different features or limitations of the data may render the interpretation of estimates too complicated for practical use. Changes in external datasets used for benchmarking or in creating projections may affect the interpretability of the data over time. In the case of the NAOMS data, sampling, design, and implementation decisions complicate straightforward estimates for either system counts or rates.

For a full analysis to account for issues related to questionnaire design, sampling, and implementation, the NAOMS air carrier data would require multiple adjustments and imputation. Additional analyses would be required to determine the nature and effect of these adjustments. Before the project's end, NAOMS researchers analyzed potential biases that they believed resulted from the filter used to identify air carrier pilots from the sampling frame. These analyses are critical for determining the appropriate uses of the data. We believe that the first priority for further analysis is to estimate the effect of the sampling frame. That is, however appropriate NAOMS's use of the publicly available Airmen Registration Database may have been for cost and programmatic considerations, it has not yet been established whether the frame sufficiently represented air carrier pilots in general, especially in light of pilots' ability to opt out of the registry.

Potential analytic approaches to assessment include but are not limited to the following:

- *Comparing pilots' reported airline fleet characteristics in the survey with outside data on the size of air carrier fleets.* NAOMS project staff added a question on airline fleet size to the survey expressly to be able to

gauge whether the pilots in the Airmen Registration Database flew in fleets similar to the air carrier fleet distribution as a whole. While this analysis might provide compelling information about how representative the frame was, it is insufficient to demonstrate that the frame fully represented air carrier pilots of interest or air carrier pilots covered by the full frame. For example, it is conceivable that the distribution of pilots' airline fleet characteristics correspond between NAOMS data and data derived from other sources, but that the distribution of pilot characteristics within each fleet size was systematically biased toward more experienced pilots who were better able to foresee and avoid safety-related events.

- *Comparing pilot characteristics from the publicly available frame or the sample (as a random subset of the frame) with the full database that the Federal Aviation Administration (FAA) maintained.* Ideally, the comparison would have been made with files used for survey fielding. However, Battelle has reported that it does not have enough data to make such a comparison. A NAOMS team member suggested that for an alternative, one could compare the full FAA database with the publicly available registry on a range of characteristics both relevant and external to NAOMS's concerns. Without knowing whether the nature of the opt-out registry had changed over time, this analysis would help determine whether pilot characteristics in the public frame can be generalized to those in the full frame. However, because neither database contains information on pilots' employment or union membership, this analysis would be insufficient to determine whether the frame used for NAOMS data collection was systematically biased to include or exclude pilots from certain airlines or unions. Thus, this approach would complement, not replace, the analysis comparing fleet characteristics discussed in the previous bullet.
- *Conducting something like a nonresponse bias assessment.* Analysts would take random samples of pilots within the filtered frame as it would be constructed from the publicly available database and from the full FAA-maintained database and would use a survey to compare pilot characteristics for these two samples. Ideally, this would have been done during the survey field trials; however, in the absence of compelling evidence that the nature of the two databases had changed over time, the comparison could still provide insight on whether pilots in the opt-out frame were sufficiently similar to those in the full database to treat the opt-out frame as representative of the population. Depending on its design, a study such as this would allow analysts to focus on characteristics that were most relevant to NAOMS, such as career flying hours or experiences of safety events, and would also provide a means of gauging potential bias

in terms of employers, union membership, and other factors that are not expressly collected in the certificate database.

In any case, analysts of NAOMS data must pursue additional research to determine the existence and nature of potential biases from using the public database rather than the full database, and determine whether and which analytic strategies will ensure that the results adequately represented safety events in the population of interest.

In addition to adjustments for sampling considerations, other analyses may be useful in generating estimates and necessary adjustments. For example, to mitigate the effect of coverage bias in systemwide event count estimates, the NAOMS team advocated using Bureau of Transportation Statistics data related to operational size categories, carrier size, flight hours, and flight legs as benchmarks for weighting these data. The feasibility of using exogenous information to weight NAOMS data depends heavily on achieving a consensus on the appropriate and inappropriate uses of the survey regarding measuring risk exposure and safety events in the national airspace system (NAS).

Battelle recommended statistical modeling—in particular, generalized linear modeling—to develop “more refined rate estimates.”² Generalized linear models would have allowed estimates of safety event rates, while controlling for the independent effect of factors such as season and operational aircraft size.³ Battelle conducted preliminary modeling with generalized linear regression models on grouped sets of data. The utility of such models is contingent on the goals of the analysis and the nature of bias or patterns of missing data; adjusting for independent factors may not be appropriate when generating rate estimates to project to the population. One Battelle statistician noted that NAOMS data lacked important explanatory factors, and that statistical models could suffer from omitted variable bias (which is unrelated to whether these data can be projected to the population of interest). This criticism did not account

²L. J. Rosenthal, “An Overview of NAOMS Decisions Relating to Sampling Approach,” memorandum (Battelle, June 18, 2008), 4.

