

Human Factors Considerations for Flight Deck Safety Assessments

Kim Cardosi
Sarah Yahoodik
Kayli Battel
Tracy Lennertz

December 2025

DOT-VNTSC-FAA-25-04

Prepared for:
US Department of Transportation
Federal Aviation Administration
800 Independence Avenue SW
Washington DC, 20591

Notice

This document is disseminated under the sponsorship of the Department of Transportation in the interest of information exchange. The United States Government assumes no liability for the contents or use thereof.

The United States Government does not endorse products or manufacturers. Trade or manufacturers' names appear herein solely because they are considered essential to the objective of this report.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) December 2025		2. REPORT TYPE Final report		3. DATES COVERED (From - To)	
4. TITLE AND SUBTITLE Human Factors Considerations for Flight Deck Safety Assessments				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Kim Cardosi, Sarah Yahoodik, Kayli Battel, Tracy Lennertz				5d. PROJECT NUMBER FG04C324	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) US Department of Transportation, Volpe National Transportation System Center 220 Binney Street Cambridge, MA 02141				8. PERFORMING ORGANIZATION REPORT NUMBER DOT-VNTSC-FAA-25-04	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) US Department of Transportation Federal Aviation Administration 800 Independence Avenue SW Washington DC, 20591				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT Functional hazard assessments (FHAs) and other system safety assessments are critical evaluations that can inform aircraft design and operational requirements. Aspects of these assessments have often relied on assumptions regarding flightcrew behavior and experience. The purpose of the current study was to identify existing FAA Aircraft Evaluation Division (AED) processes for identifying human factors assumptions, identify gaps in existing processes, and identify and expand access to searchable aviation safety datasets that may support independent review of operational human factors assumptions. This report consolidates findings from four separate, yet related tasks: documentation review, interviews with AED human factors specialists, literature review, and an analysis of publicly available aviation datasets that could help inform the validation of human factors assumptions.					
15. SUBJECT TERMS Aircraft Evaluation Division, Flight Standardization Board, Functional Hazard Assessment, Human Factors					
16. SECURITY CLASSIFICATION OF: Unclassified			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified	Unlimited	79	19b. TELEPHONE NUMBER (include area code)

SI* (MODERN METRIC) CONVERSION FACTORS				
APPROXIMATE CONVERSIONS TO SI UNITS				
Symbol	When You Know	Multiply By	To Find	Symbol
LENGTH				
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
AREA				
in ²	square inches	645.2	square millimeters	mm ²
ft ²	square feet	0.093	square meters	m ²
yd ²	square yard	0.836	square meters	m ²
ac	acres	0.405	hectares	ha
mi ²	square miles	2.59	square kilometers	km ²
VOLUME				
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft ³	cubic feet	0.028	cubic meters	m ³
yd ³	cubic yards	0.765	cubic meters	m ³
NOTE: volumes greater than 1000 L shall be shown in m ³				
MASS				
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
oz	ounces	28.35	grams	g
TEMPERATURE (exact degrees)				
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
ILLUMINATION				
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m ²	cd/m ²
FORCE and PRESSURE or STRESS				
lbf	poundforce	4.45	newtons	N
lbf/in ²	poundforce per square inch	6.89	kilopascals	kPa
APPROXIMATE CONVERSIONS FROM SI UNITS				
Symbol	When You Know	Multiply By	To Find	Symbol
LENGTH				
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
AREA				
mm ²	square millimeters	0.0016	square inches	in ²
m ²	square meters	10.764	square feet	ft ²
m ²	square meters	1.195	square yards	yd ²
ha	hectares	2.47	acres	ac
km ²	square kilometers	0.386	square miles	mi ²
VOLUME				
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m ³	cubic meters	35.314	cubic feet	ft ³
m ³	cubic meters	1.307	cubic yards	yd ³
mL	milliliters	0.034	fluid ounces	fl oz
MASS				
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
g	grams	0.035	ounces	oz
TEMPERATURE (exact degrees)				
°C	Celsius	1.8C+32	Fahrenheit	°F
ILLUMINATION				
lx	lux	0.0929	foot-candles	fc
cd/m ²	candela/m ²	0.2919	foot-Lamberts	fl
FORCE and PRESSURE or STRESS				
N	newtons	0.225	poundforce	lbf
kPa	Kilopascals	0.145	poundforce per square inch	lbf/in ²

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380. (Revised March 2003)

Preface

This report was prepared by the Transportation Human Factors Division of the Safety Management and Human Factors Technical Center at the United States (U.S.) Department of Transportation (DOT), John A. Volpe National Transportation Systems Center. This work was funded by the Federal Aviation Administration (FAA) Human Factors Division (ANG-C1), in support of technical sponsors in the Aviation Safety (AVS) organization. We thank Dr. Kathy Abbott, Eddie Austrian, Robert Bourke, Clark Davenport, John Grammer, Christy Helgeson, Dr. Heidi Kim, Pam Munro, Stephen Plishka, Dr. Victor Quach, Dr. Eric Roberts, and Kevin Siragusa, for technical guidance on this task.

The views expressed herein are those of the authors and do not necessarily reflect the views of the Volpe National Transportation Systems Center or the United States Department of Transportation.

For questions or comments, please contact Dr. Kim Cardosi, kim.cardosi@dot.gov.

Contents

Preface	5
List of abbreviations	iv
Executive summary	2
1. Introduction.....	4
1.1 FHAs and other safety assessments.....	5
1.2 FAA guidance on FHAs and other safety assessments.....	6
2. Gap analysis: AED guidance, processes, and methods for substantiating operational human factors assumptions	7
2.1 Document Review.....	7
2.1.1 Joint Authorities Technical Review	8
2.1.2 The National Transportation Safety Board	12
2.2 Contributions from AED personnel	14
2.2.1 Common assumptions identified.....	14
2.2.2 Challenges to assessing assumptions of flightcrew behavior	15
3. Literature review	18
3.1 Methods and tools.....	18
3.2 A key reference document.....	20
3.3 Recommended practices.....	21
3.4 Assessing assumptions of flightcrew behavior.....	23
3.4.1 Would a consolidated database assist in assessing assumptions of pilot error?	23
3.4.2 Additional considerations in assessing assumptions of flightcrew behavior.....	24
3.5 Identification and Verification of Assumptions of Flightcrew Behavior in Safety Assessments ..	27
3.5.1 The EASA framework	29
4. Analysis of available datasets.....	30
4.1 ASRS	31
4.1.1 Data source	31
4.1.2 Currency.....	32
4.1.3 Strengths	32

4.1.4	Limitations.....	33
4.2	NTSB CAROL aviation investigation search	34
4.2.1	Data source	34
4.2.2	Currency.....	34
4.2.3	Strengths	35
4.2.4	Limitations.....	35
4.3	SKYbrary Accidents and Incidents Reports browser	36
4.3.1	Data source	36
4.3.2	Currency.....	37
4.3.3	Strengths	37
4.3.4	Limitations.....	37
5.	Discussion and next steps.....	38
5.1	Validation of assumptions of flightcrew behavior.....	38
5.2	Increased standardization in the certification and operational evaluation processes.....	39
5.3	Tools and methods needed	39
5.4	Establish consistent documentation.....	40
5.5	Known human factors principles	40
5.6	Next steps.....	41
6.	References	43
Appendix A: Interview questions		48
Appendix B: ASRS Tutorial.....		50
Search execution		50
Techniques for searching and extracting relevant data		50
Search tutorial		52
Example search		53
Objective.....		53
Process		53
Appendix C: NTSB CAROL tutorial.....		56
Search execution		56
Techniques for searching and extracting relevant data		56
Search tutorial		56
Example search		57

Objective.....	57
Process	57
Appendix D: SKYbrary Accident and Incident Browser Tutorial	58
Search execution	58
Techniques for searching and extracting relevant data	58
Search tutorial	58
Example search	59
Objective.....	59
Process	59
Appendix E: Identified database summary.....	61

List of abbreviations

Abbreviation	Term
AIDS	Accident and Incident Data System
AC	Advisory circular
ACT ARC	Air Carrier Training Aviation Rulemaking Committee
ACSAA	Aircraft Certification, Safety, and Accountability Act
AED	Aircraft Evaluation Division
AIR	Aircraft Certification Services
ANC	Air Navigation Commission
AR	Aircraft registry
AOA	Angle of attack
ASRS	Aviation Safety Reporting System
ASIAS	Aviation Safety Information Analysis and Sharing
CAROL	Case Analysis and Reporting Online
CFR	Code of Federal Regulations
FAA	Federal Aviation Administration
FGS	Flight Guidance System
FHA	Functional Hazard Assessment
FSB	Flight Standardization Board
HR	House of Representatives
ICAO	International Civil Aviation Organization
JATR	Joint Authorities Technical Review
MCAS	Maneuvering Characteristics Augmentation System
NASA	National Aeronautics and Space Administration
NMACS	Near Mid Air Collision System
NPRM	Notice of Proposed Rulemaking
NTSB	National Transportation Safety Board
OEM	Original Equipment Manufacturer
QRH	Quick Reference Handbook
RWS	Runway Incursions (RWS) database
SME	Subject Matter Expert
US DOT	United States Department of Transportation

Executive summary

The Federal Aviation Administration (FAA) Aircraft Evaluation Division (AED) assists and coordinates with aircraft manufacturers, Aircraft Certification Offices, and other Flight Standards Offices. As part of these duties, AED may review data related to functional hazard assessments (FHAs) and other system safety assessments. Typically created at the start of the certification process, FHAs are evaluations that can inform aircraft design, safety objectives, and operational requirements. FHAs functions include predicting failure conditions, along with their possible causes and safety impacts. A critical aspect of these assessments are the assumptions regarding expected flightcrew behavior. The purpose of the current study was to expand FAA knowledge of public data sources, and thereby access to searchable aviation safety data that is useful to AED personnel who are responsible for recognizing and substantiating an applicant's assumptions regarding pilot behavior.

To address this objective, researchers identified existing AED processes for identifying human factors assumptions, gaps in existing processes, and reviewed publicly available datasets. The work was limited to operational human factors assumptions within the context of system safety assessments (e.g., FHAs) and Part 121/135 flight operations¹. This report consolidates findings from separate, yet related tasks: gap analysis consisting of a documentation review and interviews with AED human factors specialists, a literature review, and an analysis of publicly available aviation datasets that could help inform the validation of human factors assumptions.

Section 1 of this report provides an introduction and describes existing FAA guidance on FHAs and other safety assessments, such as those found in Advisory Circular (AC) 25.1329-1C, *Approval of Flight Guidance Systems*. Notably, existing guidance states that if flightcrew action is used to mitigate the effect of a failure condition, the assumed flightcrew response should be identified. However, no guidance is provided on how assumptions of flightcrew behavior can be identified or substantiated.

Section 2 identifies gaps in existing AED processes, guidance, data, methods, and tools for substantiating human factors assumptions in FHAs. Researchers reviewed key documents that analyzed human factors assumptions in the context of FHAs. In addition, researchers interviewed AED personnel to further identify gaps in existing procedures, guidance, and methods for identifying and substantiating human factors assumptions. Findings from the gap analysis highlighted the need for consistent documentation around the FHA process (e.g., the applicant providing technical details, failure modes, and design changes in a repeatable format;

¹ Part 121 refers to regularly scheduled air carrier operations. Part 135 refers to commuter or on-demand operations.

document traceability), in addition to establishing methods, tools, and data sources that can assist in the validation of explicit and implicit human factors assumptions in FHAs.

Section 3 builds on findings from the gap analysis by reviewing relevant literature, focusing on possible tools, methods, and information that may be useful to AED in their identification and assessment of human factors assumptions in safety assessments. Although few studies directly apply to the activities of the AED, previous work noted the need for human factors assumptions to be quantified. Existing literature, research, and reports can provide a starting point for validating assumptions of flightcrew behavior (e.g., providing information on pilot response time in a specific circumstance or likely errors), but this information may not necessarily be generalizable, and performance will vary with the complexity and specifics of the event. In addition, a framework was identified that provides high-level structure and guidance on the with proposed validation methods for assumptions. While this framework did not provide details on said methods or criteria, it may provide structure for the AED to develop their own tools and methods for substantiating human factors assumptions.

Finally, Section 4 details the review for searchable aviation-related databases that are available to the public and could be useful to AED personnel when substantiating human factors assumptions. Over fifty databases were identified with three evaluated in detail based on quality and relevance to AED: the Aviation Safety Reporting System (ASRS), the National Transportation Safety Board (NTSB) Case Analysis and Reporting Online (CAROL) aviation investigation database, and the SKYbrary Accident and Incident browser. Although it is important to note that findings and examples from accidents or incident reports may not be generalizable (and therefore may not necessarily be suitable for directly evaluating a specific human factors assumption), information from these databases can help inform AED decisions.

I. Introduction

The Federal Aviation Administration (FAA) Aircraft Evaluation Division (AED) assists and coordinates with both aircraft manufacturers and operators², Aircraft Certification Offices, and other Flight Standards Offices (FAA, 2024). As part of these duties, AED may review data related to functional hazard assessments (FHAs) and other system safety assessments.

FHAs are evaluations that can inform aircraft design, operational requirements, and future safety assessments. FHAs are system- or airplane-level assessments that can focus on operations and may rely on assumptions regarding flightcrew behavior and experience. These assessments, therefore, may implicitly or explicitly include flightcrew behavior as a mitigating action for a particular failure condition. However, given the lack of a central data repository on pilot behavior, AED personnel may have limited ability to substantiate applicants' assumptions in assessments (see Cardosi et al., 2024)³.

The objective of this work is to build FAA knowledge of publicly available aviation safety data that can inform the work of the AED, with a focus on data sources that may help to understand or inform assumptions of pilot behavior. The scope focuses on human factors assumptions related to FHA processes and documentation for 14 CFR Part 121/135 flight operations. This work can inform new and existing criteria for operational evaluations, as well as potential updates relevant FAA Orders and guidance material.

For this report, researchers from the U.S. DOT Volpe Center conducted several tasks. First, we conducted a gap analysis by reviewing existing findings, observations, and recommendations relevant to FHA methodologies. In addition, we conducted semi-structured discussions with human factors personnel in the AED to identify common examples of human factors assumptions made by applicants and challenges that they encounter in their assessments. Second, we conducted a review of relevant scientific literature and technical reports to identify research that could inform AED methods in identifying human factors assumptions and the process for tracking changes to these assumptions in FHA documentation. Finally, we reviewed and analyzed publicly available aviation datasets that could inform the suitability of human factors assumptions in FHAs.

² In the context of certification, aircraft manufacturers, original equipment manufacturers (OEMs), and operators are also referred to as "applicants".

³ Section 115 of ACSAA had previously identified the need to substantiate applicants' assumptions regarding pilot behavior in assessments, with the stated requirement of applicants of amended type certificates to "provide appropriate employees of the [Federal Aviation] Administration with the data and assumptions underlying each assessment and amended assessment" and for appropriate AIR and Flight Standards Service (AFX) employees to review applicant-provided data and assumptions.

1.1 FHAs and other safety assessments

Before 1988, aircraft systems could be evaluated without the use of structured means for conducting safety assessments (FAA, 1988). However, as the complexity of aircraft systems and their integration increased, so did difficulties in assessing the hazards that could result from failures of such systems or identifying potentially adverse interactions among them. These difficulties led to the recommendation to utilize structured safety assessments.

In general, the purpose of safety assessments is to evaluate the design of systems performing aircraft functions, identify and classify failure conditions, and validate safety requirements (SAE, 2023). FHAs are a subset of safety assessments that involve a systematic, comprehensive evaluation of functions to identify and classify failure conditions according to their severity (i.e., no safety effect, minor, major, hazardous, or catastrophic; Advisory Circular (AC) 25.1309-1B). FHAs are a “systematic, comprehensive examination of airplane and system functions” and can be conducted using inputs such as service experience, engineering and operational judgement, and/or a top-down deductive qualification examination of specific functions (AC 25.1309-1B). In the past, accident and hazardous incident data were analyzed to determine quantitative probabilities of failure. These probabilities were used to evaluate the quantitative analyses that supplement qualitative analyses and tests that assess the risks that the occurrence of any failure condition would prevent the continued safety of the flight or reduce the capability of the crew to cope with adverse operating conditions (AC 25.1309-1A).

FHAs can be created either at the aircraft or system level. While aircraft-level FHAs are initially performed early in the design stage to shape safety objectives of the aircraft or system (FAA, 2024; SAE, 2023), FHAs are an iterative process and should be updated as changes are made to the design or operational assumptions. FHA inputs can be derived from a variety of sources and there are several acceptable methods to substantiate failure conditions’ effects (SAE, 2023). Two methods described in SAE ARP4761A (an AIR-focused document which provides guidance on conducting FHAs and other safety assessments) with operational implications are (a) pilot and human factors evaluations and (b) published historical data. When a failure condition involves flightcrew action to mitigate the effects of the failure, experienced pilots and human factors specialists should evaluate the failure condition (SAE, 2023). SAE ARP4761A states that the “means and timeliness of failure recognition, intuitive or procedural response and overall workload should be *substantiated* [emphasis added]” (p. 47).

Within the context of the FAA, AED is responsible for reviewing, recognizing, and substantiating the operational assumptions regarding pilot behavior contained in safety assessments such as FHAs. These assessments may rely on both implicit and explicit assumptions of pilot behavior. For example, FHAs may include human factors assumptions regarding flightcrew actions

needed to respond to a warning or otherwise mitigate a failure. However, specific guidance or criteria, including methods or searchable databases, to substantiate flightcrew responses are lacking.

1.2 FAA guidance on FHAs and other safety assessments

As noted above, operational aspects of new or derivative aircraft such as training, checking and currency requirements, and other operational approvals, are under the purview of the AED⁴ and other offices within FAA Flight Standards. In 2013, FAA published specific guidance on AED operational evaluations in which a Flight Standardization Board (FSB) is convened. As stated in FAA AC 120-53B, *Guidance for Conducting and Use of Flight Standardization Board Evaluations*, all operational evaluations conducted by AED are done at the request of an applicant.

AED operational assessments can employ a variety of methods. Decisions as to operational suitability, training, testing and currency requirements may be made based on subject matter expertise, analysis of data, or testing. AED may review the results of part-task testing or full-mission simulations conducted by the applicant as part of the assessment. AED personnel have noted a need for increased standardization in FSB processes and specific guidance in terms of how to determine when an evaluation is required, the type of assessment that is needed, and criteria for pass/fail/acceptable/unacceptable performance (Cardosi, Yahoodik, & Lennertz, 2024).

While there is limited published FAA guidance on safety assessments as they directly relate to AED activities, there are several insights in documents written from an AIR perspective that may be potentially relevant to AED. In a safety assessment, failure conditions may be assumed to be resolved or mitigated by actions taken by the flightcrew. In 2006, as aircraft systems continued to advance in complexity, AC 25.1329-1B, *Approval of Flight Guidance Systems* (2006) restated the importance of a structured safety assessment and expanded on the structure and human factors issues that should be included in the assessment of Flight Guidance Systems (FGS). AC 25.1329-1B specified that “A safety assessment should be performed to identify the failure conditions, classify their hazard level according to the guidance of AC 25.1309-1A, and establish that the failure conditions occur with a probability corresponding to the hazard classification or are mitigated as intended” (p. 75). The additional guidance stressed the risks introduced with complex and highly integrated avionics and the need to assess cascading failures modes. It also discussed the importance of the flightdeck alerts and warnings, specifically to ensure that the

⁴ The predecessor to the Aircraft Evaluation Division (AED) was the Aircraft Evaluation Group (AEG). For simplification, all references to AEG have been changed to AED which is the current office that assists and coordinates with aircraft manufacturers, operators, Certification Branches, and other Flight Standards District Offices with product-specific Continued Operational Safety issues.

information presented to the flightcrew is clear and unambiguous to support flightcrew awareness and timely action. Guidance in AC 25.1329-1C emphasizes that any assumptions of flightcrew action to mitigate the effect of a failure condition should be identified. In addition, applicants should identify any information that is available to the flightcrew that would initiate the appropriate response (e.g., motion, alerts, and displays).

In terms of guidance on assessment methods, of the documentation identified and reviewed, AC 25.1329-1C, *Approval of Flight Guidance Systems*, is the only FAA source found to be relevant. AC 25.1329-1C states that the appropriate assessment method will “depend on the effect of the condition, any assumptions made, and any associated risk” (p. 61) and that “Failure conditions that are projected to be less probable than 10^{-7} per flight hour, independent of effect severity, need not be demonstrated with a flight test” (p. 61). The importance of assessing normal and non-normal conditions in “operationally relevant scenarios” using a range of workload conditions and a representative set of pilots is stressed in the document. AC 25.1329-1C also specifies that “the assumptions about pilot training and skill level should be documented” (p. 86). While this AC only applies to the AIR certification of Flight Guidance Systems, the assessment principles described could potentially be applicable to other FAA assessments. For example, all evaluations need to utilize the appropriate method of assessment and consider all anticipated users and operational conditions.

2. Gap analysis: AED guidance, processes, and methods for substantiating operational human factors assumptions

As part of this project, a gap analysis of AED guidance, processes, available data and methods for substantiation operational human factors assumptions was conducted. Researchers used two methods to conduct this analysis: review of documentation and recommendations from publicly available reports and discussion with AED human factors specialists. This analysis highlighted several key gaps in terms of methods, procedures, and available information related to the identification and substantiation of human factors assumptions in FHAs.

2.1 Document Review

For the first part of the gap analysis, two documents were reviewed: the *Joint Authorities Technical Review (JATR) report on Boeing 737 Max Flight Control Systems*, and the *National Transportation Safety Board (NTSB) Safety Recommendation Report on Assumptions Used in the Safety Assessment Process and the Effects of Multiple Alerts and Indications on Pilot*

Performance. This gap analysis yielded relevant findings and recommendations related to FHA methodologies, processes, and documentation, as well as operational human factors assumptions.

2.1.1 Joint Authorities Technical Review

In June 2019, FAA established a JATR to review the type certification of the flight control system on the B737 MAX. This team was comprised of representatives from the FAA, National Aeronautics and Space Administration (NASA), and civil aviation authorities from Australia, Brazil, Canada, China, Europe, Indonesia, Japan, Singapore, and the United Arab Emirates. They focused on the fact that as aircraft systems become more complex, so do the certification and operational approval processes and the recommendation that FAA processes should evolve in kind. As summarized in the report's cover letter, a key recommendation was to examine the current standards to determine if they are appropriate for the complex integrated systems in today's airplanes.

The following is not intended to be a review of the entire report. Instead, only the observations and recommendations that may be of interest to AED are excerpted below. While only a few of the JATR recommendations specifically reference AED activities, the intent of many of those focused on AIR and flight testing are directly transferable to AED activities.

One overarching theme of the recommendations was the criticality of human factors throughout the processes of certification and operational evaluations. Specifically, the observations and findings in this area can be classified as the following recommendations:

- Establish closer collaboration between AIR and AED to allow AED to make informed decisions on operational suitability issues that may be affected by certification details; specifically that FAA require the integration of [AIR] certification and [AED] operational functions during the certification process.
 - Whereas pilots working in AIR may not have complete knowledge of operational issues, pilots working in AED may not have complete knowledge of certification issues.
- Review and clarify documentation pertaining to the type certification and operational approval processes with particular emphasis on early FAA involvement to ensure the FAA is aware of all design assumptions, the aircraft design, and all changes to the design in cases where a changed product process is used.
- Assess aircraft functions holistically in an operational context.
- Allow the FAA to examine all “major hazards” where flightcrew action is considered a mitigation.

- Review and revise existing FAA guidance material to both emphasize the importance of human factors and human system integration throughout the certification and operational approval processes and to include more realistic expectations of pilot response times.
- Establish appropriate pilot recognition times and reaction times, based on substantive scientific studies which consider the operational environment, the circumstances under which malfunctions may occur, and the effect of surprise.
- Establish feedback mechanisms to ensure continued compliance, system safety, and human factors considerations throughout the entire development and certification process; this feedback mechanism could also be useful to AED in assessing operational suitability.
- Review and, if necessary, revise guidance material to ensure that operational considerations associated with an applicant's design change are adequately risk-assessed to minimize the potential for flight crew error.

Regarding specific recommendations on AED tasks, the report recommended that “Where the assessment of the effectiveness of differences training is not conducted in an aircraft, the FAA should require the AED to use operational flight crew complements (e.g., line captains and line first officers), with a range of flight experience, as part of the assessment” (p. 44). This recommendation stemmed from the recognition that the behavior of flight test pilots or AED pilots alone cannot be considered representative of the behavior of a line pilot in actual operations. Furthermore, the JATR recommended that FSB reports explicitly document the flight experience level and qualification of the flightcrew used to assess the effectiveness of the differences training.

Specific recommendations for changes in FAA guidance material included reviewing and revising AC 25.1302-1, *Installed Systems and Equipment for Use by the Flightcrew*, to ensure that failures of related systems are assessed by taking into account human performance and the operational environment in consultation with an AED operational specialist. They also noted that AC 120-53B, *Guidance for Conducting and Use of Flight Standardization Board Evaluations*, should be revised to ensure that the AED and FSB are provided with all the system differences between related aircraft irrespective of engineering determination of the safety significance.

It should be noted that in 2020, FAA committed to improving the use of system safety data in the certification process and the refinement of operational policy and guidance (FAA, 2020b). Specifically, “This includes validating assumptions made in the system safety assessments and review of flight crew recognition in single and multiple failure scenarios. The focus and overarching goals will ensure the FAA more effectively communicates System Safety

Assessment data to operational personnel to reduce risk through training and procedures mitigations” (p.5).

2.1.1.1 JATR comments on assumptions of flightcrew behavior

In JATR’s review of FHAs and the relevant FAA guidance material, particular emphasis was placed on appropriate assumptions of flightcrew behavior. Specifically, JATR recommended that the FAA “should establish appropriate pilot recognition times and reaction times, based on substantive scientific studies which take into account the operational environment, the circumstances under which malfunctions may occur, and the effect of surprise” (p. 14).

The JATR team noted that in the accidents reviewed, the flightcrew did not appear to exhibit the same behaviors as were assumed in the FHAs provided to the FAA. Specifically, the assumption that flightcrews would take “immediate action to reduce or eliminate high control forces by re-trimming or changing configuration or flight conditions” was not evidenced in the accidents. The assumption that pilots would take immediate action when presented with a given situation was based on FAA certification and training requirements. However, how training, experience, and certification requirements for flightcrews vary outside of the FAA was not considered in the assessments.

The JATR team noted that no studies could be found to substantiate the current FAA guidance for flight tests in AC 25-7D, *Flight Test Guide for Certification of Transport Category Airplanes*, and AC 25.1329-1C, *Approval of Flight Guidance Systems*. Specifically, they remarked that “Often, recognition time is assumed to be 1 second, and reaction time is assumed to be 3 seconds. Thus, test pilots are told that “Recovery action should not be initiated until 3 seconds after the recognition point (AC 25.1329-1C)” (p. 14).

The specific recommendations regarding pilot response behavior included:

- The FAA should establish appropriate pilot recognition times and reaction times, based on substantive scientific studies which take into account the operational environment, the circumstances under which malfunctions may occur, the effect of cascading failures, multiple alarms, and the effect of surprise.
- The FAA should require applicants to provide validated and justified pilot recognition and reaction times for any given failure, considering all associated flight deck effects within the expected operational environment.
- The FAA should provide guidance to test pilots to initiate recovery action only once the combined recognition time and reaction time validated for the given failure being tested have elapsed.

- As part of the certification and operational evaluation processes for transport category airplanes, the FAA should examine all “major hazards” (as classified by an FHA) where a key mitigation is flightcrew action.

The JATR team also noted that while this guidance was written for test pilots, applicants appear to use the guidance as justification of the assumption that a typical pilot would be able to respond correctly to a malfunction within 4 seconds of its occurrence. Furthermore, they noted that applicants are only required to prepare specific justification of their assumed recognition time if it is less than 1 second. This was seen as problematic given that previous studies (e.g., Casner et al., 2013) had demonstrated substantially longer recognition and reaction times to abnormal events, even with qualified, current, and active airline pilots than those described in AC 25-7D and AC 25.1329-1C. The JATR team also points out that beyond a potentially delayed response when handling a sudden abnormal event, flightcrew responses are more likely to be inappropriate.

2.1.1.2 Relevant FAA guidance material

AC 25.1329-1C, *Approval of Flight Guidance Systems*, breaks down the assessment of failure conditions that are expected to be mitigated by flightcrew action into component elements: the onset of the failure, pilot recognition (of the effects of the failure), the time between pilot recognition of the failure condition and initiation of the recovery (“pilot reaction time”) and “pilot recovery”. Pilot response times are normally estimated in the design stages of development and then validated in flight tests. The flight tests are conducted by the manufacturer (applicant) and then, as needed by the FAA. In part-task simulations, assumed representative pilot response times are used for portions of the recognition and recovery process that are not included in the test. The AC offers guidance on acceptable times to be used.

Within the context of AC 25.1329-1C, the recognition point is defined as a point at which “a pilot operating in non-visual conditions may be expected to recognize the need to take action” (p. 77). While there is an expectation that the recognition time in any assessment is dependent on the cues available to the pilot, AC 25.1329-1C states that “The recognition time should not normally be less than 1 second. If a recognition time of less than 1 second is asserted, specific justification will be required” (p. 77). Pilot reaction time is defined as the “time between recognition of the failure condition and initiation of the recovery” (p.76), such as making an input into the controls. The time required to initiate a recovery action will be dependent on several factors. AC 25.1329-1C delineates acceptable response times based on phase of flight, with pilot attentiveness and pilot duties varying depending on the current flight phase. In the climb, cruise, or descent phase of flight or while holding, the pilot reaction time is assumed to

be three seconds or more. Thus, the guidance is that recovery action should not be initiated in a flight test for three seconds after the failure or malfunction has been recognized. In contrast, while maneuvering, recovery can be initiated after one second. The rationale, as stated in the AC is that “The pilot can be assumed to be carefully monitoring the airplane performance and will respond rapidly once the malfunction has been recognized. A reaction time of one second between recognition point and initiation of recovery is appropriate for this phase of flight” (p. 78).

In addition to the assumption that the pilot will be attending to the critical information, there is also an underlying assumption that the action taken by the pilot will be the action intended by the applicant to correct the problem. The potential for flightcrew error, confusion or indecision (in the operation of the FGS) is to be examined in the assessment of the flightcrew interface. This assessment is to be conducted in normal and non-normal conditions that include a range of workload conditions with a representative set of pilots. AC 25.1329-1C recognizes that the effects of a combination of failure conditions, either because of a common failure or cascading failures, may be more severe than the effect of any individual failure. It also recognizes the importance of the alerting system to “provide clear and unambiguous information to the flightcrew” (p. 28 and p. 56) to identify and resolve the failure(s). It further states that “it may be necessary to address the effects of FGS failures with respect to fault propagation, detection, and isolation within other systems” (p. 56) but does not identify the conditions under which such testing would be necessary.

2.1.2 The National Transportation Safety Board

In September 2019, the NTSB identified ways in which assumptions of pilot behavior and other aspects of the safety assessment process contributed to unidentified hazards that contributed to the Lion Air flight 610 and Ethiopian Airlines flight 302 accidents (NTSB ASR-19-01). Working from these findings, they provided general recommendations to improve the FHA process and specific recommendations for changes in FAA guidance.

In their review of the two fatal accidents, the NTSB identified that pilot responses to unintentional (i.e., uncommanded) Maneuvering Characteristics Augmentation System (MCAS) operation “were not consistent with the underlying assumptions about pilot recognition and response that Boeing used, *based on FAA guidance* [emphasis added], for flight control system functional hazard assessments, including for MCAS, as part of the 737 MAX design” (p. 1).

NTSB specified four assumptions of pilot behavior that had been used across all Boeing models when performing FHAs of flight control systems:

- “• Uncommanded system inputs are readily recognizable and can be counteracted by

overriding the failure by movement of the flight controls ‘in the normal sense’ by the flight crew and do not require specific procedures.

- Action to counter the failure shall not require exceptional piloting skill or strength.
- The pilot will take immediate action to reduce or eliminate increased control forces by re-trimming or changing configuration or flight conditions.
- Trained flight crew memory procedures shall be followed to address and eliminate or mitigate the failure” (p. 6).

One aspect of pilot performance that was identified by NTSB, but not addressed by the JATR team, was the effect of multiple alerts. NTSB recognized that multiple alerts increase pilots’ workload, making it more difficult for the flightcrew to prioritize tasks and discern the most appropriate action. The combination of the alerts and indications that were activated in the two fatal accidents examined did not trigger the pilots to immediately perform the runaway stabilizer procedure as was assumed in the safety assessment. Furthermore, the effects of multiple alerts and indicators were not considered in the safety assessment, in part because of the assumption that pilots would respond immediately, and correctly, to the uncommanded MCAS operation. The other reason was that not all the potential alerts and indications that could accompany a failure that resulted in uncommanded MCAS operation were included in the functional hazard assessment or simulated in validation tests. Specifically, the possibility of an erroneous high Angle of Attack (AOA) input to the MCAS was not considered. As a result, additional flight deck effects (IAS DISAGREE, ALT DISAGREE, and stick shaker activation) resulting from the same erroneous AOA were not simulated.

The fact that the assumptions of pilot behavior that had been used in the FHA of the 737 MAX were consistent with both requirements in 14 *CFR* 25.671 and 25.672 and guidance in AC 25-7C, *Flight Test Guide for Certification of Transport Category Airplanes* led NTSB to identify a need for FAA to provide guidance to applicants on the consideration of multiple flight deck alerts and indications when evaluating pilot recognition and response behavior. Specifically, NTSB recommended that for all type-certificated transport-category airplanes, FAA require applicants: “(1) ensure that system safety assessments for which they assumed immediate and appropriate pilot corrective actions in response to uncommanded flight control inputs consider the effect of all possible flight deck alerts and indications on pilot recognition and response; and (2) incorporate design enhancements (including flight deck alerts and indications), pilot procedures, and/or training requirements, where needed, to minimize the potential for and safety impact of pilot actions that are inconsistent with manufacturer assumptions” (p. 9).

As NTSB noted, incorrect assumptions of pilot behavior can affect all aspects of certification and

operational suitability including hazard classification and resulting system design, procedural mitigations, and training mitigations.

NTSB concluded that standardization was needed in the (manufacturers') evaluation and validation of assumptions about pilot recognition and response to failure conditions, particularly those that result in multiple flight deck alerts and indications. Therefore, the NTSB recommended that the FAA develop robust tools and methods, (with the input of industry and human factors experts), for validating assumptions about pilot recognition and response to safety-significant failure conditions. Further, the NTSB recommended that once the tools and methods have been developed as recommended, the FAA revise existing FAA regulations and guidance to incorporate their use and documentation as part of the design certification process and to re-examine pilot recognition and response time assumptions permitted in FAA guidance.

2.2 Contributions from AED personnel

To complement the gap analysis of technical documentation, researchers from the US DOT Volpe Center individually met with six AED human factors personnel. The purpose of these interviews was to understand how AED personnel use FHAs to identify human factors assumptions, examples of human factors assumptions that AED personnel had previously encountered, and available methods to aid in the identification of human factors assumptions. The semi-structured 30-45 minute discussions focused on assumptions of flightcrew behavior and AED processes. The questions for discussion were provided to FAA in advance (see Appendix A) and used as a general guide. Interviewees had the option of skipping any questions and discussing other relevant topics not on the question list. In what follows, we present a summary of the findings, across discussions.

2.2.1 Common assumptions identified

The consensus of the human factors personnel interviewed was that assumptions of flightcrew behavior made in safety assessments have, and will continue to, vary by applicant, depending on the applicants' human factors expertise and experience in FAA processes. For example, while an experienced applicant may attribute an expected action in response to a system failure by flightcrew to "crew training", less experienced applicants may not even associate a hazard mitigation with an underlying assumption of flightcrew behavior.

In general, those interviewed said that they were more likely to encounter *implicit* assumptions of flightcrew behavior compared to *explicit* assumptions, although the line between implicit and explicit is not well-defined. For example, while it may be explicitly stated that specific flightcrew behavior is due to "basic airmanship", this has underlying implicit assumptions, since

the term “basic airmanship” is broad and undefined. Specific, explicit assumptions (e.g., specifying that an alert will be addressed within four seconds) were noted as being unlikely to be included in applicant documentation since such specific assumptions would logically require supporting justification. In fact, assumptions attributed to “basic airmanship” and related terms such as pilot “experience” and “knowledge” were identified as being the most common and that the term “basic airmanship” is often used as a “catch all” for presuming the expected flightcrew responses. As one interviewee noted, the term effectively transfers responsibility from the system design (and applicant) to the pilot. For example, an interviewee noted that in a discussion of an FHA, an Original Equipment Manufacturer (OEM) representative pointed to “basic airmanship” as an explanation as to why pilots could be assumed to execute a specific action. Another interviewee noted that recent pushback on the use of the term “basic airmanship” as a catch-all justification has led to the use of other synonymous, yet equally generic, terms.