³An alternative modeling approach, such as zero-inflated Poisson regression or negative binomial regression, would be useful in assessing the effect of explanatory factors, including risk exposure, on the likelihood of having experienced one or more safety events.

for the fact that NAOMS's data were not designed to be used for an investigative process or to establish causation.⁴

Estimates from NAOMS are further complicated by the need to distinguish between risk based on time exposure and risk related to the number of takeoffs and landings. Analysts using NAOMS data might want to compare various approaches to calculating trends and exposure rates to see if different analyses result in similar substantive conclusions. They should also clarify whether and how estimates relate to the overall system—for example, if they can address only crew-based risk exposure, one might ask whether this is sufficient for characterizing the NAS.

Outlier Detection and Mitigation

Outliers can greatly influence the interpretation of statistical analyses. Outlier detection and cleaning, which should consider both statistical and operational concerns, require help from subject matter experts who can identify whether a given data point seems “reasonable” in context. Researchers may also consider whether data follow statistical distributions, such as binomial or Poisson distributions, in deciding how to identify or exclude outliers. Additionally, researchers should consider whether the unit of analysis (whether counts or rates) leads to identifying different cases of outliers and the effect of various methods of outlier detection and cleaning on the substantive interpretation of the analysis.⁵

Causes of outliers can be respondents' mishearing or misinterpreting a question or deciding not to respond truthfully. Outliers may also reflect accurate data that do not correspond with the preponderance of cases. For example, one Battelle researcher cited the “cowboy theory” of aviation safety—the notion that the vast majority of accidents are caused by a small proportion of pilots. Battelle also suggested that some pilots might report events that they had not experienced in order to deliver a message about safety.

⁴The ability of statistical modeling to mitigate the effect of bias as a result of coverage or noncoverage is limited and depends heavily on how cases enter or fail to enter the sample. Other analysis to determine if missing cases relate to the dependent or independent variables of interest (regardless of their availability in the NAOMS data), including the assessments of potential bias as a result of the choice of sampling frame and the filter, are essential in establishing the utility of statistical modeling.

⁵Battelle considered both rates and counts in its research on outlier detection and resolution strategies for NAOMS data. See Thomas Ferryman and others, *Refined Outlier Detection and Resolution Process* (Richland, Wash.: Battelle, December 2002).

Survey research data collected by CATI methods are also subject to several types of outliers. An interviewer may mistype a response—for example, entering 3 as 33. CATI systems often use range checks to prevent such errors: that is, if what is typed exceeds a numerical threshold, the interviewer is prompted to ask the question again or to key the data again. Few hard range checks were incorporated into the NAOMS CATI program, because NASA had instructed the contractor not to question the veracity of pilots' responses by having interviewers re-ask questions if a response seemed unusual. The lack of range checks makes it more difficult to distinguish between outlying answers that were mistyped and those that represented accurate respondent answers. The use of free-text fields to record aircraft type may also have complicated the identification of unreasonable answers for air carrier pilots.

For most questions, the contractor developed an outlier cleaning method that was thought to be both appropriate and objective.⁶ This method was used to identify and remove cases of “doubtful quality” (such as whether the ratio of flight hours to flight legs was unreasonable or whether a pilot had “unreasonable” values on multiple questions), cases lacking information in the questionnaire's fields on flight activity, and additional outliers flagged as “not applicable.” Although the method provided a consistent means of approaching outliers for each question, it did not account for whether reported values made sense in an operational context. Furthermore, the method was developed only midway through data collection. Had the method been developed farther along, more data might have helped clarify whether a distribution-based approach to outlier detection would have been appropriate. To more thoroughly consider statistical and operational concerns, further strategies for data cleaning and outlier detection would benefit from using the full data.

Allocation Strategies

The NAOMS survey has the potential to collect multiple reports of safety events witnessed by more than one crew member or involving multiple aircraft. Several NAOMS researchers believe that the effect of this issue has been overstated, particularly in light of potential analytical strategies

⁶This method, called the “Chebyshev multiple outlier detection method,” was thought to be appropriate in that adequate information to generate distributionally driven cut-off values were lacking and to be objective in that the method did not require judgment regarding the appropriateness of any given answer. The method was nonparametric and based on the Chebyshev inequality. Drafts of analysis plans show that the NAOMS team initially planned to use distributionally based outlier cleaning.

to remedy this problem. Additionally, such concerns do not apply to analyses that determine per-crew member risk exposure (as compared with systemwide projections of event counts), if each individual crew member had an equal chance of being selected.