Similar to the challenges associated with the use of “basic airmanship” as an implicit assumption is the assumption of flightcrew training content. If a flightcrew procedure for a hazard mitigation is stated as being covered in training, the location of the specific portion of the training or evidence of the efficacy of such training may or may not be provided. When assessing an operators’ training program, or conducting an operational suitability assessment, excerpts or specific references should be provided to AED to identify exactly *where* in the training program the information is included, and how it is conveyed. This would facilitate AED’s assessment as to whether the training is sufficient.

Pilot response time was identified as another issue in assumptions of flightcrew behavior in safety assessments, but rarely in the context of a specific or prescriptive number. For example, one interviewee recalled hearing an applicant state that a pilot would be required to “immediately” respond to an alert, despite the fact that the alert had been classified as a “caution”, not a “warning”. Other examples of human factors assumptions cited included flightcrew executing the correct procedures without error and/or in sufficient time to correct the problem, or quickly identifying and running the correct checklist, recall of memory items, and immediate detection of alerts (in all expected operational conditions).

2.2.2 Challenges to assessing assumptions of flightcrew behavior

One challenge to assessing assumptions of flightcrew behavior in any safety assessment of an aircraft or its components is that they can change over time. As designs evolve throughout the certification process, so too, can assumptions of flightcrew behavior. While there are project management tools used by FAA to track program changes, they are not designed specifically for the tracking of human factors assumptions. In addition, AED does not control the information

uploaded to the system. AED typically only receives changes that are deemed relevant to human factors from an applicant or AIR perspective. While it is important to track assumptions of flightcrew behavior throughout the certification process, there is currently no standardized process or guidance to applicants.

FHAs are fundamental components of the FAA certification process. It is the first type of safety assessment in which assumptions of flightcrew behavior could arise. These FHAs are usually prepared by engineers and presented to FAA (AIR) in the initial stages of certification process. AED may or may not be involved in reviewing these assessments. Even when AIR coordinates AED involvement, AED has only an observational role; except for minor clarifications, they have no authority to directly query the applicant. Since the initial stages of certification are completely under the purview of AIR, AED would discuss any potential issues they identify in an FHA or other discussion only with AIR. With no formalized coordination, AED may or may not be kept informed of changes to design or documentation that are made prior to certification.

In general, interviewees stated that FHAs are not currently used in AED operational activities and interviewees were divided on whether FHAs could feasibly be useful to AED operational assessments, such as suitability assessments and T tests. FHAs are designed to describe hazards, categorize their worst credible outcomes (e.g., minor or catastrophic) and point to mitigation strategies. Applicants are not required to identify the source of underlying assumptions of flightcrew behavior, even when a flightcrew behavior or action is identified as the hazard mitigation (e.g., an applicant may assume that a pilot will respond to an alert within a predetermined amount of time, but fail to substantiate the source of the assumption)⁵. FHAs are usually constructed and reviewed by safety system engineers, not human factors specialists; thus, information relevant to assessing underlying assumptions may not be included in the FHA. This issue is compounded by the previously noted limitation that AED needs to coordinate with their AIR counterpart if they have follow-up questions or concerns on any aspect of flightcrew behavior. According to interviewees, reviewing FHAs are more relevant to AIR activities, however, AED will still need to identify assumptions of flightcrew behavior in FHAs and other safety assessments. AED human factors specialists will continue to be needed to identify and assess assumptions of flightcrew behavior in the context of other assessments, such as operational suitability assessments and T-Tests.

2.2.2.1 Lack of protocol for identification and assessment of assumptions of flightcrew behavior

There is no AED-wide structure for, or guidance on, identifying or assessing assumptions of flightcrew behavior. While all interviewees noted that such guidance would be helpful, different

⁵ Note, however, that AC 25.1309-1B offers guidance on documenting assumptions, stating that “any analysis used as evidence that a failure condition is extremely improbable *should* [emphasis added] include justification of any assumptions made, data sources, and analytical techniques to account for the variability and uncertainty in the analytical process (p. 3-3).

opinions were offered on the type of guidance and structure that would be most effective. Some felt that a checklist of issues to consider would be helpful; others were skeptical, stating that such a checklist could be misinterpreted as being an adequate substitute for early inclusion of human factors expertise. A compendium of issue papers used in AED was also suggested as potentially useful for future assessments. While they serve an important purpose, issue papers do not provide division-wide guidance.

Standardized guidance on the identification, tracking, assessment and resolution of assumptions of flightcrew behavior could aid both AED and applicants in several ways. First, guidance could help to inform FAA personnel not familiar with human factors of the benefits of early human factors involvement and application of human factors methods. Guidance could also add needed structure for how assumptions of flightcrew behavior are identified and assessed. Guidance for applicants could identify needed documentation (e.g., training materials) for AED to validate applicant assumptions. Such guidance would not only help inform applicants as to what is expected, it would preclude AED from having to search for critical information.

Just as there is no formal guidance on how assumptions of flightcrew behavior are to be identified, there is also no formal guidance as to how such assumptions are to be assessed. Acceptable methods range from judgements as to what is reasonable made by Subject Matter Experts (SMEs) to simulation studies. Several interviewees noted that OEM test pilots' evaluations and opinions tend to be weighted heavily, whether an assumption is reasonable or not. Because OEM test pilots are familiar with the systems they are testing, they may not recognize when they are making inaccurate assumptions of operational line pilot behavior and erroneously regard a desired flightcrew response as "basic airmanship", as discussed above. Formal guidance could also have the collateral benefit of helping to codify roles and responsibilities of human factors specialists within AED, helping to ensure that the available human factors expertise is utilized more consistently.

2.2.2.2 *Lack of available data and tools*

All interviewees stated that there are no formal reference sources, data repositories, or tools, available to support their assessment of assumptions of flightcrew behavior. Many interviewees stated that they maintain their personal archive of potentially relevant research. While journal articles and other research sponsored by FAA or NASA were cited as being potentially relevant to reviewing HF assumptions, one interviewee stated that military standards and research studies beyond the aviation domain (e.g., alerting research in the medical field) could also be helpful.

3. Literature review

To address identified gaps from the documentation review and AED HF personnel interviews, a search of relevant scientific literature and reports was conducted. The literature review focused on possible methods, tools, processes, and information that could be useful to AED in their identification and assessment of implicit and explicit assumptions of flightcrew behavior in FHAs.

3.1 Methods and tools

Specialists with human factors expertise are educated and experienced in the scientific methods that are used to assess multiple aspects of flightcrew behavior. Their training typically includes techniques to analyze and interpret the results of such assessments. This knowledge of methods and analysis used in research positively transfers to the assessments of flightcrew behavior in operational tests. AED human factors specialists are likely aware of the value of scientific principles in maintaining the operational integrity of an assessment. For example, in both academic and operational research, the first step is to define the objective of the test and select the appropriate methods of studying the issue. While a description of each feasible method is outside the scope of this effort, human factors specialists are well versed in the various methods and the appropriate matching of the method to the issue being investigated.

Another tenet of scientific investigation is that it is important that the subject pool used is representative of the target population in terms of the areas being studied. Utilizing a representative subject pool is one of the steps that need to be taken to have confidence that the results of the assessment can be applied to target population. In terms of AED assessments, this means that the pilots included in the assessment are representative in terms of training, experience level, and any other variables that would be expected to affect their performance on the tasks included in the assessment.

A similar tenet is that the test conditions used need to be representative of the “real world” conditions to which you wish to apply the results. While the engineer looks at a system to determine whether it performs its intended function, the human factors specialist determines whether the pilot can use the system effectively under the operational conditions under which it will be used. Operational suitability assessments determine if a particular procedure and/or equipment is suitable for use by a specified user in a specified operational environment.

In the past, applicants have been allowed to, for example, attribute accurate responses to a failure condition to “basic airmanship” or “training”. These and other terms to which applicants attribute flightcrew behavior would need to be operationally defined in order to assess

whether the assumptions are valid. Whenever flightcrew behaviors are attributed by an applicant to training, the applicant should be required to provide the supporting training material. Such a requirement would serve AED assessments by negating the need for AED personnel to search through the training material for the relevant material.

Applicants have also asserted that flightcrew actions required to identify and resolve failures would be correct and timely due to the pilots' knowledge of "basic airmanship". FAA has stated that airmanship "is a broad term that includes a sound knowledge of and experience with the principles of flight; the knowledge, experience, and ability to operate an aircraft with competence and precision both on the ground and in the air; and the application of sound judgment that results in optimal operational safety and efficiency" (FAA, 2022, p, 1-1). This is a descriptive and intuitive, yet untestable, definition. Perhaps a more workable definition for AED assessment purposes would be an automatic behavior as a result of foundational training. An "automatic" behavior is one that when first learned is "controlled", but with continued use becomes able to be performed quickly, effortlessly, and sometimes without conscious attention, while "controlled processing" requires focused attention and conscious control (Shiffrin & Schneider, 1984). This perspective on a definition of any given task to be performed correctly as a function of "basic airmanship" allows for the definition in terms of relative workload, accuracy and speed.

While any issue that is being assessed must be specified in terms that are measurable, there are a variety of approaches. For example, when measuring mental workload, metrics can be subjective such as perceptions of workload on a specified scale, or objective such as measures of accuracy or response time (Stanton et al., 2013). In both academic and operational studies, the results of a well-designed assessment are then (ideally) analyzed and documented. In academic pursuits, these analyses usually include statistical tests to determine the probability that the observed results could be due to chance. Operational tests, such as those to determine operational suitability, have different purposes, different constraints, and must interpret the results in different ways than an academic study. For example, it is not feasible to include as many pilots in an assessment conducted by AED, such as an operational suitability assessment, as would typically be included in an academic study of pilot behavior. Because of the limited sample size, statistical tests of significance are not appropriate. This does not limit the assessment of operational significance.

Any assessment should pre-define how the results will be interpreted and applied. This could include, but does not dictate, a pass/fail criteria. Operational assessments usually involve trade-offs that are often fluid and can be used to progress toward the desired outcome. In any event, documentation of the methods (i.e., what was done), the purpose of the assessment (i.e., why it was done), the results, the interpretation of the results and the application of those results

(i.e., how they were used) would be useful when informing future assessments.

A structured approach to AED assessments that includes adherence to these scientific tenets would strengthen the assessment in the following ways. First, it would increase the standardization with which decisions are made as to what needs to be assessed and how it is assessed. Second, it would help ensure that individual assessments are valid and reliable. Third, it would specify and help to institutionalize the critical role of human factors specialists in the assessments. Human factors personnel are uniquely qualified to identify underlying assumptions of flightcrew behavior. A search of scientific research yielded nothing that could directly aid AED in identifying or even defining human factors assumptions made by manufacturers or operational personnel. The development of a structured approach would help to fill this void.

3.2 A key reference document

The most comprehensive relevant human factors reference document is FAA's *Human Factors Considerations in the Design and Evaluation of Flight Deck Displays and Controls Version 2.0* (Yeh, Swider, Jo, & Donovan, 2016). It presents a compilation of human factors regulatory and guidance material for flight deck displays and controls. While the document is written from an AIR perspective, the text recognizes the importance of operational considerations:

“Manufacturers specify the intended function of the system based on their viewpoint of how the system should be used. However, the system may be used operationally in ways beyond what was envisioned, introducing the potential for error” (p. 249).

Yeh et al. (2016) reference FAA regulations and guidance on human factors issues to consider in the design and evaluation of avionics displays and controls for all types of aircraft (14 CFR parts 23, 25, 27, and 29). The intent of the document was to facilitate the identification and resolution of typical human factors issues that are frequently reported by FAA Aircraft Certification Specialists. However, the same human factors issues that affect certification can be expected to be relevant for other evaluations, such as T tests and operational suitability assessments. The general topics of displays, alerts/annunciations, controls, error management, workload, and automation contain important information on issues that have been problematic in the past and why such issues are important from a human performance standpoint.

The document specifies a few reference points relevant to assumptions of flightcrew responses:

“Alerts located on displays outside the pilot’s primary field-of-view may be missed. In an examination of problems associated with the use of LORAN-C/GPS receivers, pilots

reported that they sometimes took up to four minutes to notice and respond to an alert, which may not have been indicated in the primary field-of-view. The alerts included red and yellow warning lights, blinking, and freezing the information on the display. (Adams et al., 1993)” (p. 37).

Additionally, Yeh et al. (2016) describes the range of typical response times when addressing a warning:

“Approximately 8 – 11 seconds may be needed for a pilot to respond to a warning message. It is estimated that 2 - 3 seconds may be necessary to allow the pilot to process a visual or auditory warning message, 5 - 6 seconds may be needed to determine the appropriate response, and 1 - 2 seconds to execute the response (Cardosi and Huntley, 1993; Veitengruber, 1979)” (p. 133).

The document also includes several informative examples of operational assessments and several broad recommendations for evaluations such as:

- New systems should be evaluated both in isolation and in combination of other flight deck systems to ensure flightcrew performance is not negatively impacted.
- While there is no “right” number of evaluator pilots in usability evaluations, the number and composition of evaluators should ensure a range of knowledge, experience, and abilities.
- Tasks used for evaluations should be designed and/or selected to reflect the intended purpose and range of operational situations that the technology will be used.
- Task scenarios will vary depending on the context of the usability evaluation (e.g., lab environment, test bench, simulator, or actual flight).
- Selected metrics should reflect the scope and purpose of the evaluation.
- Evaluation methods, assumptions, and findings should be documented in detail to ensure clear communication and replicability.
- When communicating issues that need to be addressed by the applicant, the FAA should clearly describe the identified issue, why it is important, and what action the FAA expects the applicant to take to address the issue.

None of the above guidance would be new information to AED human factors specialists. The document does not, nor was it intended to, contain the specificity needed to guide an AED assessment. For example, in the summary of steps for conducting a usability evaluation, the first step identified is to “Determine whether a usability evaluation is needed,” yet no guidance is offered as to how to make this determination.

3.3 Recommended practices

Billman, Mumaw, and Feary (2019) describe best practices for evaluating flight deck interfaces for transport category aircraft with a particular focus on attention and decision-making. In this context, they review the strengths and limitations of several subjective and objective measures used in evaluations of prototype designs through flight testing in a simulator and present an outline of key issues in such an assessment. For example, Billman et al. (2019) note the relative ease with which eye fixations and movements can be tracked and the utility of this measure to complement other measures in determining whether an alert attracts attention in various configurations. However, as they note, fixation is neither necessary, nor sufficient, for awareness, and the depth of processing of the viewed information cannot be determined from eye movement data alone.

While the document was intended for applicants who design, build, and test flight deck interfaces and the FAA Certification personnel who evaluate them, many of the human factors issues discussed and recommendations for a general approach are relevant to AED assessments. For example, both Certification tests and Operational Suitability assessments need to determine the appropriate situations, tasks, and number and type of pilots to be included. In each case, determinations will be made on the nature and criticality of the investigation.

As the interviews with AED personnel pointed out, reliable and valid data from line pilots in addition to FAA pilots are necessary for a determination of whether a design is certifiable or operationally suitable. As Billman et al. note:

“It is increasingly recognized that test pilots are a very different population than line pilots and the difference is particularly critical for attention, awareness, and understanding. Test pilots may be able to generalize from their own experience to the experience of less skilled pilots on some topics, say, handling difficulty in manual flight; such projection, however, is not reliable for many issues in attention, awareness, or understanding (p. 39).

Similarly, data from pilots who have an *a priori* knowledge of the system, such as the pilots who helped develop the system or were involved in testing an earlier prototype could not be acceptable as representative of a line pilot in actual operations.

Additionally, both AIR and AED need to determine the evaluation criteria to be used for an acceptable level of performance as the outcome of the assessment. Such determinations must be made for each assessment in the context of the type and criticality of the issues that need to be examined. There is no template for assessing whether a design adequately supports a flightcrew’s information processing needs for problem identification and resolution. Rather, as Billman et al (2019) point out, the quality of the evaluation depends on identifying the most

important issues and selecting appropriate methods to examine the issues. They note that, “People have very poor ability to assess what they are or are not aware of, and even less to project from their own experience how a device might support attention, awareness, and understanding of a different type of person” (p. 39). This is one of the reasons that human factors expertise is needed to help determine the critical issues and appropriate means of testing. How the issues are assessed will determine whether the data obtained actually measure what they purport to measure and will be useful in an operational assessment. For example, Casner, Geven, and Williams (2013) found that when airline pilots were presented with abnormal events in scenarios where they were taught to handle them during airline training, pilots’ responses were acceptable and varied little from pilot to pilot. However, when they were presented with the same abnormal events under less predictable circumstances, the responses showed greater variability and did not always result in acceptable operational performance.

As noted, some of the information and recommendations compiled by Billman et al. (2019) for applicants who design, build, and test flight deck interfaces and the FAA Certification personnel who evaluate them could be useful to AED personnel. However, AED would be better served with guidance tailored to the types of operational assessments that they conduct.

3.4 Assessing assumptions of flightcrew behavior

3.4.1 Would a consolidated database assist in assessing assumptions of pilot error?

Noting that “one of the main factors in all aviation accidents is human error”, Rantannen, Wiegmann, Palmer, and Musiorski (2003, p. 782) sought to create a single database that could be used to explore the effects of different factors, such as tasking and technologies, on human error. This combination of several publicly accessible databases into a single searchable platform that was accessible from the web was envisioned to allow for a directed literature search and review of both empirical and theoretical research designed to advance the exploration of the effects, in terms of causes and remedies of technology on human error. At the time, the authors noted that this database would have several gaps as there are large areas of research on human-technology interaction that have not yet been conducted.

The goals of this work were laudable, summarizing available data in a way that attempted to address the types of errors that would be most likely given the implementation of a new technology and what technologies could potentially alleviate the specific type of error identified. There are good reasons for the fact that over 20 years later, no progress has been made in this endeavor. While the authors attribute the difficulties to the lack of published task analyses, the problem space is much more complex.