Strategies that researchers have suggested for addressing the potential for multiple reports of the same event include proportionally allocating events by the likely number of crew members on each aircraft. However, because the number of crew members varies by aircraft size and flight—for example, long international flights require relief crews—this strategy is complicated by the inability to determine for certain which aircraft was involved in a specific incident when a pilot flew more than one aircraft during the recall period.⁷

An alternative strategy would be to calculate events reported by pilots who flew as captains separately from those events reported by other pilots—that is, first officers, flight engineers, and relief pilots. However, this approach might also be complicated by the possibility that pilots flew in more than one capacity over the recall period and the questionnaire does not allow pilots to identify whether they were the captain when experiencing a reported safety event. Furthermore, to the extent that sampling techniques resulted in bias related to the likelihood of flying in a given capacity—that is, the so-called “left-seat bias” that resulted in disproportionate sampling of captains thought to have resulted from the sample filter—segregated analysis of different crew members would require adjustments to project event counts systemwide.

The inability to link reported safety events for pilots who flew more than one aircraft type to a specific aircraft (and, by implication, to a crew size) or day requires developing allocation strategies for other aspects of the data. Before settling on the nonproportional allocation strategies that we describe in this report, Battelle explored alternatives for allocating aircraft among operational size categories and seasons in its preliminary analyses of NAOMS data. For both size category and season, Battelle first attempted to allocate reported safety events and hours flown proportionally across the number of days in a given season or according to the percentage flown per aircraft. Both allocations proved unsatisfactory as it became administratively infeasible for the NAOMS team to maintain

⁷As we have previously discussed, not linking specific planes and a reported safety event was one of the NAOMS team's strategies for maintaining pilot confidentiality.

either system as data collection continued. Additionally, the allocations resulted in fractional degrees of freedom, in that reports from pilots that were split across seasons or aircraft were treated as less than a full case. Similarly, treating proportionally allocated safety events entails theoretical difficulties—for example, was it legitimate when calculating rates to count one-half or one-third of a bird strike?

While proportional allocation or segregated analysis of different types of crews may help to account for potential reports of the same event, these strategies may be difficult to implement because pilots could have flown more than one aircraft type or in multiple crew capacities during the recall period and because of seasonal patterns in the data. As with other weights and adjustments, researchers need to consider their analytical goals—for example, whether they are looking for per-crew member risk estimates or system counts—and should be prepared to compare the sensitivity of their estimates with different strategies and different assumptions. Analysts should also assess whether and how the necessity of multiple adjustments and allocations limits the utility of the data for characterizing trends in air carrier aviation safety.

Appendix II: Comments from the National Aeronautics and Space Administration

National Aeronautics and Space Administration
Headquarters
Washington, DC 20546-0001



February 19, 2009

Reply to Attn of: Aeronautics Research Mission Directorate

Dr. Gerald L. Dillingham
Director
Physical Infrastructure Issues
U.S. Government Accountability Office
Washington, DC 20548

Dear Dr. Dillingham:

NASA appreciates the opportunity to comment on the National Aviation Operations Monitoring Service (NAOMS) draft report. We wish to express our appreciation to the staff of the GAO for their courtesy and for the effort they expended in acquiring a high level of understanding of this complex project over a very short period of time.

The comments below apply to the report as a whole. In addition, enclosed are the specific comments listed in sequential order as they relate to items in the report, as requested by your team.

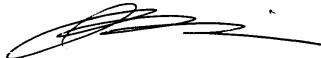
General Comments:

1. This document should emphasize that NAOMS was a NASA research and development (R&D) project to develop a methodology, and not a formally-adopted operational survey "product." That is, its purpose was to evaluate the feasibility of developing a methodology for assessing safety-related trends or changes over time on a system-wide basis. Although a great deal of initial effort was expended to get as much correct as possible, an objective of the project was to see not only what worked but what did not work, what biases evolved, and if and/or how they might be addressed. Towards the end of the report, it does acknowledge that the purpose of NAOMS was to develop a methodology. However, it would be helpful if this were also stated earlier in the report and in the Executive Summary.
2. The report makes numerous assertions that the goals of the project changed (from trending or changes over time to 20-percent changes from one year to the next and from event counts to trends). The primary goal of the project from the beginning was to develop a methodology to assess trends or changes over time (e.g., the Aviation System Monitoring and Modeling project plan of 2000 and the early presentations in 1999 and 2000).

3. The report compares the NAOMS data with the Aviation Safety Reporting System (ASRS) data and states that there is less risk of respondent confidentiality being comprised within NAOMS. Even though the authors of the report recognized improved data mining techniques that are available today, they did not see this as a great risk to NAOMS. The NASA work on the "Assessment of Probability of Disclosure" definitely showed that there were risks to pilot disclosure depending on the available information of an event or the motivation to identify a particular pilot. NASA contends that it is impossible to predict factors that may motivate someone to try to identify a pilot so that it is always necessary to protect a pilot's identity.
4. The last recommendation of the report emphasizes the need for consensus among team members before exposing controversial data. However, as pointed out with reference to the Federal Aviation Administration and the working group, such consensus--for many reasons--is difficult to achieve. This recommendation does, however, serve to point out the complex and difficult decisions that had to be made throughout the project.

In closing, NASA would again like to thank you for the opportunity to provide comments on this draft report.

Sincerely,



Jaiwon Shin
Associate Administrator for
Aeronautics Research Mission Directorate

Enclosure

Appendix III: GAO Contacts and Staff Acknowledgments

GAO Contacts

Nancy R. Kingsbury, Ph.D., (202) 512-2700, or kingsburyn@gao.gov
Gerald L. Dillingham, Ph.D., (202) 512-2834, or dillinghamg@gao.gov

Staff Acknowledgments

In addition to the persons named above, H. Brandon Haller, Assistant Director; Teresa Spisak, Assistant Director; Carl Barden; Ron LaDueLake; Maureen Luna-Long; Grant Mallie; Erica Miles; Charlotte Moore; Anna Maria Ortiz; Dae Park; Penny Pickett; Mark Ramage; Carl Ramirez; Mark Ryan; and Richard Scott made key contributions to this report.

Bibliography

Many publicly available documents on the National Aviation Operations Monitoring Service (NAOMS) are at the National Aeronautics and Space Administration's (NASA) Web site dedicated to the NAOMS project (www.nasa.gov/news/reports/NAOMS.html, last accessed Mar. 1, 2009) or at other NASA Web sites where materials on NAOMS and the Aviation Safety and Security Program are archived and searchable. The Committee on Science and Technology of the House of Representatives maintains additional information related to its October 31, 2007, hearing on NAOMS through its Web site at <http://science.house.gov/publications/> (last accessed Mar. 1, 2009).

Battelle Memorial Institute. *NAOMS Reference Report: Concepts, Methods, and Development Roadmap*. Prepared for the National Aeronautics and Space Administration Ames Research Center. November 30, 2007.

Connell, Linda. *NAOMS Workshop: National Aviation Operations Monitoring Service (NAOMS)*. Washington, D.C.: National Aeronautics and Space Administration, March 1, 2000.

Connell, Linda. *Workshop on the Concept of the National Aviation Operational Monitoring Service (NAOMS)*. Alexandria, Va.: National Aeronautics and Space Administration, May 11, 1999.

Connors, Mary, and Linda Connell. "The National Aviation Operations Monitoring Service: A Project Overview of Background, Approach, Development and Current Status." Presentation to the NAOMS Working Group 1. Seattle, Wash.: National Aeronautics and Space Administration, December 18, 2003.

Dodd, Robert S. Statement on the National Aviation Operations Monitoring Service, October 28, 2007. Statement on the National Aviation Operations Monitoring Service. Statement before the Committee on Science and Technology, House of Representatives, U.S. Congress. Washington, D.C.: October 31, 2007.

Griffin, Michael D., Administrator, National Aeronautics and Space Administration. Letter to National Aeronautics and Space Administration employees on NAOMS. Washington, D.C.: January 14, 2008.

Griffin, Michael D., Administrator, National Aeronautics and Space Administration. Statement on the National Aviation Operations Monitoring Service. Statement before the Committee on Science and Technology,

House of Representatives, U.S. Congress. Washington, D.C.: October 31, 2007.

Griffin, Michael, Administrator, and Bryan D. O'Connor, Chief, Safety and Mission Assurance, National Aeronautics and Space Administration. "Release of Aviation Safety Data." Media briefing moderated by J. D. Harrington, National Aeronautics and Space Administration Office of Public Affairs. Washington, D.C.: December 31, 2007.

Krosnick, Jon A. Statement on the National Aviation Operations Monitoring Service, October 30, 2007. Statement before the Committee on Science and Technology, House of Representatives, U.S. Congress. Washington, D.C.: October 31, 2007.