Although a repository of data on pilot performance in a variety of tasks and situations may sound appealing, our search of the literature could not find much useful information. This is, in part, due to the limited applicability of data obtained in simulation studies to operational situations. While manufacturers sometimes collect data on pilot response times and error rates in the development and testing of new systems, these data are considered proprietary and are rarely published. Even if error rates were to be established for basic flightcrew tasks for pilots of a given experience level, the context within which the tasks are being performed will influence performance, as will training, experience, corporate culture, workload, fatigue, and a host of other factors. Still, this review of the literature has found there are some foundational and guiding principles, supported by data, that need to be considered in any assessment of flightcrew behavior. These are discussed below.

3.4.2 Additional considerations in assessing assumptions of flightcrew behavior

3.4.2.1 Flightcrew assumptions are made across manufacturers but may vary between applicants.

In 2022, the Air Navigation Commission (ANC) of the International Civil Aviation Organization (ICAO) tasked the Personnel Training and Licensing Panel to address concerns about automation and its use in commercial air transport flight operations. This task included identifying assumptions of flightcrew training and behavior made by aircraft manufacturers and operators. Three manufacturers responded to an international survey. The survey found that “Manufacturers made many assumptions about manual flight, use of automated systems, and pilot training that were integrated into equipment design, manufacturer-recommended training programs, procedures, and documentation” (ICAO, 2024, p. 20). Specifically, the survey found that assumptions vary across manufacturers, but highlighted that “prior piloting experience was assumed for all aircraft models” and manufacturers consistently assumed that “the flight crew shall use an appropriate level of automation and that the automation is intended to assist the pilot”, and that “situation awareness will be maintained at all times” (p. 18).

3.4.2.2 Situations in which pilots do not have the knowledge to understand what is happening and why it can result in pilots responding inappropriately.

Dehais, Peysakhovich, Scannella, Fongue, and Gateau (2015) examined pilot responses to a change in altitude commanded by an automated system. While the purpose of the study was to demonstrate a proof of concept for a tool that would indicate an inconsistency between the pilot’s settings and the autopilot behavior, some performance data were described. This simulation study used pilots from the Institut Supérieur de l’Aéronautique et de l’Espace (a French aerospace engineering school) with a wide range of experience with automated functions (range of 15-700 hours with a mean of 200 hours). In the “impossible descent automation surprise” scenario, pilots in a simulator were instructed by ATC to increase speed

and descend 5,000' with a minimum of 1000'/min. To prevent an unacceptable speed, the automated function leveled the aircraft. This was unexpected for 9 of the 16 pilots who completed the simulation. Only 7 pilots were able to correctly assess the situation and reduce the speed with a mean reaction time of 18 seconds (the standard deviation was not reported). Unfortunately, the study did not examine pilot performance as a result of experience. Cues that identified a pilot input that would conflict with an automated function reduced the response time and resulted in all pilots responding appropriately. This could be useful as an example that situations that result in an 'automation surprise' take longer to resolve than cued situations.

3.4.2.3 *Performance will vary with the typicality and complexity of the event, even within a uniform cohort of line airline pilots.*

The more unusual or complex the event, the more time will be needed for the pilot to diagnose the problem and respond appropriately. Specifically, research has shown that atypical events result in more pilot errors and greater response times than a random selection of safety events in simulated flight conditions (Clewley & Nixon, 2023).

3.4.2.4 *Even data from the most sophisticated simulation studies cannot be assumed to be representative of pilot behavior in actual operations*

Results of a series of flight simulation studies on one aircraft indicated that with an executive system (that is, one that requires immediate action), it will take approximately two to three seconds to detect that the message is there, five to six seconds to decide what to do about it, and one to two seconds to initiate a response (Boucek et al., 1980; Boucek et al., 1982; see also Berson et al., 1981). This led to the conclusion that a total of eight to eleven seconds should be allotted for a pilot to respond to an (intuitive) alert that requires immediate action.

Perhaps the most widely studied flightcrew response has been to the Traffic Alert and Collision Avoidance System (TCAS). In a high-fidelity flight simulation study, Chappell, Billings, Scott, Tuttell, Olsen, and Kozon (1989) found that the range of pilot response times for making an input to the controls from the onset of the Resolution Advisory (RA) was one to seven seconds (standard deviation = 1.3 seconds). When testing a later prototype in which the Traffic Advisory (TA) alerted pilots to the conflict for 40 seconds before the RA and multiple RAs occurred in the simulation, the average response time was less than 2 seconds (standard deviation = 1.4 sec; Chappell, 1990). Observations of responses to TCAS in actual operations display a very different profile of pilot performance. Eurocontrol (2022) analyzed over 1,000 RAs that occurred in European airspace. They examined different ways to analyze the response, but regardless of the method, they found that "whether the assessment was at 8, 12 or 16 seconds after the RA, the compliance never exceeded 33%" (p. 79). Clearly, the flight deck behaviors observed in flight simulation studies cannot directly translate to expected behaviors in actual line

operations.

3.4.2.5 Cascading failures present additional complexity to identifying underlying assumptions, of flightcrew behavior.

While the task of assessing the validity of assumptions of flightcrew behavior in the case of a cascading failure is primarily an AIR certification concern, cascading failures also need to be considered in AED evaluations (e.g., operational suitability assessments and T-tests). As early as 2016, a Boeing report noted that, “As airplane systems have become more complex and interdependent, the challenge of building well-behaved systems becomes more difficult” (DeSalvo & Fogarty, 2016, p. 4). Due to highly integrated architectures, the failure of a single component can propagate, resulting in a variety of failure effects. The authors recognized that this increases the challenges involved in safety assessments. While the focus of the Boeing paper was the need for improved validation and verification of the engineering assumptions, it mirrors the need for improved validation of assumptions of flightcrew responses to failures. Consistent criteria would also provide a benefit to applicants by providing predictability and reducing the need for issue papers. The same types of benefits would be expected by developing consistent criteria for T-tests and operational suitability assessments. FAA’s Notice of Proposed Rulemaking (NPRM) titled, “System Safety Assessments,” (FAA, 2022) and subsequent rulemaking (FAA, 2024) sought to “improve aviation safety by making system safety assessment (SSA) certification requirements more comprehensive and consistent”.

3.4.2.6 Flightcrew error is too complex to be reliably quantified in a probabilistic way.

In an examination of the utility of probability models of flightcrew error, Thomas and Abbott (2024) discuss the common assumption made by manufacturers, and considered acceptable in early SAE guidance, that any action that will be needed to be taken by a flightcrew to identify and correct a failure condition will be correct:

“The safety assessment process ... assumes that flightcrews, cabin crews, maintenance crews, and other individuals participating in the operation of the aircraft follow documented procedures in foreseeable operating conditions (normal, malfunction or abnormal, and emergency). Intentional or unintentional deviation from these procedures is not considered in the safety assessment process described herein” [SAE, 2023, p. 22 as in Thomas and Abbott (2024) p. 2].

In response to the fact that flightcrews have not always responded as assumed in safety assessments, Thomas and Abbott (2024) report that some have suggested that probabilistic human error analyses should be incorporated into system safety assessments, to rectify this issue. For example, Guo and Sun (2020) offer a tool for quantifying human error probability and assessing flight safety in the aircraft manipulation process. While the proposed tool is a

refinement of a previous model of human error, by the authors' own description, it is not sufficient to estimate flightcrew error in situations that would be assessed by AIR or AED. The authors state that more data are needed, under a diverse set of flight conditions to make the data reliable and reduce uncertainty in the model. The author also notes that tools need to be developed for the data to be usable.

Thomas and Abbott (2024) describe several reasons why models or schemes that purport to predict flightcrew error will never be able to deliver valid predictions of the type needed by FAA Certification and AED personnel. Such models are limited to what they can validly and reliably predict based on the data which they are built and the limited operational contexts included to train the model. FAA Certification and AED personnel, however, need the flexibility to assess new designs, new applications, and new operational procedures.

Avionics are expected to continue to advance to support increasingly sophisticated air traffic control operations that result in increased airspace efficiency, reduced operator costs, and reduced emissions. Such operations will likely have reduced error tolerance and increased flightcrew task complexity. As task complexity increases, the variability in flightcrew responses tasks – that is, how abnormal situations are assessed and addressed by flightcrews, will likely vary with different operational scenarios. Flightcrew response to complex novel situations involving new equipment and/or procedures will need to be assessed on a case-by-case basis in the context within which the equipment, functions and/or procedures intend to be used.

3.5 Identification and Verification of Assumptions of Flightcrew Behavior in Safety Assessments

Only one document reviewed by the research team specifically addressed the issue of assumptions of flightcrew behavior in equipment design and safety assessments - ICAO Circular 361, Personnel Training and Licensing Panel Automation Study Report (ICAO, 2024). While the focus of this work was the flightcrew's use of automated functions, many of the results speak more broadly to the issue of the assumptions made by OEMs of flightcrew skills and expected behavior. The report notes that, "Manufacturers made many assumptions about manual flight, use of automated systems, and pilot training that were integrated into equipment design, manufacturer-recommended training programs, procedures, and documentation" (p. 20).

The ICAO team found that the assumptions varied among manufacturers. However, they noted two common, yet critically important, assumptions of pilot behavior that are clearly design-dependent. One common assumption was that the flightcrew would always maintain situation awareness. Another common assumption was that the flightcrew would always use an appropriate level of automation, although they noted that there was no uniform interpretation

of the levels of automation across manufacturers (ICAO 2024).

In 2023, the European Union Aviation Safety Agency (EASA) implemented an approach to guide the identification and verification of assumptions of flightcrew behavior in safety assessments for large aircraft. EASA noted that, “Most applicants do not conduct any systematic and structured activity to demonstrate the validity of FHA assumptions” (p. 5). They sought to address the problem of inaccurate assumptions of flightcrew behavior made in FHAs that would negate the intent of the FHA – that is, to ensure that the identification of failure conditions is complete, that the severity classification of the failure condition is correct, and that the identified remedies to negate the hazard are adequately substantiated:

Recent experience has shown that a disparity may exist between: the actually observed flightcrew behaviours, and the underlying assumptions about flightcrew recognition, interpretation, and response that applicants have made during the design and certification process. These disparities may invalidate the assumptions made in aircraft and system FHA and ultimately the validity of other assessments. (p. 5)

They also noted that the necessary prerequisites for effective flightcrew corrective actions – adequate recognition and correct interpretation of the failure condition, construction of an appropriate plan of action, and sufficient time to assess the situation and perform the corrective action(s) necessary to address the failure condition are “usually considered by applicants in aircraft and system FHA as implicitly given and fulfilled by default. These assumptions may be indirectly confirmed, and their associated requirements verified in other processes that are not directly connected to the FHA” (p. 5).

As an aid to remedy this safety concern, EASA offered a framework for aiding in the identification, assessment, and tracking of assumptions of the behavior of flightcrews in large aircraft in safety assessments. This framework will be discussed in detail in a later section of this paper. It should be noted, that while the framework intended to provide applicants with a “structured human factors process that may be used to confirm the assumptions made about expected flight crew behaviours” (p. 3), some manufacturers did not consider the guidance to be sufficient as a validation technique to airplane level functional hazard assessments. In submitted comments regarding the EASA framework, Boeing noted that “Airplane level functional hazard assessments are more general than system level assessments and inherently cannot have discrete actionable failure conditions due to the integrated aircraft functional level of the conditions. These more general failure conditions cannot be related to discrete crew actions and tasks that can be validated or tested unless the failure condition are further decomposed into sets of system level failure conditions.” (EASA – Certification Memorandum

“Flight Crew Human Factors Assumptions in Aircraft and System Safety Assessments - CM-SA-002 Issue 01 – Comment Response Document dated 10 May 2023). As part of their comments, Boeing requested more guidance on methodology for airplane level crew assumption validation. EASA’s response was very telling, stating “The level of details of the Aircraft FHA varies from one applicant to the other. The certification memo has been written in such way that it would be compatible with all types of safety assessment process.”

In conclusion, the variability of the details presented by applicants is seen in Europe as well as in the US. While EASA offered a solid framework for the identification, assessment and tracking of critical assumptions of flightcrew behavior in safety assessments, it cannot be considered a complete method. It does, however, provide a foundation that could be enhanced to aid AED personnel.

3.5.1 The EASA framework

As an aid to remedy this safety concern, EASA offered a framework to aid the identification, assessment, and tracking of assumptions of flightcrew behavior that was to be used in the evaluation of new type designs, significant changes to a type design or “any major change that introduces new failure conditions or significantly affects existing failure conditions (change in flight deck effects or in assumed flight crew reaction) on large aeroplane”. For each system failure condition where expected flightcrew action is part of the severity determination and based on the agreed task analysis framework, the applicant is expected to provide a set of prescribed information about the means and methods used to assess and confirm the assumptions about flightcrew behaviors.

Part one of this framework requires the applicant to submit responses to a set of questions for each failure condition where expected flightcrew action is part of the severity determination. It includes questions designed to address the cues available to the flightcrew to identify the hazard. It also includes questions as to the specific flightcrew behaviors that are expected to assess and mitigate the failure and post-failure operational implications. It should be noted that the information that the applicant is expected to provide to EASA aligns with the recommendation in AC 25.1329-1C (2016) that the applicant provide any information needed to validate assumed flightcrew actions used “to initiate appropriate action” and “mitigate the propagation of the effect of a failure condition” (p. 61).

Part two of the EASA framework contains a very general guide to assessing the “level of scrutiny” required to assess any of the issues identified in the assessment of the answers to the questions in Part one. If flightcrew action is identified in the recognition or mitigation of the hazard in the safety assessment, and the “expected behavior can be predicted with a high degree of confidence”, then the assumptions can be validated by the expert judgement of a

multidisciplinary team of engineering, flight test, safety and human factors experts. This team may seek additional data to support their judgement. If the team decides that the expected crew behavior cannot be “predicted with a high degree of confidence”, then “the assumption validation requires a full human factors process” (p. 11).

Part three of the framework requires the applicant to document the expected flightcrew behavior as an assumption within the safety assessment process and provide traceability to the supporting evidence that the assumption is justified.

EASA was the first organization to develop an approach to identifying, assessing, and tracking critical assumptions of flightcrew behavior in safety assessments. It should be noted that the goals of the framework proposed by EASA are consistent with many of the goals of FAA guidance material. Yeh et al (2016) states that the applicant must show that the flightcrew can be expected to complete their tasks successfully “in both normal and non-normal operational conditions, within the bounds of acceptable workload and without requiring undue concentration, exceptional skill or strength, or causing undue fatigue”. However, since the focus is on the regulatory guidance, there is no attempt to provide operational definitions for terms such as “undue concentration”, “exceptional skill” or “undue fatigue”. Similarly, both the EASA and the FAA guidance identify the importance of the applicant demonstrating that the flightcrew can detect failures and intervene appropriately. The EASA approach, while laudable, is not sufficient for AED (or AIR) purposes. The framework could, however, be used to develop such tools, given sufficient AED input and available resources.

4. Analysis of available datasets

To support AED identification, review, and substantiation of operational human factors assumptions, a search for publicly available databases and datasets relevant to aviation safety and AED personnel was conducted. Information in these databases could inform AED operational recommendations for repeatable methods, processes and checklists.

Over fifty databases were identified, however, most did not meet the requirements of being accessible to the general public, active, and searchable on relevant information. Of the relevant databases, three datasets were chosen for the quality, currency, and relevance of the information: the Aviation Safety Reporting System (ASRS), the National Transportation Safety Board (NTSB) Case Analysis and Reporting Online (CAROL) aviation investigation database, and the SKYbrary Accident and Incident browser. These three databases were chosen due to the potential relevance for informing operational assumptions of flightcrew behavior, breadth of information available, search functionality, and currency of information. The data source, currency, strengths, and limitations for each selected data set (ASRS, NTSB, & CAROL) are

described below.

For an overview of all the searchable public aviation databases identified, please refer to Appendix E. It should be noted that the Aviation Safety Information Analysis and Sharing (ASIAS) maintains a rich compilation of both public and private databases. According to FAA (2020), ASIAS combines rich data sources including:

- Surveillance radar data,
- Digital Flight Operational Quality Assurance (FOQA) data,
- Aviation Safety Action Program (ASAP) reports,
- Air Traffic Safety Action Program (ATSAP) reports,
- Air Traffic Mandatory Occurrence Reports (MORs),
- Air Registry (AR),
- FAA Accident and Incident Data Systems (AIDS),
- FAA Near Midair Collision System (NMACS).

While some of the databases that are part of the ASIAS system are publicly accessible and searchable (e.g., AR, AIDS, and NMACS), the ASIAS program accepts specific inquiries encompassing the entire database system if the inquiry is sponsored by the stakeholders.

4.1 ASRS

Managed by NASA, the ASRS database is the world's largest repository of voluntary, confidential safety information with one million reports having been submitted. Although anyone can submit a report, they are primarily submitted by pilots and air traffic controllers. These reports are deidentified before they are included in the database. The strengths and limitations of the database are described below. A complete tutorial on how to use the ASRS database can be found in Appendix B.

4.1.1 Data source

ASRS reports are derived from self-reports that can be submitted via either electronic submission or mail. In addition, reports can be transferred from organizations, e.g., airlines' internal reporting system. ASRS collects voluntary narratives using the NASA ASRS forms 277A (air traffic), 277B (general use), 277C (cabin crew), 277D (maintenance), and 277U (unmanned operations). Requested information varies depending on the type of report being submitted; in general, reporters are prompted to record information related to the event such as:

- Position held (e.g., captain, first officer, pilot flying, single pilot)
- Flying time

- ATC experience
- Airspace
- Weather conditions
- Visibility
- Details of involved aircraft(s)
- Location
- Free text narrative to provide a description of the event

All event reports are reviewed by two aviation safety analysts. When analyzing the reports, analysts identify any hazards that are discussed and can flag information relevant to areas of interest identified by FAA. In addition, analysts classify the report and identify underlying causes for the event. Analysts have backgrounds as pilots, air traffic controllers, and airplane mechanics in a variety of aviation environments such as air carrier, military, and general aviation.

4.1.2 Currency

Reports in the ASRS database range from 1988-present. The currency of uploaded reports varies and can depend on the backlog of reports for processing. The database website notes that reports are usually not available until at least 60 days from the date of receipt.