McVenes, Terry, Executive Air Safety Chairman, ALPA International. Statement on the National Aviation Operations Monitoring Service. Statement before the Committee on Science and Technology, House of Representatives, U.S. Congress. Washington, D.C.: October 31, 2007.

Miller, Brad, Chairman, Subcommittee on Investigations and Oversight, Committee on Science and Technology, House of Representatives, U.S. Congress. Letter to Robert Sturgell, Acting Administrator, Federal Aviation Administration. Washington, D.C.: July 23, 2008.

National Aeronautics and Space Administration. *National Aviation Operations Monitoring Service Application for OMB Clearance*. Moffett Field, Calif.: Ames Research Center, June 12, 2000.

National Aeronautics and Space Administration. "National Aviation Operational Monitoring Service (NAOMS): Development and Proof of Concept." Presentation to the Aviation Safety Reporting System Advisory Subcommittee. Washington, D.C.: November 13, 1998.

National Aeronautics and Space Administration. "Creation of a National Aviation Operational Monitoring Service (NAOMS): Proposed Phase One Effort." Presentation to the Flight Safety Foundation Icarus Committee Working Group on Flight Operational Risk Assessment. Washington, D.C.: March 5, 1998.

National Aeronautics and Space Administration, Office of Safety and Mission Assurance. "Final Report of the National Aeronautics and Space Administration (NASA) National Aviation Operations Monitoring Service (NAOMS) Information Release Advisory Panel (2008)." Memorandum to

the Associate Administrator, National Aeronautics and Space Administration. Washington, D.C.: May 12, 2008.

National Aeronautics and Space Administration, Office of Inspector General, Assistant General for Auditing. "Final Memorandum on the Review of the National Aviation Operations Monitoring Service (Report No. IG-08-014; Assignment No. S-08-004-00)," to the Associate Administrator for Aeronautics Research, National Aeronautics and Space Administration. Washington, D.C.: March 31, 2008.

Statler, Irving C. *Aviation Safety and Security Program (AvSSP): 2.1 Aviation System Monitoring and Modeling (ASMM) Sub-Project Plan, Version 4.0*. Washington, D.C.: National Aeronautics and Space Administration, February 2004.

Statler, Irving C., ed. *The Aviation System Monitoring and Modeling (ASMM) Project: A Documentation of Its History and Accomplishments 1999–2005*. Washington, D.C.: National Aeronautics and Space Administration, June 2007.

White House Commission on Aviation Safety and Security. *Final Report to President Clinton*. Washington, D.C.: The White House, February 12, 1997.

GAO's Mission

The Government Accountability Office, the audit, evaluation, and investigative arm of Congress, exists to support Congress in meeting its constitutional responsibilities and to help improve the performance and accountability of the federal government for the American people. GAO examines the use of public funds; evaluates federal programs and policies; and provides analyses, recommendations, and other assistance to help Congress make informed oversight, policy, and funding decisions. GAO's commitment to good government is reflected in its core values of accountability, integrity, and reliability.

Obtaining Copies of GAO Reports and Testimony

The fastest and easiest way to obtain copies of GAO documents at no cost is through GAO's Web site (www.gao.gov). Each weekday afternoon, GAO posts on its Web site newly released reports, testimony, and correspondence. To have GAO e-mail you a list of newly posted products, go to www.gao.gov and select "E-mail Updates."

Order by Phone

The price of each GAO publication reflects GAO's actual cost of production and distribution and depends on the number of pages in the publication and whether the publication is printed in color or black and white. Pricing and ordering information is posted on GAO's Web site, <http://www.gao.gov/ordering.htm>.

Place orders by calling (202) 512-6000, toll free (866) 801-7077, or TDD (202) 512-2537.

Orders may be paid for using American Express, Discover Card, MasterCard, Visa, check, or money order. Call for additional information.

To Report Fraud, Waste, and Abuse in Federal Programs

Contact:

Web site: www.gao.gov/fraudnet/fraudnet.htm

E-mail: fraudnet@gao.gov

Automated answering system: (800) 424-5454 or (202) 512-7470

Congressional Relations

Ralph Dawn, Managing Director, dawnr@gao.gov, (202) 512-4400
U.S. Government Accountability Office, 441 G Street NW, Room 7125
Washington, DC 20548

Public Affairs

Chuck Young, Managing Director, youngc1@gao.gov, (202) 512-4800
U.S. Government Accountability Office, 441 G Street NW, Room 7149
Washington, DC 20548