4.1.3 Strengths

The ASRS narratives can provide a rich source of information to researchers, including elements that might not be included in other types of reports such as an individual's decision-making process when encountering a hazard and effect of specific experience on a particular response. These topics and findings can directly inform human factors assumptions.

Most useful to AED specialists is the ability to easily search for reports filed only by specific populations such as pilots operating under Part 121 and/or Part 135. In this way, a search could be matched to the target population of an assessment. Searches can also be constrained by type of equipment, by group, or a specific model. For example, searches can be conducted for reports involving B727 all series or specific models such as B737 NextGeneration. Finally, the search allows for specific text contained in the narrative or synopsis. A complete list of search categories is presented in Appendix B.

The fact that ASRS guarantees confidentiality (except in cases of criminal activity or accidents) for reporters and that the FAA will often waive civil penalties or certificate suspension (assuming the violation was inadvertent and not deliberate) when an ASRS report is made within 10 days of the violation means that people may have a strong incentive to submit the

report. As a result, ASRS can provide insights into events which tend to be underreported. In addition, since data are relatively recent, they can point to current gaps in aviation safety before they result in an accident, and may even describe error mitigation strategies. For example, 20 years ago safety analysts were perplexed as to why even highly trained air carrier pilots were involved in a multitude of runway incursions. Most commonly, these incidents involved crossing the hold short lines, but not entering the runway. According to FAA data, in 40% of these errors, the pilots had correctly readback the ‘hold short’ instruction (note that these data were gathered *before* there was a requirement to readback ‘hold short’ instructions). No contributing factors, or insights into why these errors occurred could be derived from the FAA reports. However, an analysis of ASRS reports showed that in cases in which the flightcrew intended to hold short of the runway and crossed the hold short lines it was because both pilots were heads down performing other tasks, such as entering data into the Flight Management System or conducting a checklist (DiFiore and Cardosi, 2006). Once the hold short lines were crossed, there was no indication of the deviation until they could see the runway edge. As a result of this finding, hold short lines were enhanced, both in contrast and distance, and guidance was given to pilots not to conduct checklists in the vicinity of a runway.

4.1.4 Limitations

Although the fact that the ASRS database consists of self-reports is in many ways a strength, it also means that reports are subject to limitations of the reporter’s memory. Further, a reporter submitting to ASRS may be biased or have incomplete information regarding the incident. These reports are not investigated and are only “corroborated” when more than one crewmember reports the same event. ASRS reports cannot be interpreted in the same way as other reports (such as Mandatory Occurrence Reports). For example, if there are many reports in the ASRS database that describe a given problem, we can look to those reports for a better understanding of the reporters’ (pilots’ and/or controllers’) perspective on the issue, but the number of reports cannot be extrapolated into the frequency of such issues. ASRS reports are in narrative form, which can potentially lead to subjective and inconsistent descriptors or level of detail. This is especially relevant when using the “Text” search function in the narrative portion of the report (described below). Even with two analysts, there may be inconsistencies on the codes used to classify the data. It is also the case that acronyms, and even words, can have different meanings depending on the context. For example, the acronym “MCAS,” which may refer to the aircraft’s “Maneuvering Characteristics Augmentation System,” is also used as an abbreviation within ASRS for “Marine Corps Air Station.” Similarly, “response time” can refer to humans or engine components.

In addition, searches can be difficult to replicate over time due to the complex search algorithm. There is the potential for “closed-box searching,” defined here as confusing or

inconsistent search results with limited or no ability for the user to determine why the search is behaving the way it is. For example, using the search terms for Event assessment > Event type > Aircraft Equipment Problems > Critical & Less Severe provided conflicting numbers of results, depending on whether or not the search was new, or had been edited and re-run. This suggests that there is some behind-the-scenes operation leading to these results, but the user has no way of knowing why certain reports are found based on the order of their search. This example also suggests that there may be other, similarly inconsistent searches that users could encounter. The number of entries in the ASRS database may also mean that searches return report quantities that are simply unmanageable for research or analysis purposes. For example, if one was to select Event assessment > Human Factors > Time pressure as a search criterion, 13,907 results would be returned. For this reason, the more specific the search can be, e.g., with respect to the type of operation and specific text, the more useful the results will be.

4.2 NTSB CAROL aviation investigation search

Managed by NTSB, the CAROL database consists of all NTSB recommendations (1967-present), aviation investigation data (1962-present) and surface mode investigations (highway, marine, railroad, pipeline, and hazmat) (2010-present). The discussion below concerns the aviation investigation search function, a subset of the overall CAROL database. The strengths and limitations of the CAROL aviation database are described below. A complete tutorial on how to use the database can be found in Appendix C.

4.2.1 Data source

Records from the Census of U.S. Civil Aviation Accidents are included in the CAROL database. All civil aviation accidents (and selected incidents) within the U.S. and international waters are recorded. In addition, foreign investigations in which the NTSB participated as an accredited representative are also incorporated into the database. Over 170,000 accidents reports are available for review.

Accident investigations generally consist of an overall analysis of the event, probable cause and findings, list of factual information (history of flight, aircraft owner/operator information, meteorological information and flight plan, airport information, wreckage and impact information, and administrative information). In addition, accident investigations (once the final report is published) have a corresponding docket of information, which can provide additional information regarding the accident. The content of the docket varies depending on the investigation, but can contain items such as audio recordings, pilot/operator accident reports, flight data recorder information, interview transcripts, and investigator testimony.

4.2.2 Currency

The database is updated with aviation accident information on the first day of each month.

4.2.3 Strengths

A key strength of the NTSB CAROL database is the fact that it provides a comprehensive report on every accident or serious adverse event investigated by NTSB. This can allow users to look at patterns according to various factors such as location, operation type, weather condition, and phase of flight. Further, the docket information associated with completed reports can provide researchers an opportunity to review raw data and first-hand accounts.

To facilitate the search for aviation accidents, NTSB has developed summary statistics/dashboards, safety research reports, and published searches on specialty topics. Four of the 11 published searches relate to aviation, although none of the searches directly relate to pilot behavior (Child Restraints in Aircraft, All open aviation safety recommendations, Turbulence accidents since 2008, and TWA 800 Safety Recommendations).

4.2.4 Limitations

Despite the large scope of information available on the NTSB CAROL database, the ability to find useful information on a given topic (e.g., human factors assumptions) can largely depend on previous knowledge of relevant accidents and incidents. Although users can search accident reports based on factors such as operations type and weather, there are no search criteria for categories related to human factors such as probable cause or pilot action. It is possible to search for probable cause via the text search functionality to search for certain phrases within the accident report, but there is no published taxonomy of probable causes or human factors, nor does the language/coding appear to be standardized. For example, entering the phrase “pilot response time” into the search function returns only one report⁶ with a brief discussion regarding the effect of an unexpected event on pilot response time. However, there are well-known examples of accidents that related to “pilot response time” that *should* have been returned. For example, the “Miracle on the Hudson”⁷ accident report relates to pilot response time, stating

Simulation flights were run to determine whether the accident flight could have landed successfully at LGA or TEB following the bird strike. The simulations demonstrated that, to accomplish a successful flight to either airport, the airplane would have to have been turned toward the airport immediately after the bird strike. The immediate turn did not reflect or account for real-world considerations, such as the time delay required to recognize the extent of the engine thrust loss and decide on a course of action. The one

⁶ NTSB Accident No: CEN19LA056, 9/23/2020

⁷ <https://www.nts.gov/investigations/Pages/DCA09MA026.aspx>

simulator flight that took into account real-world considerations (a return to LGA runway 13 was attempted after a 35-second delay) was not successful. Therefore, the NTSB concludes that the captain's decision to ditch on the Hudson River rather than attempting to land at an airport provided the highest probability that the accident would be survivable.

However, because it did not include the exact phrase "pilot response time", it was not flagged in the CAROL query.

The text search functionality is also limited in that it does not allow operator functionality (AND, OR), so when searching for words or phrases, the search will only return reports with all words in the search, in order. In addition, the search functionality will only search final reports and titles, not landing webpages if the report is not yet in the "Final PDF" form or docket attachments.

The results page does not have an "update search" option, so users must refresh the search and gradually refine their search terms through re-selecting and re-submitting all their search terms. This is particularly challenging because the results return as PDFs so users must download PDFs to assess whether their search has returned the desired results before altering search terms.

4.3 SKYbrary Accidents and Incidents Reports browser

Housed on the SKYbrary website, the Accidents and Incidents report browser allows users to filter for aviation accidents and incidents that meet specific criteria. As of January 2025, the repository had 1,514 reports. These reports represent international accidents/incidents and include events from 1969-present. Included are the official accident reports that are summarized in the accident and incidents reports browser. SKYbrary also houses Bookshelf, a collection of international safety related material (such as rules, regulations, analyses, studies, investigations, and presentations) produced by civil aviation organizations.

4.3.1 Data source

Following a Wiki model, summary reports in the repository are collaboratively written and edited. The database relies on input of SKYbrary users, and anyone has the ability to submit drafts, comments, or suggested changes to articles and summary reports. However, any changes would need to go through a content management process.

The content can vary depending on the incident, but in general, these reports include narrative information about what happened, the investigation, probable causes, and safety recommendations. Additionally, these reports include event details, flight details, location,

SKYbrary indexing tags, the outcome, causal factor groups, safety recommendations, and investigation type.

Summary reports are derived from official reports from the corresponding civil aviation organization in charge of conducting the investigation. Links to official reports are included in each summary page, making it easy for users to obtain the primary source of information.

4.3.2 Currency

SKYbrary states that their content can be updated by the Editor. Because the SKYbrary Accidents and Incidents browser entries are derived from preliminary or final reports of aviation investigations, there will inevitably be a delay from the time the event occurred to the time there is a summary entry uploaded. However, in January 2025, there were at least two entries from events occurring in 2024, suggesting that summaries are uploaded in a timely manner. SKYbrary Bookshelf entries are even more timely, with some documents being uploaded weeks or days after publication.

4.3.3 Strengths

All reports in the Accidents and Incidents browser have been tagged, making filtering for relevant reports easier than relying on key word searches. In most cases, the filters show the number of reports with that specific tag, meaning users can understand the scope of relevant reports before executing their search. In addition, the SKYbrary database is the most comprehensive database of aviation accidents and incidents that occur outside the USA, allowing for an international perspective on aviation data.

In addition, there is a strong emphasis on human factors and the influence of human behavior on SKYbrary. For example, the database includes a comprehensive set of human factors search terms, including, but not limited to, terms such as the following: fatigue, distraction, inappropriate crew response (skills deficiency, automatics, or technical fault), ineffective monitoring, and pilot startle response. The breadth of these search terms may allow users to pare down their search without too many search iterations.

4.3.4 Limitations

Since the database includes data from other countries, the certification of equipment and pilot training requirements can be expected to differ. Additionally, the use of vocabulary may differ from report to report (for example, “Commander” vs. “Pilot Flying”).

While the pre-defined filters and search terms may be beneficial for some users, the lack of custom word string searches and “OR” operator means that search capabilities are limited. In

addition, the categorization of summary reports can be subject to human error. Some events were found to not be categorized correctly which could lead to the erroneous inclusion or exclusion of relevant data from search results. For example, on 22 July 2011 an Airbus A340-300 in night Instrument Meteorological Conditions (IMC) encountered moderate turbulence which led to an overspeed annunciation which was followed by the aircraft abruptly pitching up and a gain of over 3000 feet in less than a minute before it was returned to the previous cruise level over the Atlantic Ocean. The report noted that without the high angle of attack protection, the airplane would have kept its ascending trajectory until the triggering of the stall warning. The autopilot disconnect warning “had been rendered inaudible to both pilots by the concurrent sounding of higher category alerts, a situation which although in accordance with the 1995 type certification of the A340 would no longer meet current certification requirements.” The report also noted that the pilot not flying was probably surprised by the OVERSPEED warning as well as by the unanticipated turbulence and experienced a “startle effect.” As of January 2025, this report incorrectly includes the tag “Malicious Interference” rather than “Inadequate Monitoring” or “Pilot Startle Response.”

5. Discussion and next steps

After reviewing key documentation, interviewing AED human factors personnel, literature review and database analysis, several themes on human factors assumptions in the context of FHAs became evident. First, there is a need (stated by the JATR and NTSB reports and verified by AED human factors personnel) that the human factors assumptions, whether implicit or explicitly stated, should be validated and justified. Second, there is a need for increased standardization in the certification and operational evaluation processes. Third, the development of tools and methods could help with the validation of said assumptions and contribute to the standardization of assessments, especially given the lack of specific tools and methods found in a literature review process. Fourth, consistent documentation regarding human factors assumptions and a repository of previous decisions on human factors assumptions would further increase standardization and help to ensure that the review process is repeatable. Finally, leveraging known human factors principles would ensure that the above points are implemented based on existing methodology and best practices.

5.1 Validation of assumptions of flightcrew behavior

At the highest level, review of existing documentation and interviews with AED personnel revealed the need for assumptions of flightcrew behavior to be validated and justified in safety assessments. Specific findings and recommendations from published documents include:

- Assumptions of flightcrew behavior used by Boeing in FHAs for flight control systems did

not match actual flightcrew behavior (JATR, NTSB). Further, assumptions of flightcrew behavior used by Boeing were based on FAA guidance (JATR, NTSB).

- Assumptions of flightcrew response to a failure did not account for factors that could negatively affect flightcrews' ability to intervene in an accurate and timely manner (NTSB).
- There are gaps in existing FAA guidance regarding human factors assumptions (e.g., AC 25.1329-1C) and such gaps should be addressed (ACSAA Sec. 115, JATR, NTSB).
- Assumptions of realistic flightcrew behavior should be validated by applicants (ACSAA Sec. 119, JATR, NTSB).

Human factors specialists within AED confirmed the need for validation of assumptions of flightcrew behavior in safety assessments and state the need to uncover implicit assumptions for validation as well. While the analysis of available aviation databases did not yield well-organized, specific results that could directly support or challenge applicant assumptions included in safety assessments, they could still be useful to AED personnel in certain circumstances. For example, if an applicant documents an assumption regarding flightcrew response time, AED personnel could use the available databases to identify accidents or incidents where that assumption may be challenged, encouraging the applicant to provide more data or context.

5.2 Increased standardization in the certification and operational evaluation processes

Both applicants and the FAA would benefit from increased standardization in safety assessments. Specific recommendations included:

- Identify specific procedures that are to be applied consistently in the safety assessments (JATR).
- Provide for document traceability and clarity of explanations for changes to aircraft type designs and system safety assessment certification documents (ACSAA Sec. 115).
- Development and use of standardized criteria for determining the criteria for changes identified as substantial and significant; tools need to be developed to assist in this process (ACSAA Sec. 117).

5.3 Tools and methods needed

To facilitate the validation of human factors assumptions, more tools and methods need to be developed. This was noted in both the NTSB report and ACSAA (Section 126). While a literature review of potentially useful tools and methods identified sources that offered general guidance

(e.g., Billman et al., 2019; EASA framework), there was limited information that would be directly relevant to AED activities. As noted in the documentation review, NTSB recommended that industry and human factors provide input into the proposed tools and methods; once developed, said tools should be incorporated into FAA guidance. AED human factors personnel echoed the need for the development and use of standardized methods and tools in safety assessments. It will also be important to ensure that such tools are developed with the proper expertise and used by the proper personnel, in the proper context.

5.4 Establish consistent documentation

Consistent documentation is needed for improvement in standardization. Related findings and recommendations include:

- Clear documentation of technical details, failure modes, and effects of design changes should be included in FAA guidance (ACSAA Sec. 115). This would ensure the review of any change that results in an FHA classification of major or greater.
- Members of Flight Standardization Boards (Cardosi et al., 2024) and AED human factors specialists interviewed stated that a repository of completed assessments, FSB decisions, and rationale for said decisions should be established to provide guidance on future assessments.
- Standardizing format, content, and application of FAA Issue Papers would benefit both applicants and the AED (ACT ARC Recommendation 20-5, 2021).

5.5 Known human factors principles

The NTSB and JATR have identified human factors gaps in the FAA certification and operational assessment processes and recommended ways to infuse human factors principles and add more objectivity into FAA procedures. However, some of the issues pointed out by NTSB and JATR have been well-known to human factors specialists and practitioners for years, namely:

- Pilot response time depends on several factors including the nature of the malfunction, the operational circumstances under which it occurs, the corrective action required, and the individual pilot. In short, as aircraft systems and operational environments become increasingly complex, so do pilot tasks; these complexities affect pilot responses to malfunctions.
- Subtle failure modes will take longer for a pilot to recognize than obvious failures.
- Neither the responses of flight test pilots, nor of any pilots who test a particular malfunction with foreknowledge of the malfunction to be tested and the proper

response to be initiated, can be interpreted as indicative of the response of a line pilot on a revenue flight.

- Automated functions can mask some failures and significantly delay the opportunity for the pilot to recognize the malfunction.

AED human factors specialists are well-aware of these characteristics of flightcrew behavior and fundamental human factors principles. In interviews, AED human factors specialists agreed that these issues speak to the importance of assessing both aircraft systems and pilot behavior in a holistic way, rather than on a component-by-component basis.

Work is needed to increase the validity and reliability of flight deck safety assessments by identifying assumptions of flightcrew behavior in aircraft design and determining if such assumptions are operationally acceptable. Such validation is particularly critical when flightcrew actions are proposed by an applicant to mitigate hazards identified as “major” or greater. With the proper combination of operational and human factors expertise, known human factors principles, data, and methods could provide the foundation for developing tools for the identification and assessment of assumptions of flightcrew behavior in safety assessments.

5.6 Next steps

“Safety assessment” is a broad term. It includes FHAs that serve as foundational steps for identifying hazard level and mitigations and operational suitability assessments to determine, for example, whether given equipment is suitable for a specific use in certain operations and/or airspace. Different types of assessments will be best served by the development of specialized tools. The first step toward the development of tools for use by AED would be to specify the types of assessments that AED conducts or reviews to determine the types of tools that would be beneficial and the information that should be required from an applicant to support these assessments.

A formal AED framework for identifying, assessing, and documenting assumptions of flightcrew behavior in safety assessments would be beneficial in several ways. First, it would increase the standardization with which such assessments are conducted. Second, it could require that an applicant document and present any identified assumptions of flightcrew behavior for each failure condition in which expected flightcrew action is part of the severity determination and the disposition of that assumption. The documentation of the justification of the assumptions (e.g., with data from a simulation study or specific excerpts from training material) could facilitate the FAA’s review as well as support their identification of underlying assumptions that the applicant may have overlooked. The fact that tools and methods need to be developed to validate critical assumptions made by manufacturers of flightcrew recognition and response has been noted by NTSB (2019) and in ACSAA (Section 126). Further, NTSB recommended that

industry and human factors provide input into the proposed tools and methods; once developed, said tools should be incorporated into FAA guidance. AED human factors personnel echoed the need for the development and use of standardized methods and tools in safety assessments. To realize expected benefits, such tools will need to be developed with the proper expertise and used by the proper personnel in the proper context. This would involve working with AED to develop such tools and ensuring that the tools are useful to human factors specialists within AED.

6. References

- Air Carrier Training Aviation Rulemaking Committee (ACT ARC) (2021). *ACT ARC recommendations: Recommendation 20-5 Standard Issue Papers*.
https://www.faa.gov/sites/faa.gov/files/about/office_org/headquarters_offices/avs/ACT_ARC_Rec_20-5.pdf.
- Aircraft Certification, Safety, and Accountability Act, Division V of P.L. No. 116-260 (2020).
<https://www.congress.gov/116/plaws/publ260/PLAW-116publ260.pdf>.
- Beringer, D., Ball, J., Brennan, K., & Taite, S. (2005) *Comparison of a typical electronic attitude-direction indicator with terrain-depicting primary flight displays for performing recoveries from unknown attitudes: Using difference and equivalence tests*. (Report No. DOT/FAA/AM-05/23). Federal Aviation Administration, Office of Aerospace Medicine.
https://www.faa.gov/sites/faa.gov/files/data_research/research/med_humanfacs/oamtechreports/0523.pdf
- Beringer, D. & Harris, H. (1997). *Automation in general aviation: Two studies of pilot responses to autopilot malfunctions*. (Report No. DOT/FAA/AM-97/24). Federal Aviation Administration, Office of Aviation Medicine.
https://www.faa.gov/sites/faa.gov/files/data_research/research/med_humanfacs/oamtechreports/AM97-24.pdf
- Berson, B.L, Po-Chedley, D.A., Boucek, G.P., Hanson, D.C., Leffler, M.F., & Wasson, R.L. (1981). *Aircraft alerting systems standardization study volume II: Aircraft alerting systems design guidelines* (Report No. FAA/RD-81/38/1). Boeing Commercial Airplane Co., Seattle, WA.; Federal Aviation Administration, Washington, DC. [Aircraft Alerting Systems Standardization Study. Volume II. Aircraft Alerting System Design Guidelines. | National Technical Reports Library - NTIS](#)
- Billman, D., Mumaw, R., & Feary, M. (2019) Best practices for evaluating flight deck interfaces for transport category aircraft with particular relevance to issues of attention, awareness, and understanding CAST SE-210 Output 2 Report 6 of 6. (Report No. NASA/TM—2019–220390). National Aeronautics and Space Administration, Ames Research Center. <https://ntrs.nasa.gov/citations/20200001615>
- Boucek, G. P., Erickson, J. B., Berson, B. L., Hanson, D. C, Leffler, M. F., & Po-Chedley, D. A, (1980) *Aircraft alerting systems standardization study phase I. Final report*. (Report No. FAA/RD-80-68). Federal Aviation Administration, Systems Research and Development Service. <https://apps.dtic.mil/sti/tr/pdf/ADA090834.pdf>
- Boucek, G. P., Po-Chedley, D. A., Berson, B. L., Hanson, D. C, Leffler, M. F., & White, R. W. (1981) *Aircraft Alerting Systems Standardization Study, Volume I: System Validation and Time-Critical Display Evaluation* (Report No. FAA-RD-81-38-1). Federal Aviation

- Administration, Systems Research and Development Service.
<https://apps.dtic.mil/sti/tr/pdf/ADA107225.pdf>
- Boucek, G., White, R., Smith, W., & Kraus, J. (1982) *Traffic alert and collision avoidance system*. (Report No. DOT/FAA/RD-82/49). Federal Aviation Administration, Program Engineering & Maintenance Service. <https://apps.dtic.mil/sti/tr/pdf/ADA126342.pdf>
- Cardosi, K., Yahoodik, S., & Lennertz, T. (2024) *Human factors information and processes for FAA's Aircraft Evaluation Division – A review of the literature*. (Report No. DOT-VNTSC-FAA-24-04). Federal Aviation Administration. <https://rosap.ntl.bts.gov/view/dot/74664>
- Casner, S. M., Geven, R. W., & Williams, K. T. (2013). The effectiveness of airline pilot training for abnormal events. *Human Factors*, 55(3), 477-485. DOI:10.1177/0018720812466893
- Chappell, S.L. (1990) Human factors research for TCAS. *Proceedings of the Royal Aeronautical Society Conference on the Traffic Alert and Collision Avoidance System (TCAS) -Its Development, Testing, Installation and Operational Use*.
- Chappell, S. L., Billings, C. E., Scott, B. C., Tuttell, R. J., Olsen, M. C., & Kozon, T. E. (1989). *Pilots' use of a traffic alert and collision-avoidance system (TCAS 2) in simulated air carrier operations. Volume 1: Methodology, summary and conclusions* (Report no. NASA-TM-100094-VOL-1). NASA Ames Research Center. [Wayback Machine \(archive.org\)](https://www.nasa.gov/pubs/1989001main_pilots_1989001main.pdf)
- Clewley, R., & Nixon, J. (2023). A new facet of category theory: cognitive disadvantage and its implications for safety in the cockpit. *Ergonomics*, 66 (6), 762-771.
<https://doi.org/10.1080/00140139.2022.2125587>
- Dehais, F., Peysakhovich, V., Scannella, S., Fongue, J., & Gateau, T. (2015). “Automation Surprise” in aviation: Real-time solutions. Proceedings of the 33rd annual conference on human factors in computing systems, New York, NY, 2525–2534.
<https://dl.acm.org/doi/abs/10.1145/2702123.2702521>
- DeSalvo, P. & Fogarty, D. (2016). *Safety issues and shortcomings with requirements definition, validation, and verification processes*. (Report No. DOT/FAA/TC-16/39).
https://www.faa.gov/sites/faa.gov/files/aircraft/air_cert/design_approvals/air_software/TC-16-39.pdf
- Equipment, Systems, and Installations, 14 CFR § 25.1309 (2024).
<https://www.ecfr.gov/current/title-14/chapter-I/subchapter-C/part-25/subpart-F/subject-group-ECFR9f24bf451b0d2b1/section-25.1309>.
- European Organisation for the Safety of Air Navigation (2022). *The assessment of pilot compliance with TCAS RAs, TCAS mode selection and serviceability using ATC radar data*.
<https://skybrary.aero/sites/default/files/bookshelf/5842.pdf>
- European Union Aviation Safety Agency (2023). *Flight Crew Human Factors Assumptions in*

Aircraft and System Safety Assessments Final Certification Memorandum ref. CM-SA-002 Issue 02. <https://www.easa.europa.eu/en/document-library/product-certification-consultations/final-certification-memorandum-ref-cm-sa-002>

Federal Aviation Administration (1988). *Advisory Circular (AC) 25.1309-1A – Systems Design and Analysis.* https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_25_1309-1A.pdf

Federal Aviation Administration (2006). *Advisory Circular (AC) 25.1329-1B – Approval of Flight Guidance Systems.* https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC%2025.1329-1B.pdf

Federal Aviation Administration (2012). *Advisory Circular (AC) 25-7C– Flight Test Guide For Certification Of Transport Category Airplanes.* https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_25-7C.pdf

Federal Aviation Administration (2013). *Advisory Circular (AC) 120-53B - Guidance for Conducting and Use of Flight Standardization Board Evaluations - With Change 1.* https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_120-53B_Chg_1.pdf

Federal Aviation Administration (2013). *Advisory Circular (AC) 25.1302-1- Installed Systems and Equipment for Use by the Flightcrew.* https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_25_1302-1.pdf

Federal Aviation Administration (2014). *Advisory Circular (AC) 25.1329-1C – Approval of Flight Guidance Systems, Including Change 1.* https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_25_1329-1C.pdf

Federal Aviation Administration (2018). *Advisory Circular (AC) 25-7D– Flight Test Guide for Certification of Transport Category Airplanes.* https://www.faa.gov/documentLibrary/media/Advisory_Circular/25-7D.pdf

Federal Aviation Administration (2020). *Report to Congress: Report on the Status of Aviation Safety Information Analysis and Sharing (ASIAS) Capability Acceleration.* https://www.faa.gov/sites/faa.gov/files/2021-11/FAA_Report_on_Aviation_Safety_Information_Analysis_and_Sharing_ASIAS_03312020.pdf

Federal Aviation Administration (2020b). *Response to Official Report of the Special Committee on the Federal Aviation Administration’s Aircraft Certification Process. Aviation Safety April 2020 Final Report.* <https://www.faa.gov/sites/faa.gov/files/2022-08/FAAActionPlanResponseToSpecialCommitteeReport.pdf>

Federal Aviation Administration (2022). *Airplane Flying Handbook (FAA-H-8083-3C)* https://www.faa.gov/regulations_policies/handbooks_manuals/aviation/airplane_hand

book

- Federal Aviation Administration *Advisory Circular (AC) 25.1309-1B – System Design and Analysis*. https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_25.1309-1B.pdf
- Flightcrew Alerting, 14 CFR § 25.1322 (2010). <https://www.ecfr.gov/current/title-14/chapter-I/subchapter-C/part-25/subpart-F/subject-group-ECFR3043e8882f16d6e/section-25.1322>.
- Guo, Y., & Sun, Y. (2020). Flight safety assessment based on an integrated human reliability quantification approach. *PLoS ONE*, 15(4): e0231391. <https://doi.org/10.1371/journal.pone.0231391>.
- International Civil Aviation Organization (2024) Circular 361 Personnel training and licensing panel: Automation study report.
- Installed Systems and Equipment for Use by the Flightcrew, 14 CFR § 25.1302 (2013). <https://www.ecfr.gov/current/title-14/chapter-I/subchapter-C/part-25/subpart-F/subject-group-ECFR9f24bf451b0d2b1/section-25.1302>.
- Joint Authorities Technical Review (2019). *Boeing 737 MAX Flight Control System – Observations, Findings, and Recommendations*. https://www.faa.gov/sites/faa.gov/files/2021-08/Final_JATR_Submittal_to_FAA_Oct_2019.pdf
- National Transportation Safety Board (2014). *Descent Below Visual Glidepath and Impact with Seawall, Asiana Airlines Flight 214, Boeing 777-200ER, HL7742, San Francisco, California, July 6, 2013*. Aircraft Accident Report NTSB/AAR-14/01. Washington, DC. <https://www.ntsb.gov/investigations/accidentreports/reports/aar1401.pdf>
- National Transportation Safety Board (2019). *Safety Recommendation Report on Assumptions Used in the Safety Assessment Process and the Effects of Multiple Alerts and Indications on Pilot Performance*. (Report No. NTSB ASR-19-01). <https://www.ntsb.gov/investigations/accidentreports/reports/asr1901.pdf>
- Rantanen, E. M., Wiegmann, D. A., Palmer, B. O., & Musiorski, K. M. (2003). *A human factors database for matching human errors and technological interventions*. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 47 (4). <https://doi.org/10.1177/154193120304700426>
- Roy, K., & Beringer, D. (2002). *General aviation pilot performance following unannounced in-flight loss of vacuum system and associated instruments in simulated instrument meteorological conditions*. (Report No. DOT/FAA/AM-02/19). Federal Aviation Administration, Office of Aerospace Medicine. https://www.faa.gov/sites/faa.gov/files/data_research/research/med_humanfacs/oamt

- SAE International (2023). *Aerospace Recommended Practice: Guidelines for Conducting the Safety Assessment Process on Civil Aircraft, Systems and Equipment*. Warrendale, PA: SAE International, ARP4761TM Revision A.
- Shiffrin, R. M., & Schneider, W. (1984). Automatic and controlled processing revisited. *Psychological Review*, 91(2), 269–276. <https://doi.org/10.1037/0033-295X.91.2.269>
- Stanton, N. A., Salmon, P. M., Rafferty, L. A., Walker, G. H., Baber, C., & Jenkins, D. P. (2013). *Human factors methods: a practical guide for engineering and design*. CRC Press.
- Sumwalt, R. L., Landsberg, B., & Homendy, J. (2019). *Assumptions used in the safety assessment process and the effects of multiple alerts and indications on pilot performance*. National Transportation Safety Board. <https://www.nts.gov/investigations/accidentreports/reports/asr1901.pdf>
- System Safety Assessments, Notice of Proposed Rulemaking (2022). Docket No.: FAA-2022-1544; Notice No. 23-04. <https://www.federalregister.gov/documents/2022/12/08/2022-26369/system-safety-assessments>.
- Thomas, L. J., & Abbott, K. H. (2024). The devil between the details: Limitations of probability-based approaches to human error. *Proceedings of 21st International Engineering Psychology and Cognitive Ergonomics Conference*, Washington, DC.
- Yeh, M., Swider, C., Jo, Y. J., & Donovan, C. (2016). *Human factors considerations in the design and evaluation of flight deck displays and controls version 2.0*. (Report No. FAA/TC-16/56 DOT-VNTSC-FAA-17-02)

Appendix A: Interview questions

Background

We are Human Factors researchers at the U.S. DOT, Volpe Center and have been tasked with documenting public data sources and other information that may be useful to AED when reviewing and validating OEM assumptions of expected pilot behaviors. We are conducting semi-structured interviews to better align our research tasks with the needs of AED personnel. Toward that goal, we are looking for (a) de-identified and generic examples (or categories) of human factors assumptions that you may need to assess and (b) how these assumptions are tracked and communicated to AED. Specifically, we are interested in assumptions that might be made in relation to functional hazard assessments (FHAs) and other system safety assessments (SSAs).

These assumptions can be stated explicitly or implicitly. A hypothetical example of an explicit assumption is that the pilot will make a correctional input into the controls within 4 seconds of alert activation. An implicit assumption is one that is not *stated* by an OEM but identified by human factors specialists. An example of an implicit assumption is that an OEM has assumed (only) the typical training of US Part 121 carriers, or that the pilot would be as likely to be able to recognize a fault if the alert or automated behavior was intermittent as if it was continuous. We are interested in the types of assumptions that AED may encounter and how such assumptions are likely to be conveyed. The responses obtained will be used to help guide an examination of searchable databases for potential use by AED.

Discussion questions

The questions below will be used as a high-level guide for the discussion. Any question can be skipped, and we understand that some questions may not be applicable to your specific responsibilities or job description. In addition, we know specific information may be proprietary and cannot be shared. Discussion findings or statements will not be attributed to an individual. We welcome any information that would make the resulting product more useful to the AED.

Assumptions of pilot behavior

- In your experience, what *explicit* assumptions of pilot behavior have you encountered, or could you envision?
- In your experience, what *implicit* assumptions of pilot behavior have you encountered, or could you envision?
- Which do you think you're more likely to encounter, explicit or implicit assumptions of pilot behavior?
- Are there some HF assumptions (explicit or implicit) that are more likely to be made than others?

- Can you provide an example (real or hypothetical) of someone unknowingly making a human factors assumption?
- Is there any AED guidance on how such assumptions could be identified or assessed? If so, please describe.

AED Processes and tools

- How could HF assumptions be identified and assessed?
 - Are there any specific tools or methods that could be used to assist in these assessments?
 - What guidance, tools, or methods would you like to see developed or refined to aid you in reviewing applicant assumptions?
- Are there data sources that you could turn to in your review of such assumptions?
 - If so, what are they?
 - What type of data source might be helpful if it were to exist in the future?
- How could changes in HF assumptions be tracked from the beginning of AED involvement until operational approval?
- How should the results of an applicant's FHAs or other SSAs used in operational suitability assessments, T-tests, and/or other AED tasks?

Are there any additional comments or ideas that you would like to share with us?

Appendix B: ASRS Tutorial

Search execution

Techniques for searching and extracting relevant data

- Review the provided ASRS Coding Taxonomy prior to searching to choose useful search terms https://asrs.arc.nasa.gov/docs/dbol/ASRS_CodingTaxonomy.pdf
- Review the links under “New Users”
<https://asrs.arc.nasa.gov/search/dbol/newusers.html>
 - Program Briefing <https://asrs.arc.nasa.gov/overview/summary.html>
 - About ASRS Data <https://asrs.arc.nasa.gov/search/dbol/aboutdata.html>
 - Search Strategies <https://asrs.arc.nasa.gov/search/dbol/strategies.html>
 - Search Examples <https://asrs.arc.nasa.gov/search/dbol/samples.html>
- Use increasingly narrow searches with criteria specific to the desired report type to avoid overwhelming numbers of results in a final search.
- The algorithm sometimes conflates previous searches with new ones. It is a good idea to create a “new search” for new topics, rather than editing a previous search.
- If using narrative (Text) searches:
 - Become familiar with the ASRS Text Reserved Word Commands and Operating Characters (OR, AND, NEAR, NOT, %, and “”) found on [this website page](#).
 - Note: If copying and pasting search terms from a separate document, users may need to re-insert characters such as quotation marks if the font is not recognized by the search system.
 - Note: Even though the [website](#) states in the “Textual Fields and Text Reserved Word Commands and Operating Characteristics” section that quotation marks are used to keep phrases together, this was found to not be the case. From our observations, the “text” search will assume that phrases are connected unless “OR” or parentheses are used. See Table 1 below.
 - Note: ASRS will automatically interpret users’ text input. See Table 1 for an illustration of how elements such as parenthesis and operators impact search results.
 - Include both the full term, and any potential acronyms, found in [this Abbreviations PDF](#), using the OR word command.
 - Ex. (Distract% OR DISTR) AND (“Flight Management Computer” OR FMC)
 - Check searches for misspellings and paired parentheses.

- Go back to the previously run search by selecting “Go back and edit current search,” clicking the link for the words, and looking at how the algorithm interpreted the search terms (i.e. the column for interpretation in Table 1).
- Confirm that the search has the desired selection of “Narrative” and/or “Synopsis” (i.e. For more comprehensive searches, ensure that both checkboxes are selected).
- Review the ASRS Textual Fields and Text Reserved Word Commands and Operating Characters section of the ASRS Search Strategies webpage, found [here](#).
- To avoid closed-box mistakes, run multiple searches with different combinations of the search terms and ensure that search results are consistent.

Table 1. Example "text" searches and results

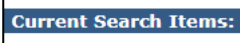
Text (as typed)	Text (as interpreted) ⁸	Number of ACNs	Interpretation
MCAS	(MCAS OR {MARINE CORPS AIR STATION})	52	There are 52 reports that refer either to the word “MCAS” or to the “Marine Corps Air Station.” “MCAS” may refer to either the “Marine Corps Air Station” or the “Maneuvering Characteristics Augmentation System”
“MCAS”	("MCAS" OR {MARINE CORPS AIR STATION})	52	There are 52 reports that refer either to the word “MCAS” or to the “Marine Corps Air Station.” “MCAS” may refer to either the “Marine Corps Air Station” or the “Maneuvering Characteristics Augmentation System”
wing	(WING)	13775	There are 13775 reports that contain the word “wing”
“wing”	(WING)	13775	There are 13775 reports that contain the word “wing”
(MCAS) OR (wing)	(MCAS) OR (WING)	13821	There are 13821 reports that contain either “MCAS” or “wing”
MCAS wing	(MCAS WING)	0	There are no reports that have the phrase “MCAS wing” in them. Without the OR, or AND, only reports with the two terms adjacent are returned.
(MCAS) and (wing)	(MCAS) AND (WING)	5	There are 5 reports that use both the word “MCAS” and the word “wing”
MCAS and wing	(MCAS AND WING)	5	There are 5 reports that use both the word “MCAS” and the word “wing”
MCAS or wing	(MCAS OR WING)	13821	There are 13821 reports that contain either “MCAS” or “wing”
Maneuvering Characteristics	(MANEUVERING CHARACTERISTICS)	7	There are 7 reports that use the phrase “maneuvering characteristics”

⁸ This column refers to the way ASRS seems to interpret the user’s search. This is found by returning to the search and re-opening the “Text contains” field to observe how ASRS has altered it.

Augmentation System	AUGMENTATION SYSTEM)		augmentation system"
---------------------	----------------------	--	----------------------

- The ASRS algorithm does not recognize special characters other than specific characters mentioned in their search strategies. This includes hyphens in words such as “non-normal,” which needs to be changed to “non normal,” for example.
- Due to the ASRS algorithm, one needs to use caution when pairing parentheses to ensure that phrases maintain their intended meaning. In the complex example search below, several searches were returned with errors due to mismatched or missing parentheses.
- The reports vary in their level of detail, making it harder to distinguish which will provide a necessary amount of information without having to go through a number of filtered search items. Sorting by general narrative length, for example, may be helpful.
- Results can be exported via Excel, csv, or Word file, allowing for efficient analysis.

Search tutorial

1. In a preferred search engine, search <https://asrs.arc.nasa.gov/search/database.html>
2. Determine the search terms to be used depending on the type of operator, operating environment, and other operational factors of interest.
 - Review ASRS Coding Taxonomy for insights into search terms
 - Review ASRS Abbreviations for an understanding of abbreviations used in report narratives
3. Click “Start Search” 
4. Select applicable search terms from the lists available on the search page
 - To choose search terms, click the gray “plus” sign  beside the category
 - *Options are included but not limited to report number, date of incident, location, environmental factors, people involved, aircraft, event details, and/or specific text.*
 - To multi-select, hold down the Control key or the Shift key
 - Navigate to the bottom of the box under “Current Search Items:” 
 - Click blue underlined hyperlinks that say “Click here”  and select applicable terms from ASRS’s pre-set fillers
 - *Examples include dropdowns for dates, lists for specific terms, and fill-in-the-blanks*
5. Optional:
 - Remove search terms if necessary by clicking the green “negative” sign  beside the category

6. Click “Run Search”



7. Select options to display results

- Options include viewing reports online, exporting results, exporting search criteria, and modifying the search.

Example search

The example search below references an article by Chandra et al., which used ASRS reports “to study flightcrew response to system failures, malfunctions, and systems not functioning as expected.”⁹ For the purposes of this example, the search was altered slightly. These changes include the date range (from 2021-2022, to 2023-2024) and the FAR Part (from 121 and 135 to 121 only), but the search terms were taken directly from the paper.

Objective

The users want to better understand incidents involving alerts on the flight deck, functionality of systems, and flightcrew checklists and workload in aircraft carriers. The research is taking place in July of 2024 and the users are aiming for a sample size of $n=30$.

Process

Search Attempt 1:

The users must first select a baseline set of search terms. There is a search option for “Human Factors,” which has detailed options for specific human factors issues. However, the users in this case chose to curate their own detailed list of search terms, in part because the tags as specified did not account for the use cases that they were looking for.

1. **Date of Incident** was between January, 2019 and July 2024
2. **Federal Aviation Regs (FAR)** Part was Part 121 (Regularly Scheduled Air Carrier)
3. **Reporter Organization** was Air Carrier (i.e. pilots self-reported)
4. **Event Type** was Critical or Less Severe (Aircraft Equipment Problem)
5. **Text Contains:** (both Narrative and Synopsis)
 - a. ((ALARM% OR ALERT% OR WARN% OR CAUTION% OR ADVISORY OR ANNUNCIAT% OR FLAG% OR (MESSAGE% AND (ECAM OR ELECTRONIC CENTRALIZED AIRCRAFT MONITOR)) OR (MESSAGE% AND (EICAS OR ENGINE INDICATING AND CREW ALERTING SYSTEM)) OR (MESSAGE AND (CAS OR CREW ALERT% SYSTEM)) OR (MSG AND (ECAM OR ELECTRONIC CENTRALIZED AIRCRAFT MONITOR)) OR (MSG AND (EICAS OR ENGINE INDICATING AND CREW ALERTING SYSTEM)) OR (MSG AND (CAS OR CREW ALERT% SYSTEM))) OR (LAND AS SOON AS POSSIBLE OR TURN BACK OR (DIVERT% OR DIVERSION%) OR EMERGENCY PROC% OR ABNORMAL PROC% OR NON

⁹ (Chandra et al., 2023)

NORMAL PROC% OR STALL OR STERILE COCKPIT OR (QRH OR QUICK REFERENCE HANDBOOK) OR (QRC OR QUICK REFERENCE CHECKLIST) OR EMERGENCY CHECKLIST% OR ABNORMAL CHECKLIST% OR NON NORMAL CHECKLIST%))

- b. Interpretation of text: Search terms in blue include many options for words or phrases that may be related to alerts. Search terms in orange include many options for words or phrases related to procedures and checklists. All alerting terms are contained within one overarching set of parentheses, and all procedures/checklists terms are similarly contained within a separate set of parentheses. These parenthesized elements are connected with the operator “OR,” so results should involve either some alert term or some procedure/checklist term.

This search returns 4268 results, which is an unmanageable number for this research.

Search Attempt 2:

The large number of search results may have been due to the operator “OR” between the alert terms and the checklist terms. Since the users want to understand the interaction between alerts and procedures/checklists, it may benefit them to change this to “AND” without altering any other search terms.

6. **Date of Incident** was between January, 2019 and July 2024
7. **Federal Aviation Regs** (FAR) Part was Part 121 (Regularly Scheduled Air Carrier)
8. **Reporter Organization** was Air Carrier (i.e. pilots self-reported)
9. **Event Type** was Critical or Less Severe (Aircraft Equipment Problem)
10. **Text Contains:** (both Narrative and Synopsis)
 - a. ((ALARM% OR ALERT% OR WARN% OR CAUTION% OR ADVISORY OR ANNUNCIAT% OR FLAG% OR (MESSAGE% AND (ECAM OR ELECTRONIC CENTRALIZED AIRCRAFT MONITOR)) OR (MESSAGE% AND (EICAS OR ENGINE INDICATING AND CREW ALERTING SYSTEM)) OR (MESSAGE AND (CAS OR CREW ALERT% SYSTEM)) OR (MSG AND (ECAM OR ELECTRONIC CENTRALIZED AIRCRAFT MONITOR)) OR (MSG AND (EICAS OR ENGINE INDICATING AND CREW ALERTING SYSTEM)) OR (MSG AND (CAS OR CREW ALERT% SYSTEM))) AND (LAND AS SOON AS POSSIBLE OR TURN BACK OR (DIVERT% OR DIVERSION%) OR EMERGENCY PROC% OR ABNORMAL PROC% OR NON NORMAL PROC% OR STALL OR STERILE COCKPIT OR (QRH OR QUICK REFERENCE HANDBOOK) OR (QRC OR QUICK REFERENCE CHECKLIST) OR EMERGENCY CHECKLIST% OR ABNORMAL CHECKLIST% OR NON NORMAL CHECKLIST%))
 - b. Due to the change of the operator from “OR” to “AND,” the results should now involve both some alert term and some procedure/checklist term.

This search returns 1697 results, which is still an unmanageable number for this research.

Search Attempt 3:

To narrow the results, the users limit the time range to 1 year instead of 5 but do not alter any other search terms.

1. **Date of Incident** was between January, 2023 and July 2024
2. **Federal Aviation Regs** (FAR) Part was Part 121 (Regularly Scheduled Air Carrier)

3. **Reporter Organization** was Air Carrier (i.e. pilots self-reported)
4. **Event Type** was Critical or Less Severe (Aircraft Equipment Problem)
5. **Text Contains:** (both Narrative and Synopsis)
 - a. ((ALARM% OR ALERT% OR WARN% OR CAUTION% OR ADVISORY OR ANNUNCIAT% OR FLAG% OR (MESSAGE% AND (ECAM OR ELECTRONIC CENTRALIZED AIRCRAFT MONITOR)) OR (MESSAGE% AND (EICAS OR ENGINE INDICATING AND CREW ALERTING SYSTEM)) OR (MESSAGE AND (CAS OR CREW ALERT% SYSTEM)) OR (MSG AND (ECAM OR ELECTRONIC CENTRALIZED AIRCRAFT MONITOR)) OR (MSG AND (EICAS OR ENGINE INDICATING AND CREW ALERTING SYSTEM)) OR (MSG AND (CAS OR CREW ALERT% SYSTEM))) AND (LAND AS SOON AS POSSIBLE OR TURN BACK OR (DIVERT% OR DIVERSION%) OR EMERGENCY PROC% OR ABNORMAL PROC% OR NON NORMAL PROC% OR STALL OR STERILE COCKPIT OR (QRH OR QUICK REFERENCE HANDBOOK) OR (QRC OR QUICK REFERENCE CHECKLIST) OR EMERGENCY CHECKLIST% OR ABNORMAL CHECKLIST% OR NON NORMAL CHECKLIST%))

This search returns 361 results, which may be a manageable number of results for the users to contend with. The users may choose to randomly sample results until they find the desired number of results that satisfy their criteria (e.g., Since their $n=30$, they may randomly read reports until they find 30 reports that involve the relevant interaction between alerts and checklists).

If users wanted to continue narrowing their search to a smaller set of results, they could consider the following: further constraining their timeframe, limiting the Aircraft Equipment Problem to either “Critical or Less Severe”, or restricting the text search to more specific scenarios or terms.

Appendix C: NTSB CAROL tutorial

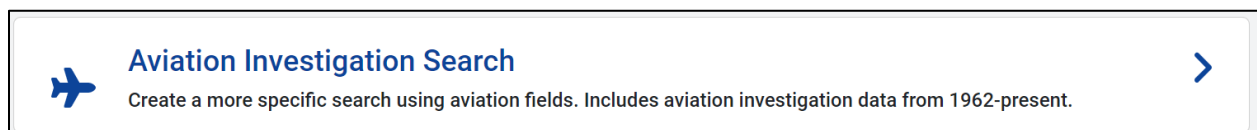
Search execution

Techniques for searching and extracting relevant data

- As a first step, it would be helpful to conduct several searches with specific search terms to attempt to find the best results.
- The database also includes predefined, downloadable data sets on the following page:
<https://data.nts.gov/avdata>.
 - This page contains data sets for 1962-1981 (PRE1982.zip, seen in Figure 1) and for 1982-present (avall.zip, seen in Figure 2) in Microsoft Access. These data sets download as tables containing information about aviation accidents, such as aircraft data, narrative data, and events sequence.
- As noted above, text searches do not allow for operator functionality (AND, OR). Therefore, users may need to conduct multiple, separate searches.
- Searches can be downloaded via CSV or JSON allowing for report contents to be analyzed via statistical software.

Search tutorial

1. In a preferred search engine, search <https://carol.nts.gov/>
2. Select "Aviation Investigation Search"



3. To choose search terms, input the relevant information in the boxes or select options from the dropdowns.

Submit Query

4. Click "Submit Query."
5. Optional:
 - Download data sets from the search by scrolling down past the results and clicking one of the following buttons: "Download Data (JSON)" and "Download Data (CSV)."



6. Results will return in a table with the following columns:

NTSB#	Public Docket	Event Date	Location	Report	Original Published Date	Event Type	Highest Injury	Aircraft Details	Status	Has Safety Rec
-------	---------------	------------	----------	--------	-------------------------	------------	----------------	------------------	--------	----------------

7. Optional:

- Click on the hyperlinked column titles (e.g. [NTSB#](#)) to sort results by that topic either ascending or descending

Example search

Objective

The users want to better understand pilot behavior and flight crew workload under time pressure. To do this, they want to study one or two highly applicable case studies.

Process

Searching FAR Part 121 returns 2592 results. However, the database is not searchable by probable cause, so there is no way to narrow the results down to be useful to the search. This means that users need to either have a very broad research goal (such as phase of flight or damage), or know of the accident before looking up the report. In this case, the users know of US Airways Flight 1549, otherwise known as “The Miracle on the Hudson,” and would like to look this accident up.

Search Attempt 1:

The users first filter by date.

1. **Event Start Date** was 01/15/2009
2. **Event End Date** was 01/15/2009

This search returns 3 results, which may be a manageable number to sift through for the correct result. However, if necessary, they would proceed by narrowing their search.

Search Attempt 2:

The users narrow the search to FAR Part and date.

3. **FAR Part** was Part 121
4. **Event Start Date** was 01/15/2009
5. **Event End Date** was 01/15/2009

This search returns 1 result, which is the report the users are looking for: [AAR1003](#): “Loss of Thrust in Both Engines After Encountering a Flock of Birds and Subsequent Ditching on the Hudson River,” which covers US Airways Flight 1549. From this summary landing page, the users can access the [docket](#), [full report](#) and links to each of the 35 recommendations.

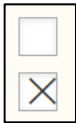
Appendix D: SKYbrary Accident and Incident Browser Tutorial

Search execution

Techniques for searching and extracting relevant data

1. This database automatically includes the number of articles included in each search filter. Users can potentially use this information to narrow their search down to a manageable quantity of results without iterative searching.
2. There is no “or” operator when selecting multiple search filters, so searches can only become more constrained. For example, once the filter “Food Poisoning” is selected, only the search terms that relate to the article in which food poisoning was a factor will be selectable. There is no way to expand a search to include articles from several causes. As such, users may benefit from:
 - Running multiple independent searches with different search filters from their research topic. For example, if there interested in how pilots respond to fatigue and distraction, it is probably best to filter for distraction and fatigue separately to provide the greatest breadth of results.
 - Selecting the search term related to the research topic with the greatest number of results before narrowing their search can help ensure relevant reports are included in the search. For example, if there is an interest in the types of mistakes pilots make while distracted, one should filter for distraction (181 results) before filtering for types of mistakes.

Search tutorial

1. In a preferred search engine, search <https://skybrary.aero/accidents-and-incidents>
2. The search filters are in a column on the right side. Select the “+” dropdowns  for categories relevant to your search.
3. Check the boxes next to your chosen search filters.  The search results will automatically update after each box has been checked.
4. Optional:
 - a. To remove an individual selected search filter, click the “X” in the checkbox.
 - b. To remove all selected search filters, scroll to the bottom of the search categories and

click the button to “Reset filters.”

Reset filters

5. View Results

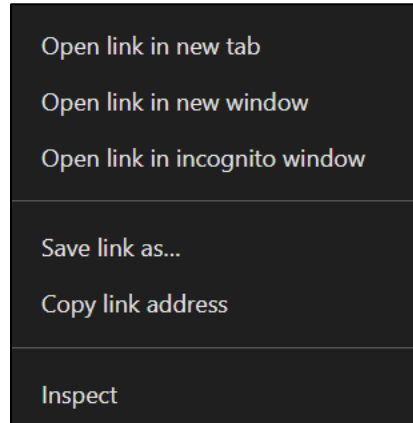
- a. Optional: Click on the link for a desired result. To view other results, select the “back

page” button on the browser.



- b. Alternative:

- i. Right-click on each desired result and select “Open link in new tab”



- ii. View results laterally in the browser tabs.

Example search

Objective

The users want to better understand incidents in the United States in which pilots make mistakes while controlling the aircraft, due to insufficient training or understanding of processes. This research is taking place in July of 2024, and the users are aiming for a sample size of $n=30$.

Process

Search Attempt 1:

The users must first select a baseline set of search filters.

1. **When:** From Date 01/01/2019 To Date 07/01/2024
2. **Human Factors:** Inappropriate Crew Response – Skills deficiency
3. **Causal Factor Group(s):** Aircraft Operation

This search returns 45 results, of which only 4 occurred in the United States. These results are too narrow for the users.

Search Attempt 2:

To broaden the search, the users eliminate the time limit, but do not alter any other search terms.

1. **When:** No time limit
2. **Human Factors:** Inappropriate Crew Response – Skills deficiency
3. **Causal Factor Group(s):** Aircraft Operation

This search returns 258 results, of which 24 occurred in the United States. 24 relevant articles may be enough for the users, who may alter their research statement to have $n=20$ samples. However, the users need to ensure that the dataset of 24 does not include more than 4 erroneous results. If the users want a greater number of results, they may consider running additional, related searches. These may include searches with human factors terms such as the following: Inappropriate Crew Response – Automatics; Inappropriate Crew Response – Technical Fault; or Manual Handling.

Appendix E: Identified database summary

Name	Description	Comments	Link (if applicable)
Airbus accident data	Provides a dashboard for some Airbus statistics, along with links to other statistics pages including Generations of Jet, Fatal Accidents, Hull Losses, Accident Categories, Accident Rates by Category & Generation, and Accidents by Flight Phase. The main page has summary information on world traffic in flight cycles per week from 2019 to 2023, and the comparison of traffic and accidents between 2022 and 2023.	Contains quantitative and summary information; however, it does not contain detailed analyses of causes of accidents or human factors.	https://accidentstats.airbus.com/#::~:~:text=Traffic%20and%20Accidents%20in%202023&text=There%20were%20no%20fatal%20accidents,turbo%20prop%20aircraft%2C%20with%20five%20fatalities.
Boeing annual safety report	Opens a PDF document for the Boeing annual safety report. The report includes statistics, accidents information, summaries and rates, fatalities, regions, etc.	This document provides quantitative information on the year's aviation accidents and incidents but does not necessarily provide insights into the contributing factors or causes.	https://www.boeing.com/content/dam/boeing/boeingdotcom/company/about_bca/pdf/statsum.pdf
BTS - Bureau of Transportation Statistics Home Page	Includes statistics on air traffic, capacity, fatalities, and accidents.	Irrelevant to AED tasks; does not discuss safety or causes of incidents or accidents.	https://www.transtats.bts.gov/
BTS Airlines, Airports, and Aviation	Includes a "quick links" button that leads to a webpage with access to United States data on air carrier performance, financials, traffic, and other items. It references the National Transportation Library repository (ROSA P) for research, policy, and historical reports.	Irrelevant for AED tasks.	https://www.bts.gov/topics/airlines-airports-and-aviation
BTS Aviation Accident Statistics	Linked by the BTS Database Finder but page no longer exists.	Inactive	http://www.nts.gov/data/aviation_stats.html
BTS Census of Fatal	Dashboard of Bureau of Labor Statistics injuries,	Irrelevant to AED tasks;	https://www.bls.gov/iif/home

Name	Description	Comments	Link (if applicable)
Occupational Injuries (CFOI)	illnesses, and fatalities; includes aviation fatalities.	information is not searchable.	.htm
BTS Database Finder	Includes links to 27 databases and descriptions which can be searched for "Aviation" and "Safety".	Links to other datasets/databases; contains several inaccessible links.	https://www.transtats.bts.gov/databases.asp?20=F&Z1qr_VQ=E&Z1qr_Qr5p=N8vn6v10&f7owrp6_VQF=D
BTS National Transportation Statistics	A webpage with a collection of links to transportation statistics. Includes access to aviation database information, primarily the accordion dropdown Section 2.B - Air Safety. Provides links to other datasets.	Only provides links to other datasets/databases.	https://www.bts.gov/topics/national-transportation-statistics
BTS Prohibited items Intercepted at Airport Screening Checkpoints	Includes an embedded dataset and access to an Excel spreadsheet with a graph of the number of firearms intercepted at airport screening checkpoints and the number of domestic enplanements from 2002 to 2022. The dataset also includes a table separating the prohibited items by type (firearms, knives, box cutters, other cutting instruments, clubs, incendiaries, and other).	Includes only quantitative data by year; irrelevant to AED tasks.	https://www.bts.gov/content/prohibited-items-intercepted-airport-screening-checkpointsa
BTS Safety Data Library	Provides background information on safety in transportation and links to sources of transportation safety data (e.g., NHTSA, FRA, FAA, NTSB, National Transportation Library Repository).	Only provides links to other databases; redundant with other entries.	https://www.bts.dot.gov/topics/safety
BTS U.S. Air Carrier Fatal Accidents by Defining Event of Operation	Includes an embedded dataset and access to an Excel spreadsheet with a table of U.S. Air Carrier Fatal Accidents by Defining Event of Operation. Table includes total fatal accidents and phase of operation separated into approach/landing, taxi/climb, cruise (in-flight), standing (static), maneuvering, pushback/tow, and other/not reported from 1990 to 2022.	Includes only quantitative data by year; irrelevant to AED tasks.	https://www.bts.gov/content/us-air-carrier-fatal-accidents-first-phase-operationb
BTS U.S. Air Carrier	Includes an embedded dataset and access to an	Includes only quantitative data	https://www.bts.gov/content

Name	Description	Comments	Link (if applicable)
Safety Data	Excel spreadsheet with a graph of U.S. Air Carrier Safety Data for all-cargo carriers and commercial operators of large aircraft. Graph includes fatalities, total serious injuries, total accidents, fatal accidents, and flight hours between 2005 and 2023.	by year; irrelevant to AED tasks.	/us-air-carrier-safety-data
BTS U.S. Commuter Air Carrier Fatal Accidents by Defining Event of Operation	Includes an embedded dataset and access to an Excel spreadsheet with a table of U.S. Commuter Air Carrier Fatal Accidents by Defining Event of Operation. Table includes total fatal accidents and phase of operation separated into approach/landing, taxi/climb, cruise (in-flight), standing (static), maneuvering, and other/not reported from 1990 to 2022.	Includes only quantitative data by year; irrelevant to AED tasks.	https://www.bts.gov/content/us-commuter-air-carrier-fatal-accidents-first-phase-operation
BTS U.S. Commuter Air Carrier Safety Data	Includes an embedded dataset and access to an Excel spreadsheet with a graph of U.S. Commuter Air Carrier Incident Per Flight Hours Rate for all-cargo carriers. Graph includes fatalities, serious injuries, total accidents, fatal accidents, and flight hours between 2005 and 2023.	Includes only quantitative data by year; irrelevant to AED tasks.	https://www.bts.gov/content/us-commuter-air-carrier-safety-data
BTS U.S. General Aviation Safety Data	Includes an embedded dataset and access to an Excel spreadsheet with a graph for U.S. General Aviation Incident Rate per Flight Hours. Graph includes fatalities, seriously injured persons, total accidents, fatal accidents, and flight hours between 2005 and 2023.	Outside of project scope (Part 121/135 flight operations); only includes statistics on fatalities and injuries by year.	https://www.bts.gov/content/us-general-aviation-safety-data
BTS U.S. On-Demand Air Taxi Safety Data	Includes an embedded dataset and access to an Excel spreadsheet with a graph of U.S. On-Demand Air Taxi Safety Data. Graph includes fatalities, seriously injured persons, total accidents, fatal accidents, and flight hours from 2005 to 2023.	Includes only quantitative data by year; irrelevant to AED tasks.	https://www.bts.gov/content/us-demand-air-taxi-safety-data

Name	Description	Comments	Link (if applicable)
FAA Accident and Incident Data	Contains both a list of links to reports on preliminary data (via ASIAs), final data (via NTSB), commercial air carrier fatalities (link broken), a search function (via NTSB's CAROL), and links to Air Traffic Control audio for accidents between 2006 and 2019.	Information outdated; database is not searchable on relevant factors.	https://www.faa.gov/data_research/accident_incident
FAA Accident and Incident Data System (AIDS)*	Contains incident records for civil aviation that do not meet the NTSB criteria of “accident”. Report details vary. Provides reports from 1978-2025.	Narrative can provide information on flightcrew behavior. Human factors-related filters are not available.	https://www.asias.faa.gov/apex/f?p=100:12:::NO
FAA Accident Investigation and Prevention Policies and Forms	A webpage listing FAA Orders and links to Order information.	Irrelevant to AED tasks; there are only two Orders, one of which was cancelled in 2018.	https://www.faa.gov/about/office_org/headquarters_offices/avs/offices/avp/orders_forms
FAA Accident Lessons Learned Libraries	Provides background information on the FAA Lessons Learned dashboard, which uses selected accidents to provide useful information to practitioners. Includes access to three libraries: Small Airplane Accident Library, Transport Airplane Accident Library, and Rotorcraft Accident Library (described below).	Provides links to other datasets/databases.	https://www.faa.gov/lessons_learned
FAA Aircraft Registry (AR)*	Contains information on current registered aircraft owners, documents, dealers, and more.	Has not been updated since 2022; irrelevant to AED tasks.	https://www.faa.gov/licenses_certificates/aircraft_certification/aircraft_registry/interactive_aircraft_inquiry
FAA Aviation Safety Information Analysis and Sharing (ASIAS) Homepage*	Home page for the ASIAs database, which links various ASIAs resources.	Only serves as a landing page to provide links to other datasets/databases.	https://www.asias.faa.gov/apex/f?p=100:1
FAA Data & Research – Safety landing page	Website with safety data, which provides additional linked resources.	Provides links to other databases but does not directly	https://www.faa.gov/data_research/safety

*Denotes database incorporated into the FAA ASIAs system.

Name	Description	Comments	Link (if applicable)
		provide information.	
FAA Dynamic Regulatory System (DRS)	Searchable database for regulatory guidance material from the Office of Aviation Safety. Includes all information found in the Flight Standards Information System (FSIMS) and the Regulatory Guidance System (RGL).	Regulatory and guidance material does not contain useful information on human factors assumptions.	https://www.faa.gov/about/office_org/headquarters_offices/avs/programs/drs
FAA Near Mid Air Collision System (NMACS)*	Searchable database tracking near midair collisions; includes search functionality for event report, date, evaluation, location, operations, and aircraft.	Irrelevant to AED tasks; has not been updated since 2021.	https://www.asias.faa.gov/apex/f?p=100:33::NO
FAA Preliminary Accident and Incident Notices (subset of ASIAs)*	A table of aircraft events and fatal accidents categorized by aircraft type. Includes links to simple information pages on the reported accident or incident, updated daily.	Consists of preliminary information on fatal accidents without established causes; not searchable on relevant factors.	https://www.asias.faa.gov/apex/f?p=100:93::NO
FAA Quarterly Enforcement Report	Includes quarterly compilation of enforcement actions "against regulated aviation entities that are closed with either a civil penalty or issuance of a certificate suspension or revocation," accessible as PDFs that link from the webpage.	Dashboard only contains access to quarterly reports; information is irrelevant to AED tasks.	https://www.faa.gov/about/office_org/headquarters_offices/agc/practice_areas/enforcement/reports
FAA Rotorcraft Accident Library (Lessons Learned)	Includes access to various information, including the accident library, accident statistics, and three "perspectives" with relevant accidents. The accident library includes 14 accidents from 1993 to 2018. The perspectives are "Rotorcraft Life Cycle," "Threat Categories/Industries," and "Common Themes," accessible via a link.	Out of scope for current project (Part 121/135 operations).	https://www.faa.gov/lessons_learned/rotorcraft
FAA Runway Incursions (RWS) database*	Searchable database to identify runway incursions. Maintained by the Runway Safety Office, each entry has a description of the event, basic information such as date, and aircraft involved along with the incursion severity rating.	Irrelevant to AED tasks.	https://www.asias.faa.gov/apex/f?p=100:28::NO:28::
FAA Runway Safety	Summary page with access to dashboards for	Dashboards only contains data	https://www.faa.gov/airports

*Denotes database incorporated into the FAA ASIAs system.

Name	Description	Comments	Link (if applicable)
Statistics	runway incursion totals by quarter (compared between the two most recent fiscal years), and runway incursion totals for fiscal years from 2019 to present.	snapshot for incursion totals; does not contain searchable information; irrelevant to AED tasks.	/runway_safety/statistics
FAA Safety Record of Airlines/Aircraft	Provides information on how to submit requests for records on accidents, enforcements, and maintenance difficulties. Provides links to an FAA Aviation Safety Data Page and NTSB Aviation Accident Database.	Irrelevant to AED tasks; redundant with other entries.	Link unavailable
FAA Small Airplane Accident Library (Lessons Learned)	Includes access to statistics and a database of accidents and lessons learned reports for commercial and cargo accidents. Small Airplane Lessons Learned Library includes 13 reports from 1959 to 2019.	Very limited number of accidents; requires signing in via Tableau; out of scope for current project (Part 121/135 operations).	https://www.faa.gov/lessons_learned/small_airplane
FAA Transport Airplane Accident Library (Lessons Learned)	Includes access to various information, including the accident library, accident statistics, and three "perspectives" with relevant accidents. The accident library includes a list of selected accidents from around the world, but no search function. The perspectives are "Airplane Life Cycle," "Threat Categories/Groupings," and "Common Themes," accessible via a link.	Themes are grouped into five broad categories that may be relevant from a human factors perspective ("Flawed Assumptions," "Human Error," "Organizational Lapses," "Pre-existing Failures," and "Unintended Effects"). Themes cannot be adjusted; database is not searchable.	https://www.faa.gov/lessons_learned/transport_airplane
FAA UAS Accident and Incident Preliminary Reports*	A table and downloadable CSV file of Unmanned Aircraft System accident and incident preliminary reports with information on the sponsor category (academia or NASA in most cases), sponsor, date, location, type, aircraft, and initial report description.	Irrelevant to AED tasks.	https://www.asias.faa.gov/apex/f?p=100:446:::NO:446
FAA UAS Quarterly Scorecard	Originally designed to provide quarterly reports on UAS topics and events, in the USA and abroad. Has	Website is not updated.	https://www.faa.gov/about/plans_reports

*Denotes database incorporated into the FAA ASIAs system.

Name	Description	Comments	Link (if applicable)
	not been updated since 2021.		
IATA annual safety reports	Page developed by the International Air Transport Association (IATA) to provide information and access to the IATA Interactive Safety Report page. The Interactive Safety Report has tabs with information including a summary, accidents, regional information, IOSA and non-IOSA, accident counts, threat and error management, accident investigation reports, and a general list of accidents. Each tab includes relevant quantitative data and graphics, and many are searchable.	Quantitative information useful for understanding trends in aviation safety but do not t provide insights into human factors contribution to accidents or incidents.	https://www.iata.org/en/publications/safety-report/
ICAO library of e-reports	Links to a page for the International Civil Aviation Organization (ICAO). The e-Library opens a search page on applicable accident information including the date, location, responsible entity, occurrence class, manufacturer, registration, operator, injury level, and other information.	Challenging user interface for submitting queries and accessing reports.	https://www.icao.int/safety/airnavigation/aig/pages/e-library-of-final-reports.aspx
Interstate Aviation Committee (IAC) Accident Investigations	Contains information and reports on aviation accident investigations relevant to the States in the interstate Agreement on Civil Aviation and Use of Airspace. The countries involved in the IAC are Azerbaijan, Armenia, Belarus, Kazakhstan, Kyrgyzstan, Russian Federation, Tajikistan, Turkmenistan, and Uzbekistan.	This database contains information about the countries that are a part of the IAC and may not have information relevant to the U.S. Limited search and filter functionality.	https://mak-iac.org/en/
NASA Aviation Safety Reporting System: General Aviation Flight Training Reports	A webpage linking to a PDF document with a list of sampled reports involving aviation flight training; page was last updated in 2023.	Outside project scope (Part 121/135 flight operations); non-searchable information; outdated information.	https://catalog.data.gov/data/set/aviation-safety-reporting-system-general-aviation-flight-training-reports
NASA/FAA Flight Test Safety Database	Provides information on "Test Hazard Analyses" documented by NASA for industry knowledge. Creates a table with Maneuver, Reference Number,	Scope limited to flight testing; last updated in 2019. Link is no longer functional and may have	Link unavailable

Name	Description	Comments	Link (if applicable)
	Hazard, and more. Has built-in search features to allow browsing.	been decommissioned.	
National Wildlife Aircraft Strike Database	Includes records of wildlife strikes since 1990, with a searchable database.	Wildlife strikes are too narrow a context to provide useful data.	Link unavailable
NLR Air Safety Database	The Netherlands Aerospace Center (NLR) Air Safety Database has a broken link on SKYbrary, which says that it contains information on accidents and incidents of fixed wing aircraft and helicopters from 1960. SKYbrary says the database contains information on 37,000 accidents and incidents worldwide and an additional 310,000 commercial airline safety reports for fixed wing aircraft and helicopters and non-accident data.	Unable to access the database (likely decommissioned).	
NLR Safety Methods Database	A living document created by the Netherlands Aerospace Centre (NLR) with the intention of use for safety assessments. Includes information on safety techniques, methods, databases, and models. Has three parts: Overview of Safety Methods, Statistics, and References.	Contains information on safety assessment tools (e.g., ACAT - Air Carrier Assessment Tool). Consists of a single PDF; not searchable. Link is no longer functional and may have been decommissioned.	https://www.nlr.nl/wp-content/uploads/2021/12/Safety-Assessment-Techniques-Database_1.2_3-Nov-2020.pdf
NTSB Aviation Accident Database and Synopsis	Linked by the BTS Database Finder but page no longer exists.	Inactive	http://www.nts.gov/aviationquery/index.aspx
NTSB Investigation/Accident Reports - Aviation	A list of Accident Reports produced by the NTSB including information about the accident, factual data, conclusions and probable cause, and safety recommendations.	Unlike NTSB's CAROL, the links are in a non-searchable list. Several links to Accident Reports listed on the webpage are inaccessible or contain broken links.	https://www.nts.gov/investigations/AccidentReports/Pages/Reports.aspx?mode=Aviation
NTSB Safety Recommendations to	Searchable database with NTSB's Recommendations to the FAA and corresponding	Only provides information on safety recommendations;	https://www.asias.faa.gov/apex/f?p=100:38:::NO::

Name	Description	Comments	Link (if applicable)
the FAA with FAA Responses (NTSFE)	FAA responses; search criteria include a narrative search, recommendation number, date, subject, and status.	limited scope may exclude important information relevant to AED.	
PHMSA Hazardous Material Incident Reporting System (HMIRS)	Home page for the Pipeline and Hazardous Materials Safety Administration, including news, data, safety initiatives, and links to other relevant pages.	Irrelevant to AED tasks.	https://www.phmsa.dot.gov/
Society of Experimental Test Pilots (SETP) Technical Library	Collection of the SETP's publications, documents, and websites.	Only accessible with a membership.	https://www.setp.org/technical-library/technical-library.html
Society of Flight Test Engineers Papers Database	Collection of the Society of Flight Test Engineers' publications and documents.	Only accessible with a profile.	https://sfte.org/sfte_papers_database.php
U.S. DOT Repository & Open Science Access Portal (ROSA P)	A collection of open access research articles and reports. Users can complete searches in their collections.	Research articles may provide useful insights for AED analysts; high quality government reports.	https://rosap.ntl.bts.gov/
World Aircraft Accident Summary (WAAS) – Subset	A subset of the WAAS purchased by FAA to be made public; includes data and descriptive information about all known airline accidents with passenger fatalities for the last 10 years. Can only be searched by narrative and operator name.	Limited scope; redundancy with NTSB CAROL database.	https://www.asias.faa.gov/apex/f?p=100:44:::NO

U.S. Department of Transportation
John A. Volpe National Transportation Systems Center
220 Binney St.
Cambridge, MA 02142-1093

617-494-2000
www.volpe.dot.gov

DOT-VNTSC-FAA-25-04